



APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA COMO ESTRATÉGIA OPERACIONAL NO MERCADO DE CAPITAIS BRASILEIRO

LUCAS YUKI NISHIMOTO RA: 211024678

MATEUS RODER

ANDRE LUIS DEBIASO ROSSI

SUMÁRIO

- **INTRODUÇÃO**
 - **TRABALHOS RELACIONADOS**
 - **METODOLOGIA**
 - **RESULTADOS EXPERIMENTAIS**
 - **CONCLUSÃO**
-

MERCADO DE AÇÕES

01 DEFINIÇÃO

- O mercado de ações é um ambiente onde são negociadas partes do capital de empresas, denominadas ações.
- Essas ações representam uma fração do valor de uma empresa, e ao adquiri-las, os investidores se tornam "sócios" da companhia.

02 OBJETIVO

- Permitir que empresas captem recursos para financiar suas atividades e crescimento.
 - Oferecer aos investidores uma oportunidade de retorno financeiro por meio da valorização das ações e dividendos.
-

MERCADO DE AÇÕES

03 PRINCIPAIS PARTICIPANTES

- Empresas (companhias abertas): Emitem ações para captar recursos.
- Investidores: Podem ser individuais (pessoa física) ou institucionais (fundos de investimento, bancos, etc.).
- Bolsas de Valores: Ambiente regulado para a negociação de ações, como a B3 (Brasil), NYSE (EUA), etc.
- Corretoras de Valores: Facilitam a compra e venda de ações.

PROBLEMA

01 DESAFIOS NO MERCADO DE AÇÕES

- Risco Elevado
- Complexidade dos Dados

02 PROBLEMAS COMUNS PARA INVESTIDORES

- Tomada de Decisões Baseada em Dados
- Ruído no Mercado

PROBLEMA

03 MOTIVAÇÃO PARA USAR APRENDIZADO DE MÁQUINA

- Automação da Análise
- Adaptação Rápida
- Previsões Baseadas em Dados

04 PROBLEMAS ABORDADOS NO ESTUDO

- Previsão de Comportamento dos Ativos
- Redução do Risco
- Otimização de Portfólios

OBJETIVO

- Investigar o uso de técnicas de aprendizado de máquina na composição de portfólios de ativos financeiros, com foco na maximização de retornos a médio e longo prazo (a partir de 3 anos).

REVISÃO BIBLIOGRÁFICA

TRABALHOS CORRELATOS

01 ZHANG, WANG E ZHAO (2020)

- Objetivo: Utilizar Redes Neurais Convolucionais (CNN) para classificação de sentimentos em notícias financeiras e prever movimentos de preços de ações.
- Metodologia:
 - Indicadores: Índice de sentimento e variáveis macroeconômicas.
 - Métricas: Precisão e F1-score.
- Resultados: A combinação de CNN com análise de sentimentos melhorou a capacidade de previsão dos movimentos dos preços.

02 SBRANA E CASTRO (2023)

- Objetivo: Previsão de preços de criptomoedas usando técnicas aplicáveis ao mercado de ações.
- Metodologia:
 - Modelo: N-BEATS com camadas convolucionais, camadas de atenção e função de ativação Mish.
 - Arquitetura: N-BEATS Perceiver (com Transformer).
 - Indicadores: Preço histórico e volume de negociação.
 - Métricas: Erro Médio Absoluto (MAE) e Erro Médio Quadrático (RMSE).
- Resultados: O N-BEATS Perceiver demonstrou superioridade em previsão de preços em nível de portfólio.

METODOLOGIA

ETAPAS

- Coleta de dados
- Cálculo dos indicadores
- Gerar classe
- Pré-processamento
- Treinamento dos modelos
- Escolha do portfólio

AQUISIÇÃO DE DADOS

- Dados históricos de ações obtidos da API do Yahoo Finance (YFinance).
- Período de Análise: 01/2010 a 10/2024.
- Características Coletadas:
 - Preços de abertura, fechamento, máxima, mínima.
 - Volume de negociação.

CÁLCULO DOS INDICADORES

- Uso de um arquivo python separado
- 13 indicadores
- Entre eles: RSI, MACD, DIR, DIR_P...

GERANDO A CLASSE

- Uma outra coluna é adicionada ao Data frame, chamada “Classe”, que poderá ter 3 valores. -1, 0 e 1

GERANDO A CLASSE

Situação 1:

Considerando:

Período: 5 dias

Percentual: 1%.

No dia 1 o valor para uma ação é R\$ 100.

No dia 6 o valor para a mesma ação é R\$102.

O preço variou 2% em relação ao dia 1.

Logo houve uma variação maior que o percentual estabelecido, e portanto será gerada uma classe "1" (comprar) para o dia 1.

GERANDO A CLASSE

Situação 2:

Considerando:

Período: 5 dias

Percentual: 1%.

No dia 1 o valor para uma ação é R\$ 100.

No dia 6 o valor para a mesma ação é R\$100,5.
5.

O preço variou 0,5% em relação ao dia 1.

Logo houve uma variação menor que o percentual estabelecido, e portanto será gerada uma classe "0" (manter) para o dia 1.

GERANDO A CLASSE

Situação 3:

Considerando:

Período: 5 dias

Percentual: 1%.

No dia 1 o valor para uma ação é R\$ 100

No dia 6 o valor para a mesma ação é R\$97

O preço variou 3% em relação ao dia 1. (mas negativamente)

Por ter hávido uma variação negativa acima do percentual, será gerada uma classe “-1” (vender) para o dia 1.

PRÉ-PROCESSAMENTO

Os dados foram limpos para remover valores ausentes ou não numéricos gerados durante o processo de criação dos indicadores.

TREINAMENTO E VALIDAÇÃO DOS MODELOS

- Para este trabalho foram utilizados 3 modelos de aprendizado de máquina.
 - Random Forest
 - SVM
 - XGBoost
- No total foram avaliados 72 papéis, retirados da bolsa brasileira
- Qualquer coluna que não fosse ser utilizada foi deletada do DF, deixando apenas as variáveis desejadas
- Para o treinamento dos modelos foi definido como target a classe gerada anteriormente, e como variáveis as demais colunas do DF.
- Período destinado para treino foi de 10 anos (01/2010 – 12/2019) e para validação foi de 2 anos (01/2020 – 12/2021).

