

UNIVERSIDADE PRESBITERIANA MACKENZIE

CARLOS SUETIRO KAERIYAMA

GABRIELA GERMANA DA SILVA

JONATHAN HENRIQUE GOMES DA PAZ FACCHINETTI

VANESSA TEÓFILO RIBEIRO DULTRA

**Explorando o Impacto das Taxas de Inflação e SELIC no
Mercado de Ações Brasileiro**

SÃO PAULO

2024

Sumário

1.	Introdução	3
2.	Fonte de dados.....	4
3.	Referencial teórico	4
4.	Diagrama de soluções.....	6
5.	Modelo Base	7
6.	Modelo	10
7.	Discussão	11
8.	Conclusão.....	12
9.	Referências.....	12

Explorando o Impacto das Taxas de Inflação e SELIC no Mercado de Ações Brasileiro

Carlos Suetiro, Gabriela Germana, Jonathan Facchinetti, Vanessa Teófilo

Universidade Presbiteriana Mackenzie

Faculdade de Computação e Informática - SP - Brazil

Abstract. *The Ibovespa is the main stock index in Brazil, composed of the most traded and representative stocks on the São Paulo Stock Exchange (B3). Calculated in real-time, it reflects the market capitalization of listed companies. It serves as a benchmark for investors and analysts to assess the performance of the Brazilian stock market. The index is influenced by economic and political factors, international events, and company performance, making it volatile. The project aims to understand how inflation and the basic interest rate (SELIC) affect the Ibovespa, using advanced analytical techniques and statistical models to forecast future trends, evaluated by performance metrics such as MSE, MAE, and RMSE.*

Resumo. *O Ibovespa é o principal índice de ações do Brasil, composto pelas ações mais negociadas na Bolsa de Valores de São Paulo (B3). Calculado em tempo real, reflete a capitalização de mercado das empresas listadas. Ele serve como referência para investidores e analistas avaliarem o desempenho do mercado de ações brasileiro. O índice é influenciado por fatores econômicos, políticos, eventos internacionais e desempenho das empresas, podendo ser volátil. O projeto busca entender como a inflação e a taxa básica de juros (SELIC) afetam o Ibovespa, utilizando técnicas analíticas avançadas e modelos estatísticos para prever tendências futuras, avaliados por métricas de desempenho como MSE, MAE e RMSE.*

1. Introdução

Antes da criação da Bovespa em 1968, o mercado de ações no Brasil era fragmentado, com várias bolsas regionais como São Paulo, Rio de Janeiro e Porto Alegre, cada uma com suas próprias regras. A negociação de ações era infrequente e com menor volume, tornando o mercado menos líquido e mais volátil, e o acesso era restrito, especialmente para pequenos investidores. A regulamentação fraca aumentava a insegurança jurídica e o risco de fraudes. Em 1968, como parte do Plano Trienal para modernizar a economia, o governo criou a Bovespa, visando democratizar o acesso ao capital e estimular o investimento privado.

Nos anos 70, o Ibovespa cresceu rapidamente devido ao Plano Nacional de Desenvolvimento e ao Milagre Econômico, mas a crise do petróleo em 1973 e a instabilidade política causaram quedas significativas no final da década. Nos anos 80, o índice enfrentou grande volatilidade com a crise da dívida externa e a hiperinflação, com recuperações temporárias durante os Planos Cruzado e Bresser. Nos anos 90, o Plano Real estabilizou a economia, levando o Ibovespa a níveis recordes. No novo milênio, crises

como a bolha da internet em 2001 e a crise financeira global de 2008 impactaram o índice, que perdeu mais de 30% de seu valor em 2008, mas se recuperou gradualmente posteriormente.

O projeto utilizará técnicas analíticas avançadas para analisar o mercado de ações, como diferenciação para estacionarizar séries temporais, análise exploratória de dados, identificação de correlações e decomposição das séries. Modelos estatísticos como ARIMA, SARIMA, Regressão Linear e RandomForestRegressor serão aplicados para prever tendências futuras, avaliando a eficácia com métricas de desempenho como MSE, MAE e RMSE para garantir previsões precisas.

