



FIAP + alura

Turma - 2TDAT

**FASE 2 - MACHINE LEARNING AND TIME
SERIES**
Grupo 82

Jonathan Abner Jerônimo de Freitas

jonathanabner2015@gmail.com

Osvaldo Caio Oliveira dos Santos

osvaldocaio@hotmail.com

Ranielli Santos

raniellidos@hotmail.com

Thalita Mendes Maina Begliomini

thalitamaina@gmail.com

_ROTEIRO

01 - Definições

02 - Análise Exploratória dos Dados

03 - Modelagem

04 - Prophet

05 - Forecasting

06 - Considerações Finais

07 - Refências





DEFINIÇÕES

PROBLEMA

B3

IBOVESPA

ECONOMIA BRASILEIRA

01



_ DEFINIÇÃO DO PROBLEMA

Imagine que você foi escalado para um time de investimentos e precisará realizar um modelo preditivo com dados da IBOVESPA (Bolsa de valores) para criar uma série temporal e prever diariamente o fechamento da base. Para isso utilize a base de dados contida no site da investing e selecione o período “diário”, com o intervalo de tempo que achar adequado. Você precisará demonstrar para o time de investimentos:

- ✓ O modelo com o storytelling, desde a captura do dado até a entrega do modelo;
- ✓ Justificar a técnica utilizada;
- ✓ Atingir uma acuracidade adequada (acima de 70%).

Fonte de dados: <https://br.investing.com/indices/bovespa-historical-data>

_ DEFINIÇÃO DE B3

B3 sigla para Brasil, Bolsa e Balcão é o resultado da fusão que ocorreu em março de 2017 entre a BM&FBOVESPA e a Cetip. B3 é a bolsa de valores do mercado de capitais brasileiro, a maior da América Latina. Responsável por intermediar negociações de ações, títulos públicos, Fundos Imobiliários, derivativos, entre outros, além de oferecer serviços de custódia, registro, compensação e liquidação de operações financeiras. Desempenha um papel de suma importância no mercado financeiro, fornecendo transparência, segurança e liquidez para investidores e meios para as empresas se capitalizarem.





— INVESTIDORES

A bolsa de valores brasileira atingiu uma marca recorde de mais de 18 milhões de investidores em 2023, um número significativo quando comparado ao ano de 2021 na qual possuía um pouco mais de 13 milhões de investidores, isso representa um aumento de cerca de 40% em apenas 2 anos. Este é um resultado extremamente favorável que foi impulsionado pela popularização dos assuntos relacionados as finanças por meio de canais do Youtube e cursos, o que gerou uma necessidade de simplificação do mercado, que foi aplicada nas corretoras e bancos (principalmente digitais) que facilitam a entrada de novos clientes e possuem investimentos com valores extremamente acessíveis. A preocupação com as finanças vem se tornando discussão também na educação do Brasil, tendo em vista que o assunto vem ganhando grandes forças para ser incluído na grade de estudos obrigatória da rede de ensino básico. Apesar das evoluções na quantidade de investidores, apenas cerca de 8% de toda população brasileira efetua investimentos na bolsa de valores.

Em Dezembro de 2023 o Índice Ibovespa atinge sua maior marca histórica superando 134 mil pontos!!

_ IBOVESPA

Conhecido como IBOV ou Ibovespa, o índice funciona como um termômetro que indica o desempenho do mercado acionário do Brasil. Em resumo, o índice Ibovespa é calculado através de um sistema de pontos (baseado em reais) que acompanha o desempenho médio de uma carteira teórica com as ações mais representativas e negociadas na Bolsa.



Quando se fala que o “Ibovespa subiu” significa que as empresas do índice em média valorizaram, com isso os investidores em geral possuem uma visão mais otimista do mercado.

Quando se fala que o “Ibovespa caiu” significa que em média houve uma desvalorização das principais ações, gerando um cenário de cautela entre os investidores.



PRINCIPAIS ACONTECIMENTOS ECONOMIA BRASILEIRA 2000 - 2023

O Brasil enfrentou uma **crise cambial devido à incerteza eleitoral e à desconfiança dos investidores sobre a capacidade do país de pagar suas dívidas**, devido as quedas de liquidez do mercado de capitais internacional. Entre os meses de janeiro e setembro de 2002, a cotação do dólar, em reais por dólar, chegou a subir aproximadamente 70% e encerrou o ano com inflação de 12,5%. Isso levou a uma desvalorização do real e a medidas emergenciais do governo para evitar um colapso financeiro.

A partir de 2014, o Brasil **enfrentou uma combinação de recessão econômica, alta inflação** e escândalos de corrupção. A queda nos preços das commodities, como o petróleo, afetou a economia brasileira. Além disso, investigações de corrupção envolvendo empresas estatais contribuíram para a instabilidade política, impactando negativamente a confiança dos investidores e a economia como um todo.

2000 2002 2006 2008 2009 2011 2014 2016 2018 2020 2023

A **crise financeira global iniciada nos EUA (crise do subprime)** teve um impacto significativo no Brasil. Embora o país estivesse em crescimento antes da crise, a queda na demanda internacional por commodities afetou a economia brasileira, levando a uma desaceleração econômica. As instabilidades do mercado fizeram com que os investidores removessem seu dinheiro da bolsa, gerando um dos piores resultados da história. O governo precisou executar diversas ações como a redução da taxa básica de juros, a fim de promover a liquidez no mercado.

A crise fiscal já existente no país foi agravada pela **pandemia de COVID-19 em 2020**. Medidas de isolamento social para conter a propagação do vírus impactaram severamente a atividade econômica. Houve aumento do desemprego, queda do PIB e um grande aumento nos gastos do governo para lidar com os efeitos da pandemia. A incerteza sobre o futuro afetou todo o mercado e investidores. Em 5 de maio de 2023, a OMS anunciou o fim da pandemia de coronavírus.



ANÁLISE EXPLORATÓRIA DOS DADOS

SÉRIE TEMPORAL

MÉDIA

DESVIO PADRÃO

AJUSTES DE DADOS

02

_ SÉRIE TEMPORAL

Período de Análise: 27-04-1993 até 28-12-2023

Tamanho da Base: 2 colunas com 7593 linhas



```
[106] df_ibov.head()
```

	ds	y
0	1993-04-27	24.500000
1	1993-04-28	24.299999
2	1993-04-29	23.700001
3	1993-04-30	24.100000
4	1993-05-03	24.100000


```
df_ibov.tail()
```

	ds	y
7588	2023-12-21	132182.0
7589	2023-12-22	132753.0
7590	2023-12-26	133533.0
7591	2023-12-27	134194.0
7592	2023-12-28	134185.0

_ SÉRIE TEMPORAL

Visualização Média Móvel para que possamos analisar com suavização as flutuações.

Média Móvel - Global (toda série de dados)



