

A2 Séries Temporais

Alunos: Mateus Mourão, Miguel Mediano, Adrian Castro, Gustavo Reis, Luiza Schneider e Thiago Melchiors

Baselines

Para avaliar os diferentes métodos de previsão, os dados de consumo foram divididos em 75% para treino e 25% para teste. Foram analisados quatro métodos baselines: **método da média**, o método **ingênuo simples**, o **ingênuo sazonal** e o **método do desvio**. Além disso, foram aplicadas transformações aos dados para investigar possíveis melhorias no desempenho.

O **método da média** calcula a média dos valores do conjunto de treino e utiliza esse valor como previsão constante para o teste. Embora seja um método simples, mostrou-se competitivo para séries sem sazonalidade ou tendências marcantes.

O método **ingênuo simples** utiliza como previsão o último valor observado no conjunto de treino, o que resulta em uma previsão constante para todos os períodos de teste. Por outro lado, o método **ingênuo sazonal** se baseia no padrão sazonal da série temporal, replicando os valores do último ciclo sazonal do treino para o conjunto de teste. Esse método apresentou um desempenho superior em séries com sazonalidade clara, com métricas ligeiramente melhores em relação ao ingênuo simples.

Já o **método do desvio**, que prevê incrementos ou decrementos constantes com base na média das diferenças observadas no treino, teve desempenho inferior, especialmente devido à ausência de uma tendência definida nos dados.

Buscando aprimorar as previsões, foram testadas transformações nas variáveis, como a **raiz quadrada**, **Box-Cox** e **Yeo-Johnson**. A transformação **raiz quadrada** demonstrou melhorias modestas, com resíduos ligeiramente mais próximos da normalidade. A transformação **Box-Cox**, ajustada com $\lambda = 1.347$, apresentou os melhores resultados gerais, com melhorias significativas nas métricas de erro e uma distribuição de resíduos mais adequada. Já a transformação **Yeo-Johnson** teve desempenho similar ao da raiz quadrada, com uma redução na autocorrelação dos resíduos, mas sem ganhos expressivos nas métricas de erro.

Os resultados finais indicam que, embora as transformações (especialmente a Box-Cox) melhorem a adequação dos modelos aos dados, os métodos baselines, em particular o ingênuo sazonal, continuam sendo alternativas eficientes devido à sua simplicidade e capacidade de capturar padrões sazonais básicos. As métricas detalhadas e os gráficos relacionados confirmam essas observações e destacam as diferenças entre os métodos analisados. A visualização dos modelos e avaliação com as métricas avaliadas estão exibidas no repositório do github.

Suavização Exponencial

Com o objetivo de comparar o modelo com os baselines, separamos os primeiros 75% dos dados de consumo para treino e os últimos 25% para fazer essa comparação.

Assim, ajustando o modelo de Suavização Exponencial Simples, definido por:
 $\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t$, onde α é o parâmetro de suavização e \hat{y}_t representa o valor ajustado

em t , chegamos ao valor otimizado do parâmetro de suavização $\alpha^* = 0.2814302652890589$

Como esse valor é próximo de 0, o modelo se aproxima à previsão pelo método da média simples, o que de fato ocorre, já que a primeira previsão da suavização exponencial é de **0.78376579**, e a média até aquele momento é igual a **0.83642965**. Por fim, o método ingênuo prevê o último observado, dado por **0.96309962**. A comparação dessas previsões pode ser vista de forma mais detalhada no seguinte gráfico: (ver imagem 1)

Observe que, como os valores analisados são muito pequenos (entre -2 e 2 na maior parte), as previsões do modelo de suavização exponencial são praticamente as mesmas para todos os dados do teste, tendo uma diferença de ordem inferior a 10^{-8} .

