

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS - CAMPUS BAURU
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

GABRIEL ZANFORLIN PARIZOTTO

**PREDIÇÃO DE TENDÊNCIAS NO MERCADO AMERICANO COM
RANDOM FOREST**

BAURU
Novembro/2024

GABRIEL ZANFORLIN PARIZOTTO

PREDIÇÃO DE TENDÊNCIAS NO MERCADO AMERICANO COM RANDOM FOREST

Trabalho de Conclusão de Curso do Curso de Bacharelado em Ciência da Computação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências, Campus Bauru.

Orientador: Profa. Dra. Simone das Graças Domingues Prado

BAURU
Novembro/2024

P234p

Parizotto, Gabriel Zanforlin

Predição de Tendências no Mercado Americano com
Random Forest / Gabriel Zanforlin Parizotto. -- Bauru, 2024
56 p. : il., tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Ciência da
Computação) - Universidade Estadual Paulista (UNESP),
Faculdade de Ciências, Bauru

Orientadora: Simone das Graças Domingues Prado

1. Ciência da Computação. 2. Floresta Aleatória. 3.
Aprendizado de Máquina. 4. Indicadores Financeiros. 5.
Mercado Financeiro Americano. I. Título.

Gabriel Zanforlin Parizotto

Predição de Tendências no Mercado Americano com Random Forest

Trabalho de Conclusão de Curso do Curso
de Bacharelado em Ciência da Computa-
ção da Universidade Estadual Paulista "Jú-
lio de Mesquita Filho", Faculdade de Ciên-
cias, Campus Bauru.

Banca Examinadora

Profa. Dra. Simone das Graças Domingues Prado
Orientador

Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Computação

Profa. Dra. Andréa Carla Gonçalves Vianna

Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Computação

Prof. Dr. Douglas Rodrigues

Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Computação

Bauru, 11 de Novembro de 2024.

Resumo

Este estudo explora a aplicação de técnicas avançadas de inteligência artificial no contexto da análise e predição de movimentos no mercado financeiro, com ênfase na implementação de modelos de aprendizado de máquina para aprimorar a tomada de decisões de investimento. Dentre as abordagens desenvolvidas, destaca-se o modelo *Random Forest*, estruturado como um *ensemble* de múltiplas árvores de decisão para combinar de forma eficaz dados provenientes de indicadores financeiros amplamente utilizados. Tal abordagem visa à identificação precisa de sinais de compra e venda, capturando padrões complexos presentes nos ativos analisados. A análise de séries temporais foi aplicada em ativos selecionados do mercado financeiro americano, utilizando dados históricos para avaliar a capacidade preditiva do modelo, bem como sua eficácia na mitigação de riscos inerentes à volatilidade dos mercados. O processo de desenvolvimento incluiu técnicas de validação cruzada e tratamento dos dados, com o objetivo de maximizar o desempenho do modelo, reduzindo o sobreajuste e aprimorando sua generalização. Ademais, foi desenvolvida uma solução *web* que abrange tanto o treinamento de modelos quanto a visualização das oportunidades de compra e venda no mercado americano, permitindo ao usuário buscar ações específicas e obter resultados do modelo em tempo real, bem como se inscrever para receber notificações automáticas sempre que surgirem oportunidades de compra ou venda. Como resultado, este trabalho apresenta uma aplicação automatizada e de alta confiabilidade que sugere oportunidades de investimento, evidenciando o potencial da inteligência artificial na análise preditiva do mercado financeiro.

Palavras-chave: Inteligência Artificial, Aprendizado de Máquina, Floresta Aleatória, Indicadores Financeiros, Mercado Financeiro Americano.

Abstract

This study explores the application of advanced artificial intelligence techniques in the context of analyzing and predicting movements in the financial market, with an emphasis on the implementation of machine learning models to enhance investment decision-making. Among the developed approaches, the Random Forest model stands out, structured as an ensemble of multiple decision trees to effectively combine data from widely used financial indicators. This approach aims to accurately identify buy and sell signals, capturing complex patterns present in the analyzed assets. Time series analysis was applied to selected assets from the American financial market, using historical data to assess the model's predictive capacity as well as its effectiveness in mitigating risks inherent to market volatility. The development process included cross-validation techniques and data treatment, aiming to maximize the model's performance by reducing overfitting and enhancing generalization. Furthermore, a web solution was developed to encompass both model training and visualization of buy and sell opportunities in the American market, allowing users to search for specific stocks and obtain real-time model results, as well as subscribe to receive automatic notifications whenever buy or sell opportunities arise. As a result, this work presents an automated and highly reliable application that suggests investment opportunities, highlighting the potential of artificial intelligence in predictive analysis of the financial market.

Keywords: Artificial Intelligence, Machine Learning, Random Forest, Financial Indicators, American Financial Market.

Lista de figuras

Figura 1 – Médias Móveis Exponenciais no gráfico de AAPL	17
Figura 2 – MACD no gráfico de AAPL	18
Figura 3 – RSI no gráfico de AAPL	19
Figura 4 – ATR no gráfico de AAPL	20
Figura 5 – Bandas de Bollinger no gráfico de AAPL	22
Figura 6 – Estocástico no gráfico de AAPL	24
Figura 7 – Exemplo de árvore de decisão final em uma RF	26
Figura 8 – Importância dos parâmetros da Random Forest	39
Figura 9 – Gráfico comparando resultados do RF com outros ativos	41
Figura 10 – Arquitetura do servidor principal do sistema	45
Figura 11 – Arquitetura do <i>workflow</i> de análise de ativos	46
Figura 12 – Arquitetura do <i>workflow</i> de subscrição de ativos	47
Figura 13 – Arquitetura do componente de eventos	48
Figura 14 – Tela principal de busca de ativos	51
Figura 15 – Tela de subscrição a eventos de recomendação	52
Figura 16 – Notificação de compra enviada via email	52
Figura 17 – Notificação de compra enviada via SMS	53

Lista de tabelas

Tabela 1 – Dados hipotéticos para treinamento de uma Random Forest	25
Tabela 2 – Dados simulados para <i>Bootstrap Aggregation</i> com sinais de compra, venda e manutenção	27
Tabela 3 – Amostras bootstrap dos dados de treino para o modelo RF	27
Tabela 4 – Exemplo didático de dados com sinal de pressão de mercado . . .	35
Tabela 5 – Métricas de desempenho da estratégia	36
Tabela 6 – Parâmetros otimizados do modelo Random Forest	37
Tabela 7 – Indicadores de desempenho para diferentes ações	38
Tabela 8 – Resultados das operações por ativo	40
Tabela 9 – Comparação de retorno e <i>drawdown</i> máximo entre modelos e ativos	41

Lista de abreviaturas e siglas

API	<i>Application Programming Interface</i>
ATR	<i>Average True Range</i>
AWS	<i>Amazon Web Services</i>
ECS	Elastic Container Service
EFS	Elastic File System
EMA	<i>Exponential Moving Average</i>
IA	Inteligência Artificial
KMS	Key Management Service
LSTM	<i>Long Short-Term Memory</i>
MACD	<i>Moving Average Convergence Divergence</i>
RF	<i>Random Forest</i>
RSI	<i>Relative Strength Index</i>
SMA	<i>Simple Moving Average</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
SNS	<i>Simple Notification Service</i>
SDK	<i>Software Development Kit</i>
SR	<i>Sharpe Ratio</i>
VPC	Virtual Private Cloud

Sumário

1	INTRODUÇÃO	11
2	PROBLEMA	11
3	OBJETIVOS	12
3.1	Objetivo Geral	12
3.2	Objetivos Específicos	12
4	FUNDAMENTAÇÃO TEÓRICA	14
4.1	Mercado Financeiro	14
4.2	Séries Temporais	14
4.3	Indicadores financeiros	15
4.3.1	Médias Móveis (<i>Moving Averages</i>)	16
4.3.2	<i>Moving Average Convergence Divergence</i> (MACD)	17
4.3.3	RSI	18
4.3.4	ATR	19
4.3.5	Bandas de Bollinger	21
4.3.6	Estocástico	22
4.4	<i>Random Forest</i>	24
4.4.1	Exemplo de Funcionamento	25
4.4.1.1	Árvore de Decisão Final	25
4.4.2	<i>Bootstrap Aggregation</i>	26
4.4.3	Modelos <i>Ensemble</i>	28
4.5	Métricas Utilizadas	28
4.5.1	<i>Sharpe Ratio</i>	29
4.5.2	Acurácia	29
4.5.3	<i>F1 Score</i>	30
4.5.4	Retorno Percentual	30
4.5.5	<i>Drawdown</i>	30
4.6	Escolha do Mercado Americano	31
4.6.1	CrITÉRIOS de Escolha das Ações	31
4.6.1.1	Indicadores Financeiros Chave	31
4.6.1.2	Relevância da Marca	32
4.6.1.3	Diversificação Setorial	32
4.6.2	Análise por Setor	32

5	IMPLEMENTAÇÃO	34
5.1	Estrutura e Configuração do Modelo	34
5.1.1	Tratamento dos Dados	34
5.1.2	Otimização e Validação dos Modelos	35
5.2	Performance	36
5.2.1	Resultado do Modelo	38
5.2.2	Resultado no mercado financeiro	39
6	ARQUITETURA	43
6.1	Arquitetura do Servidor da Aplicação	44
6.1.1	<i>Workflow</i> de Análise de Ativos	45
6.1.2	<i>Workflow</i> de Subscrição de Ativos	47
6.2	Arquitetura do Componente de Eventos	48
6.3	Tecnologias Utilizadas	49
6.4	Produto Final	50
7	CONCLUSÃO	54
	REFERÊNCIAS	55

1 Introdução

O mercado financeiro, que por muito tempo foi algo muito restrito e nichado a especialistas, ascendeu cada vez mais ao cidadão comum como uma promessa de riqueza rápida e abundante através da compra e venda de diferentes ativos. No entanto, essa busca por fortuna fácil e rápida esbarra em uma realidade complexa, onde a lógica e a razão se entrelaçam com a emoção e a subjetividade.