Média Móvel - Últimos 300 dias



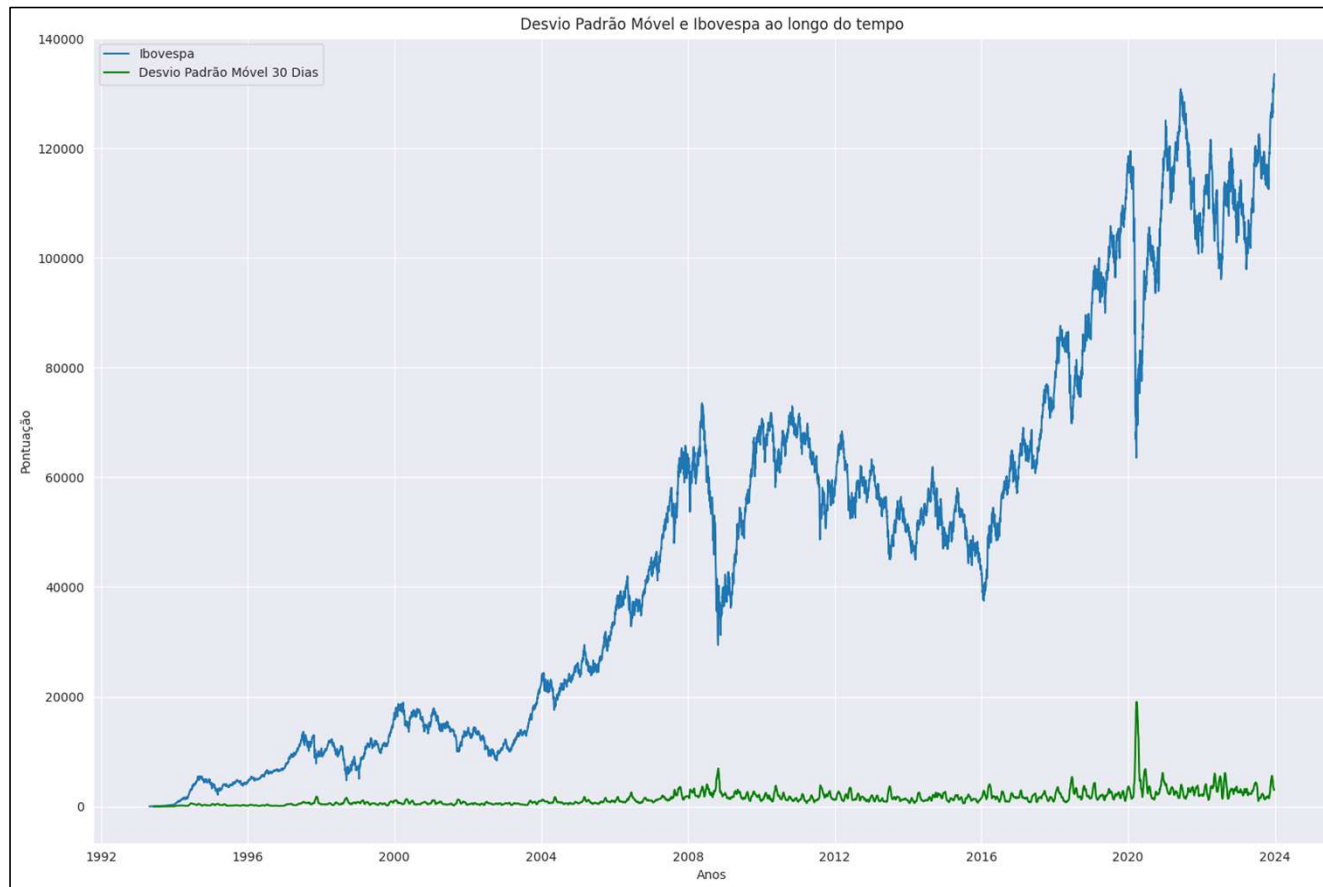
_ SÉRIE TEMPORAL

Previsão de 10 dias baseado nos 10 dias anteriores. Os resultados apresentados são descontínuos e não servem para fazer previsão futura.



_ SÉRIE TEMPORAL

Análise de Desvio Padrão Móvel dos dados Globais. Visualização da Volatilidade ao longo do tempo.

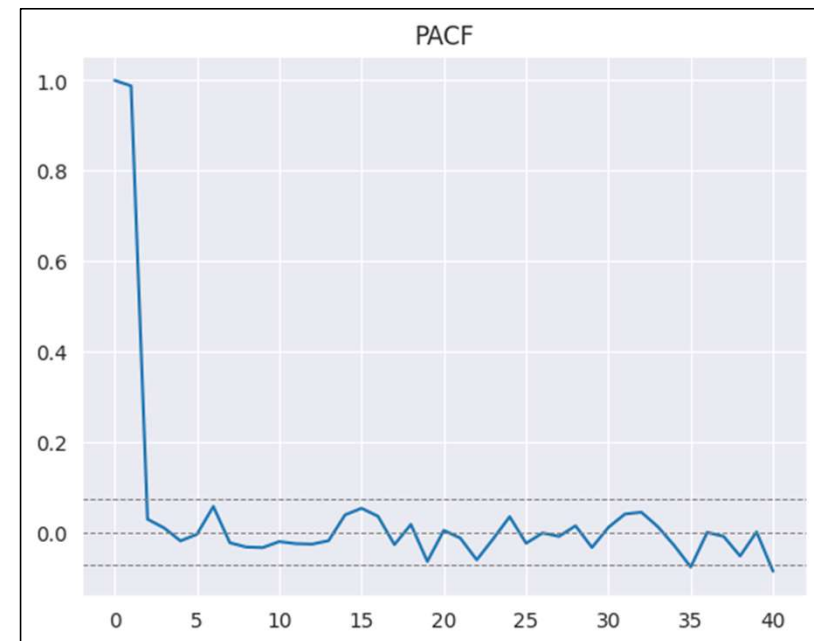
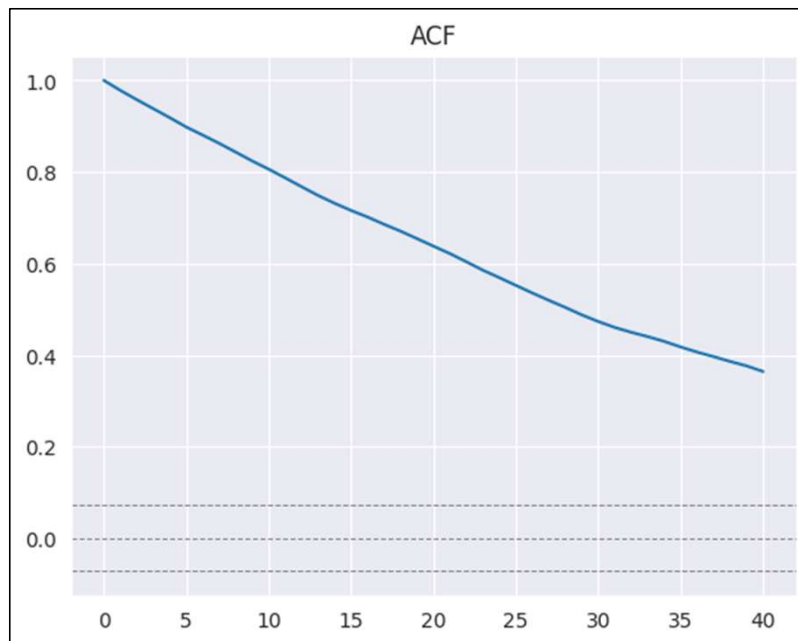


_ SÉRIE TEMPORAL

ARIMA - Aplicação da primeira redução da série temporal por meio de ACF e PACF

Premissas:

- Restrição da Série Temporal em um intervalo que vai de 04-01-2021 à 28-12-2023.
- Base mais estável e recente para análise.



```
[750] #Determinando o range para trabalhar, depois de 2020 da pandemia
      start_date_new = '2021-01-01'
      end_date_new = '2023-12-31'

      # Filtrar as datas desejadas
      df_ibov2021 = df_ibov[(df_ibov.index >= start_date_new) & (df_ibov.index <= end_date_new)]

      df_ibov2021.head()
```