Analisando o RMSE e o MAE dessas previsões, vemos que a suavização exponencial de fato possui o menor erro nas duas métricas. (ver imagem 2)

Analisando os gráficos da distribuição dos resíduos e das funções de autocorrelação e autocorrelação parcial, observou-se que, embora quase todos os valores de autocorrelação caíam dentro do intervalo de confiança e os resíduos pareçam seguir uma distribuição normal, é possível que haja uma melhora na adequação dos modelos através do uso de transformação de variáveis. (ver imagem 3)

Então, foram aplicadas diferentes transformações (raiz quadrada, box-cox e yeo-johnson) e os modelos foram ajustados com todos os dados para ver se elas apresentavam alguma melhoria. Dessa forma, verificou-se que a transformação Box-Cox acarretava em uma melhoria não tão significativa em relação às mesmas métricas de comparação de erro, mas ao analisar a distribuição dos resíduos, percebe-se que ela está mais próxima de uma distribuição normal, revelando uma melhor adequação do modelo. (ver imagens 4 e 5)

Além disso, modelos mais complexos, como Holt (com tendência) e Holt-Winters (com sazonalidade), foram testados e suas previsões não apresentaram melhorias em relação ao modelo ajustado apenas com a suavização exponencial simples, o que pôde ser verificado analisando as métricas de erro do modelo com tendência (imagem 6) e o gráfico do modelo com sazonalidade (imagem 7). Isso sugere que os dados analisados possuem uma tendência e sazonalidade pouco pronunciadas, tornando desnecessária a decomposição nesses componentes para este caso específico.

Assim, o modelo final foi ajustado da seguinte forma: $\hat{y}_{t+1} = \alpha y_t^\lambda + (1 - \alpha)\hat{y}_t - 2.277$, sendo $\lambda = 1.347$ parâmetro da transformação Box-Cox, $\alpha = 0.288$ o parâmetro otimizado e -2.277 o offset adicionado para deixar os valores positivos para fazer a transformação.

Regressão Linear

Ao rodar o modelo de regressão linear na base de dados obtivemos o resultado que pode ser visto resumidamente em 10. Analisando-o, vemos que o modelo tem um bom ajuste (R^2 de 0,768), o que indica que aproximadamente 76,8% da variação na variável dependente "Consumption" pode ser explicada pelas variáveis independentes incluídas no modelo.

Os coeficientes do modelo sugerem que "Income" ($\beta = 0.6668$) tem um impacto positivo significativo sobre o consumo, enquanto "Production" ($\beta = 0.0718$) também influencia positivamente, embora com menor significância estatística. Por outro lado, "Savings" ($\beta = -0.6329$) apresenta uma forte relação negativa com o consumo, sugerindo que aumentos na poupança levam a reduções no consumo. O coeficiente para "Unemployment" ($\beta = -0.0642$) indica que um aumento no desemprego está associado a uma redução no consumo, contudo, o p -valor próximo ao limiar de significância (0.069) requer cautela na interpretação.

A Estatística Omnibus quase zero e o Prob(Omnibus) indicam que os resíduos do modelo não seguem uma distribuição normal, um ponto de atenção que pode afetar a validade das inferências estatísticas. A estatística Durbin-Watson de 2.218 sugere ausência de autocorrelação significativa nos resíduos, e o teste de Jarque-Bera reforça a não normalidade dos resíduos ($p = 1.71 \times 10^{-8}$).

Portanto, apesar da capacidade do modelo de explicar significativamente a variação no consumo, as questões relacionadas com a distribuição dos resíduos podem necessitar de uma investigação mais aprofundada para garantir a adequação do modelo.

Modelo SARIMA

Teste de Estacionariedade (ADF): O teste de Dickey-Fuller Aumentado (ADF) foi aplicado para verificar se a série temporal é estacionária. A estatística calculada foi de -4.610, e o p-valor foi 0.0001, muito abaixo do nível de significância comum de 5% (0.05). Isso significa que rejeitamos a hipótese nula (H_0) de que a série não é estacionária. Em outras palavras, os dados já são estacionários ou foram transformados de forma adequada (por exemplo, com diferenciação) para atender a essa condição. Isso é crucial, pois a estacionariedade é um requisito fundamental para a modelagem SARIMA.