Para muitas pessoas, o mercado financeiro é visto como um grande jogo de xadrez no qual cada movimento é estrategicamente calculado para gerar um evento específico na partida, mas essa filosofia vai contra a natureza do sistema, que é profundamente influenciado por fatores emocionais e subjetivos. Estudos mostram que as emoções humanas, como medo, ganância e otimismo excessivo, desempenham um papel significativo nas decisões de investimento e no comportamento do mercado (KAHNEMAN; TVERSKY, 1979).

Os mercados financeiros são altamente influenciados pela psicologia dos investidores, que muitas vezes são levados por emoções e comportamentos irracionais ((SHILLER, 2015). Nesse sentido, é possível analisar padrões comportamentais anteriores do mercado para tentar prever tendências futuras. A análise técnica, por exemplo, busca identificar e interpretar padrões gráficos e tendências históricas que refletem a psicologia coletiva dos investidores.

Portanto, ao considerar que o mercado financeiro é influenciado por fatores emocionais e comportamentais, percebe-se a necessidade de ferramentas mais avançadas para lidar com essa complexidade. Nesse cenário, a Inteligência Artificial (IA) tem se destacado como uma ferramenta poderosa de análise e previsão. Por meio de algoritmos de *machine learning* e técnicas de processamento de linguagem natural juntamente com indicadores financeiros e padrões históricos, a IA é capaz de identificar padrões e tendências nos dados do mercado com eficiência e sem interferências sentimentais ou emocionais.

2 Problema

Compreender e prever o comportamento do mercado financeiro apresenta um desafio multifacetado devido à sua natureza irracional e emocional. A interação com-

plexa de fatores subjetivos e irracionais, como medo, ganância e otimismo, obscurece a compreensão clara dos elementos que impulsionam as flutuações do mercado. Além disso, a vasta quantidade de dados e informações consideradas pelos participantes do mercado, desde indicadores econômicos até notícias e eventos geopolíticos, adiciona uma camada de complexidade ao processo de tomada de decisão. Essa diversidade de dados torna desafiador o desenvolvimento de modelos preditivos precisos e eficazes.

Diante desse cenário, a tarefa de treinar a IA para identificar padrões e tomar decisões financeiras eficazes torna-se ainda mais complexa. A capacidade de identificar e interpretar padrões significativos nos dados do mercado, em meio ao ruído e à volatilidade característicos do ambiente financeiro, representa um desafio crítico. Assim, a necessidade de compreender a complexidade do mercado financeiro, considerando tanto fatores emocionais quanto a diversidade de informações disponíveis, emerge como um problema fundamental a ser abordado na integração de indicadores e padrões na tomada de decisão de IA.

3 Objetivos

3.1 Objetivo Geral

O objetivo deste trabalho é aprofundar os estudos sobre a aplicação de algoritmos de *machine learning*, em especial o *Random Forest* (RF), para a implementação de estratégias lucrativas no mercado financeiro, utilizando indicadores financeiros como base. Além de desenvolver essas estratégias baseadas em IA, o estudo também busca avaliar a rentabilidade das mesmas, utilizando dados históricos para verificar sua eficácia e viabilidade.

3.2 Objetivos Específicos

- Analisar os principais indicadores e padrões do mercado financeiro que são relevantes para a tomada de decisão em investimentos.
- Desenvolver modelos de IA que integrem diversos indicadores e padrões para auxiliar na tomada de decisão no mercado financeiro.
- Avaliar a precisão e a eficácia dos modelos de IA na identificação de oportunidades de investimento e na mitigação de riscos associados.

- Propor recomendações para otimizar a integração de indicadores e padrões financeiros nos modelos de IA para tomada de decisão.
- Implementar uma aplicação capaz de analisar dados de empresas com base nos modelos desenvolvidos e fornecer recomendações de investimento aos usuários.

4 Fundamentação Teórica

4.1 Mercado Financeiro

O mercado financeiro é um ambiente onde se realizam transações de compra e venda de ativos financeiros, como ações, títulos e derivativos. Ele é composto por diversas instituições, incluindo bolsas de valores, bancos, corretoras e outras entidades, cujo objetivo é facilitar a troca de recursos entre aqueles que desejam investir e aqueles que precisam de capital. O mercado financeiro desempenha um papel crucial na economia, ajudando a alocar recursos de forma eficiente, promovendo o crescimento econômico e permitindo que empresas e governos captem fundos para seus projetos e operações (MISHKIN, 2021).

A econometria aplica ferramentas estatísticas para modelar fenômenos econômicos, utilizando variáveis econômicas para prever comportamentos futuros, como lucros e perdas. No contexto do mercado financeiro, a previsão de tendências de preços é essencial para investidores que buscam maximizar lucros e evitar perdas. Com o auxílio de softwares e planilhas, analisa-se uma vasta quantidade de dados históricos na tentativa de identificar padrões de mercado. No entanto, o mercado é sensível a diversos fatores externos, como mudanças políticas, econômicas e sociais, além de fenômenos imprevisíveis, o que dificulta a criação de previsões totalmente precisas. Esses desafios remetem à teoria da eficiência de mercado, que sugere que os preços dos ativos refletem todas as informações disponíveis, tornando difícil obter vantagens com base em previsões (WOOLDRIDGE, 2016).

O uso de conhecimentos estatísticos para prever movimentos financeiros não é recente; essas filosofias já foram exploradas e documentadas por influentes personagens do mercado financeiro, que defendem a ideia de previsibilidade de ativos. Grandes nomes do mercado, como Paul Tudor Jones e George Soros, apresentaram teorias sobre a possibilidade de lucrar com mudanças de tendência, especialmente nos momentos de reversão do mercado: “Eu acredito que o melhor dinheiro é feito nas viradas do mercado” (SCHWAGER, 1989).

4.2 Séries Temporais

A análise de séries temporais no contexto de investimentos e *trading* no mercado financeiro, especialmente o americano, é uma técnica amplamente utilizada para estudar o comportamento de ativos ao longo do tempo. Séries temporais são definidas

como uma sequência de observações coletadas em intervalos regulares de tempo, como os preços diários de ações, volumes de negociação ou taxas de câmbio. No mercado financeiro, essa análise é empregada para identificar padrões passados, prever tendências futuras e desenvolver estratégias de investimento baseadas em dados históricos (HAMILTON, 1994; TSAY, 2010).

A aplicação de séries temporais permite a decomposição de componentes importantes, como tendência, sazonalidade e ciclos. A tendência refere-se à direção geral de longo prazo dos dados, que pode ser crescente, decrescente ou estável. A sazonalidade capta padrões que se repetem em intervalos fixos, como as flutuações sazonais nas vendas de determinados setores ou aumentos em épocas específicas, como o final de ano. Por sua vez, ciclos são flutuações que ocorrem em intervalos irregulares, muitas vezes influenciados por fatores econômicos ou eventos globais, como crises financeiras ou mudanças de políticas econômicas (BROCKWELL; DAVIS, 2016; HAMILTON, 1994).

No campo do *trading*, a análise de séries temporais é fundamental para a previsão de preços e identificação de oportunidades de compra ou venda. Técnicas como modelos de médias móveis, suavização exponencial e modelos de volatilidade são frequentemente empregadas para analisar o comportamento futuro de ativos. Por exemplo, ao identificar uma tendência de alta em uma série temporal, os *traders* podem decidir manter suas posições compradas, enquanto uma queda sustentada pode ser um sinal para venda (BOX; JENKINS, 1976).

Além disso, técnicas modernas como redes neurais do tipo *Long Short-Term Memory* (LSTM) e RF têm sido incorporadas à análise de séries temporais para lidar com a não linearidade e a alta volatilidade dos dados financeiros. Esses métodos têm demonstrado eficácia ao capturar padrões complexos, proporcionando previsões mais precisas e informando estratégias de *trading* mais robustas (TSAY, 2010).

4.3 Indicadores financeiros

Indicadores financeiros são ferramentas matemáticas utilizadas para analisar dados históricos de preço e volume no mercado, ajudando *traders* a identificar padrões, tendências e a volatilidade dos ativos. Eles fornecem informações que facilitam a tomada de decisões, oferecendo uma visão clara de quando entrar ou sair de uma operação. Indicadores como as médias móveis, o *Relative Strength Index* (RSI) e o *Average True Range* (ATR) são amplamente usados para suavizar flutuações e medir a força ou volatilidade do mercado. Esses indicadores são essenciais para transformar dados complexos em sinais práticos, e têm sido amplamente adotados por *traders* e investidores para aprimorar suas estratégias e mitigar riscos (MURPHY, 1999; WILDER,

1978).

4.3.1 Médias Móveis (*Moving Averages*)

As Médias Móveis são um dos indicadores mais comuns usados na análise técnica para suavizar as flutuações de preços e identificar tendências subjacentes. Existem dois tipos principais: a *Simple Moving Average* (SMA) e a *Exponential Moving Average* (EMA). A SMA calcula a média aritmética dos preços de um ativo em um determinado período, enquanto a EMA dá mais peso aos preços mais recentes, tornando-a mais sensível às últimas mudanças do mercado (MURPHY, 1999; PRING, 2014).

A SMA é calculada como a média aritmética dos preços de fechamento de um ativo ao longo de um período n , conforme a Fórmula (4.1):

$$SMA = \frac{P_1 + P_2 + \dots + P_n}{n} \quad (4.1)$$

Onde:

- P_1, P_2, \dots, P_n : preços de fechamento do ativo nos períodos considerados;
- n : número de períodos.

Já a EMA atribui um peso maior aos preços mais recentes, tornando-a mais sensível às mudanças de mercado. Sua fórmula é apresentada na Fórmula (4.2):

$$EMA_t = P_t \cdot \frac{2}{n+1} + EMA_{t-1} \cdot \left(1 - \frac{2}{n+1}\right) \quad (4.2)$$

Onde:

- EMA_t : valor da média exponencial no período atual;
- P_t : preço de fechamento no período atual;
- EMA_{t-1} : valor da EMA calculada no período anterior;
- n : número de períodos utilizados para o cálculo.