MODELAGEM

BANDAS DE BOLLINGER

TRANSFORMANDO A SÉRIE EM ESTACIONARIA

ARIMA

AUTOCORRELAÇÃO

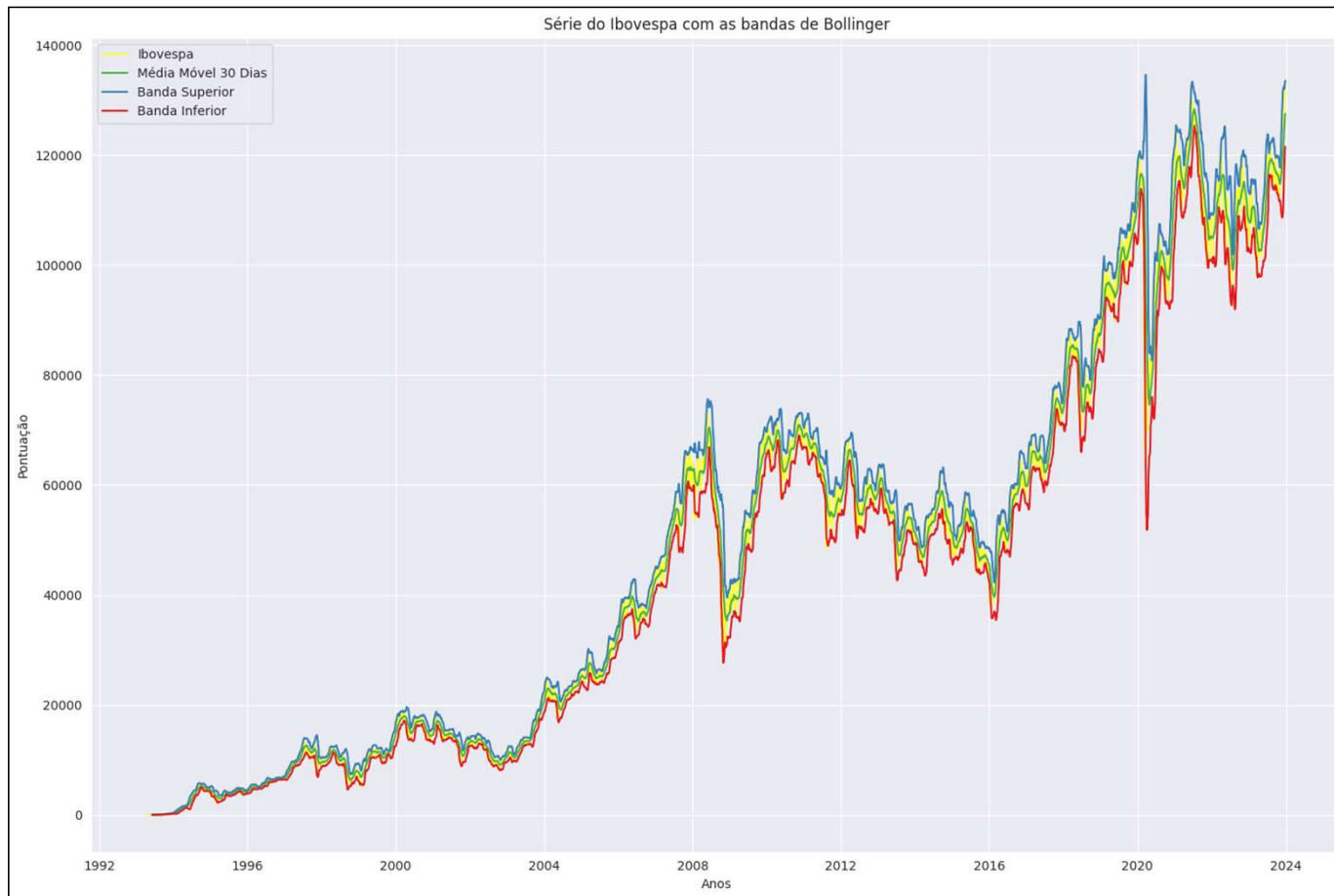
AUTOCORRELAÇÃO PARCIAL

ANALISANDO RESÍDUOS, MÉDIA MÓVEL E DESVIO PADRÃO

03

_ MODELAGEM

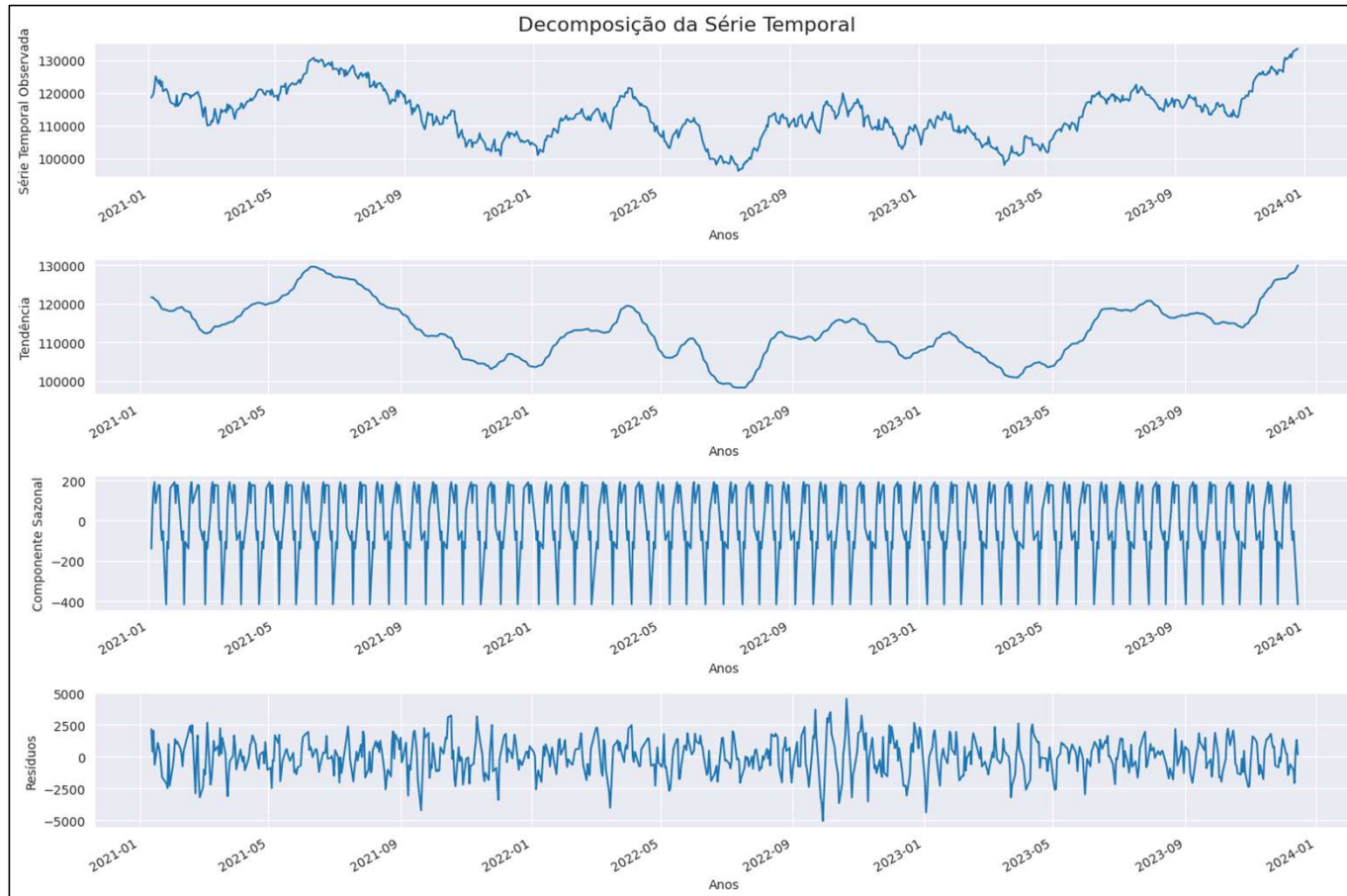
Análise por meio das Bandas de Bollinger que é uma ferramenta de análise técnica popular e útil utilizada principalmente no mercado financeiro para analisar a volatilidade e os possíveis pontos de reversão em um preço de ativo.



```
y          99770.000000
MM30       94380.766667
MM180      98511.455556
MM30_forward_30  95280.179544
DesvioPadrao  3253.615884
Banda Superior  100887.998435
Banda Inferior  87873.534898
Name: 2020-07-08 00:00:00, dtype: float64
```


MODELAGEM

Decomposição da série temporal e teste ADF



Teste ADF
Teste estatístico: -1.8177677809316526
P-Value: 0.3716768428097197
Valores críticos:
1%: -3.4391698996357687
5%: -2.8654325580580204
10%: -2.568842816582842

- Valor p alto (de acordo com a hipótese nula) (Valor p no máximo pode ser 1;
- Os valores críticos em intervalos de confiança de 1% 5% 10% devem ser o mais próximo possível das Estatísticas de Teste. Portanto, podemos concluir estatisticamente que nossa Série Temporal no momento **Não é Estacionária.**

