

Análise de Séries Temporais para Previsão do Valor Máximo das Ações ELET3

Pedro Henrique Rodrigues da Silva

December 14, 2024

Contents

1	Resumo	2
2	Introdução	2
2.1	Objetivos	2
3	Referencial Teórico	2
3.1	Séries Temporais e Previsão	2
3.2	Modelo ARIMAX	2
3.3	Variáveis Exógenas	3
3.4	Transformada Rápida de Fourier (FFT)	3
4	Metodologia	3
4.1	Coleta de Dados	3
4.2	Pré-Processamento	3
4.3	Remoção da Tendência	4
4.4	Teste de Estacionaridade	4
4.5	Primeira Diferença	4
4.6	Análise de Frequência	4
4.7	Análise de Autocorrelação	6
4.8	Modelo ARIMAX	7
5	Resultados	7
6	Discussão	8
7	Conclusão	8
8	Referências	8

1 Resumo

Neste trabalho, investigamos a aplicação de técnicas de séries temporais para prever o valor máximo das ações da ELET3. Foi realizada uma análise exploratória dos dados, incluindo a remoção da tendência, testes de estacionaridade, aplicação da Transformada de Fourier (FFT) e análise de autocorrelação. Em seguida, implementamos um modelo ARIMAX com variáveis exógenas para previsão do valor máximo. Os resultados indicam que o modelo ajusta-se aos dados, embora com certas limitações. Concluimos com sugestões para pesquisas futuras na área.

2 Introdução

A análise de séries temporais desempenha um papel crucial no mercado financeiro, permitindo identificar padrões, tendências e realizar previsões sobre o comportamento de ativos. Neste contexto, este trabalho visa aplicar técnicas de modelagem de séries temporais para prever o valor máximo das ações da ELET3, uma empresa de grande relevância no setor energético brasileiro. O objetivo principal é desenvolver um modelo de previsão capaz de capturar a dinâmica do valor máximo das ações e fornecer informações úteis para tomadas de decisão.

2.1 Objetivos

Os objetivos deste trabalho incluem:

- Realizar uma análise exploratória dos dados, incluindo a remoção da tendência e testes de estacionaridade.
- Aplicar a Transformada Rápida de Fourier (FFT) para análise de frequência.
- Modelar a série temporal do valor máximo das ações ELET3 utilizando um modelo ARIMAX com variáveis exógenas.
- Avaliar o desempenho do modelo implementado através do RMSE.

3 Referencial Teórico

3.1 Séries Temporais e Previsão

Uma série temporal é uma sequência de dados coletados em intervalos de tempo regulares. A análise de séries temporais envolve o uso de técnicas estatísticas para entender os padrões e as dependências nos dados e, assim, realizar previsões sobre o comportamento futuro da série. Modelos como ARIMA e suas variações, incluindo ARIMAX, são frequentemente empregados para esse fim.

3.2 Modelo ARIMAX

O modelo ARIMAX (AutoRegressive Integrated Moving Average with Exogenous Variables) é uma extensão do modelo ARIMA que incorpora variáveis exógenas para melhorar a capacidade de previsão. O modelo ARIMAX é útil em situações em que a série temporal de interesse é afetada por outras variáveis que não são parte da série.

3.3 Variáveis Exógenas

Variáveis exógenas são aquelas que influenciam a série temporal de interesse, mas que não são influenciadas por ela. No contexto deste trabalho, as variáveis exógenas incluem abertura, variação, mínimo, máximo e volume das ações.

3.4 Transformada Rápida de Fourier (FFT)

A Transformada Rápida de Fourier (FFT) é uma técnica que decompõe um sinal em suas componentes de frequência. Esta técnica é útil para identificar padrões periódicos em séries temporais, e pode auxiliar na criação de modelos de previsão mais precisos.

4 Metodologia

4.1 Coleta de Dados

Os dados utilizados neste trabalho foram obtidos da plataforma InfoMoney e correspondem a uma série histórica das ações da ELET3 no período entre 2023 e 2024. As variáveis coletadas incluem:

- Abertura
- Fechamento
- Variação
- Mínimo
- Máximo
- Volume

4.2 Pré-Processamento

Os dados coletados passaram por um processo de pré-processamento que incluiu:

- Conversão da coluna de "DATA" para o formato de data.
- Remoção das letras "M" e "B" da coluna de volume e conversão para valores numéricos.
- Substituição de valores "n/d" por NaN.
- Conversão de vírgulas para pontos nas colunas numéricas.
- Definição da coluna de data como índice do DataFrame.