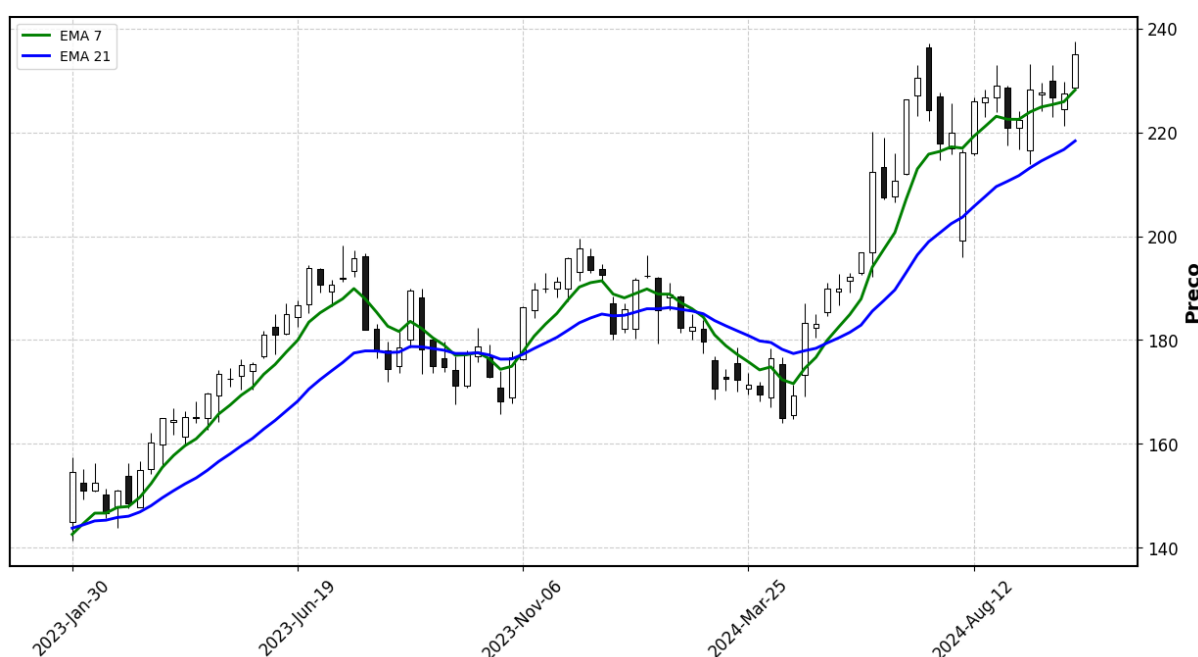
De acordo com Murphy (1999), as médias móveis desempenham um papel importante na suavização de séries de preços, ajudando a reduzir ruídos e identificar a direção das tendências. Pring (2014) também ressalta que as médias móveis são ferramentas essenciais para detectar potenciais sinais de compra ou venda.

Além de utilizar as médias móveis para identificar a direção predominante do mercado em determinado período, será implementada a técnica de cruzamento de

médias móveis. Essa técnica envolve o uso de duas médias móveis: uma rápida, calculada com um período menor, e uma lenta, com um período maior. Quando a média móvel rápida cruza acima da média lenta, isso pode indicar um aumento na pressão compradora, sugerindo o início de uma tendência altista no mercado. Por outro lado, uma vez que a média rápida cruza abaixo da média lenta, isso normalmente indica que a pressão vendedora está se sobressaindo no mercado.

A Figura 1 ilustra um exemplo de médias móveis exponenciais aplicadas ao gráfico de preços da AAPL, destacando os cruzamentos que indicam possíveis pontos de compra e venda.

Figura 1 – Médias Móveis Exponenciais no gráfico de AAPL



Fonte: Elaborada pelo autor

4.3.2 *Moving Average Convergence Divergence (MACD)*

O MACD é um indicador de momento amplamente utilizado na análise técnica, que mostra a relação entre duas médias móveis de diferentes períodos. Ele ajuda os *traders* a identificar mudanças no impulso, cruzamentos de médias e divergências, que são sinais antecipados de reversão de tendência. A fórmula básica do MACD, apresentada na Equação (4.3), envolve subtrair uma EMA de 26 períodos de uma EMA de 12 períodos, enquanto a linha de sinal é calculada como a EMA de 9 períodos da linha MACD, conforme a Equação (4.4).

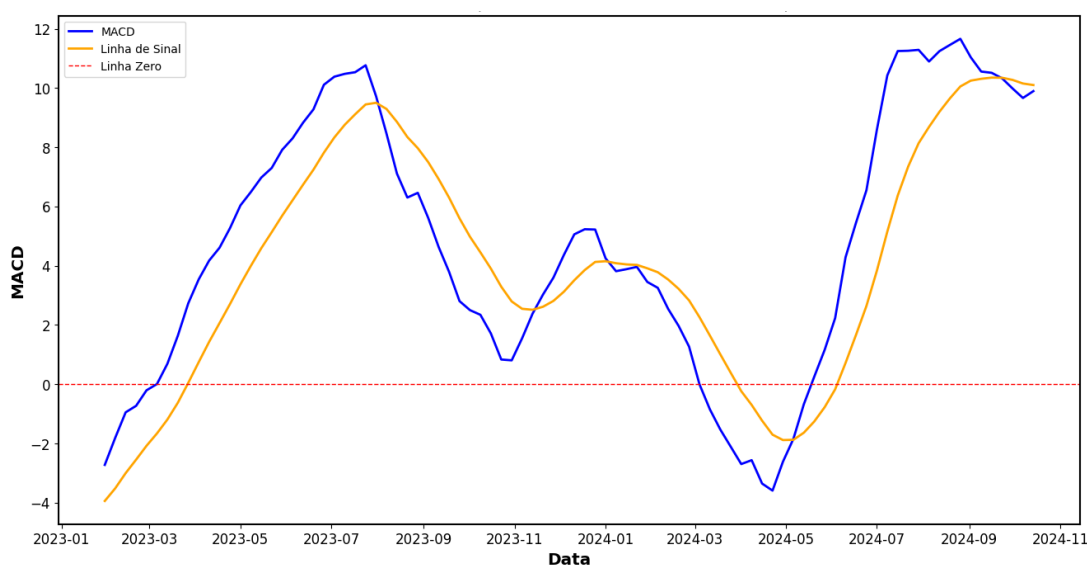
$$\text{MACD} = \text{EMA}_{12} - \text{EMA}_{26} \quad (4.3)$$

$$\text{Linha de Sinal} = \text{EMA}_9(\text{MACD}) \quad (4.4)$$

Na prática, esses cálculos permitem que os *traders* identifiquem pontos em que a linha MACD cruza a linha de sinal, o que pode indicar potenciais oportunidades de compra ou venda. Segundo Appel (2005), "o MACD é uma ferramenta poderosa para detectar mudanças de tendência antes que estas se consolidem". Estudos também indicam que o MACD pode ser eficaz em mercados de forte tendência, embora seja menos útil em mercados laterais (MURPHY, 1999).

A Figura 2 ilustra a aplicação do MACD no gráfico da AAPL, destacando os cruzamentos entre a linha MACD e a linha de sinal, que podem ser usados para interpretar sinais de compra ou venda.

Figura 2 – MACD no gráfico de AAPL



Fonte: Elaborada pelo autor

4.3.3 RSI

O RSI é um indicador de momento que mede a velocidade e a mudança dos movimentos de preço. Ele varia de 0 a 100, com leituras acima de 70 indicando uma condição de sobrecompra e abaixo de 30 indicando uma condição de sobrevenda (WILDER, 1978). O RSI é amplamente utilizado para identificar pontos de reversão no mercado, com base na hipótese de que movimentos extremos de preço são insustentáveis a longo prazo. A fórmula do RSI, apresentada na Equação (4.5), é dada por:

$$RSI = 100 - \frac{100}{1 + \frac{\text{Média dos Ganhos}}{\text{Média das Perdas}}} \quad (4.5)$$

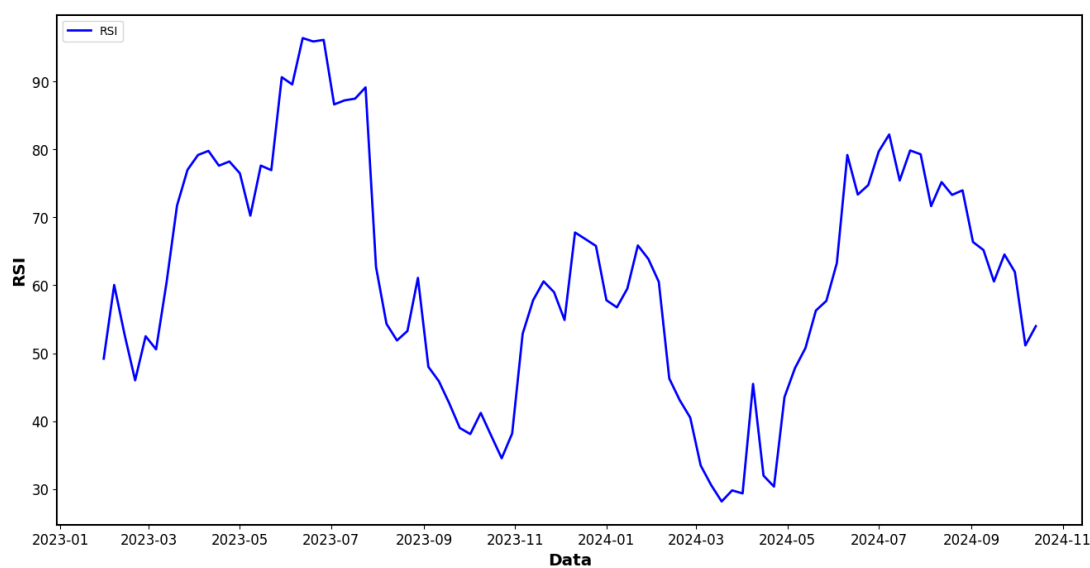
Sendo que:

- Média dos Ganhos: média dos movimentos positivos no período de cálculo.
- Média das Perdas: média dos movimentos negativos no mesmo período.