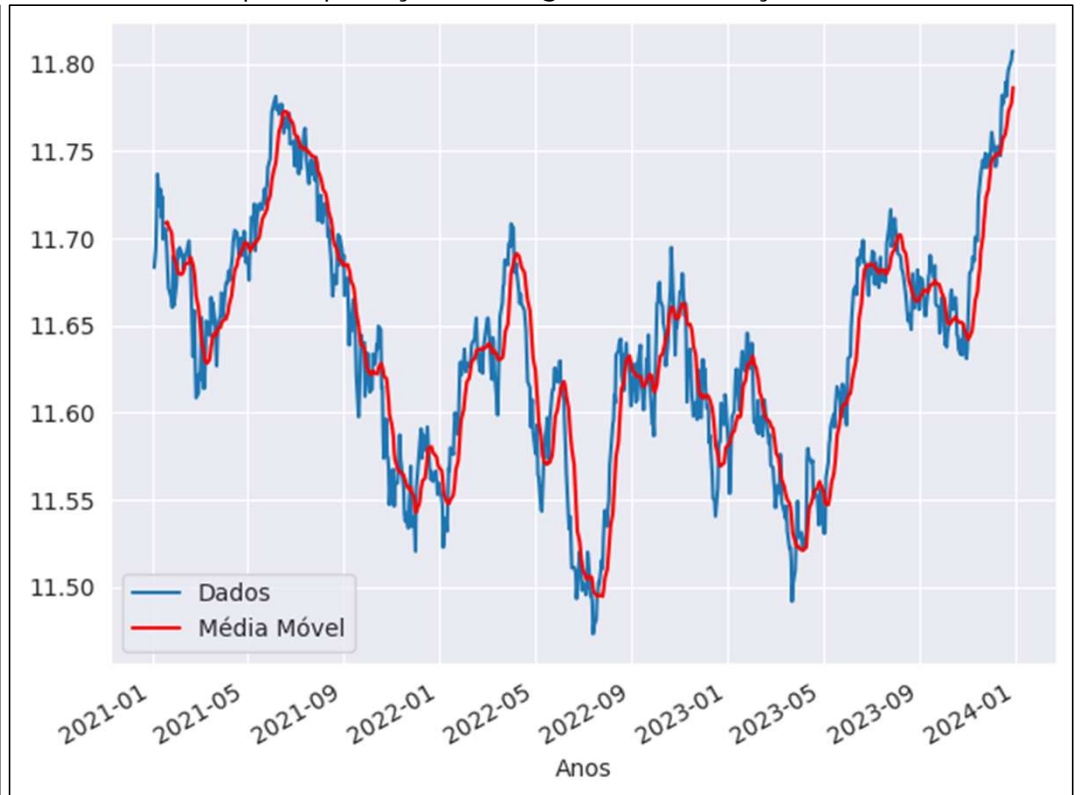
_ MODELAGEM

Análise Média Móvel

Média Móvel - Base original



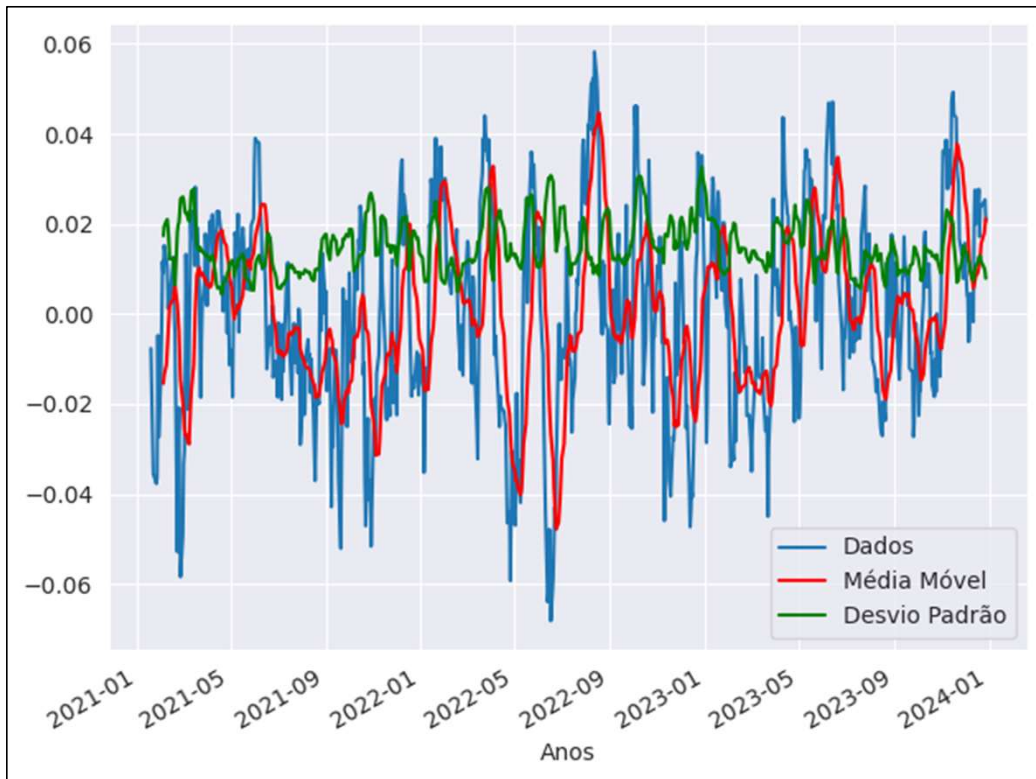
Média Móvel após aplicação de Log - Transformação de escala



_ MODELAGEM

Análise Média Móvel, Desvio Padrão e o teste ADF)

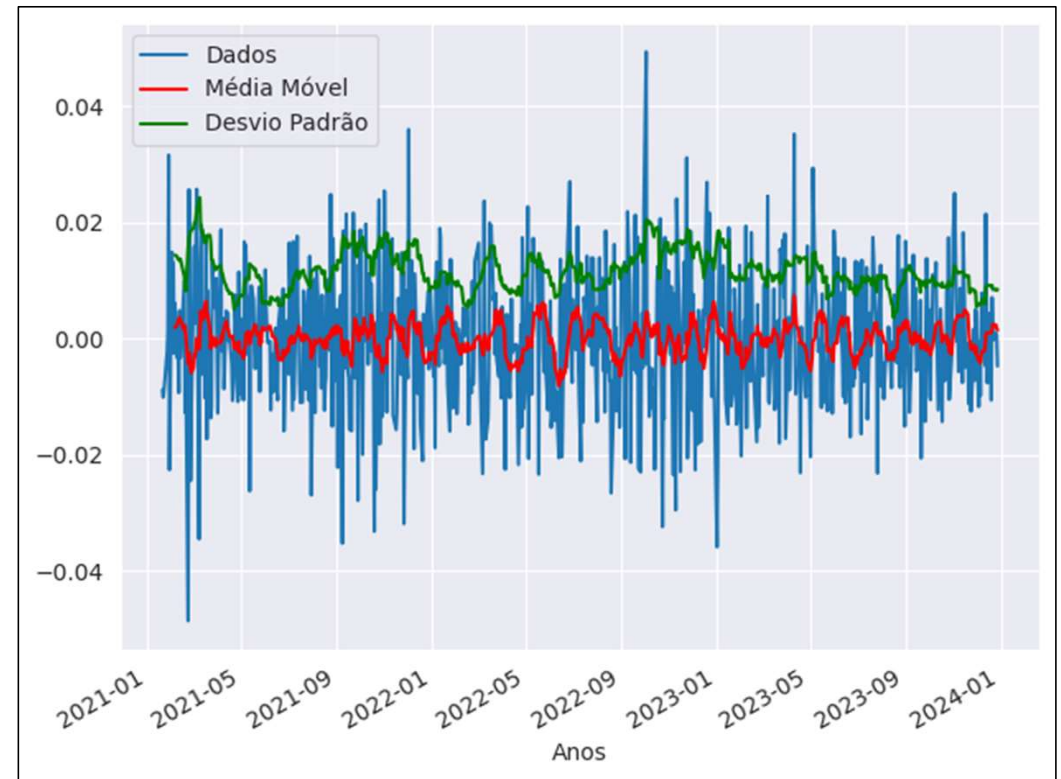
Média Móvel e Desvio Padrão



Teste ADF
Teste estatístico: -7.396533473983272
P-Value: 7.751743499450267e-11
Valores críticos:
1%: -3.439302726125018
5%: -2.8654910924750547
10%: -2.5688740002680124

(P-value quanto mais próximo de zero mais confiança possui)

Média Móvel e Desvio Padrão - Após aplicação de diferenciação



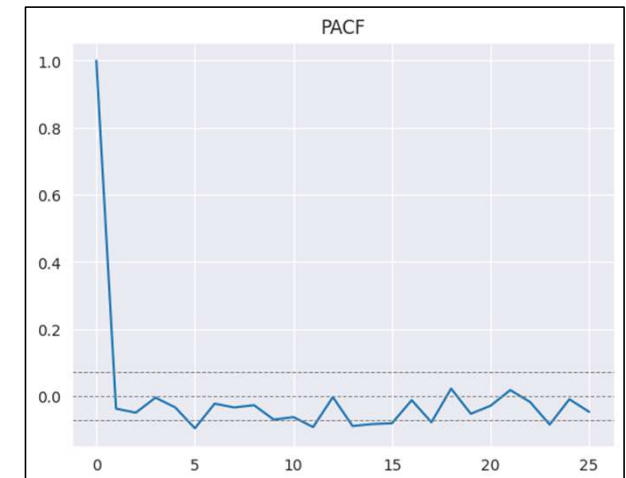
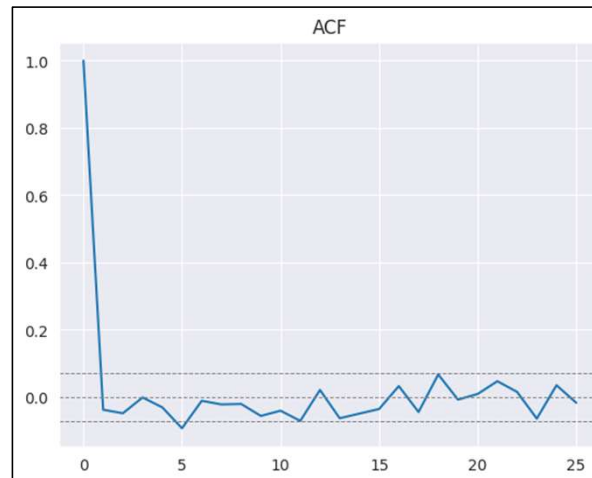
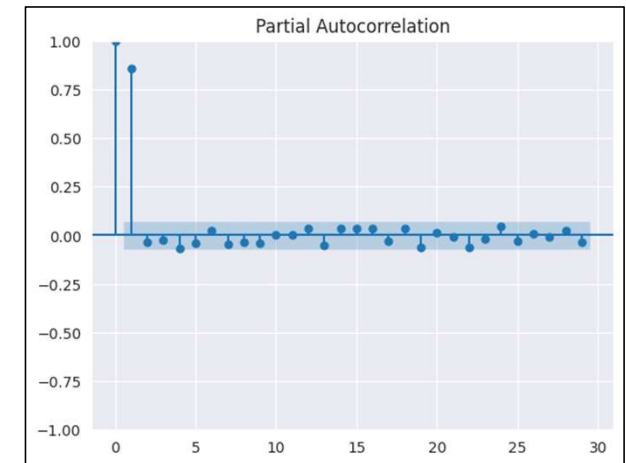
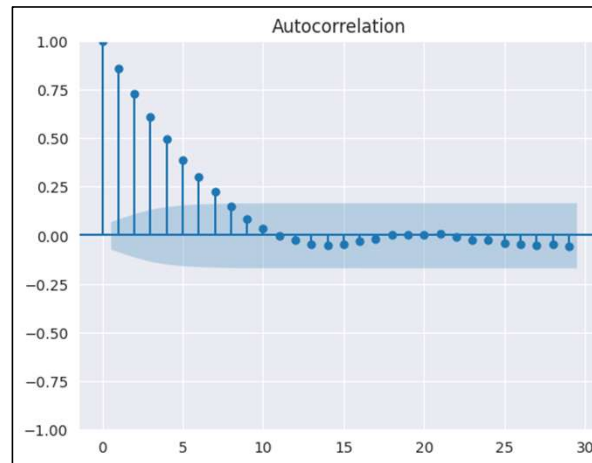
Teste ADF
Teste estatístico: -9.989749622071283
P-Value: 2.0107788496876173e-17
Valores críticos:
1%: -3.439516060164992
5%: -2.8655850998755263
10%: -2.5689240826597173