AR(autoregressivo): O modelo considera uma dependência com o valor imediatamente anterior da série. **I (diferenciação):** Indica que a série foi diferenciada uma vez para atingir estacionariedade. **MA (média móvel):** Inclui um termo de média móvel para modelar erros passados. (1, 1, 1, 12) captura os componentes sazonais com periodicidade de 12, o que sugere que os dados possuem um padrão sazonal anual (por exemplo, meses do ano).

AIC = 372.483 e **BIC** = 388.585. Essas métricas penalizam modelos mais complexos para evitar overfitting, com valores menores indicando melhor ajuste. O AIC e BIC mostram que o modelo é relativamente bem ajustado, mas podem ser usados para comparar com outras

AR.L1 (componente autoregressivo): O coeficiente é 0.2918, e o p-valor associado é menor que 0.05, indicando significância estatística. Isso implica que o valor atual da série tem uma correlação positiva com o valor imediatamente anterior. **MA.L1 (componente de média móvel):** O coeficiente é -0.8090, também estatisticamente significativo. Isso sugere que os erros passados influenciam os valores atuais de forma negativa. **AR.S.L12 e MA.S.L12 (componentes sazonais):** Esses parâmetros capturam a sazonalidade da série com periodicidade de 12 (mensal). Ambos os coeficientes são significativos, indicando que o modelo considera corretamente o padrão sazonal anual. Foram realizadas previsões para os próximos 12 períodos. Os valores previstos mostram uma oscilação que segue a tendência e sazonalidade observadas nos dados originais, o que sugere que o modelo conseguiu capturar bem os padrões da série temporal. Essa previsão pode ser útil para planejamento ou decisões baseadas no consumo projetado.

Erro absoluto médio (MAE) = 0.383: Este valor representa o erro médio absoluto entre os valores previstos e os valores reais. **Raiz do erro quadrático médio (RMSE) = 0.437:** Esta métrica penaliza erros maiores, e um valor baixo indica que o modelo tem um bom desempenho preditivo. Ambos os valores são baixos, o que mostra que o modelo conseguiu ajustar bem os dados e produzir previsões precisas.

Os resíduos do modelo foram analisados para verificar a adequação do ajuste: O gráfico dos resíduos mostra que eles oscilam em torno de zero, sem grandes padrões aparentes,

indicando que o modelo capturou bem as dinâmicas da série. O gráfico de autocorrelação dos resíduos (ACF) mostra que a maioria das autocorrelações está dentro dos limites de confiança, sugerindo que os resíduos são essencialmente ruído branco. No entanto, há pequenas autocorrelações que podem ser investigadas. O teste de Ljung-Box foi aplicado para verificar a independência dos resíduos. O p-valor obtido foi 0.0027, indicando que ainda há alguma correlação nos resíduos. Isso sugere que o modelo pode ser refinado, mas o impacto no desempenho geral é provavelmente pequeno.

O modelo SARIMA ajustado captura de maneira eficiente as características da série temporal, incluindo tendências e sazonalidades. As previsões são precisas, como mostram as métricas de erro (MAE e RMSE). Apesar disso, o teste de Ljung-Box indica que ainda há alguma correlação nos resíduos, o que pode ser explorado com ajustes adicionais, como a inclusão de variáveis exógenas ou revisão dos parâmetros sazonais. No geral, o modelo fornece uma boa base para previsão e análise da série temporal estudada.

Modelos de Convolução 1D

Além dos modelos tradicionais e da modelagem SARIMA, foi explorado o uso de Redes Neurais Convolucionais (CNN) para prever a variável dependente. A abordagem utilizou uma arquitetura unidimensional (1D), considerando a natureza sequencial dos dados temporais. O modelo foi avaliado com diferentes combinações de parâmetros, incluindo o tamanho do kernel (3, 5, 7, 9), o número de canais ocultos (8, 16, 32, 64) e o número de camadas convolucionais (1, 2, 3, 4) e o melhor resultado foi obtido com tamanho do kernel 7x7, 16 número de canais ocultos e 4 camadas convolucionais, tendo um erro quadrático médio de 0.083202 com desvio padrão 0.005191.