Compreender essas relações permite tomar medidas preventivas e corretivas diante de mudanças no mercado financeiro, alocando recursos eficientemente, comunicando-se proativamente com investidores e melhorando a infraestrutura para cenários adversos. O projeto está relacionado ao ODS 8 (Trabalho Decente e Crescimento Econômico), promovendo um entendimento mais profundo do crescimento econômico e das condições de trabalho no país, analisando a influência das taxas de inflação e da Selic no mercado de ações brasileiro.

2. Fonte de dados

Os dados utilizados são a evolução mensal do índice IBOVESPA (https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-estatisticas-historicas.htm), a taxa SELIC (<https://www3.bcb.gov.br/sgspub/consultarvalores/consultarValoresSeries.do?method=getPagina>) e a taxa de inflação IPCA (<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=series-historicas>) do IBGE. Os dados são formados pelo mês e seus respectivos valores.

O IBOVESPA é o principal indicador do mercado de ações brasileiro, refletindo o desempenho das ações negociadas na Bolsa de Valores de São Paulo (B3). Por sua vez, a SELIC desempenha o papel de taxa básica de juros, servindo como referência para todas as demais taxas de juros no país. Enquanto isso, o Índice Nacional de Preços ao Consumidor Amplo (IPCA), calculado pelo IBGE, destaca-se como um dos principais índices de inflação do Brasil, com ampla utilização em análises econômicas e tomadas de decisão.

3. Referencial teórico

A análise de problemas complexos frequentemente se beneficia da inspiração em trabalhos correlacionados, os quais fornecem uma base sólida de conhecimento e diferentes abordagens para resolver problemas semelhantes. Nesta seção, exploram-se alternativas de solução empregadas em problemas relacionados, destacando suas vantagens e limitações, bem como definindo resumidamente os principais conceitos envolvidos em cada solução.

Montgomery, D. C., Peck, E. A., e Vining, G. G. (2015), em seu livro "Introdução à Análise de Regressão Linear", oferecem um guia abrangente e aprofundado sobre os conceitos e técnicas fundamentais da regressão linear. A obra abrange desde os princípios básicos da regressão linear simples até modelos mais sofisticados, como a regressão múltipla e modelos não lineares. Além de explorar os aspectos teóricos, o livro fornece orientações práticas para a aplicação dos conceitos em análises reais, abordando a seleção

de variáveis, interpretação dos resultados e validação dos modelos.

"Applied Predictive Modeling" de Kuhn, M. e Johnson, K. (2013), é um livro prático e detalhado que aborda técnicas essenciais de modelagem preditiva. Inicia com conceitos fundamentais e avança para tópicos mais avançados, cobrindo o pré-processamento de dados, limpeza e transformação de dados, e explicando uma variedade de modelos de regressão e classificação, como regressão linear, árvores de decisão e redes neurais. Discussões detalhadas sobre métricas de avaliação de modelos, como MSE e RMSE, e técnicas de validação cruzada e bootstrap são apresentadas. O livro inclui estudos de caso práticos de várias áreas, demonstrando a aplicação real das técnicas discutidas, sendo amplamente elogiado por sua abordagem clara e aplicada à modelagem preditiva.

Albuquerque, F. (2024), em seu artigo "Modelo Preditivo com base na IBOVESPA (bolsa de valores)", publicado no Medium, apresenta um modelo preditivo para o índice IBOVESPA da Bolsa de Valores de São Paulo, utilizando a série histórica de cotações diárias entre 2002 e 2023. A pesquisa utiliza a metodologia de séries temporais, especificamente modelos ARIMA (Autoregressive Integrated Moving Average), para modelar o comportamento do IBOVESPA. Os modelos ARIMA capturam a relação entre valores passados e futuros de uma série temporal, considerando a sazonalidade do índice. Os resultados mostram que modelos ARIMA com sazonalidade são mais adequados para prever os movimentos do índice do que modelos sem sazonalidade, embora a performance diminua para previsões de longo prazo.