indica que podemos rejeitar a hipótese nula de não estacionariedade. Isso sugere que a sua série temporal é estacionária.

MODELAGEM

ARIMA após diferenciação - Ao observar os gráficos de ACF e PACF, notamos que ambos representam consistentemente o intervalo de confiança em todos os lags. Isso sugere que, após a diferenciação, a série temporal não apresenta autocorrelações significativas em seus lags. Esse é um bom indicativo de que a diferenciação foi eficaz em remover padrões e tendências da série, tornando-a estacionária e, assim, mais adequada para modelagem ARIMA.

- A cone azul é a largura banda de confiança
- A correlação parcial, tem a negativa



Escolhendo os melhores hiperparâmetros para ARIMA

x = é onde o ponto encosta no limite superior na primeira vez do ACF que seria o número 1.

y = é onde o ponto passa o limite superior no PACF na primeira vez - que seria 0.

$(x, y, z) \rightarrow$ ACF, PACF

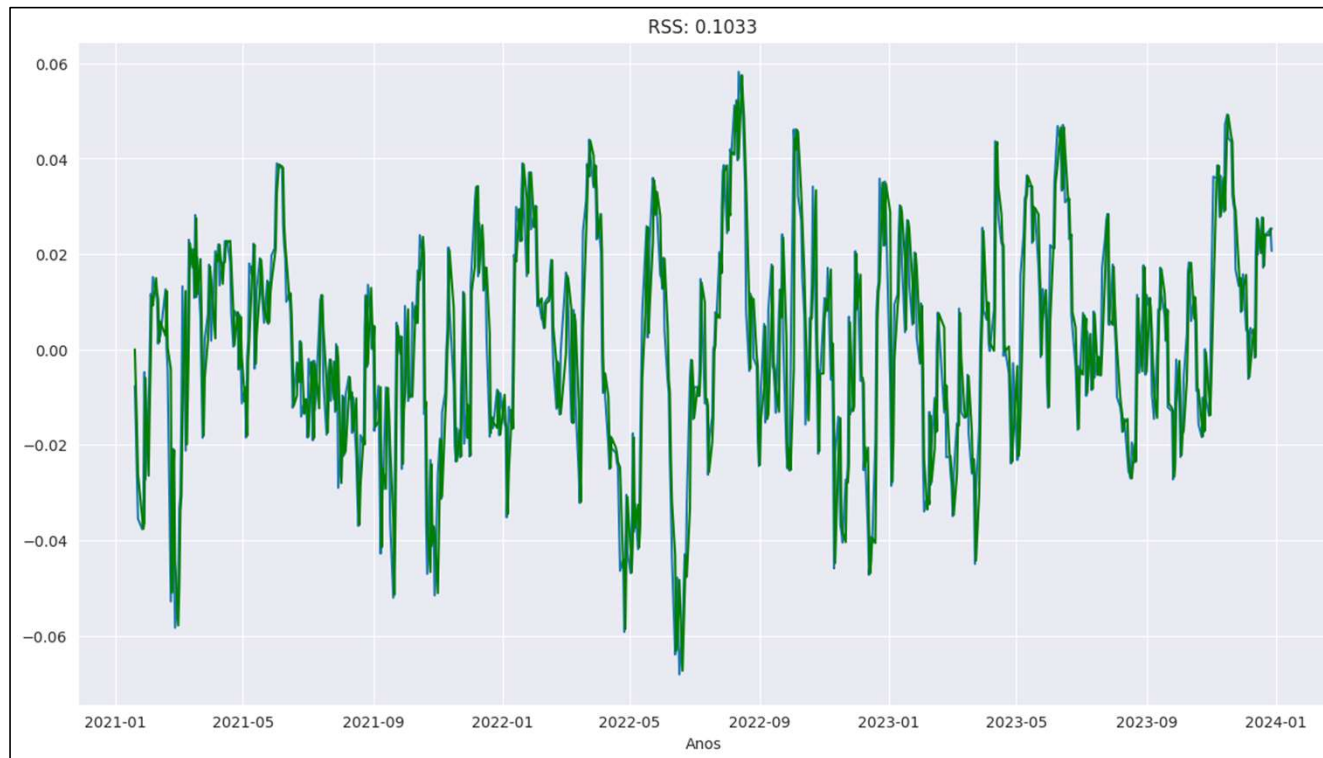
ACF = Autocorrelação - em 1 variável

PACF = Autocorrelação parcial

MODELAGEM

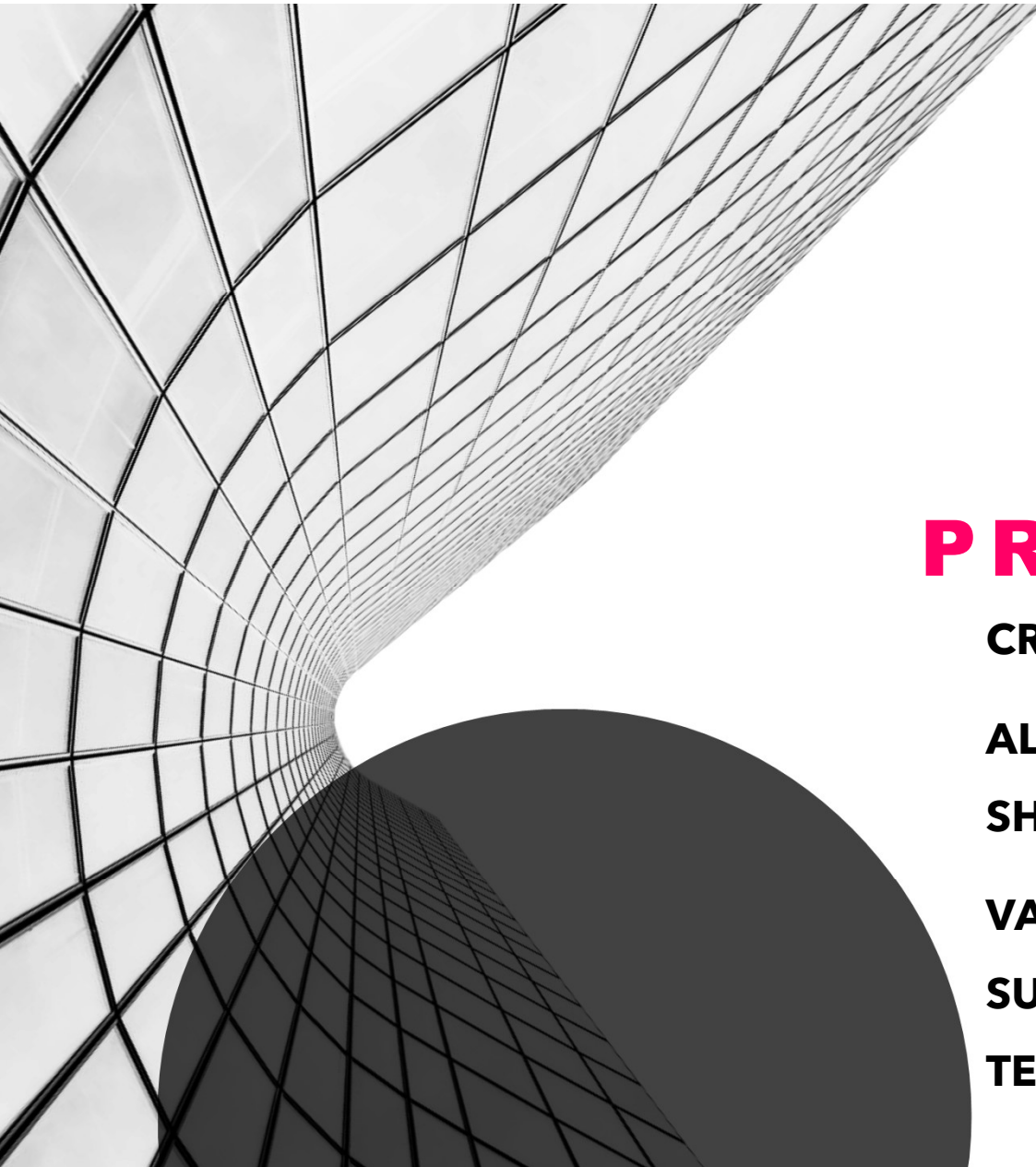
RSS é uma medida que representa a quantidade de variação não explicada pelo modelo.