TREINAMENTO E VALIDAÇÃO DOS MODELOS

- Foi utilizada a técnica de grid search com validação cruzada, com 5 partições, para encontrar uma boa combinação de hiper-parâmetros
- A escolha da melhor combinação é feita através de uma métrica de avaliação que pode ser a acurácia, acurácia balanceada, índice sharpe e maximum drawdown

TREINAMENTO E VALIDAÇÃO DOS MODELOS

Tabela 3 – Combinação de hiper-parâmetros para a Random Forest.

Random Forest	<i>n_estimators</i>	<i>max_depth</i>	<i>ccp_alpha</i>	percentual lateralização
	[80, 90, 100, 110, 120, 130]	[3, 4, 5, 6, 7, 8]	[0.05, 0.01, 0.015, 0.001, 0]	[0, 75, 1, 1, 25]

Tabela 4 – Combinação de hiper-parâmetros para a SVM.

SVM	<i>C</i>	γ	percentual lateralização
	[0, 1, 0, 5, 1, 5, 10, 50, 100]	[$1e-4$, $5e-4$, $1e-3$, $5e-3$, $1e-2$, $5e-2$, $1e-1$]	[0, 75, 1, 1, 25]

Tabela 5 – Combinação de hiper-parâmetros para o XGBoost.

XGBoost	<i>learning rate</i>	<i>n_estimators</i>	<i>max_depth</i>	percentual lateralização
	[0.05, 0.04, 0.03, 0.02, 0.01, 0.005, 0.001]	[80, 90, 100, 110, 120, 130]	[3, 4, 5, 6, 7, 8]	[0, 75, 1, 1, 25]

TREINAMENTO E VALIDAÇÃO DOS MODELOS

- A acurácia balanceada foi utilizada na busca em grade para escolher o melhor conjunto de hiper-parâmetros, uma vez que a distribuição de classes pode ser desbalanceada dependendo do percentual de lateralização.

ESCOLHA DO PORTIFÓLIO

Funcionamento

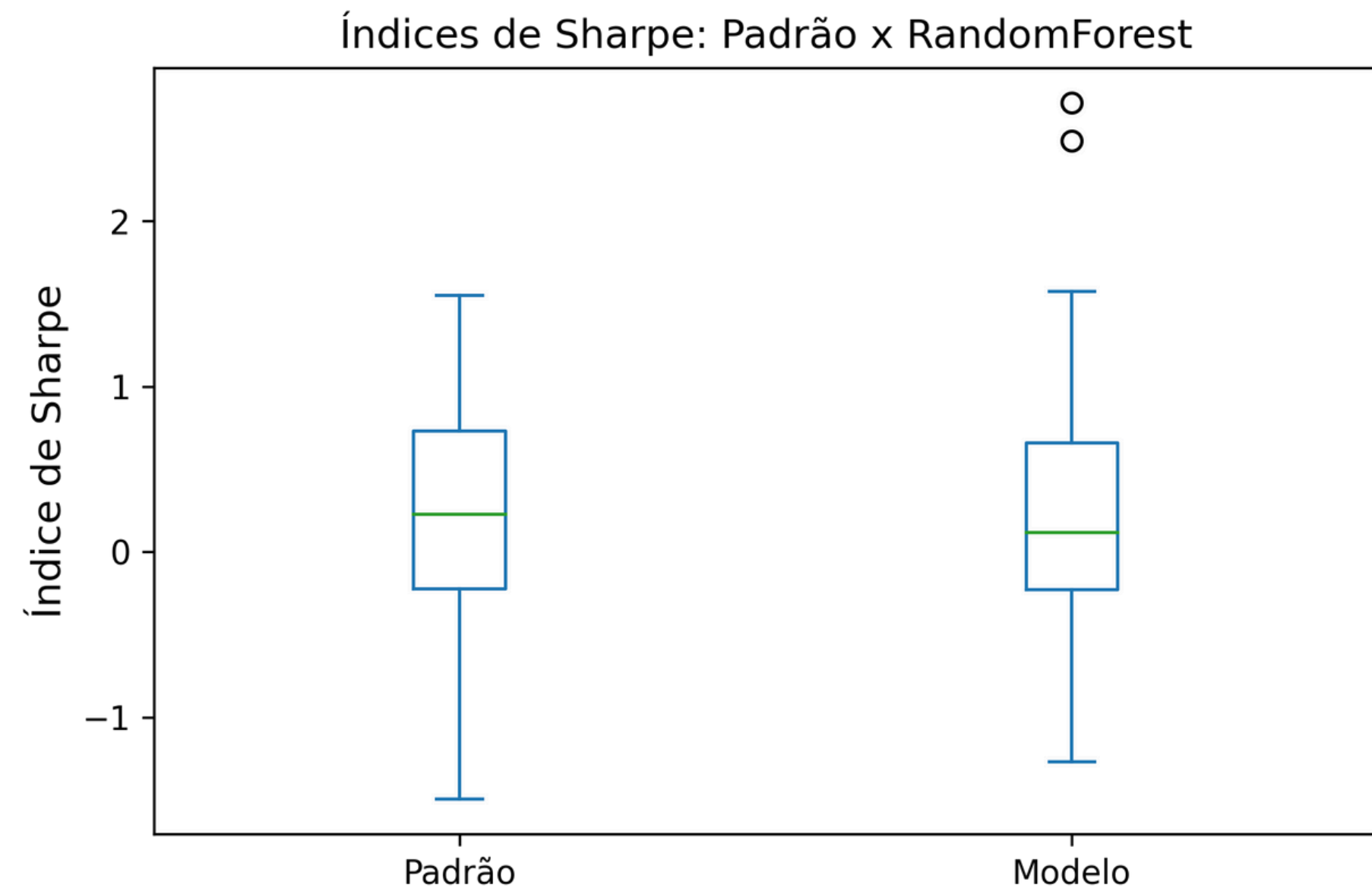
Dado o conjunto de previsões feitas pelos três modelos, para todos os papéis disponíveis no conjunto de validação:

- Foi selecionado um limiar (threshold) inferior para o índice de Sharpe considerando cada algoritmo.
- Após isso, foram filtrados os papéis com índice de Sharpe maior que o limiar
- Por fim, foram selecionados os papéis cujo rebaixamento máximo, MDD, dos algoritmos foi maior que o rebaixamento dos papéis selecionados (modelo versus real).

ESCOLHA DO PORTIFÓLIO

Sharpe

Sharpe entre os 25% superiores foram selecionados para o portfólio.