Santos, G. C. (2020), em sua dissertação de mestrado "Algoritmos de Machine Learning para Previsão de Ações da B3", Universidade Federal de Uberlândia, utiliza dados históricos de preços e indicadores financeiros de 33 ações da B3 entre 2010 e 2019. O estudo aplica cinco algoritmos de Machine Learning: Regressão Linear, K-Nearest Neighbors (KNN), Random Forest, Support Vector Machines (SVM) e Redes Neurais Artificiais (RNA). O algoritmo de RNA apresentou o melhor desempenho geral, seguido por SVM, Random Forest, KNN e Regressão Linear. Os modelos foram capazes de prever retornos futuros das ações com precisão razoável, especialmente para períodos de curto prazo (30 dias).

Os fundamentos fornecidos pelos trabalhos e livros referenciados proporcionam uma base teórica e prática sólida para conduzir uma análise robusta e produzir resultados significativos. Aplicando técnicas analíticas avançadas e modelos estatísticos, e avaliando a eficácia dos modelos com métricas relevantes, é possível obter uma compreensão mais profunda de como a taxa de inflação e a SELIC influenciam o mercado de ações brasileiro.

4. Diagrama de soluções

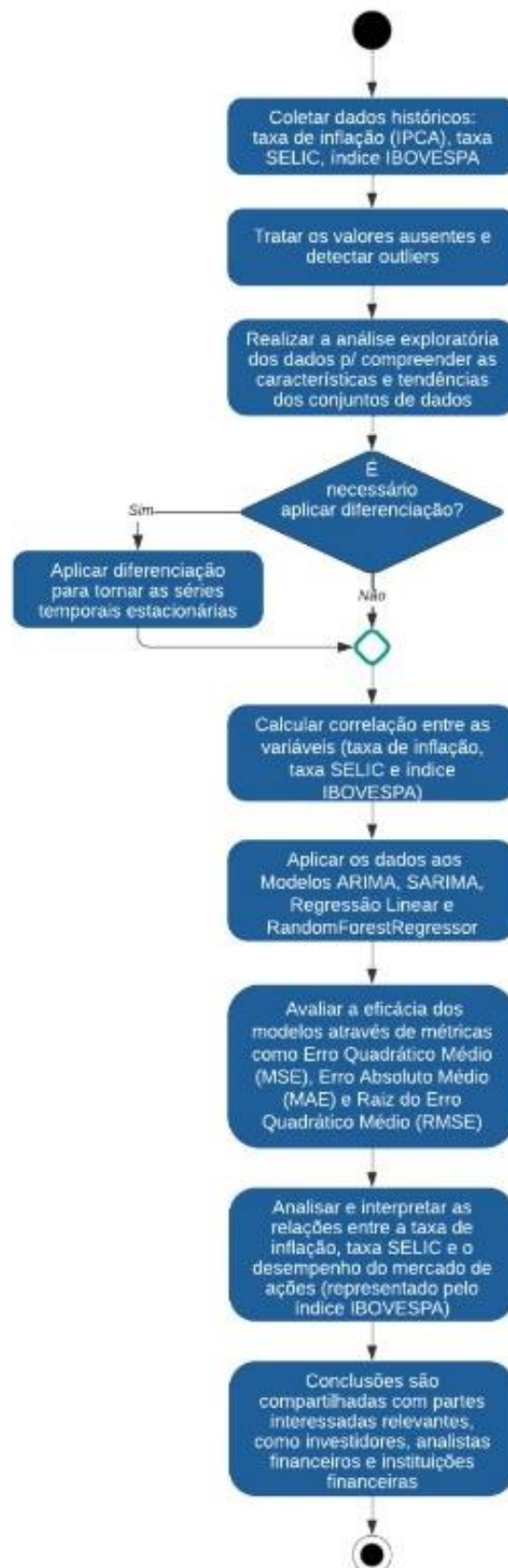


Figura 1. Diagrama de Solução

5. Modelo Base

Para prever os próximos índices da Ibovespa, inicialmente, utilizou-se ferramentas estatísticas de análise da série Ibovespa.