Em termos simples, o RSS é calculado como a soma dos quadrados das diferenças entre os valores reais e os valores previstos pelo modelo. Quanto menor o valor do RSS, melhor o modelo está em se ajustar aos dados.



MAPE: 1126.06%

Um MAPE tão alto geralmente indica que as previsões do modelo estão muito distantes dos valores reais. RUIM



PROPHET

CROSS VALIDATION

**ALGORITMO LSTM (LONG
SHORT-TERM MEMORY)**

**VALIDANDO COM
SUAVIZAÇÃO DA SÉRIE
TEMPORAL**

04

_ PROPHET

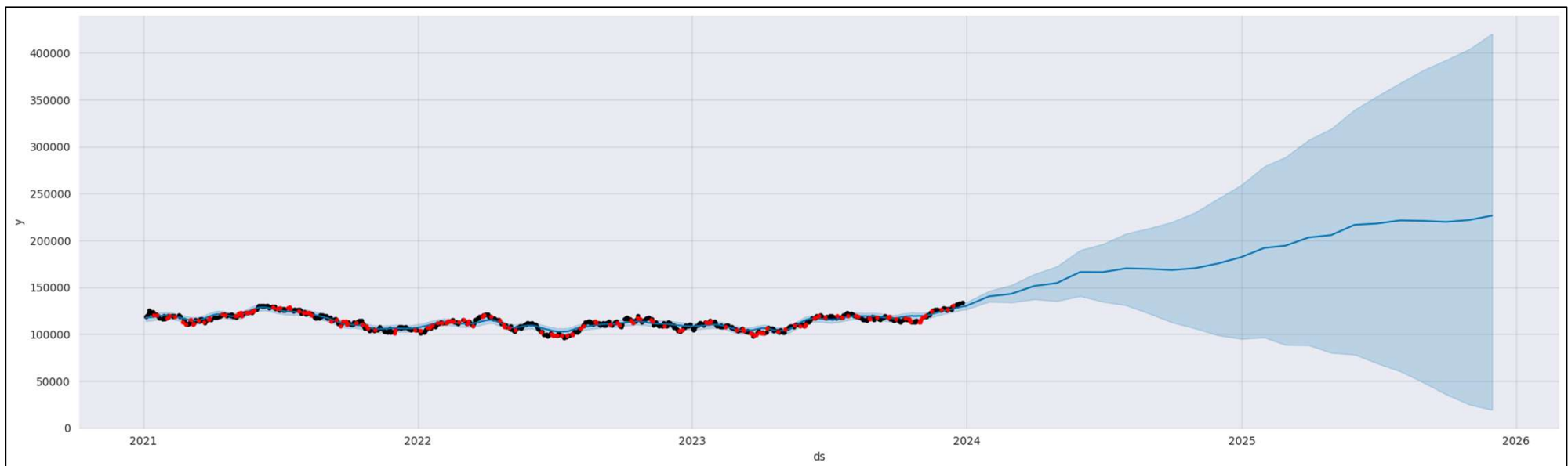
Prophet é uma biblioteca de código aberto desenvolvida pelo Facebook para previsão de séries temporais.

Divisão dos itens para treinamento:

```
training data size : (596, 4)
testing data size  : (149, 4)
```

Intervalo de datas desejado - format '%Y-%m-%d'

```
Date
2021-01-04    118558.0
2021-01-05    119223.0
2021-01-06    119851.0
2021-01-07    121956.0
2021-01-08    125077.0
...
2023-12-21    132182.0
2023-12-22    132753.0
2023-12-26    133533.0
2023-12-27    134194.0
2023-12-28    134185.0
Name: Close, Length: 745, dtype: float64
```



PROPHET

Decomposição automática em uma série temporal em seus componentes. Isso inclui a tendência geral, a sazonalidade anual.

Separação para utilizar dados à partir de 2021



MAPE: 1.97%

MAPE 1,97% = Quando calculado à partir de 2021;

Curiosidade:

MAPE 21,69% = Quando calculado com toda a base de dados desde 1993.

PROPHET

LSTM (Long Short-Term Memory) é um tipo de rede neural recorrente (RNN) especializada em lidar com sequências de dados e é conhecida por lidar de forma eficaz com problemas de dependência de longo prazo.

```
[ ] # Avaliando o modelo nos dados de teste
    mse = model.evaluate(test_generator, verbose=1)
    print("Erro Quadrático Médio", mse[0])

139/139 [=====] - 1s 4ms/step
Erro Quadrático Médio 0.0020658730063587427
```

```
139/139 [=====] - 1s 3ms/step
MAPE: 1.9674

] #O RMSE é a raiz quadrada do MSE (Mean Squared Error), que é a média dos quadrados
  rmse_value = np.sqrt(mse[0])

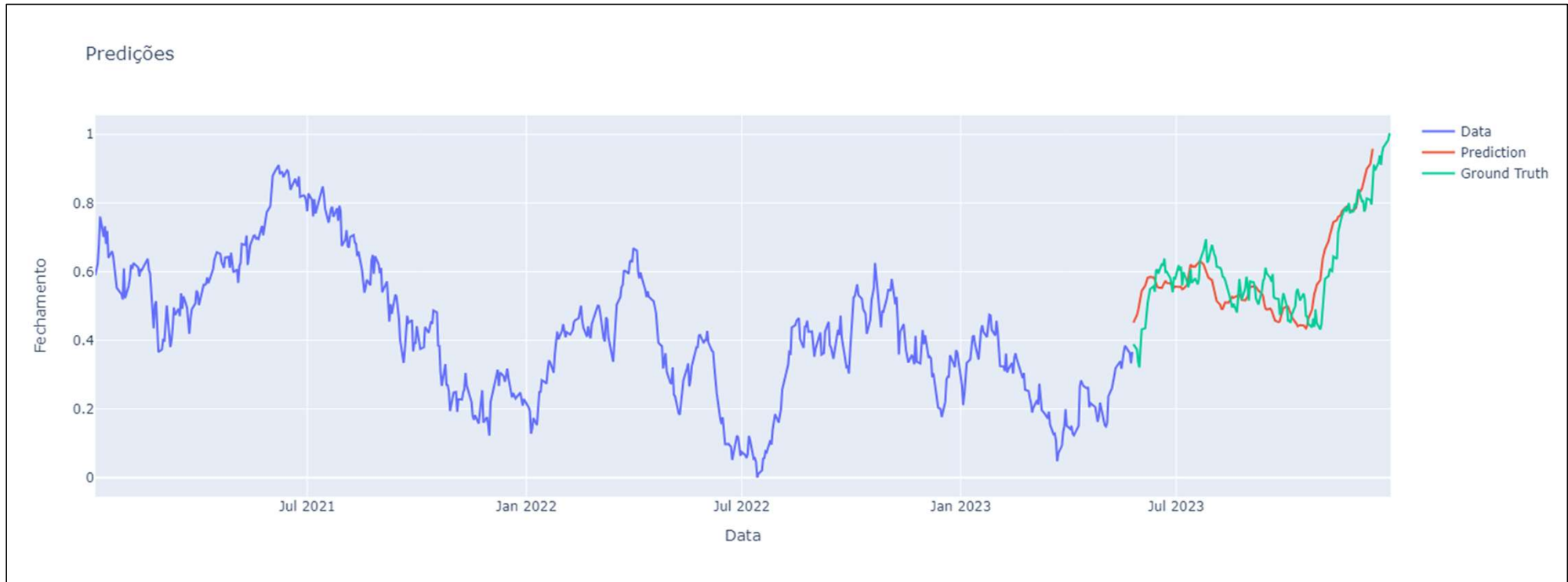
  print("RMSE:", rmse_value)

RMSE: 0.04545187571881652
```

O RMSE fornece uma métrica de erro na mesma unidade que a variável alvo (nesse caso, o preço de fechamento).