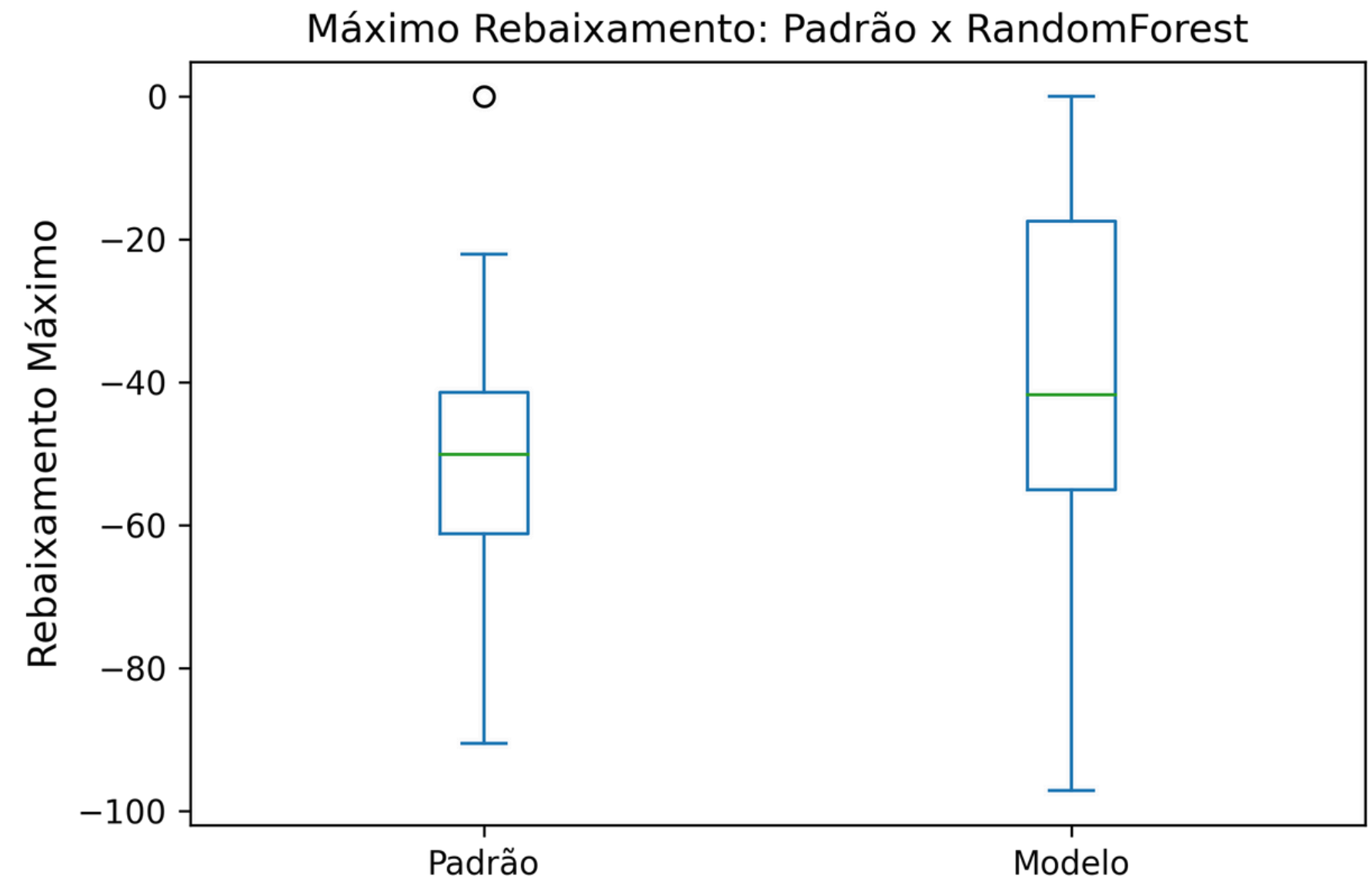


Box-plot do índice de Sharpe para a Random Forest, com terceiro quartil igual a 0,72.

ESCOLHA DO PORTIFÓLIO

MDD

Seleção dos papéis que tiveram um MDD maior que o rebaixamento dos papéis selecionados



Box-plot do drawdown máximo para o modelo Random Forest.

ESCOLHA DO PORTIFÓLIO

Algoritmo	Ativos Selecionados
Random Forest (RF)	BEEF3.SA, ABEV3.SA, TIMS3.SA, CCRO3.SA, ITSA4.SA, VIVT3.SA, CPLE6.SA, TOTS3.SA, VBBR3.SA, MRVE3.SA, ALPA4.SA, EGIE3.SA, EQTL3.SA
SVM	TIMS3.SA, EMBR3.SA, CCRO3.SA, ITSA4.SA, RAIL3.SA, CRFB3.SA, ELET3.SA, MULT3.SA, ALPA4.SA, TRPL4.SA, B3SA3.SA, AZZA3.SA, CPLE6.SA, UGPA3.SA, RADL3.SA, VIVT3.SA
XGBoost	BEEF3.SA, CCRO3.SA, BBDC3.SA, ABEV3.SA, BBDC4.SA, VIVT3.SA, ALPA4.SA, RADL3.SA, KLBN11.SA, BRKM5.SA, ELET6.SA, BRFS3.SA, PETR4.SA, PETR3.SA, CPLE6.SA, CMIG4.SA, WEGE3.SA