Durante o desenvolvimento do modelo, foram realizados testes para avaliar o impacto da normalização dos dados e da inclusão de uma variável categórica "Quarter" com apenas o número equivalente ao quarto do ano. Os resultados obtidos estão organizados na tabela 1.

Avaliação Final da CNN

Para avaliar a robustez do modelo, foi realizado um novo teste, desta vez com normalização dos dados e sem a inclusão da variável "Quarter" e com os hiperparâmetros encontrados anteriormente. Neste experimento, o conjunto de dados foi dividido entre treinamento e teste, os 80% primeiros dados foram escolhidos como treino e o restante das datas não foi visto pelo modelo. Com isso, avaliou-se a acurácia do modelo nesses dados não vistos e o erro médio quadrático (MSE) obtido foi de 0.07424131780862808. Além disso, foi feito um gráfico com a série temporal real e a previsão feita pelo modelo (apenas dos 20% não inclusos no treinamento) que pode ser visto em 11.

Modelos RNN

O modelo escolhido é uma rede neural Long Short-Term Memory (LSTM) projetada para lidar com dados sequenciais e temporais. A arquitetura específica utilizada neste caso é composta por uma camada LSTM com um número ajustável de unidades ocultas (`hidden_layer_size`) configurada para processar os dados sequenciais, seguida por uma camada de dropout (com taxa de 20%) que ajuda a regularizar a rede e prevenir que ela

se ajuste demais aos dados de treinamento (overfitting), o que é crucial dado o tamanho limitado do conjunto de dados. A camada final **fc** mapeia as unidades ocultas da LSTM para a previsão final.

A escolha da LSTM se deve à sua capacidade de capturar dependências temporais e sequenciais ao introduzir mecanismos de "gates" que controlam o fluxo de informações, permitindo que o modelo retenha ou descarte informações ao longo do tempo. No entanto, como o conjunto de dados disponível é relativamente pequeno, a rede foi projetada com poucas camadas e unidades ocultas para evitar que o modelo aprenda ruídos ou variações esparsas que não são representativas das tendências reais da série temporal.

Foram escolhidas três métricas de erro com o objetivo de fornecer uma visão abrangente do desempenho do modelo: o MSE mede a média dos quadrados das diferenças entre os valores reais e previstos, penalizando fortemente grandes erros; o R^2 avalia quanto da variabilidade dos dados pode ser explicada pelo modelo (próximo de 1 indica um bom ajuste, enquanto valores negativos indicam que o modelo está fazendo previsões piores que uma simples média dos dados); o MAPE mede o erro percentual médio, fornecendo uma visão mais intuitiva da precisão do modelo.

A estratégia adotada foi escrever um script com loops aninhados para testar diferentes combinações de estratégias e variáveis, com o objetivo de medir qual delas seria mais eficiente para prever o consumo. Isso resultou na criação de 16 modelos diferentes, nos quais foram comparadas as abordagens de normalização dos dados, remoção de tendências e sazonalidades, e diferentes tamanhos de janelas de tempo. A normalização foi feita utilizando o **MinMaxScaler** para escalar a variável **Consumption** e as variáveis de entrada. A remoção de tendências e sazonalidades foi aplicada para que o modelo focasse nos padrões de longo prazo, como ciclos econômicos, e pudesse capturar mais facilmente as variações dos dados. Além disso, o uso de diferentes tamanhos de janela temporal (**Time Step**) foi explorado para entender como a quantidade de dados históricos afeta a capacidade do modelo de fazer previsões precisas.