A primeira análise foi a decomposição da série temporal Ibovespa em tendência, sazonalidade e resíduos.

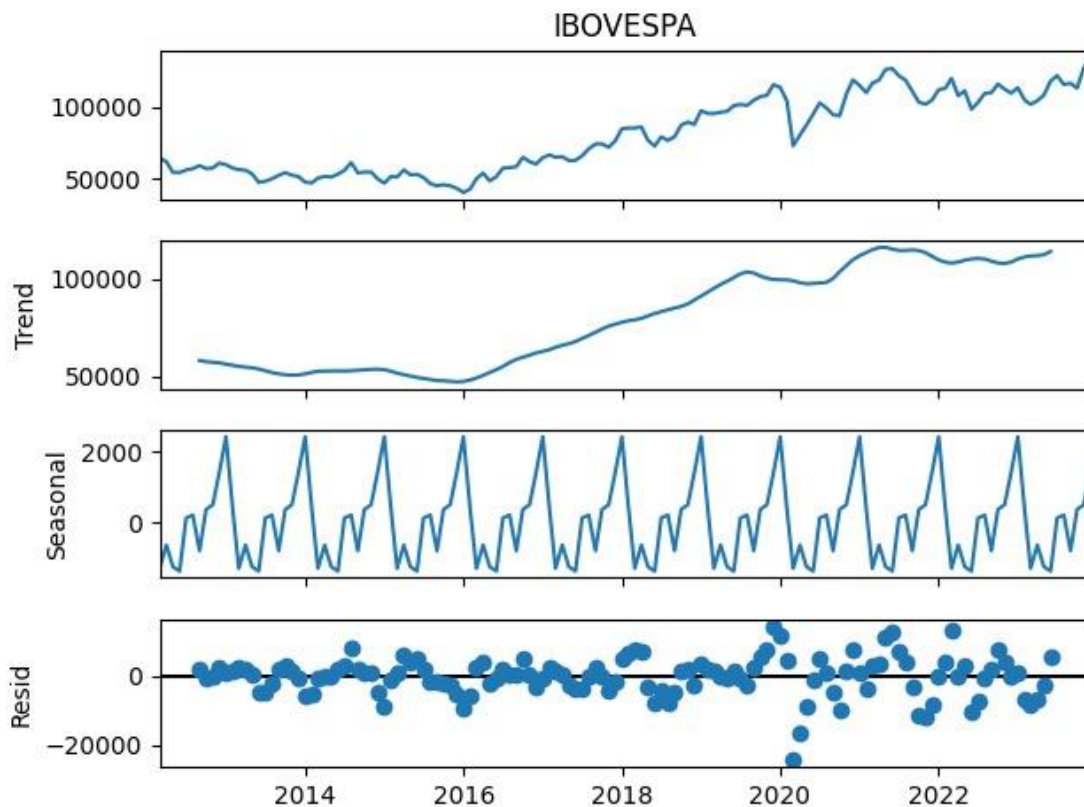


Figura 2. Gráfico com a decomposição, que utilizou-se o `seasonal_decompose()` da biblioteca `statsmodels`