Conforme Wilder (1978) descreve, "o RSI pode ser uma ferramenta poderosa para detectar a força relativa de um ativo e suas possíveis reversões, especialmente em mercados voláteis". Esse indicador é particularmente útil para avaliar se um ativo está em condições extremas de sobrecompra ou sobrevenda.

A Figura 3 apresenta o RSI aplicado ao gráfico da AAPL, destacando os momentos em que o indicador atinge valores extremos, sugerindo potenciais reversões no preço do ativo.

Figura 3 – RSI no gráfico de AAPL



Fonte: Elaborada pelo autor

4.3.4 ATR

O ATR mede a volatilidade de um ativo, mostrando a amplitude média do preço em um determinado período de tempo. Esse indicador financeiro não indica a direção do preço, mas sim a volatilidade, sendo útil para definir *stop-loss* (perda máxima aceitável em uma operação) ou identificar momentos de maior incerteza no mercado (WILDER, 1978). A fórmula do *True Range (TR)*, que é a base para o cálculo do ATR, é apresentada na Equação (4.6):

$$\text{True Range} = \max \left\{ \begin{array}{l} \text{Máximo} - \text{Mínimo}, \\ |\text{Máximo} - \text{Fechamento Anterior}|, \\ |\text{Fechamento Anterior} - \text{Mínimo}| \end{array} \right\} \quad (4.6)$$

O *ATR* é então calculado como a média do *True Range* ao longo de n períodos, conforme a Equação (4.7):

$$\text{ATR} = \frac{\sum_{i=1}^n \text{True Range}_i}{n} \quad (4.7)$$

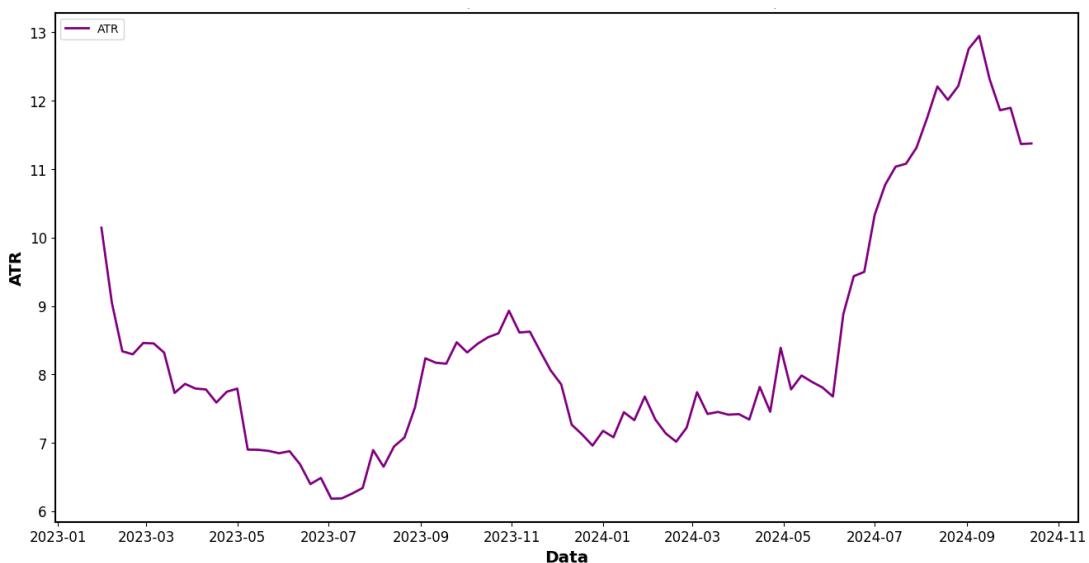
Onde:

- Máximo: maior preço do ativo no período.
- Mínimo: menor preço do ativo no período.
- Fechamento Anterior: preço de fechamento do período anterior.
- n : número de períodos considerados.

Segundo Wilder (1978), "o ATR é especialmente eficaz para medir a volatilidade em mercados instáveis, ajudando os *traders* a ajustar suas estratégias de acordo com a movimentação do mercado".

A Figura 4 demonstra o ATR aplicado ao gráfico da AAPL, destacando as variações na volatilidade do ativo ao longo do tempo, o que pode ser útil para ajustar níveis de *stop-loss* ou identificar períodos de maior incerteza no mercado.

Figura 4 – ATR no gráfico de AAPL



Fonte: Elaborada pelo autor

4.3.5 Bandas de Bollinger

As Bandas de Bollinger são um indicador de volatilidade amplamente utilizado na análise técnica, criado por John Bollinger nos anos 1980. Elas consistem em três linhas principais que ajudam a medir o grau de volatilidade de um ativo ao longo do tempo. A linha central é uma SMA, geralmente com período de 20, enquanto as bandas superior e inferior são calculadas com base em desvios padrão da média móvel, refletindo a amplitude de flutuação dos preços em relação à média (BOLLINGER, 2002; MURPHY, 1999). As fórmulas para as Bandas Superior e Inferior são apresentadas nas Equações (4.8) e (4.9), com $k = 2$, um valor comumente utilizado.

$$\text{Banda Superior} = \text{SMA} + (k \cdot \text{Desvio Padrão}) \quad (4.8)$$

$$\text{Banda Inferior} = \text{SMA} - (k \cdot \text{Desvio Padrão}) \quad (4.9)$$

Onde:

- SMA: média móvel simples, geralmente calculada com 20 períodos.
- Desvio Padrão: medida da dispersão dos preços em relação à média no mesmo período;
- k : fator multiplicador, que define a amplitude das bandas em relação à média.

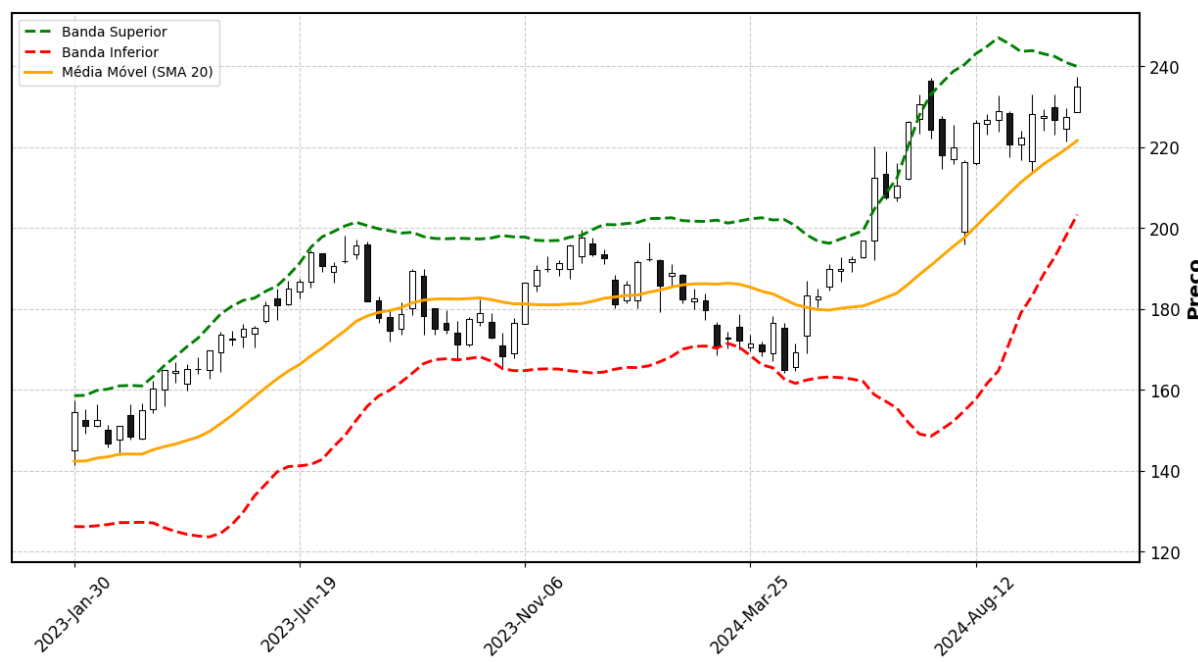
Segundo Bollinger (2002), as Bandas de Bollinger são especialmente úteis para identificar períodos de alta e baixa volatilidade, fornecendo sinais potenciais de compra e venda. Quando os preços se aproximam da Banda Superior, o ativo pode estar sobrecomprado, sugerindo uma possível reversão de tendência. Da mesma forma, quando os preços caem para perto da Banda Inferior, pode ser um sinal de sobrevenda, indicando uma possível oportunidade de compra (BOLLINGER, 2002).

Murphy (1999) reforça que, ao contrário de indicadores que apenas acompanham tendências, as Bandas de Bollinger são reativas, ajustando-se automaticamente às mudanças na volatilidade do mercado. Elas podem ser usadas em conjunto com outros indicadores para confirmar tendências e filtrar sinais de entrada e saída em momentos de maior ou menor volatilidade.

Além disso, o estreitamento das Bandas de Bollinger — uma situação chamada de "*squeeze*" — é considerado um sinal de que a volatilidade está em níveis mínimos e pode preceder movimentos de preço mais intensos. Estudos indicam que essa contração das bandas precede, muitas vezes, grandes mudanças nos preços, tornando-se um elemento essencial na análise preditiva para *traders* que buscam antecipar movimentos de mercado significativos (BOLLINGER, 2002; PRING, 2014).

A Figura 5 apresenta as Bandas de Bollinger aplicadas ao gráfico da AAPL, destacando interações com as bandas superior e inferior que podem indicar sinais de compra ou venda.

Figura 5 – Bandas de Bollinger no gráfico de AAPL



Fonte: Elaborada pelo autor

4.3.6 Estocástico

O Estocástico é um indicador de momento desenvolvido por George Lane, que mede a relação entre o preço de fechamento de um ativo e sua faixa de preço em um determinado período. Ele é amplamente utilizado para identificar condições de sobrecompra e sobrevenda, oferecendo informações sobre possíveis pontos de reversão no mercado. O Estocástico varia entre 0 e 100, onde valores acima de 80 indicam uma condição de sobrecompra e valores abaixo de 20 indicam uma condição de sobrevenda (LANE, 1984). As fórmulas para o cálculo do Estocástico envolvem duas linhas principais: %K e %D.