Os resultados obtidos indicam que a escolha do tamanho da janela temporal teve um impacto notável no desempenho do modelo. Com o **Time Step** de 4, o modelo obteve um MSE de 0.1351 e um R^2 de -1.727, o que, apesar de um MSE razoavelmente baixo, reflete uma dificuldade significativa em explicar os dados, como mostrado pelo R^2 negativo. Em comparação, ao usar janelas maiores, como o **Time Step** de 30, o modelo apresentou um MSE de 0.0704 (o melhor valor entre os 16 modelos testados) e um R^2 de -0.4349. Embora o MSE tenha sido significativamente melhor, o R^2 ainda permaneceu negativo, indicando que, apesar de uma menor média de erro, o modelo continuou a não capturar adequadamente a variação dos dados. A normalização teve um impacto variável, sendo mais eficaz em algumas configurações. Por exemplo, no **Time Step** de 4 com normalização, o MSE foi 0.1351, comparado a 0.1624 sem normalização, mas o R^2 permaneceu negativo em ambos os casos. Já no **Time Step** de 30, a normalização piorou o MSE de 0.0704 (sem normalização) para 0.2037, e novamente não houve uma melhoria significativa no R^2 , que continuou muito abaixo do esperado.

Portanto, embora o melhor modelo tenha sido o **Time Step** de 30 com normalização, com MSE de 0.0704 e R^2 de -0.4349, o desempenho geral foi insatisfatório. O R^2 negativo em todos os cenários sugere que o modelo LSTM não conseguiu aprender padrões significativos dos dados. A falta de explicabilidade por parte do modelo pode ser resultado de uma quantidade insuficiente de dados, da arquitetura da rede não sendo otimizada para esse tipo de série temporal ou da dificuldade do modelo em capturar variações de longo prazo em dados com forte variabilidade de curto prazo.

Apêndice

Link do repositório: <https://github.com/TFrankeM/series-temporais>

Figure 1: Comparação com Baselines

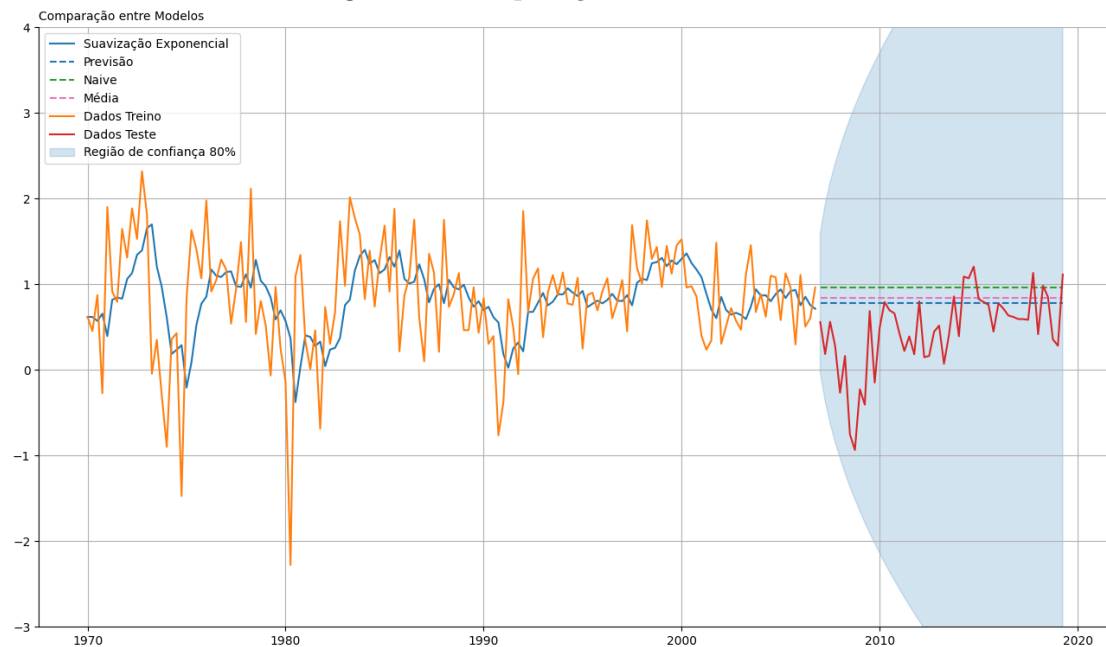


Figure 2: Baselines Errors

	RMSE	MAE	No. Observations
Naive	0.670611	0.531336	50
Média	0.582600	0.436599	50
Suav Exp	0.550465	0.403867	50