_ PROPHET

Predição dos dados

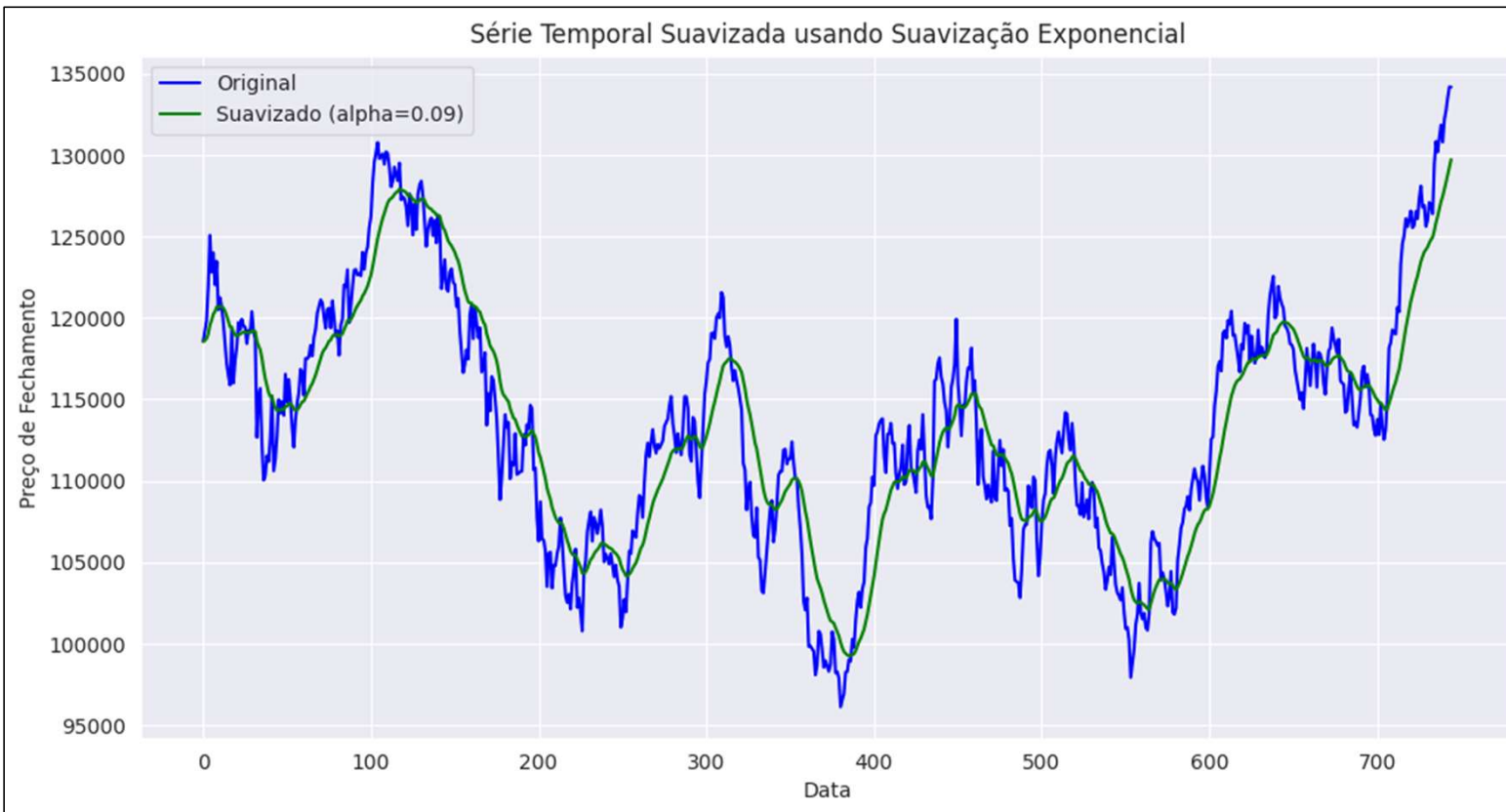


- "Ground truth" refere-se aos dados ou informações reais e conhecidos que são usados como referência para avaliar a qualidade ou a precisão de um modelo ou sistema em análises estatísticas, de aprendizado de máquina e outras áreas.

PROPHET

Suavizando a Série Temporal

Determina o quão rapidamente o impacto das observações antigas diminui à medida que você avança no tempo



```
ADF Statistic: -1.7478506904036826  
p-value: 0.40664211217870133  
Resultados do Teste de Estacionariedade:  
-----  
Teste Estatístico: -1.7478506904036826  
Valor-p: 0.40664211217870133  
Valores Críticos:  
1%: -3.439181811684251  
5%: -2.865437807603377  
10%: -2.5688456132154935
```

```
144/144 [=====] - 1s 2ms/step  
MAPE: 0.76%
```

```
# Avaliando o modelo nos dados de teste  
mse = model.evaluate(test_generator, verbose=1)  
print("Erro Quadrático Médio:", mse[0])
```

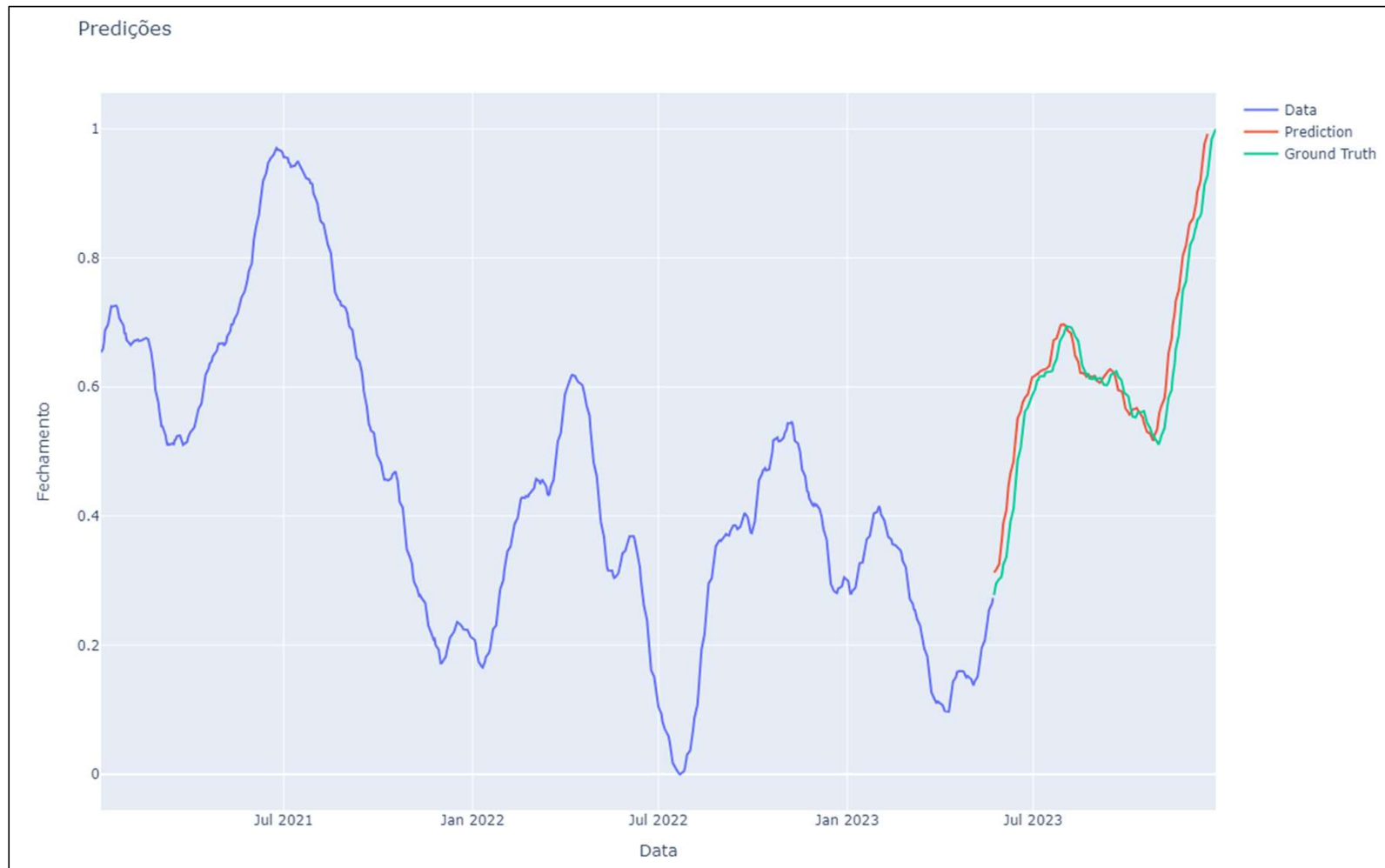
```
144/144 [=====] - 1s 3ms/step  
Erro Quadrático Médio: 0.00019008708477485925
```

```
# O RMSE é a raiz quadrada do MSE (Mean Squared Error)  
rmse_value = np.sqrt(mse[0])  
print("RMSE:", rmse_value)
```

```
RMSE: 0.013787207287005561
```