RESULTADOS

RESULTADOS OBTIDOS EM **TESTE**

Métricas

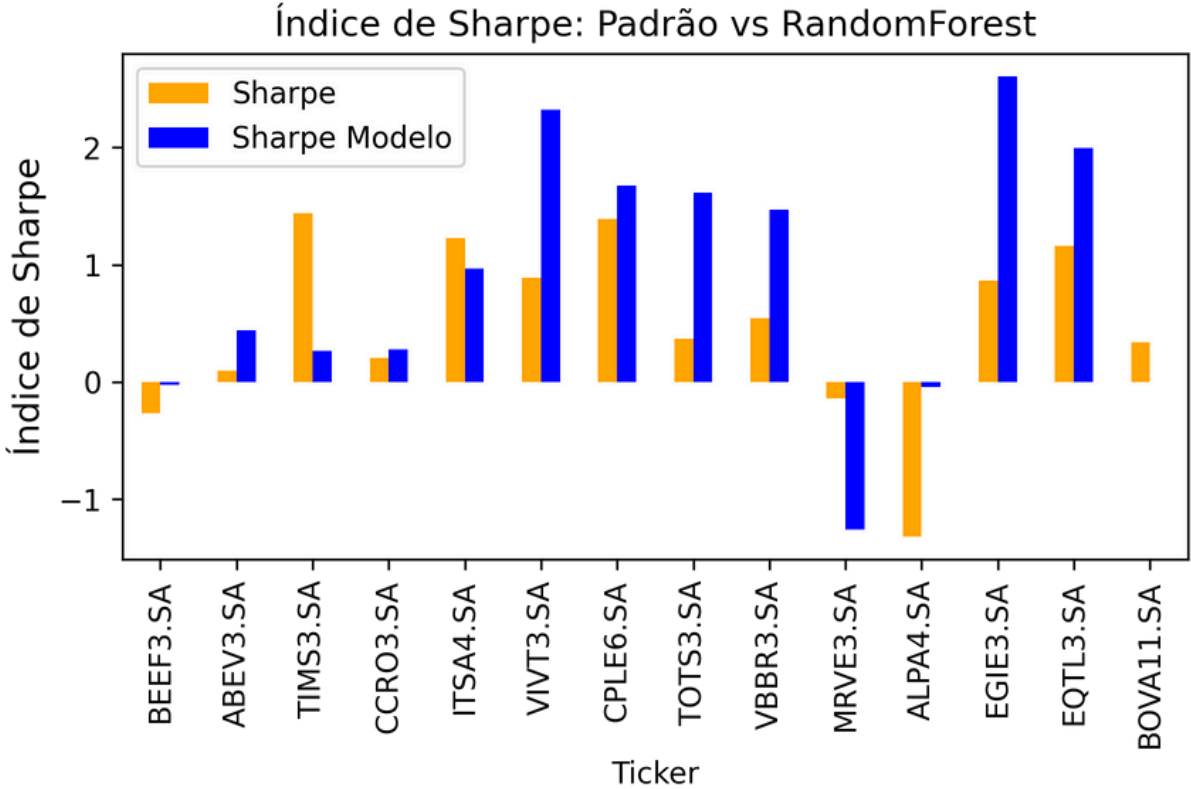
Para a análise dos resultados obtidos, foram utilizadas duas métricas.

- Sharpe
- Retorno percentual

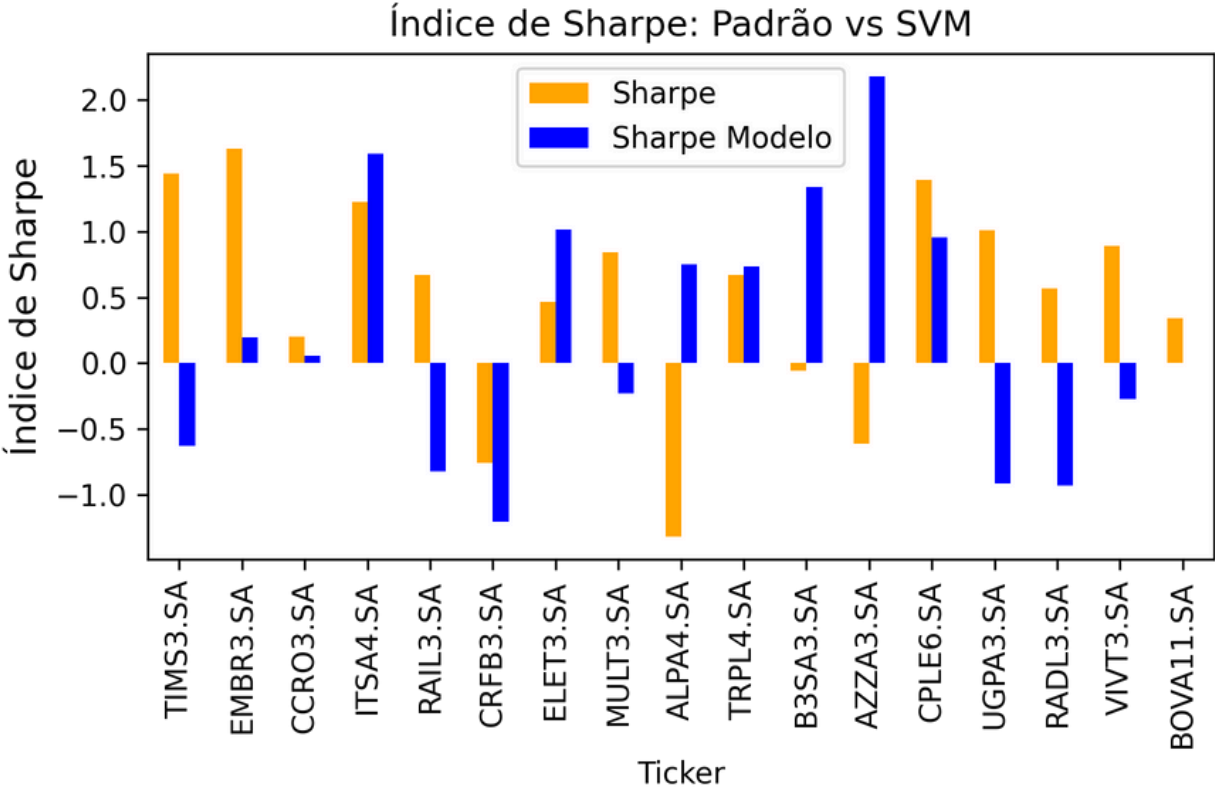
Comparando os resultados dos modelos com os resultados dos próprios papéis.
O período destinado a teste foi de aproximadamente 3 anos. (01/2022 – 10/2024)