4.3 Remoção da Tendência

A tendência da série temporal foi removida através de uma regressão linear ajustada sobre os dados, utilizando a seguinte função:

```
def regressao_linear(x, y):
    x_sum = sum(x)
    y_sum = sum(y)
    x_sum2 = x_sum ** 2
    x2 = [xi ** 2 for xi in x]
    x2_sum = sum(x2)
    xy = [xi * yi for xi, yi in zip(x, y)]
    xy_sum = sum(xy)
    n = len(x)
    b1 = (x_sum * y_sum - n * xy_sum) / (x_sum2 - n * x2_sum)
    b0 = (y_sum - b1 * x_sum) / n
    ym = [b0 + b1 * xi for xi in x]
    e2 = [(yi - ymi) ** 2 for yi, ymi in zip(y, ym)]
    e2_sum = sum(e2)
    y2 = [yi ** 2 for yi in y]
    y2_sum = sum(y2)
    y_sum2 = y_sum ** 2
    r2 = 1 - (e2_sum)/(y2_sum - y_sum2/n)
    return b0, b1, r2
```

Onde ‘x’ representa a sequência de tempo e ‘y’ representa a série original. A tendência é dada pela expressão $ym = b0 + b1 * xi$. Os resíduos, ou seja, a série sem tendência, são obtidos subtraindo os valores ajustados da série original.

4.4 Teste de Estacionaridade

O teste de Dickey-Fuller Aumentado (ADF) foi aplicado para verificar a estacionaridade da série sem tendência. O resultado do teste indicou que a série não é estacionária, necessitando diferenciação para se tornar adequada para modelagem.

4.5 Primeira Diferença

A primeira diferença da série sem tendência foi calculada através do método ‘.diff()’, que calcula a diferença entre cada valor e seu valor anterior.

4.6 Análise de Frequência

A Transformada Rápida de Fourier (FFT) foi utilizada para decompor a série temporal diferenciada em suas componentes de frequência. O gráfico a seguir ilustra a amplitude e fase dessas componentes.

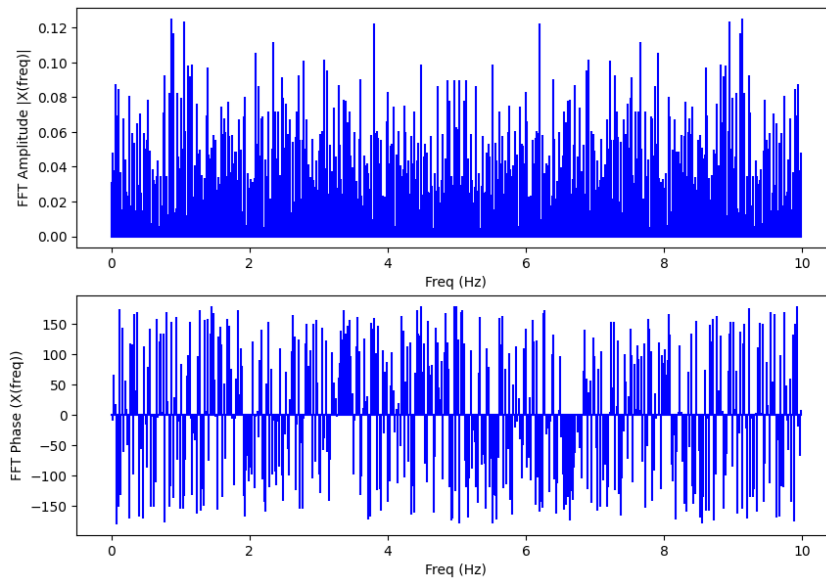


Figure 1: Gráfico da amplitude e fase das componentes de frequência da série temporal diferenciada.

As frequências com amplitudes superiores a 0.124 foram selecionadas, e seus valores, fases e amplitudes estão na tabela abaixo:

Frequências com amplitudes maiores que 0.124:

	Frequência (Hz)	Fase (radianos)	Amplitude
0	0.876494	154.43981	0.125301
1	9.123506	-154.43981	0.125301

Com base nessas frequências, amplitudes e fases, a série temporal foi reconstruída através da seguinte expressão:

$X_{fft} =$

$0.125301 * \cos(2 * \pi * 0.876494 * t - (2.695483)) + 0.125301 * \cos(2 * \pi * 9.123506 * t - (-2.695483))$

Onde 't' representa o vetor de tempo. O gráfico abaixo mostra a comparação entre a série real e a série reconstruída.