PROPHET

Predição Gerando sequências temporais para treinamento e teste em um modelo de aprendizado de máquina



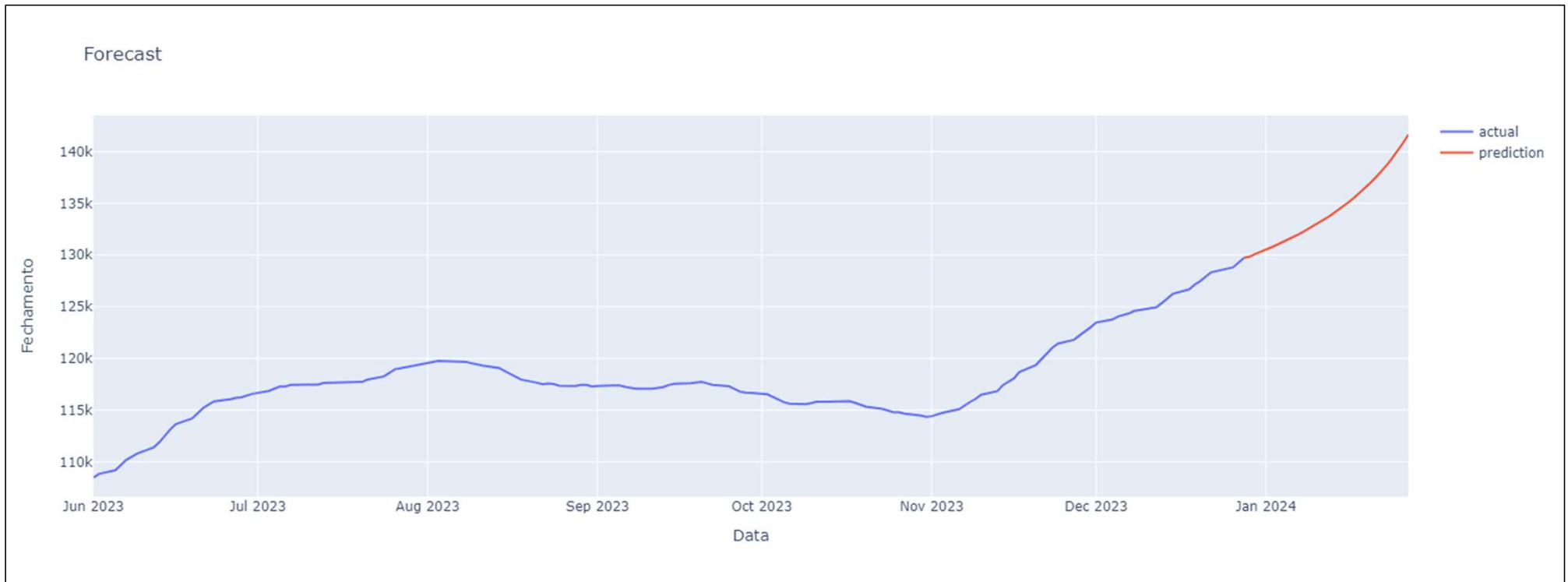


FORECASTING

05

FORECASTING

Prevendo os próximos pontos da Série Temporal



_ FORECASTING

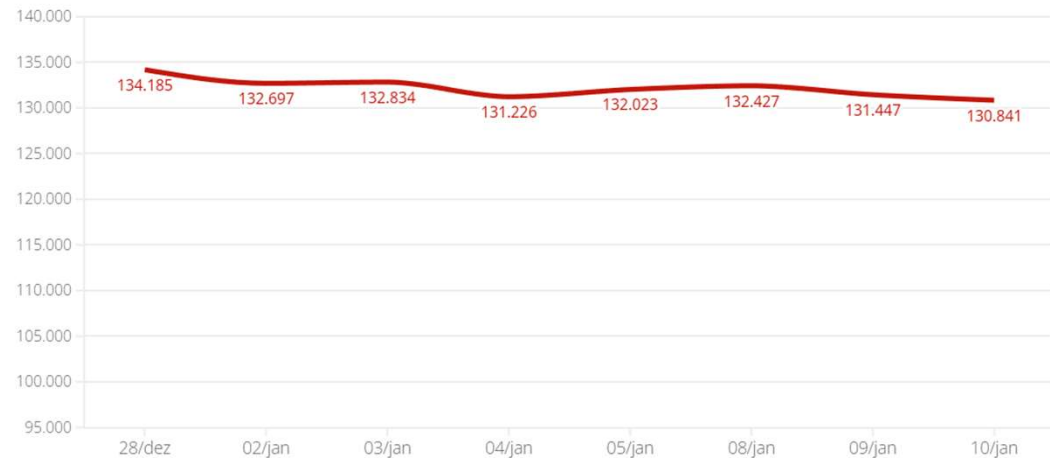
Tabela mostrando os valores previstos para todo o mês de janeiro.

Modelo rodado para testes com dados apenas até o final de 2023, não considerado informações de 2024.

Variação do Ibovespa em 2024

Pontuação de fechamento

Clique nas linhas para visualizar outros valores



Date	Actual	Forecast
01/01/2024	NaN	130528.287140
02/01/2024	NaN	130749.065610
03/01/2024	NaN	130988.329516
04/01/2024	NaN	131229.483399
05/01/2024	NaN	131482.798562
06/01/2024	NaN	131743.731788
07/01/2024	NaN	132015.848594
08/01/2024	NaN	132298.211260
09/01/2024	NaN	132592.422632
10/01/2024	NaN	132898.987916
11/01/2024	NaN	133219.168309
12/01/2024	NaN	133553.970586
13/01/2024	NaN	133904.735906
14/01/2024	NaN	134272.856309
15/01/2024	NaN	134659.927373
16/01/2024	NaN	135067.748211
17/01/2024	NaN	135498.303300
18/01/2024	NaN	135952.246863
19/01/2024	NaN	136432.424770
20/01/2024	NaN	136941.283088
21/01/2024	NaN	137482.031146
22/01/2024	NaN	138058.187208
23/01/2024	NaN	138673.971013
24/01/2024	NaN	139334.162026
25/01/2024	NaN	140044.455619
26/01/2024	NaN	140811.477620
27/01/2024	NaN	141643.125953



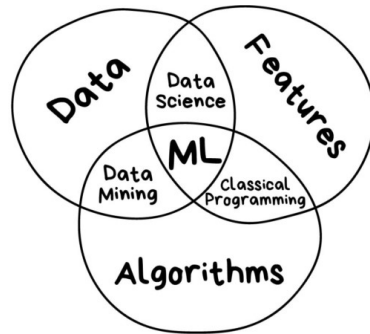
CONSIDERAÇÕES FINAIS

06

_ CONSIDERAÇÕES FINAIS

Durante o desenvolvimento da análise dos dados do Ibovespa, foram treinados e testados alguns modelos de ML para verificar a previsão do mercado futuro. Iniciando pela análise exploratória dos dados, transformação dos dados, aplicação de modelos complexos e verificação dos resultados. O objetivo é identificar um modelo que não apenas capture os padrões e tendências nos dados, mas que também consiga prever os dados futuros com a maior precisão possível.

Por fim, podemos concluir após todos esses fatores e analisando MAPE e RMSE que o Prophet é o melhor modelo.



Para continuidade e trabalhos futuros recomendamos melhorar as métricas do modelo com mais features que podem ajudar e utilizar previsões em outra escala. Cabe ressaltar que cada ativo na bolsa possui um comportamento, portanto um modelo poderia ser considerado para cada tipo de ativo.



REFERÊNCIAS

07

REFERÊNCIAS

- ❖ <https://br.investing.com/indices/bovespa-historical-data>
- ❖ <https://blog.toroinvestimentos.com.br/bolsa/bolsa-de-valores-b3/>
- ❖ <https://master.clear.com.br/o-que-e-ibovespa/>
- ❖ <https://conteudos.xpi.com.br/aprenda-a-investir/relatorios/o-que-e-ibovespa/>
- ❖ https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm
- ❖ https://www.b3.com.br/pt_br/noticias/numero-de-investidores-na-b3-cresce-34-em-renda-fixae-23-em-renda-variavel-em-12-meses.htm
- ❖ <https://www.scielo.br/j/ram/a/NHnT4cmQhZPnScfqNtnTvdm/?lang=pt&format=pdf>
- ❖ <https://www.politize.com.br/fim-da-pandemia/>
- ❖ <https://www.politize.com.br/crise-financeira-de-2008/>
- ❖ <https://www.nordinvestimentos.com.br/blog/taxa-selic-hoje-tudo-o-que-voce-precisa-saber/#:~:text=A%20rela%C3%A7%C3%A3o%20entre%20o%20Ibovespa%20e%20a%20Taxa%20Selic%20%C3%A9,precisam%20investir%20em%20seus%20neg%C3%B3cios>