RESULTADOS OBTIDOS EM TESTE

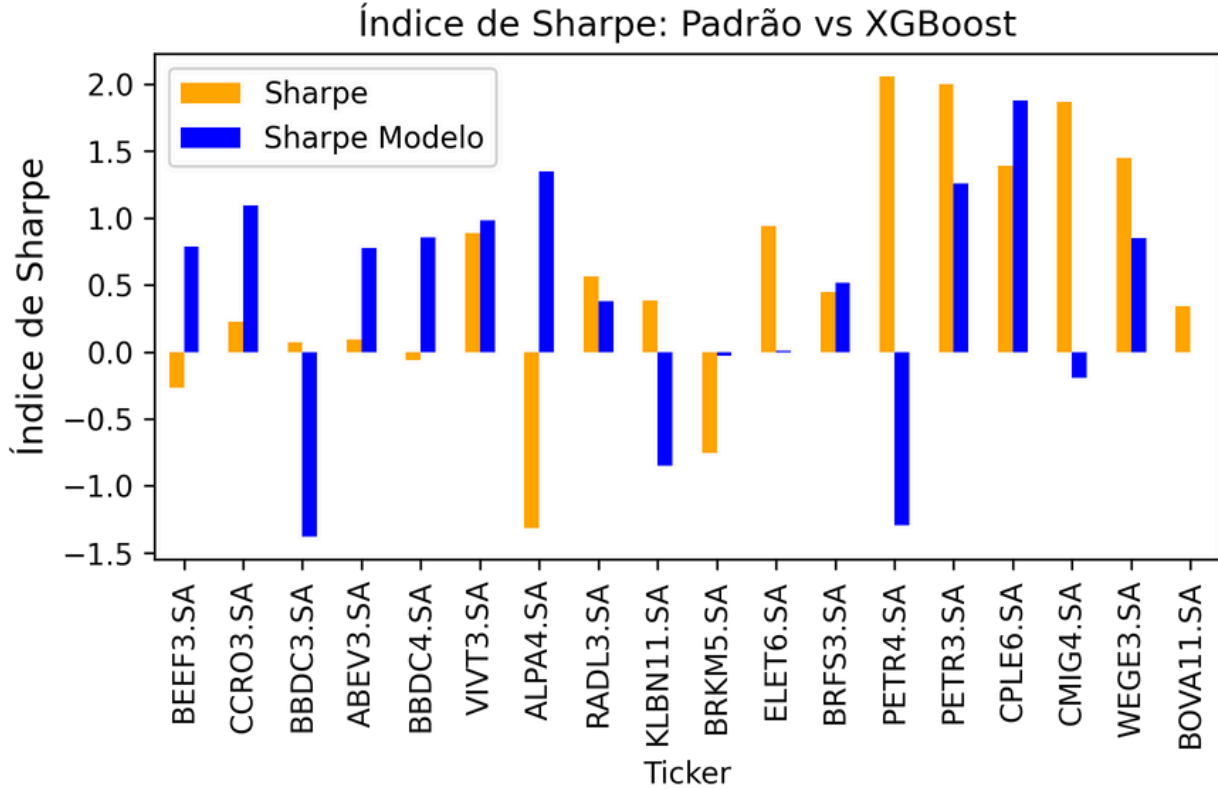
ÍNDICE SHARPE



Esse cenário pode indicar um bom resultado de retorno percentual acumulado, bem como uma volatilidade controlada,



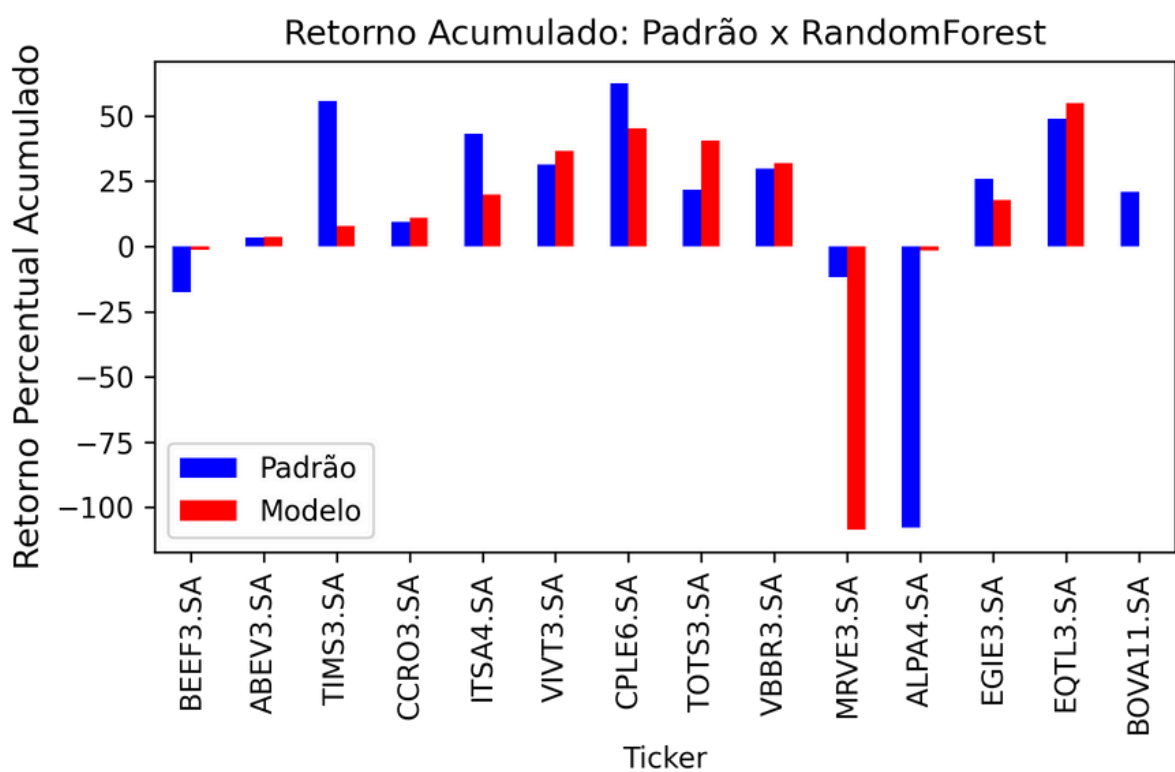
A variabilidade foi alta em sua maioria, indicando um portfólio com oscilações de retorno percentual e maior risco.