A linha %K, apresentada na Equação (4.10), mede a posição atual do preço de fechamento em relação à faixa de preços do período:

$$\%K = \frac{\text{Preço de Fechamento Atual} - \text{Mínimo do Período}}{\text{Máximo do Período} - \text{Mínimo do Período}} \times 100 \quad (4.10)$$

A linha %D, apresentada na Equação (4.11), é uma SMA da linha %K calculada ao longo de n períodos:

$$\%D = \text{SMA}_n(\%K) \quad (4.11)$$

Neste trabalho foi utilizado $n = 3$ períodos para o cálculo da %D, o que é uma prática comum na aplicação do Estocástico.

Onde:

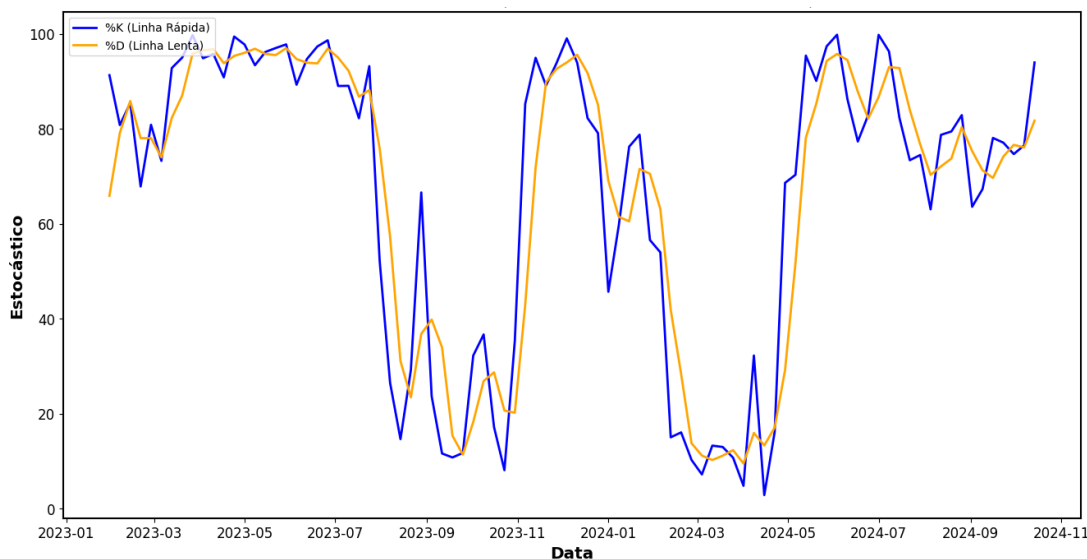
- Preço de Fechamento Atual: preço de fechamento do ativo no período mais recente;
- Máximo do Período: maior preço alcançado durante o período de análise;
- Mínimo do Período: menor preço alcançado durante o período de análise;
- %D: suavização da linha %K com uma média móvel simples de n períodos;
- n : número de períodos considerados para calcular a média móvel simples.

Quando a linha %K cruza acima da linha %D em regiões de sobrevenda (abaixo de 20), isso pode indicar uma oportunidade de compra. Por outro lado, quando a %K cruza abaixo da %D em regiões de sobrecompra (acima de 80), isso pode sugerir uma oportunidade de venda.

De acordo com Lane (1984), "o Estocástico não segue o preço ou o volume, mas a velocidade ou o momento do preço. À medida que o momento muda de direção, geralmente precede mudanças nos preços, tornando-o uma ferramenta eficaz para a detecção de reversões de tendência".

A Figura 6 apresenta o Estocástico aplicado ao gráfico da AAPL, destacando as regiões de sobrecompra e sobrevenda, bem como os cruzamentos entre as linhas %K e %D, que podem ser usados para identificar potenciais sinais de compra ou venda.

Figura 6 – Estocástico no gráfico de AAPL



Fonte: Elaborada pelo autor

4.4 *Random Forest*

A RF, ou floresta aleatória, é um modelo de aprendizado de máquina que combina diversas árvores de decisão para gerar previsões mais robustas e precisas. Esse modelo, introduzido por Breiman (2001), baseia-se na premissa de que a agregação de múltiplos modelos simples pode reduzir o erro geral de predição. No contexto da RF, as árvores de decisão individuais são construídas com variações nos dados e nas variáveis selecionadas, o que garante diversidade no conjunto.

Cada árvore de decisão é formada a partir de um processo de construção hierárquico, em que os nós internos representam condições baseadas em variáveis preditoras, e os ramos indicam os resultados dessas condições. Por exemplo, uma divisão pode ser baseada na regra " $RSI > 50$ ", onde o nó verifica se o valor do *Relative Strength Index* (RSI) de um ativo é maior que 50. Esse processo é repetido até que critérios de parada, como um número mínimo de observações em cada folha ou um limite de profundidade, sejam atingidos (LEVINE; ROSS, 2021).

O funcionamento do RF pode ser descrito em três etapas principais, combinando aspectos técnicos e exemplos práticos:

1. **Amostragem aleatória (*Bootstrap Sampling*):** Cada árvore é treinada com uma amostra gerada aleatoriamente a partir do conjunto de dados original. Esse método, conhecido como *bootstrap sampling*, permite que algumas observações sejam repetidas enquanto outras são excluídas (MAIMON; ROKACH, 2014). Por

exemplo, uma amostra pode conter linha 1, linha 2, linha 1, linha 4, enquanto outra pode conter linha 3, linha 1, linha 2, linha 4.

2. **Seleção aleatória de variáveis:** Em cada nó de decisão, apenas um subconjunto aleatório das variáveis é considerado para determinar a melhor divisão. Essa estratégia reduz a correlação entre as árvores, contribuindo para a diversidade no conjunto (HASAN; SANTOS; KUMAR, 2023). Por exemplo, uma árvore pode iniciar avaliando a variável "Variação (%) > 0", enquanto outra pode começar com "RSI > 50".
3. **Treinamento das árvores e combinação das previsões:** Com a amostra e o subconjunto de variáveis definidos, a árvore é treinada para dividir os dados em subconjuntos cada vez mais homogêneos em relação à variável alvo, até atingir os critérios de parada. Após o treinamento, cada árvore fornece uma previsão individual. Para problemas de classificação, o RF aplica o princípio de votação majoritária. Suponha que, de 100 árvores, 70 votem na classe 1 (alta) e 30 votem na classe -1 (baixa); nesse caso, a previsão final será a classe 1.

4.4.1 Exemplo de Funcionamento

Para ilustrar como uma RF funciona, considere um problema de classificação no qual o objetivo é prever se o preço de uma ação subirá (classe 1) ou cairá (classe -1). Os dados hipotéticos utilizados para este exemplo estão apresentados na Tabela 1.

Tabela 1 – Dados hipotéticos para treinamento de uma Random Forest

Variação (%)	Volume	RSI	Classe
1.2	1000	55	1
-0.8	1500	45	-1
0.5	800	60	1
-1.1	1200	40	-1

Fonte: Elaborada pelo autor

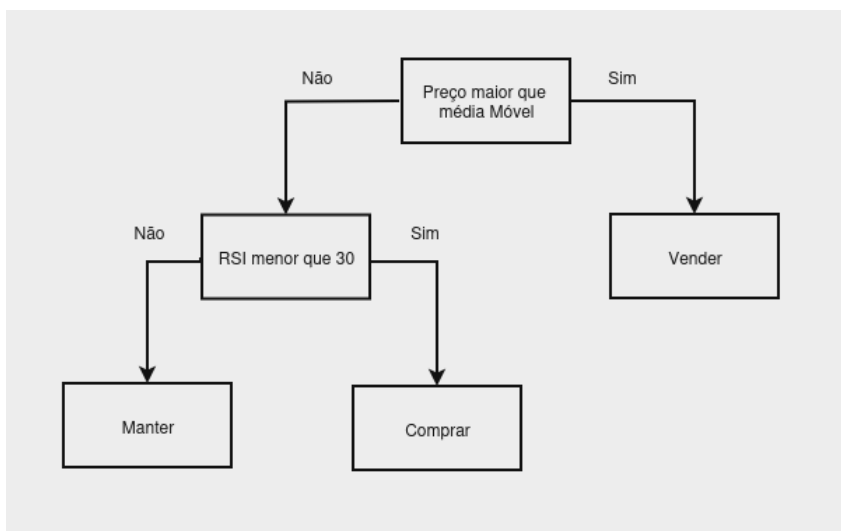
Com base nesses dados, o modelo cria várias árvores de decisão, cada uma analisando diferentes combinações de amostras e variáveis. Após o treinamento, as previsões das árvores são combinadas para formar a decisão final. Por exemplo, se a maioria das árvores indicar que o preço subirá, a classe final será definida como 1.

4.4.1.1 Árvore de Decisão Final

Considerando que o modelo foi treinado utilizando dados financeiros relevantes, incluindo indicadores técnicos como RSI, médias móveis e volumes de negociação,

uma possível árvore resultante de um modelo RF é apresentada na Figura 7. Essa árvore exemplifica como o modelo pode ser estruturado para prever se um ativo se encontra em posição de sobrecompra, sobrevenda ou manutenção no mercado, com base nas condições analisadas. Cada nó representa uma decisão baseada em uma métrica específica, enquanto as folhas indicam a previsão final associada ao ramo percorrido.

Figura 7 – Exemplo de árvore de decisão final em uma RF



Fonte: Elaborada pelo autor

4.4.2 *Bootstrap Aggregation*

O modelo de RF se destaca pela utilização de múltiplas árvores de decisão treinadas de forma independente. Essa construção usa uma técnica chamada *bootstrap aggregation*, ou *bagging*, onde cada árvore é treinada em subconjuntos aleatórios dos dados de treino e em variáveis selecionadas aleatoriamente. Esse processo reduz o *overfitting*, permitindo que o modelo seja mais sólido e generalizável (BREIMAN, 2001). Em uma RF, cada árvore contribui com uma previsão, e o resultado final é obtido por meio de uma média das previsões (em problemas de regressão) ou pela moda (em problemas de classificação).