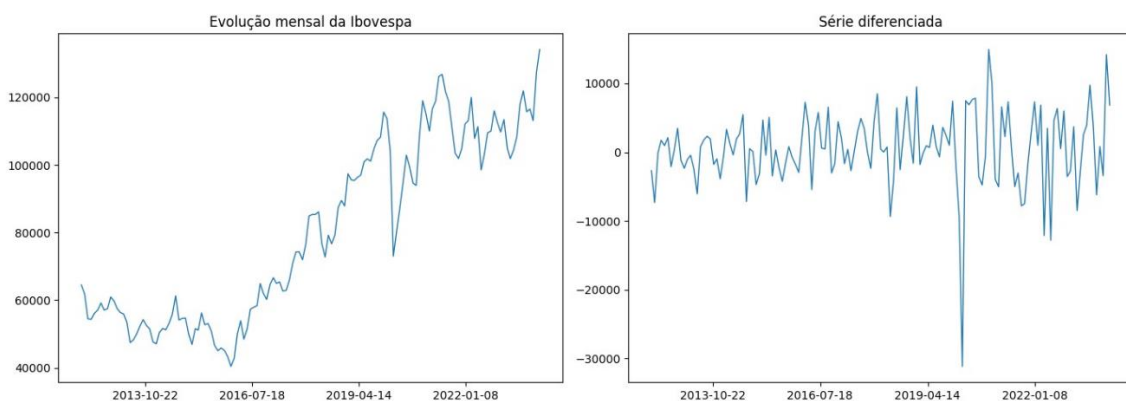


Figura 3. Gráfico que aplicando apenas uma diferenciação, a série apresenta-se como estacionária.

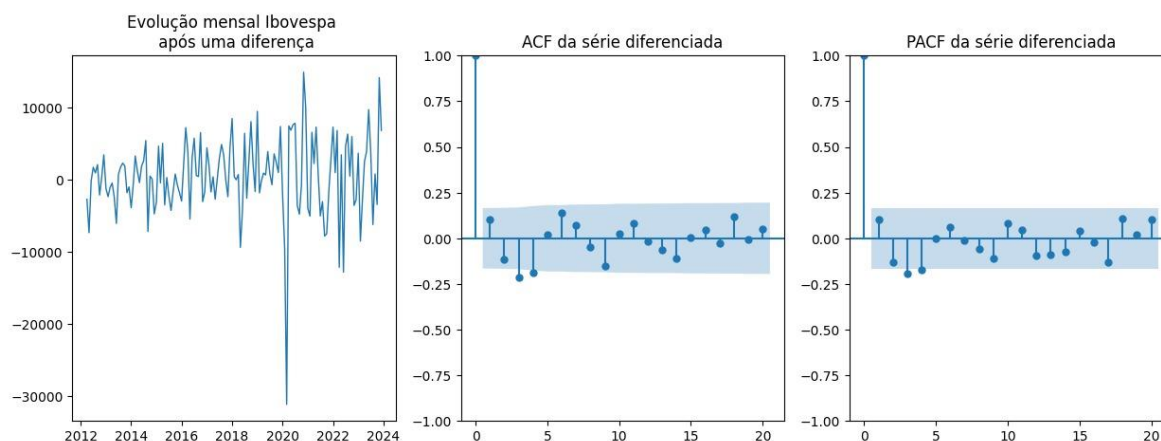


Figura 4. Gráfico, percebe-se a grande queda dos valores para dentro do intervalo, mostrando que após a aplicação de uma diferença a série se tornou estacionária.

Por causa desse fato, aplicou-se, como primeiro modelo de previsão, a técnica ARIMA. Para aplicação dessa técnica, precisa-se definir os parâmetros AR (parte Autorregressiva - p), I (parte integrada - d) e MA (Médias móveis - q).

Para aplicar a técnica ARIMA, utilizou-se a função ARIMA da biblioteca statsmodels e para descobrir os parâmetros, utilizou-se a biblioteca pmdarima.

```
Best model: ARIMA(4,1,0)(0,0,0)[1] intercept
Total fit time: 2.697 seconds
```

SARIMAX Results						
=====						
Dep. Variable:	y	No. Observations:	142			
Model:	SARIMAX(4, 1, 0)	Log Likelihood	-1407.907			
Date:	Sun, 02 Jun 2024	AIC	2827.814			
Time:	02:14:11	BIC	2845.507			
Sample:	03-01-2012	HQIC	2835.004			
	- 12-01-2023					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

intercept	750.0255	481.820	1.557	0.120	-194.324	1694.375
ar.L1	-0.0050	0.059	-0.084	0.933	-0.121	0.111
ar.L2	-0.1372	0.102	-1.348	0.178	-0.337	0.062
ar.L3	-0.1068	0.060	-1.790	0.073	-0.224	0.010
ar.L4	-0.1018	0.058	-1.742	0.081	-0.216	0.013
sigma2	2.712e+07	0.043	6.38e+08	0.000	2.71e+07	2.71e+07
=====						
Ljung-Box (L1) (Q):	0.57	Jarque-Bera (JB):	437.23			
Prob(Q):	0.45	Prob(JB):	0.00			
Heteroskedasticity (H):	5.61	Skew:	-1.42			
Prob(H) (two-sided):	0.00	Kurtosis:	11.14			
=====						

Figura 5. Resultado obtido dos parâmetros, após aplicar a função pm.auto_arima().