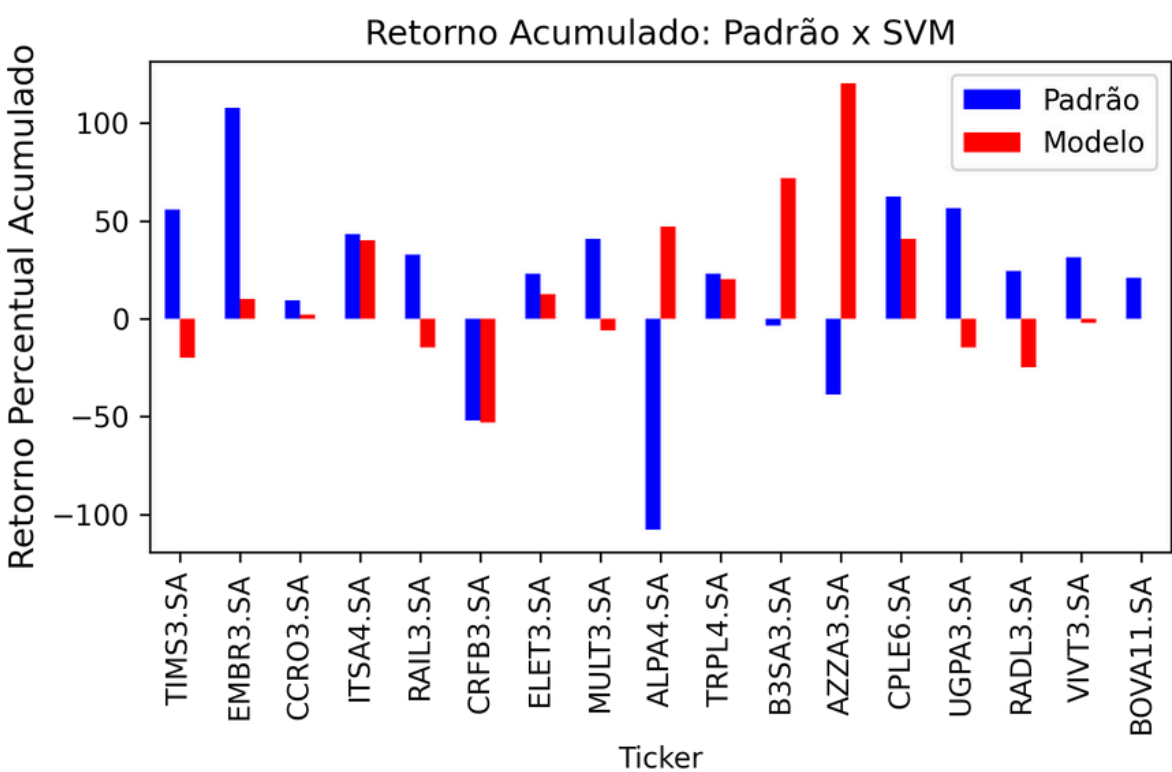
Maior possibilidade retorno percentual e variabilidade, uma vez que boa parte dos valores de Sharpe estão próximos de 1