Ainda usando o exemplo anteriormente apresentado, suponhamos que esses são os dados iniciais:

Tabela 2 – Dados simulados para *Bootstrap Aggregation* com sinais de compra, venda e manutenção

	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
Preço	100	105	102	98	97	99	101	103	100	96
Média Móvel	102	102	101	100	99	100	101	102	101	100
RSI	35	45	32	28	25	30	40	38	33	20
Sinal	0	-1	0	1	1	0	-1	-1	0	1

Fonte: Elaborada pelo autor

Os dados serão organizados em três conjuntos principais: treino ($t=1, \dots, 6$), teste ($t=7, 8$), e validação ($t=9, 10$). O algoritmo aplicará a técnica de *bootstrap* ao conjunto de treino, extraíndo aleatoriamente três amostras *bootstrap*. Esse processo resulta em um total de 15 dados, em vez dos 8 dados iniciais, permitindo que as amostras tenham elementos repetidos.

Tabela 3 – Amostras bootstrap dos dados de treino para o modelo RF

Amostra 1				
t	Preço	Média Móvel	RSI	Sinal
1	101.2	100.5	45	0
3	98.6	99.2	30	1
4	97.5	98.9	28	1
5	103.2	102.0	52	-1
8	104.3	103.5	54	-1
6	100.0	99.6	40	0

Amostra 2				
t	Preço	Média Móvel	RSI	Sinal
2	102.4	100.8	50	-1
4	97.5	98.9	28	1
7	99.1	98.5	33	0
1	101.2	100.5	45	0
3	98.6	99.2	30	1
8	104.3	103.5	54	-1

Amostra 3				
t	Preço	Média Móvel	RSI	Sinal
5	103.2	102.0	52	1
7	99.1	98.5	33	0
2	102.4	100.8	50	-1
4	97.5	98.9	28	1
6	100.0	99.6	40	0
1	101.2	100.5	45	0

Fonte: Elaborado pelo autor

Cada amostra *bootstrap* é usada para treinar uma árvore de decisão de forma independente, o que permite que cada árvore faça previsões ligeiramente diferentes devido à aleatoriedade na seleção dos dados. Após o treinamento, o modelo aplica o processo de "*aggregation*" para combinar as previsões das árvores. No caso de classificação, cada árvore "vota" em um dos sinais (por exemplo, compra, venda ou manutenção), e o sinal com a maioria dos votos é escolhido como a previsão final do modelo.

4.4.3 Modelos *Ensemble*

Modelos *ensemble* são técnicas em *machine learning* que combinam previsões de diversos modelos individuais para formar um modelo único e mais confiável, reduzindo os erros específicos de cada um e melhorando a capacidade preditiva geral (MAIMON; ROKACH, 2014; HASTIE; TIBSHIRANI; FRIEDMAN, 2009). O princípio por trás dos *ensembles* é que, ao combinar diferentes perspectivas (ou "modelos base"), é possível captar padrões e complexidades que seriam difíceis de identificar com um único modelo.

Um exemplo popular de modelo *ensemble* é justamente o RF, que utiliza várias árvores de decisão para gerar previsões mais confiáveis. Cada árvore no RF faz uma previsão independente, e o resultado final é a agregação das respostas individuais — geralmente por meio de votação majoritária, no caso de classificação, ou da média, em problemas de regressão (BREIMAN, 2001). Essa estrutura torna o modelo mais resistente a erros isolados, já que cada árvore pode captar aspectos distintos dos dados. Por exemplo, ao prever se uma ação subirá ou cairá, cada árvore de decisão fornece uma "opinião", e o modelo final adota a direção indicada pela maioria, conferindo maior precisão à previsão.

O uso de múltiplos modelos base aumenta a diversidade e reduz o sobreajuste, uma limitação comum em árvores de decisão isoladas. Isso se traduz em uma capacidade superior de generalização, o que faz do RF um dos modelos *ensemble* mais aplicados em áreas complexas, como a previsão de preços no mercado financeiro (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

4.5 Métricas Utilizadas

A avaliação de estratégias financeiras e modelos de previsão no mercado depende de métricas quantitativas que permitem compreender o desempenho em diferentes perspectivas. Entre as principais métricas utilizadas neste trabalho estão o *Sharpe Ratio*, a acurácia, o *F1 Score*, o retorno percentual e o *drawdown*.

4.5.1 *Sharpe Ratio*

O *Sharpe Ratio* (SR) é amplamente utilizado para medir o desempenho ajustado ao risco de uma estratégia ou investimento. Ele é definido como a razão entre o retorno excedente em relação a um ativo livre de risco e o desvio padrão dos retornos (Equação 4.12)

$$SR = \frac{R_p - R_f}{\sigma_p} \quad (4.12)$$

Sendo que:

- R_p : Retorno médio da estratégia ou portfólio.
- R_f : Retorno do ativo livre de risco.
- σ_p : Desvio padrão dos retornos da estratégia.

Em geral, valores de SR superiores a 1,0 indicam que a estratégia possui um bom retorno ajustado ao risco, enquanto valores entre 1,5 e 2,0 são considerados excelentes (LEVINE; ROSS, 2021). No entanto, estratégias com retornos assimétricos podem não ser plenamente avaliadas por essa métrica. Por exemplo, uma estratégia que gera lucros altos esporádicos e pequenas perdas frequentes pode apresentar um SR inferior a 1,0, mas ainda ser altamente lucrativa no longo prazo. Isso ocorre porque o SR mede a constância do retorno ajustado ao risco e pode ser enviesado por distribuições de retorno que não são simétricas.

4.5.2 *Acurácia*

A acurácia avalia a proporção de previsões corretas em relação ao total de previsões realizadas. Sua fórmula é apresentada na Equação 4.13:

$$\text{Acurácia} = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.13)$$

Onde:

- TP: Verdadeiros positivos.
- TN: Verdadeiros negativos.
- FP: Falsos positivos.
- FN: Falsos negativos.

Valores elevados de acurácia indicam um bom desempenho geral do modelo. Contudo, em cenários com classes desbalanceadas, essa métrica pode ser enganosa, já que a maioria das previsões corretas pode se concentrar na classe majoritária. Por isso, métricas como o *F1 Score* são mais indicadas em contextos desse tipo.

4.5.3 *F1 Score*

O *F1 Score* é uma métrica harmônica que combina precisão (*precision*) e *recall*, oferecendo uma visão equilibrada do desempenho do modelo (Equação 4.14):

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.14)$$

Onde:

- *Precision*: Proporção de previsões corretas entre todas as previsões positivas.
- *Recall*: Proporção de verdadeiros positivos em relação ao total de casos positivos reais.

Valores próximos de 1,0 no *F1 Score* indicam que o modelo é eficiente tanto na captura de padrões relevantes quanto na redução de alarmes falsos (HASAN; SANTOS; KUMAR, 2023).

4.5.4 Retorno Percentual

O retorno percentual mede o desempenho financeiro de uma estratégia em relação ao capital inicial. Sua fórmula está descrita na Equação 4.15:

$$\text{Retorno (\%)} = \frac{\text{Valor Final} - \text{Valor Inicial}}{\text{Valor Inicial}} \times 100 \quad (4.15)$$

Essa métrica é essencial para avaliar a viabilidade de uma estratégia no mercado financeiro, mensurando os ganhos ou perdas acumulados em termos percentuais.

4.5.5 *Drawdown*

O *drawdown* representa a maior perda acumulada de uma estratégia em um período específico, sendo uma métrica importante para medir o risco. Ele é calculado como a diferença percentual entre o pico de capital e o ponto mais baixo subsequente (Equação 4.16):

$$\text{Drawdown} = \frac{\text{Pico de Capital} - \text{Ponto Mínimo}}{\text{Pico de Capital}} \times 100 \quad (4.16)$$

Valores elevados de *drawdown* indicam que a estratégia pode enfrentar perdas significativas durante períodos de baixa, o que é um fator de risco relevante para investidores e gestores de portfólio.

4.6 Escolha do Mercado Americano

O mercado financeiro dos Estados Unidos é amplamente reconhecido como o mais influente do mundo, desempenhando um papel central na economia global. Essa preeminência deriva de sua diversificação econômica e capacidade de inovação tecnológica, além de uma resiliência histórica que lhe permitiu superar crises significativas, como o *crash* de 1929, o “*Black Monday*” de 1987 e a crise financeira de 2008. Em cada uma dessas situações, o mercado americano mostrou uma capacidade impressionante de recuperação e crescimento sustentado, características que atraem investidores globais em busca de estabilidade e oportunidades de longo prazo.

Para investidores, o mercado dos EUA não representa apenas um espaço de diversificação setorial, mas também um ambiente de segurança institucional, fortalecido por regulamentações consistentes e políticas econômicas que incentivam o crescimento. A visão positiva sobre o mercado americano é respaldada por especialistas e por investidores renomados como Warren Buffett, que destacou em várias ocasiões sua confiança na economia dos EUA, enfatizando a frase “nunca aposte contra a América.” Esta visão, que remete ao constante progresso econômico do país, reflete-se na prática de investidores que veem nas crises, muitas vezes, uma oportunidade para adquirir ativos a preços mais baixos, com o potencial de lucrar durante as fases de recuperação econômica (BUFFETT; CUNNINGHAM, 2008).

4.6.1 Critérios de Escolha das Ações

As ações selecionadas para o portfólio foram escolhidas com base em uma análise fundamentalista detalhada, considerando fatores quantitativos e qualitativos que demonstram a solidez financeira e o potencial de longo prazo de cada empresa.

4.6.1.1 Indicadores Financeiros Chave

Os indicadores financeiros foram essenciais para a escolha das empresas, proporcionando uma análise quantitativa detalhada:

- **ROI (Retorno sobre Investimento):** Eficiência da empresa em gerar lucros a partir de seus recursos.
- **P/L (Preço sobre Lucro):** Identifica empresas sobrevalorizadas ou subvalorizadas.