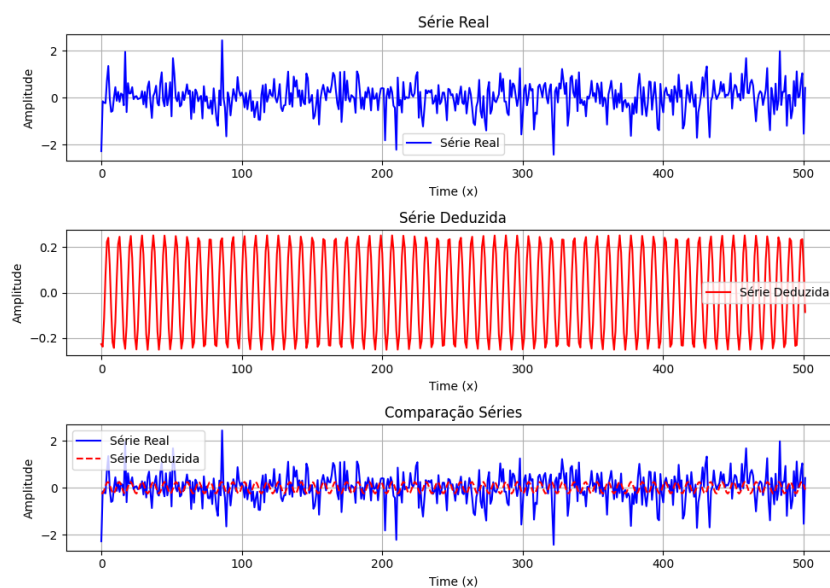


Figure 2: Gráfico comparativo entre a série real e reconstruída pela FFT.

4.7 Análise de Autocorrelação

As funções de autocorrelação (ACF) e autocorrelação parcial (PACF) foram calculadas para auxiliar na identificação da ordem do modelo ARIMAX.

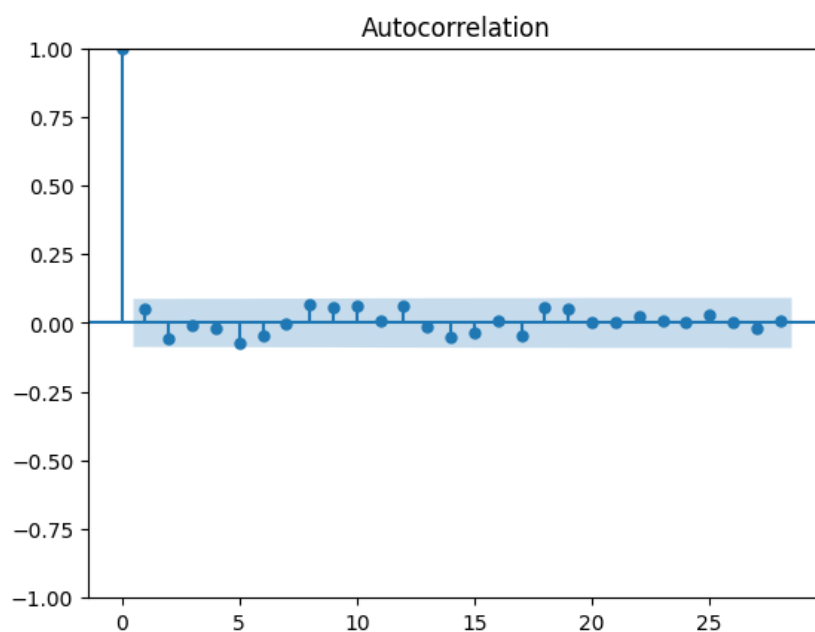


Figure 3: Gráfico da função de autocorrelação (ACF) da série temporal diferenciada.