RESULTADOS OBTIDOS EM TESTE

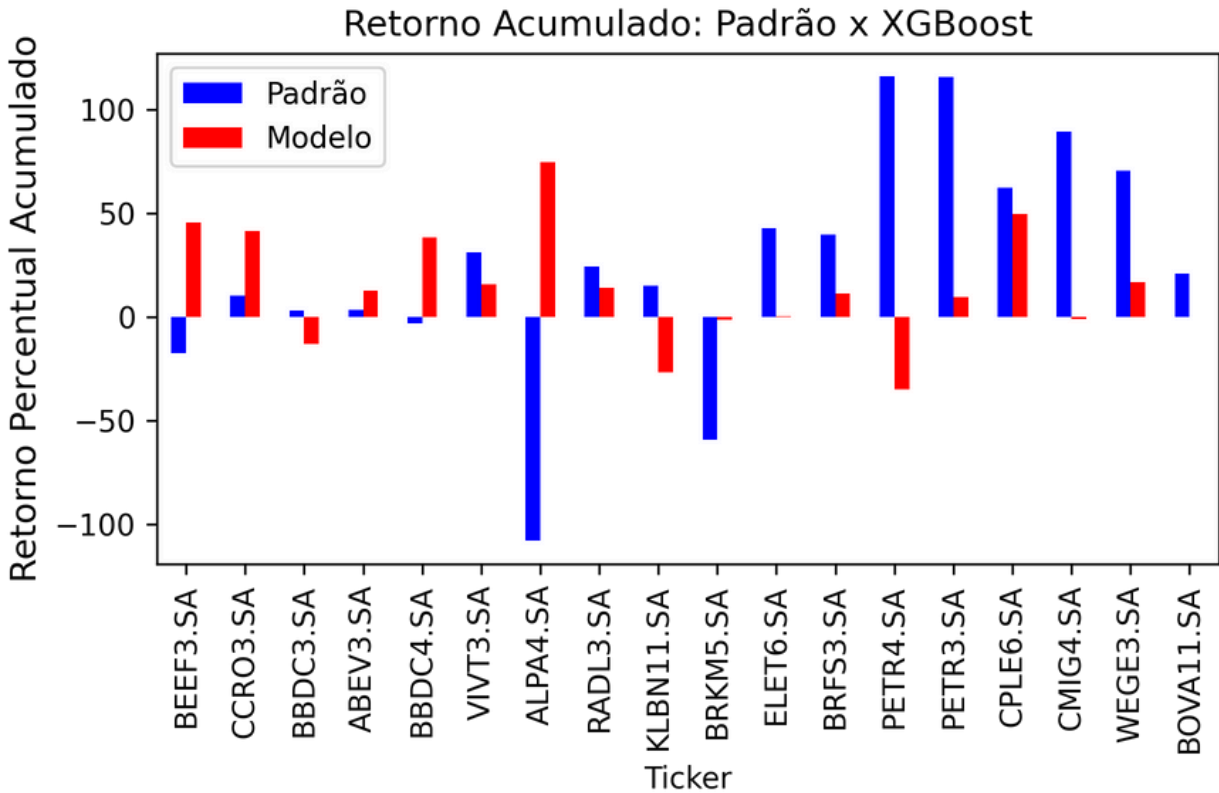
RETORNO ACUMULADO



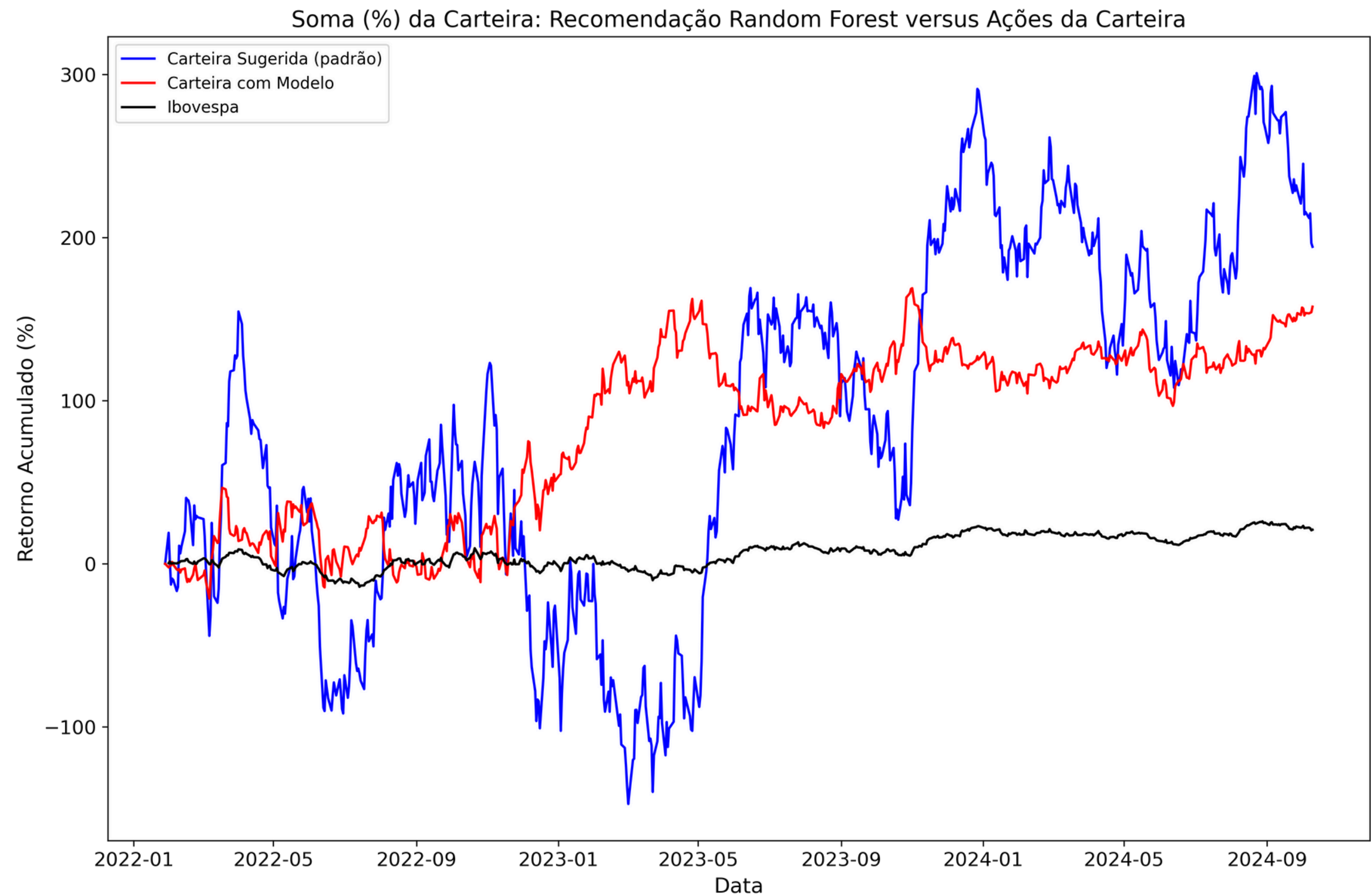
É um modelo promissor, já que possui estabilidade e consistência nos resultados com a estratégia adotada.