- **Lucro Líquido:** Crescimento constante do lucro foi um sinal de resiliência e boa gestão.
- **Dívida/Patrimônio Líquido:** Empresas com baixa relação de dívida sobre patrimônio foram priorizadas para maior segurança.
- **Margem Líquida:** Indica a eficiência das empresas em converter receita em lucro.
- **Dividend Yield:** Retorno das empresas em dividendos.
- **Crescimento da Receita:** Capacidade de expansão ao longo do tempo.

4.6.1.2 Relevância da Marca

A força da marca foi um critério qualitativo importante, especialmente para empresas como Coca-Cola (KO) e Procter & Gamble (PG). Essas marcas possuem reconhecimento global e ampla aceitação entre os consumidores, o que proporciona estabilidade mesmo durante crises econômicas.

4.6.1.3 Diversificação Setorial

A diversificação setorial foi utilizada para garantir a redução de riscos e a exposição equilibrada a diferentes setores da economia. Essa estratégia permite mitigar perdas de setores específicos e maximizar as oportunidades em segmentos em expansão.

4.6.2 Análise por Setor

Os setores abordados no portfólio incluem uma ampla gama de indústrias, cada uma representada por empresas com fundamentos sólidos:

- **Tecnologia:** Com empresas como Apple (AAPL) e Microsoft (MSFT), o setor é impulsionado pela inovação e por ecossistemas integrados de produtos, atraindo consumidores globais e sustentando o crescimento.
- **Financeiro:** JPMorgan Chase (JPM) e Bank of America (BAC) são líderes, com gestão sólida e resiliência em crises, contando com forte presença global e balanços patrimoniais robustos.
- **Saúde:** Johnson & Johnson (JNJ) e Merck (MRK) destacam-se pela inovação em biotecnologia e histórico de dividendos, além de manterem portfólios diversificados.

- **Consumo Básico:** Procter & Gamble (PG) e Coca-Cola (KO) atendem a demandas por produtos essenciais, demonstrando estabilidade e crescimento sustentável mesmo em períodos de recessão.
- **Energia:** Chevron (CVX) e Exxon Mobil (XOM) representam o setor energético, investindo tanto em petróleo quanto em fontes renováveis, equilibrando uma gestão conservadora com lucratividade.
- **Indústria:** Boeing (BA) e 3M (MMM) são relevantes, com a Boeing voltada para a aviação comercial e de defesa, e a 3M para fabricação diversificada.
- **Utilidades:** NextEra Energy (NEE) e Duke Energy (DUK) lideram em energias renováveis, com foco em energia eólica e solar, promovendo previsibilidade de receita e inovação em energia limpa.
- **Imobiliário:** Prologis (PLD) e American Tower (AMT) se beneficiam do crescimento do comércio eletrônico e da demanda por infraestrutura de telecomunicações, atendendo a logística e distribuição.

5 Implementação

Para a implementação do modelo, os ativos mencionados anteriormente foram analisados em um intervalo semanal. A escolha desse período visa minimizar os ruídos e movimentos erráticos do mercado, pois intervalos maiores tendem a ser menos suscetíveis a manipulações, envolvendo volumes de capital mais expressivos. Como argumenta o investidor Peter Lynch, o desempenho de curto prazo no mercado é muitas vezes afetado por fatores emocionais, enquanto no longo prazo são os fundamentos que exercem maior influência sobre o valor dos ativos (LYNCH; ROTHCHILD, 1989).

Além disso, os dados obtidos da biblioteca `yfinance`¹ incluem apenas os preços de abertura, fechamento, máxima e mínima. Foi realizado um tratamento adicional desses dados para calcular e incluir indicadores técnicos e estratégias no *dataframe* original, enriquecendo a análise com informações relevantes para a tomada de decisão.

5.1 Estrutura e Configuração do Modelo

Muitos estudos na área de previsão de mercado financeiro seguem a filosofia de *forecasting*, que busca prever o preço futuro de um ativo ao longo de uma série temporal (HAMILTON, 1994; BROCKWELL; DAVIS, 2016). No entanto, essa abordagem apresenta grandes desafios, pois o preço de um ativo é influenciado por uma infinidade de variáveis que muitas vezes não estão diretamente relacionadas ao ativo em questão. Fatores externos, como decisões do Banco Central Americano, sentimento do mercado, eventos geopolíticos, guerras e pandemias, podem alterar drasticamente a direção dos preços, tornando a previsão precisa uma tarefa extremamente complexa.

Nesse contexto, o modelo RF desenvolvido neste trabalho toma como base conceitos amplamente aceitos por investidores e *traders* renomados. Entre esses conceitos estão a identificação de tendências, reversões e a análise de padrões de mercado. Princípios estabelecidos por figuras como Charles Dow, Jesse Livermore e outros serviram de fundamentação para a escolha dos indicadores técnicos que foram utilizados para gerar sinais de compra e venda.

5.1.1 Tratamento dos Dados

Para o estudo, os dados obtidos da API do `yfinance` incluem os valores de abertura, fechamento, máxima e mínima de um ativo em intervalos de tempo específicos. Os dados abrangem o período de janeiro de 2000 até outubro de 2024, proporcionando

¹ Disponível em: <<https://pypi.org/project/yfinance/>>. Acesso em: 30 out. 2024.

uma base histórica sólida para análise, englobando períodos de euforia e depressão. Esse período foi dividido em dois conjuntos: 80% dos dados foram utilizados para treinamento do modelo, enquanto os 20% restantes foram reservados para teste.

Com base nesses dados iniciais, calcularam-se valores dos indicadores financeiros previamente apresentados para enriquecer o conjunto de dados, permitindo uma análise mais detalhada do comportamento do ativo. Esses indicadores foram combinados aos sinais de pressão de compra e venda no mercado, onde foi atribuído:

- Sinal de 1 para indicar pressão compradora.
- Sinal de -1 para pressão vendedora.
- Sinal de 0 quando não há uma força predominante definida.

A Tabela 4 apresenta um exemplo didático do conjunto de dados gerado, incluindo os valores de abertura, fechamento, máxima e mínima, bem como o respectivo sinal de pressão de mercado atribuído a cada linha.

Tabela 4 – Exemplo didático de dados com sinal de pressão de mercado

Data	Abertura	Fechamento	Máxima	Mínima	Sinal
16/10/2024	145.10	150.20	151.50	144.30	1
23/10/2024	150.00	149.00	152.00	148.50	0
30/10/2024	149.50	146.80	150.00	146.00	-1

Fonte: Elaborada pelo autor

5.1.2 Otimização e Validação dos Modelos

A validação cruzada é uma técnica essencial para avaliar o desempenho de modelos de aprendizado de máquina e garantir sua capacidade de generalização para dados novos. Em particular, a validação cruzada em *k-fold* é amplamente utilizada, onde o conjunto de dados é dividido em *k* partes iguais. Em cada iteração, uma parte é usada para validação e as outras para treino, repetindo o processo *k* vezes até que todas as partes tenham servido como conjunto de validação. Esse método resulta em uma média das métricas de desempenho, proporcionando uma avaliação robusta da eficácia do modelo (WONG; YEH; LIN, 2015).

A otimização de hiperparâmetros é frequentemente realizada junto com a validação cruzada para identificar a combinação de parâmetros que maximiza o desempenho do modelo. Nesse contexto, técnicas como o *Grid Search*, em conjunto com o *GridSearchCV* da biblioteca Scikit-Learn², são utilizadas para realizar uma busca exaustiva sobre

² Disponível em: <<https://scikit-learn.org/stable/>>. Acesso em: 28 out. 2024.

o espaço de parâmetros, visando minimizar a possibilidade de *overfitting* e garantir a generalização dos resultados.

5.2 Performance

Inicialmente, o modelo se baseou apenas em médias móveis, uma de 7 períodos e outra de 21, para prever possíveis entradas e saídas. Tomando a empresa NextEra Energy (NEE) como exemplo, os resultados da estratégia são apresentados na Tabela 5.

Tabela 5 – Métricas de desempenho da estratégia

Métrica	Valor
Retorno <i>Buy and Hold</i> (%)	99.23
Retorno estratégia (%)	21.59
<i>Drawdown</i> máximo da estratégia (%)	19.19
<i>Drawdown</i> máximo de <i>Buy and Hold</i> (%)	47.17
Acurácia (%)	68.07

Fonte: Elaborada pelo autor.

- **Retorno *Buy and Hold* (%)**: Valorização do ativo ao longo do período, sem operações intermediárias.
- **Retorno estratégia (%)**: Valorização obtida aplicando as regras do modelo.
- ***Drawdown* máximo da estratégia (%)**: Maior perda da estratégia em relação ao pico de capital.
- ***Drawdown* máximo de *Buy and Hold* (%)**: Maior perda percentual do *Buy and Hold*.
- **Acurácia (%)**: Percentual de operações da estratégia que resultaram em lucro.

Apesar do modelo de *machine learning* ter apresentado um retorno positivo de 21%, tal desempenho mostrou-se inferior à estratégia de *buy and hold*, tradicionalmente utilizada como *benchmark* em análises financeiras. No entanto, o modelo demonstrou eficácia ao reduzir o *drawdown* em 50% quando comparado à estratégia de *buy and hold*, evidenciando uma melhor gestão do risco. Contudo, a acurácia, ou seja, a capacidade do modelo de classificar corretamente os sinais, foi relativamente baixa, alcançando cerca de 68%. Tal desempenho sinaliza a necessidade de aprimoramentos adicionais no modelo para otimizar a sua capacidade preditiva.

Com o intuito de melhorar os resultados, foram introduzidas ao modelo métricas e técnicas amplamente utilizadas na análise de mercado, como o RSI, MACD, ATR e

as Bandas de Bollinger. Essas métricas possuem reconhecida relevância na análise técnica financeira por contribuírem para a avaliação da força das tendências, detecção de reversões e análise da volatilidade do mercado. A inclusão de indicadores técnicos, como o RSI e MACD, pode aumentar a capacidade dos modelos baseados em aprendizado de máquina de identificar padrões mais sutis e, conseqüentemente, fornecer sinais de compra e venda com maior assertividade (CHAN, 2013).