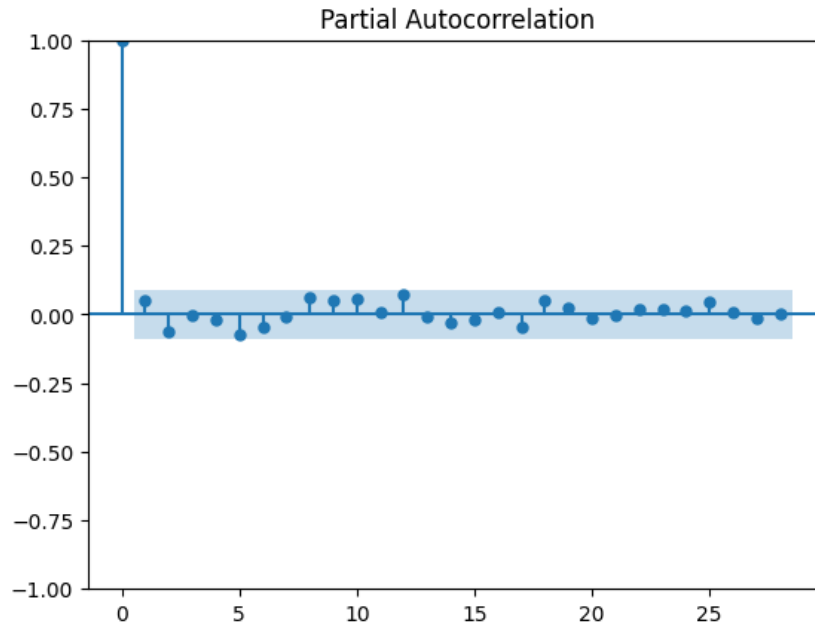


Figure 4: Gráfico da função de autocorrelação parcial (PACF) da série temporal diferenciada.

4.8 Modelo ARIMAX

O modelo ARIMAX foi implementado utilizando a biblioteca ‘statsmodels’ do Python. Para treinar o modelo foi definido a variável alvo como a coluna ‘MÁXIMO’ e as variáveis exógenas como ‘ABERTURA’, ‘VARIAÇÃO’, ‘MÍNIMO’, e ‘VOLUME’. Os dados foram divididos em conjuntos de treinamento (80

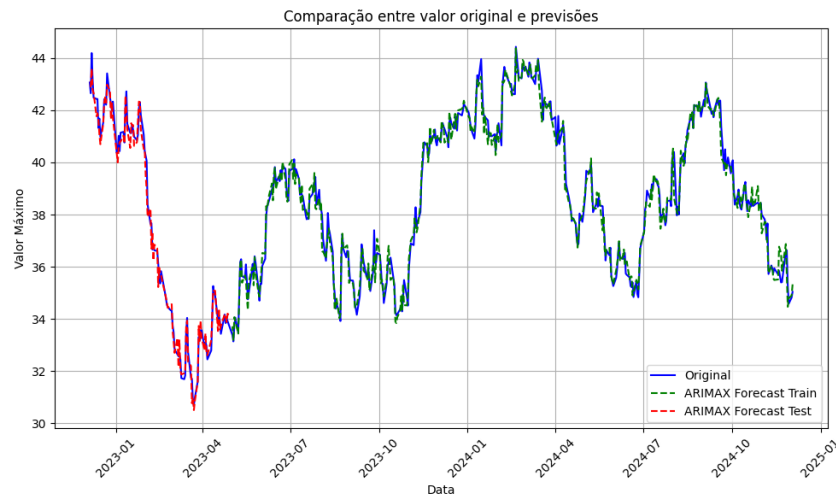


Figure 5: Gráfico comparando o valor original e as previsões do modelo ARIMAX.

5 Resultados

Os resultados obtidos através da metodologia descrita acima, indicam que o modelo ARIMAX com variáveis exógenas tem potencial para modelar e prever o valor máximo

das ações da ELET3, embora apresente algumas limitações. A remoção da tendência através da regressão linear e a diferenciação da série temporal foram passos essenciais para tornar os dados adequados para a modelagem. A FFT ajudou na identificação de frequências dominantes, mas não foi utilizada como componente de entrada no modelo ARIMAX.

6 Discussão

Apesar de o modelo ARIMAX ter apresentado um bom ajuste aos dados, com um RMSE de 0.1469, é importante considerar que a precisão da previsão pode variar em diferentes períodos de tempo. A escolha da ordem do modelo ARIMAX e das variáveis exógenas, podem ser refinadas para melhorar ainda mais o desempenho. Além disso, a análise de frequência com a FFT pode ser explorada para a extração de características que podem ser utilizadas na modelagem. É importante considerar as limitações do modelo, como a não linearidade nas relações entre as variáveis, que não foram exploradas nesse trabalho.

7 Conclusão

Neste estudo, apresentamos uma metodologia para análise e previsão do valor máximo das ações da ELET3 utilizando técnicas de séries temporais e o modelo ARIMAX. Os resultados obtidos demonstraram que o modelo possui capacidade de prever os preços de forma satisfatória, embora possam ser refinadas em trabalhos futuros. Como sugestões para pesquisas futuras, recomendamos explorar modelos mais complexos, como redes neurais ou outros modelos de machine learning, e analisar a incorporação de mais informações como variáveis exógenas. Além disso, a análise da FFT pode ser explorada em maior profundidade.

8 Referências

- Pandas Documentation: <https://pandas.pydata.org/docs/>
- Matplotlib Documentation: <https://matplotlib.org/stable/contents.html>
- Statsmodels Documentation: <https://www.statsmodels.org/stable/index.html>
- Scikit-learn Documentation: <https://scikit-learn.org/stable/index.html>