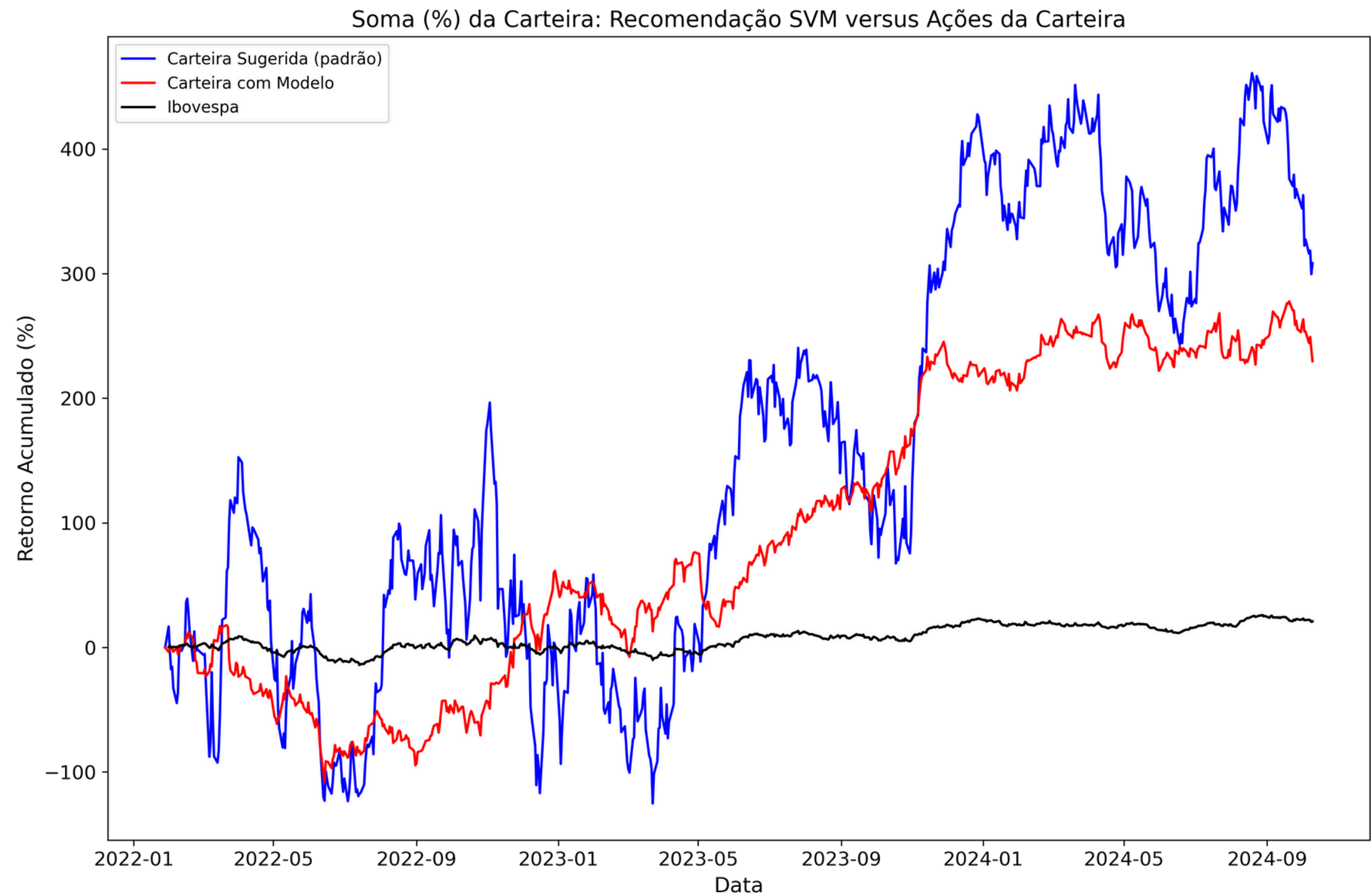


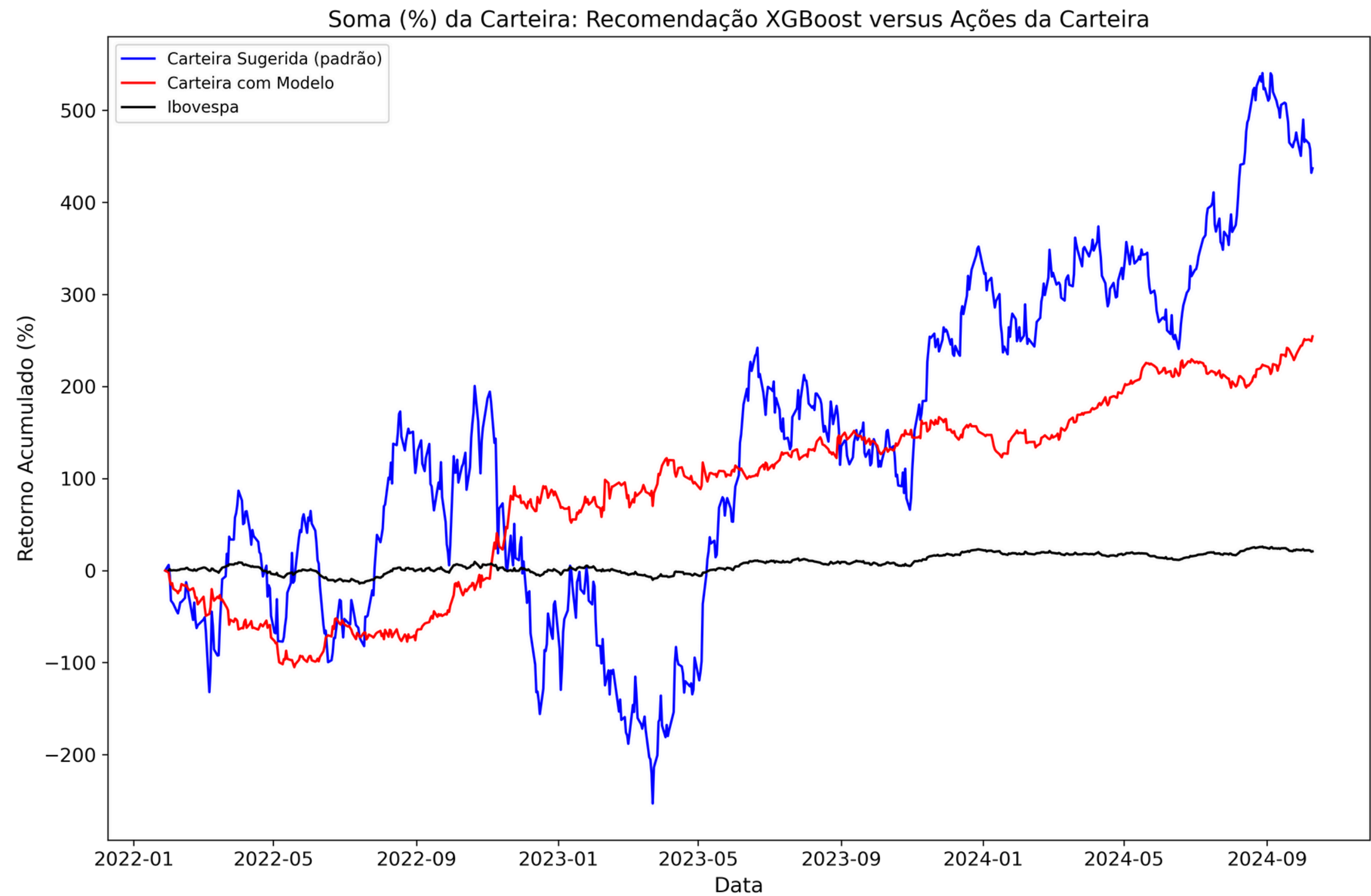
Indica um portfólio de alto risco e um desempenho de retorno instável ao longo do tempo,



Foi o portfólio que apresentou resultados com grande retorno, apesar de uma alta variabilidade







RESULTADOS OBTIDOS EM **TESTE**

Modelo / Ativo	Retorno (%)	Sharpe
Ibovespa	20,875	0,3404
RF	157,620	1,018
Ativos RF	194,354	0,477
SVM	229,709	1,357
Ativos SVM	308,472	0,560
XGBoost	254,413	1,731
Ativos XGBoost	436,996	0,892

CONCLUSÃO

CONCLUSÃO

Portanto, a análise sugere que, para investidores que buscam maximizar o retorno absoluto, a seleção direta de ativos pode ser atrativa.

Entretanto, para aqueles que preferem uma abordagem com menor risco e maior consistência, os modelos de aprendizado de máquina são uma escolha mais adequada, oferecendo uma vantagem significativa na gestão de portfólio com foco em estabilidade.

Esses insights mostram o potencial das técnicas de aprendizado de máquina para aprimorar a tomada de decisões no mercado financeiro e oferecem uma base promissora para futuras pesquisas e aprimoramentos nas estratégias de investimento.

OBRIGADO!

REFERÊNCIAS

ZHANG, L.; WANG, Y.; ZHAO, M. SENTIMENT ANALYSIS AND PREDICTION OF STOCK PRICE MOVEMENTS USING CNN. JOURNAL OF FINANCIAL STUDIES, V. 18, P. 182–195, 2020.

SBRANA, A.; CASTRO, P. A. LIMA DE. N-BEATS PERCEIVER: A NOVEL APPROACH FOR ROBUST CRYPTOCURRENCY PORTFOLIO FORECASTING. COMPUTATIONAL ECONOMICS, SPRINGER, P. 1–35, 2023.