Função pm.auto_arima(), foram:

$p = 4$; $d = 1$; $q = 0$.

O valor de $d = 1$ estava previsto, pois ao aplicar apenas uma diferença, a série adquiriu características estacionárias, conforme análise anterior.

Do resultado acima, observa-se que o critério de informação Akaike (AIC) obtido foi 2827.814.

Ainda aplicando modelos estatísticos para realização da previsão, utilizou-se a técnica SARIMA. Para aplicação do SARIMA, precisa-se definir os parâmetros de sazonalidade, além dos parâmetros descritos anteriormente para o ARIMA.

```

Performing stepwise search to minimize aic
ARIMA(2,1,2)(1,0,1)[7] intercept : AIC=2831.265, Time=0.85 sec
ARIMA(0,1,0)(0,0,0)[7] intercept : AIC=2830.418, Time=0.03 sec
ARIMA(1,1,0)(1,0,0)[7] intercept : AIC=2834.132, Time=0.08 sec
ARIMA(0,1,1)(0,0,1)[7] intercept : AIC=2834.069, Time=0.07 sec
ARIMA(0,1,0)(0,0,0)[7] : AIC=2829.575, Time=0.03 sec
ARIMA(0,1,0)(1,0,0)[7] intercept : AIC=2832.343, Time=0.06 sec
ARIMA(0,1,0)(0,0,1)[7] intercept : AIC=2832.327, Time=0.08 sec
ARIMA(0,1,0)(1,0,1)[7] intercept : AIC=2832.923, Time=0.17 sec
ARIMA(1,1,0)(0,0,0)[7] intercept : AIC=2832.197, Time=0.05 sec
ARIMA(0,1,1)(0,0,0)[7] intercept : AIC=2832.141, Time=0.05 sec
ARIMA(1,1,1)(0,0,0)[7] intercept : AIC=2834.410, Time=0.08 sec

Best model: ARIMA(0,1,0)(0,0,0)[7]
Total fit time: 1.560 seconds

```

Figura 6. Parâmetro - ARIMA

```

Out[26]: SARIMAX Results

```

Dep. Variable:	y	No. Observations:	142
Model:	SARIMAX(0, 1, 0)	Log Likelihood	-1413.787
Date:	Sun, 02 Jun 2024	AIC	2829.575
Time:	02:15:36	BIC	2832.523
Sample:	03-01-2012	HQIC	2830.773
	- 12-01-2023		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
sigma2	2.976e+07	1.71e+06	17.441	0.000	2.64e+07	3.31e+07

Ljung-Box (L1) (Q):	1.49	Jarque-Bera (JB):	332.21
Prob(Q):	0.22	Prob(JB):	0.00
Heteroskedasticity (H):	7.37	Skew:	-1.27
Prob(H) (two-sided):	0.00	Kurtosis:	10.08

Figura 7. Resultados dos parâmetros

O resultado obtido dos parâmetros, após aplicar a função `pm.auto_arima()`, foram:

$p = 0$; $d = 1$; $q = 0$; $P = 0$; $D = 0$; $Q = 0$.

O valor de $d = 1$ estava previsto, pois ao aplicar apenas uma diferença, a série adquiriu características estacionárias, conforme análise anterior.

Do resultado acima, observa-se que o critério de informação Akaike (AIC) obtido foi 2829.575. O valor obtido do AIC foi próximo ao valor do ARIMA, porém, superior. Isso representa que o modelo ARIMA tem uma maior qualidade e simplicidade.

Além dos métodos estatísticos, aplicou-se o aprendizado de máquina para tentar prever o valor do índice Ibovespa.