Figure 3: Resíduos

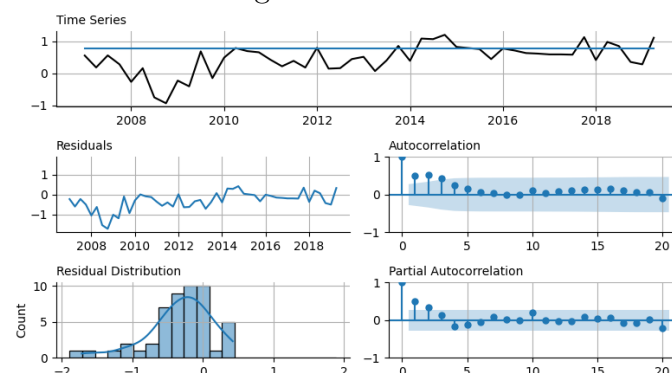


Figure 4: Comparação das Transformações

	RMSE	MAE	No. Observations
Sem Transf	0.605686	0.442510	198
Raiz Quadrada	0.624740	0.455852	198
Box-Cox	0.600860	0.440232	198
Yeo-Johnson	0.649927	0.483331	198

Figure 5: Resíduos com Box-Cox

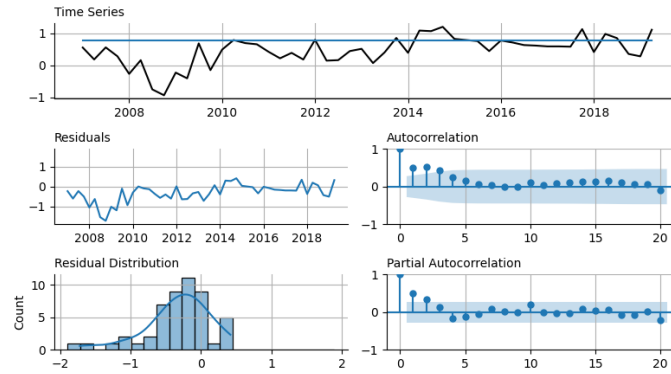
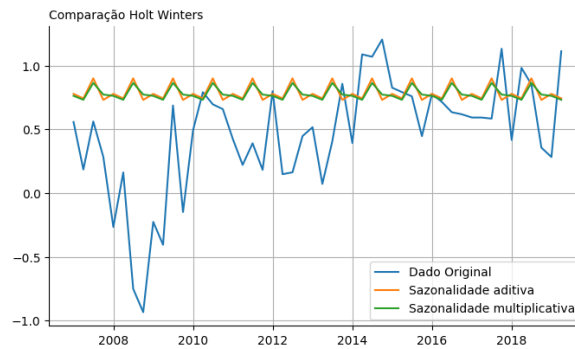


Figure 6: Comparação com Tendência

	RMSE	MAE	No. Observations
Sem Tendência	0.552619	0.405790	50
Com Tendência	0.556016	0.409538	50

Figure 7: Comparação com Sazonalidade



MSE Mean	MSE Std	Normalização	Quarter
0.059690	0.003399	Sim	Não
0.128068	0.024840	Não	Sim
0.076810	0.006900	Sim	Sim
0.083202	0.005191	Não	Não

Table 1: Resultados dos testes com normalização e quarter.

Figure 8: Comparação com Sazonalidade

OLS Regression Results						
=====						
Dep. Variable:	Consumption	R-squared:	0.768			
Model:	OLS	Adj. R-squared:	0.763			
Method:	Least Squares	F-statistic:	160.0			
Date:	Tue, 26 Nov 2024	Prob (F-statistic):	3.93e-60			
Time:	19:59:47	Log-Likelihood:	-46.660			
No. Observations:	198	AIC:	103.3			
Df Residuals:	193	BIC:	119.8			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975

const	0.7425	0.022	33.679	0.000	0.699	0.786
Income	0.6668	0.036	18.461	0.000	0.596	0.738
Production	0.0718	0.035	2.038	0.043	0.002	0.141
Savings	-0.6329	0.035	-18.088	0.000	-0.702	-0.564
Unemployment	-0.0642	0.035	-1.829	0.069	-0.133	0.005
=====						
Omnibus:	19.961	Durbin-Watson:	2.218			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	35.767			
Skew:	0.525	Prob(JB):	1.71e-08			
Kurtosis:	4.798	Cond. No.	3.05			