Como a série Ibovespa tem uma tendência linear, conforme a decomposição realizada anteriormente, aplicou-se a regressão linear para a previsão. Com o auxílio do IPCA e da Selic, aplicou-se a Random Florest Regressor para realização da previsão. Como o Random Florest Regressor realiza o agrupamento dos valores, esse método pode auxiliar na previsão do índice Ibovespa, pois quando o IPCA e a Selic tiverem determinado valor, será possível verificar qual o grupo pertence o índice Ibovespa.

6. Modelo

Dos resultados abaixo, verifica-se que a informação presente na série Ibovespa foi considerado no modelo ARIMA, pois, no gráfico ACF os valores são todos próximos de zero e os resíduos apresentam uma distribuição quase normal.

Após fazer essa verificação, realizou-se a previsão dos últimos doze meses da série temporal. Essa ação é necessária para verificar a acurácia do modelo.

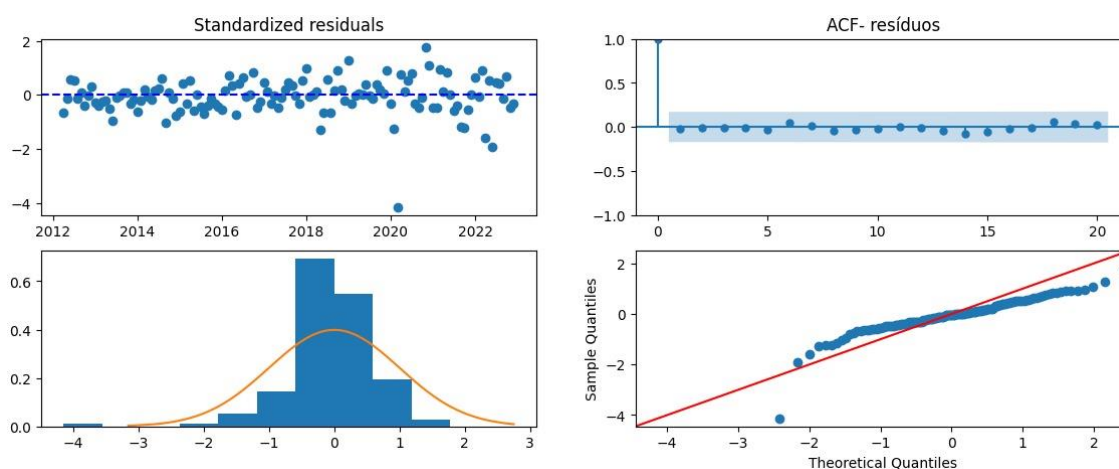


Figura 8. Previsão dos últimos doze meses da série temporal

Dos resultados abaixo, verifica-se os primeiros cinco valores previstos, dentro do modelagem, e o gráfico apresentando os resultados obtidos. As previsões foram realizadas com um intervalo de precisão de 95%.

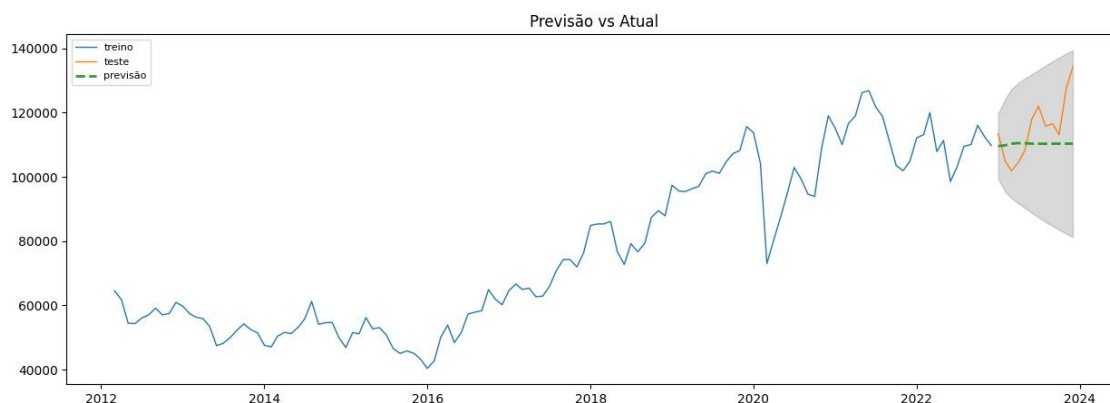


Figura 9. Os primeiros cinco valores previstos

7. Discussão

O estudo foi dividido em dois grupos, o primeiro grupo considera os modelos estatísticos e o segundo grupo os modelos de aprendizado de máquina. A seleção dos modelos ocorreu através do AIC, para os modelos estatísticos, e o MAE, para os modelos de aprendizado de máquina. As tabelas abaixo apresentam os valores obtidos.

Modelo Estatístico:

Modelo	AIC
ARIMA	2.827.814
SARIMA	2.829.575

Modelo de Aprendizado de Máquina:

Modelo	MAE
Regressão linear	7091.61
Random Forest	4433.40

Os modelos ARIMA e Random Forest foram selecionados por apresentarem os melhores parâmetros dentro dos seus grupos. Após o desenvolvimento do modelo ARIMA foi possível obter o seu MAE e compara-lo com o Random Forest, conforme tabela abaixo:

Modelo	MAE
ARIMA	8336.96
Random Forest	4433.40

Da tabela acima, percebe-se que o MAE do Random Forest é melhor que o do ARIMA, porém, o valor previsto de janeiro/23 do ARIMA é mais próximo que o valor previsto pelo Random Forest. A tabela abaixo apresenta os valores obtidos e, também, o valor real do índice Ibovespa para esse mês.

Modelo	Ibovespa Previsto	Ibovespa Real
ARIMA	109570.55	113430.54
Random Forest	107385.45	

8. Conclusão

Foram aplicados dois métodos de aprendizado de máquina e dois métodos estatísticos. Dentre os métodos testados, destacam-se o Random Forest e o ARIMA, devido aos menores erros médios absolutos (MAE) e critérios de informação de Akaike (AIC) em seus respectivos grupos. Após uma análise comparativa, optou-se por desenvolver o modelo ARIMA, dada sua consagração no mercado financeiro. Durante o desenvolvimento do ARIMA, foi possível obter um MAE e realizar previsões com um intervalo de confiança de 95%. Posteriormente, o modelo foi utilizado para prever o índice Ibovespa para janeiro de 2023, permitindo uma comparação com os valores reais. Uma melhoria potencial para o projeto seria a incorporação de redes neurais artificiais, especialmente as redes LSTM, conhecidas por se adaptarem bem a séries temporais. Além disso, ampliar o conjunto de dados utilizado e testar mais modelos, como os de Suavização Exponencial, poderia enriquecer a análise e melhorar a precisão das previsões.

9. Referências

Oliveira, R., Abarracin, O. Y. E., Silva, G. R. (2024) Introdução às Séries Temporais: Uma Abordagem Prática em Python (in printing). Editora Mackenzie.

Análise Prática de Séries Temporais: Predição com estatística e aprendizado de máquina. Aileen Nielsen/ traduzido por Cibelle Ravaglia. Rio de Janeiro: Alta Books, 2021

GRÔPPO, Gustavo de Souza. Causalidade das variáveis macroeconômicas sobre o IBOVESPA. 2004. Dissertação (Mestrado) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2004.

B3. Índice Bovespa (Ibovespa B3). São Paulo, 2024. Disponível em: https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-estatisticas-historicas.htm . Acesso em: 13 março 2024.

Banco Central do Brasil. SGS - Sistema Gerenciador de Séries Temporais - v2.1. Brasília, 2024. Disponível em: <https://www3.bcb.gov.br/sgspub/consultarvalores/consultarValoresSeries.do?method=getPagina> . Acesso em: 13 março 2024.

Instituto Brasileiro de Geografia e Estatística. IPCA - Índice Nacional de Preços ao Consumidor Amplo. Brasília, 2024. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=series-historicas> . Acesso em: 13 março 2024.