Figure 9: Resíduo SARIMA

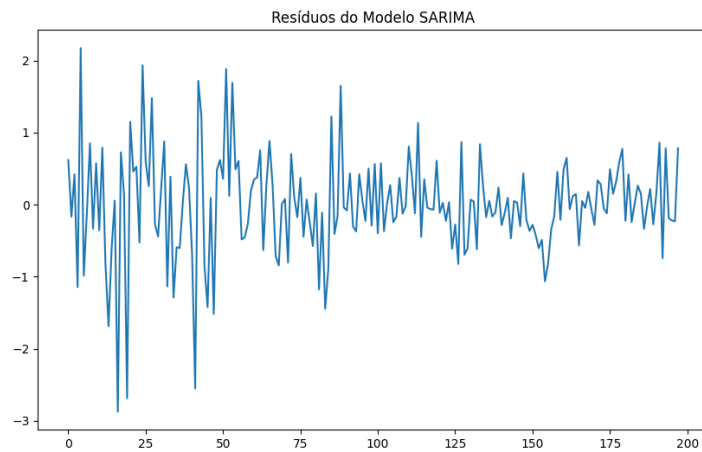


Figure 10: Resíduo SARIMA

ADF Statistic: -4.610123144902199

p-value: 0.00012373240927130775

SARIMAX Results

Dep. Variable:	Consumption	No. Observations:	198
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 12)	Log Likelihood:	-181.241
Date:	Tue, 26 Nov 2024	AIC	372.482
Time:	22:42:45	BIC	388.584
Sample:	0	HQIC	379.008
	- 198		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.2923	0.089	3.297	0.001	0.119	0.466
ma.L1	-0.8912	0.054	-16.640	0.000	-0.996	-0.786
ar.S.L12	-0.1926	0.071	-2.709	0.007	-0.332	-0.053
ma.S.L12	-0.3500	0.264	-3.601	0.000	-1.467	-0.433
sigma2	0.3503	0.083	4.207	0.000	0.187	0.513

Ljung-Box (L1) (Q):	0.43	Jarque-Bera (JB):	18.01
Prob(Q):	0.51	Prob(JB):	0.00
Heteroskedasticity (H):	0.25	Skew:	-0.19
Prob(H) (two-sided):	0.00	Kurtosis:	4.48

...

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

MAE: 0.3826761658257369

RMSE: 0.4359844084194706

Figure 11: Previsão da CNN

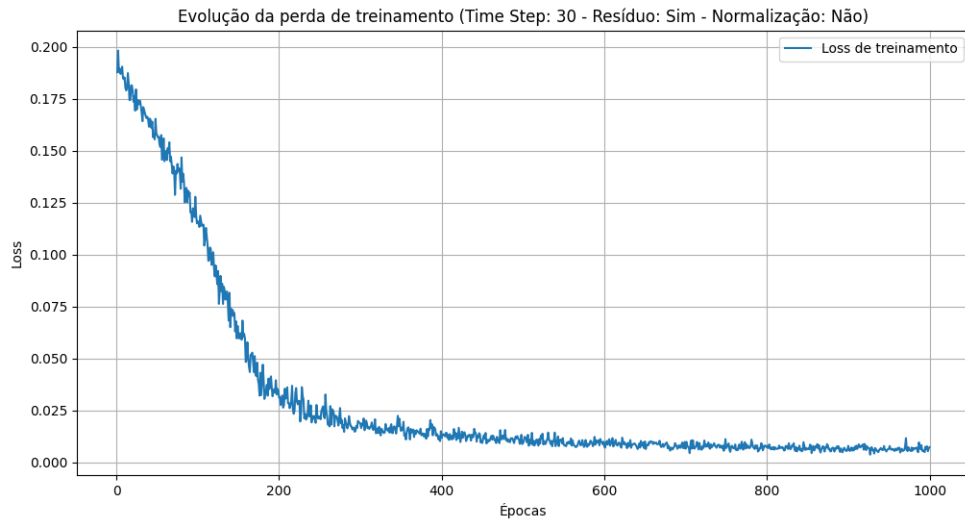
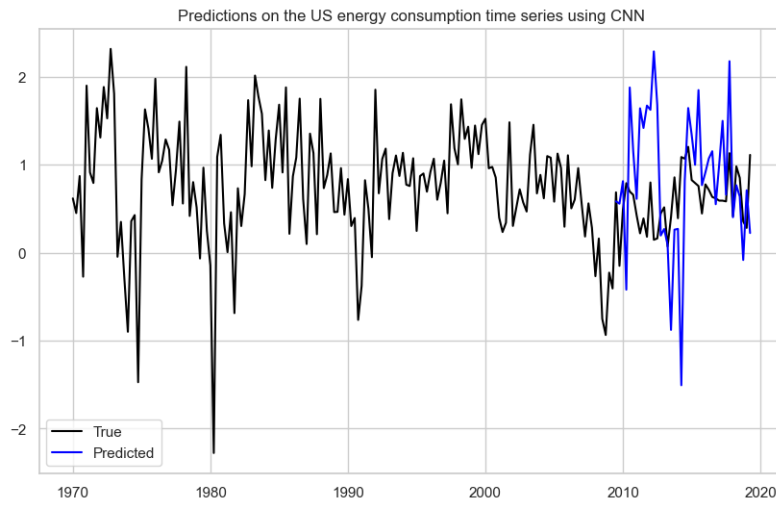


Figure 12: Resultado do treinamento com a melhor configuração de LSTM.

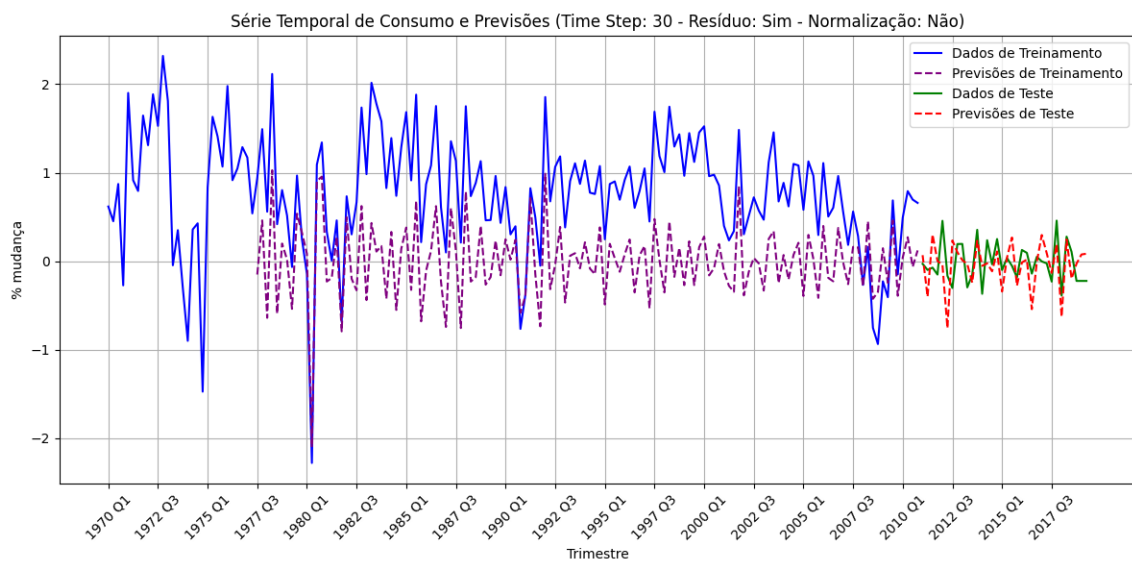


Figure 13: Resultado de previsão com a melhor configuração de LSTM.