Além da introdução dessas métricas, o modelo foi submetido a dois processos importantes: a validação cruzada e a técnica de superamostragem *SMOTE* (*Synthetic Minority Over-sampling Technique*). Ambos os métodos foram aplicados com o objetivo de refinar o desempenho do modelo e lidar com o desbalanceamento das classes no conjunto de dados.

A validação cruzada foi utilizada como técnica de avaliação e generalização do modelo, implementada através do método *GridSearchCV*, que explorou diferentes combinações de hiperparâmetros e selecionou a configuração mais apropriada. Como resultado, foram identificados os seguintes parâmetros ideais, observados na Tabela 6.

Tabela 6 – Parâmetros otimizados do modelo Random Forest

Parâmetro	Valor
Profundidade máxima (<i>max depth</i>)	10
Número mínimo de amostras na folha (<i>min samples leaf</i>)	2
Número mínimo de amostras para divisão (<i>min samples split</i>)	5
Número de árvores (<i>n estimators</i>)	100

Fonte: Elaborado pelo autor

Para tratar o desbalanceamento de classes, usou-se a biblioteca *imbalanced-learn*³ em conjunto com a validação cruzada de três *folds*. Nesse processo, o conjunto de dados foi dividido em três subconjuntos, onde, em cada iteração, um dos *folds* foi utilizado para validação e os demais para treinamento, aplicando a técnica *SMOTE* apenas nos dados de treino. Assim, o modelo foi treinado em um conjunto balanceado e validado em dados que mantinham a distribuição original, permitindo uma avaliação mais realista do desempenho e da capacidade de generalização.

O número de três *folds* foi escolhido empiricamente, buscando otimizar a relação entre o tempo de processamento e a estabilidade das métricas de desempenho.

Após as otimizações, os resultados obtidos podem ser observados na Tabela 7.

³ Disponível em: <https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html>. Acesso em: 30 out. 2024.

Tabela 7 – Indicadores de desempenho para diferentes ações

Ação	Sharpe Ratio	Acurácia	F1 Score
AAPL	0.4	0.94	0.92
JPM	0.42	0.92	0.96
JNJ	0.74	0.93	0.92
PG	0.4	0.96	0.96
CVX	0.36	0.85	0.83
BA	0.72	0.96	0.96
NEE	0.17	0.95	0.93
PLD	0.08	0.96	0.93

Fonte: Elaborada pelo autor

5.2.1 Resultado do Modelo

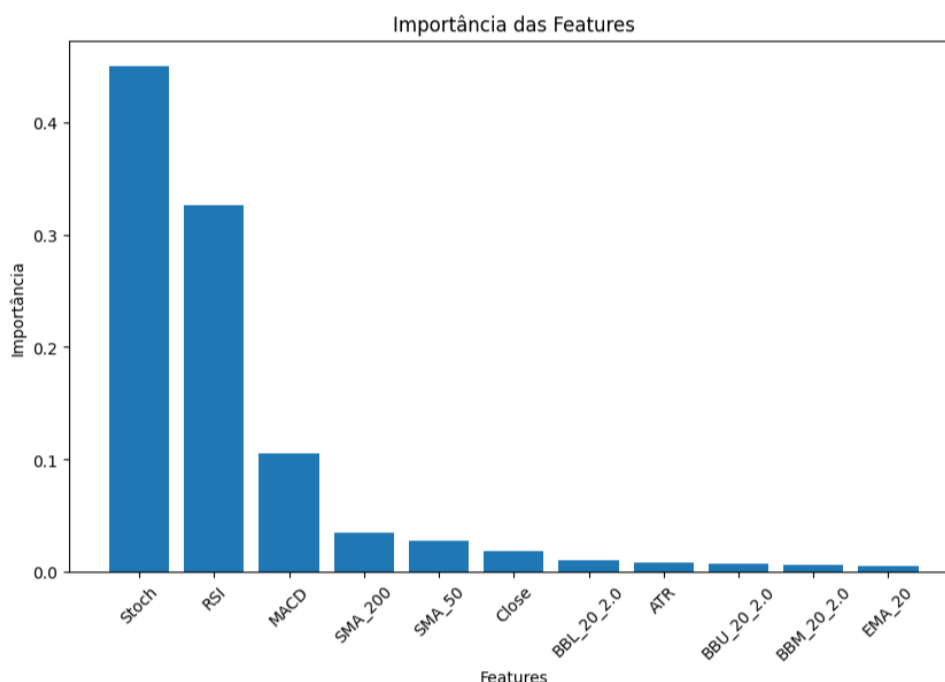
Diversos estudos sugerem que um SR superior a 1,0 é frequentemente considerado ideal para representar um bom retorno ajustado ao risco. Em contextos como fundos de hedge, valores entre 1,5 e 2,0 são vistos como excelentes, indicando desempenho consistente em relação ao risco assumido (MUNSCH, 2019; JONES; SMITH; THOMPSON, 2018). Contudo, para a estratégia específica analisada, um SR entre 0 e 1,0 é aceitável e oferece uma expectativa positiva no longo prazo. Embora o SR seja aparentemente mais baixo, a estratégia tende a gerar lucros consideráveis em operações vencedoras e pequenas perdas em operações perdedoras, o que distorce essa métrica sem prejudicar o desempenho final. Dessa forma, o SR pode não captar totalmente o valor da estratégia devido à sua natureza assimétrica de retorno (MUNSCH, 2019; LEVINE; ROSS, 2021).

A acurácia e o *F1 Score* indicaram resultados bastante positivos para a estratégia. Uma acurácia acima de 90% sugere que o modelo foi eficaz em identificar padrões de pressão compradora e vendedora ao longo do tempo, o que é fundamental em estratégias que buscam capturar movimentos de alta e baixa no mercado.

O *F1 Score* combina a precisão e o *recall*, oferecendo uma visão mais equilibrada do desempenho do modelo ao lidar com falsos positivos e negativos. Valores próximos de 1,0 no *F1 Score* indicam que o modelo mantém uma alta precisão na identificação de sinais de mercado ao mesmo tempo em que minimiza o número de sinais incorretos. Por isso, o *F1 Score* é especialmente útil para avaliar o desempenho em cenários em que é essencial equilibrar ambos os tipos de erros, garantindo que o modelo seja confiável na captura de sinais reais e na redução de alarmes falsos (HASAN; SANTOS; KUMAR, 2023). Esta métrica é particularmente valorizada em mercados financeiros, onde a identificação correta de movimentos raros pode ser um diferencial estratégico significativo (HASAN; SANTOS; KUMAR, 2023).

Por fim, foi gerado um gráfico que apresenta a importância de cada indicador na tomada de decisão para compra e venda. O indicador estocástico destacou-se como o mais determinante, representando 45% da importância total nas decisões. Em seguida, o RSI obteve uma relevância de 32%, reforçando seu papel nas entradas. Os demais indicadores apresentaram importâncias inferiores a 10%, indicando um impacto limitado na definição das decisões de compra e venda.

Figura 8 – Importância dos parâmetros da Random Forest



Fonte: Elaborada pelo autor

5.2.2 Resultado no mercado financeiro

Para a simulação, foi considerado um capital inicial de \$20.000. Esse valor foi escolhido arbitrariamente, visando mitigar o risco de que eventuais perdas na estratégia levassem a uma quebra da conta, situação comum em contas mais frágeis. Em cada sinal apontado pelo modelo, foi investido um percentual de 15% do capital. Dessa forma, caso a estratégia não apresentasse um desempenho satisfatório, o impacto negativo no capital total seria atenuado, evitando uma penalização grave.

O período do teste foi de dezembro de 2019 até novembro de 2024, englobando um cenário de aproximadamente 5 anos.

Tabela 8 – Resultados das operações por ativo

Ativo	Operações	Taxa de Acerto	Média Ganhos	Média Perdas	Drawdown Máximo	Média de Duração (semanas)
AAPL	3	100%	55%	0%	20%	58
MSFT	1	100%	22.9%	0%	3%	37
JPM	4	75%	18.24%	13.1%	6.01%	25
BAC	4	75%	12.08%	6.2%	4.64%	26
JNJ	4	100%	14.46%	0%	7.66%	44
MRK	3	100%	11.67%	0%	4.74%	28
PG	3	100%	11.78%	0%	4.78%	45
KO	2	100%	5.33%	0%	1.48%	54
CVX	4	100%	33.69%	0%	16.57%	33
XOM	3	100%	25.76%	0%	10.07%	21
BA	3	33.33%	41.03%	14.81%	19.34%	41
MMM	3	66.67%	42.02%	39.38%	41.4%	55
NEE	2	50%	46.15%	11.85%	8.76%	65
DUK	3	100%	18.46%	0%	7.33%	32
PLD	3	100%	21.17%	0%	8.31%	35
AMT	3	66.67%	29.09%	10.01%	7.83%	46

Fonte: Elaborada pelo autor

O desempenho da estratégia ao longo do período de testes apresentou resultados promissores. Conforme demonstrado na Tabela 8, a maioria dos ativos obteve uma taxa de acerto de 100%, resultando em um acerto geral de aproximadamente 85%. Isso indica que a estratégia foi bem-sucedida na identificação de oportunidades de compra e venda. Além disso, a média de ganhos foi elevada para vários ativos, destacando-se o ativo AAPL, com uma média de ganhos de 55%, e o ativo CVX, com um ganho médio de 33,69%. Esses resultados sugerem que, quando a estratégia acerta, ela consegue capturar uma valorização significativa.

Em relação ao *drawdown* máximo, ele se manteve relativamente baixo em grande parte dos ativos, o que reflete uma boa gestão de risco da estratégia. Esse ponto é essencial, pois drawdowns elevados podem indicar maior exposição a perdas significativas, e, nesse caso, a estratégia conseguiu limitar essas perdas. Por exemplo, ativos como MSFT e KO tiveram drawdowns de 3% e 1.48%, respectivamente, o que demonstra consistência na contenção de perdas.

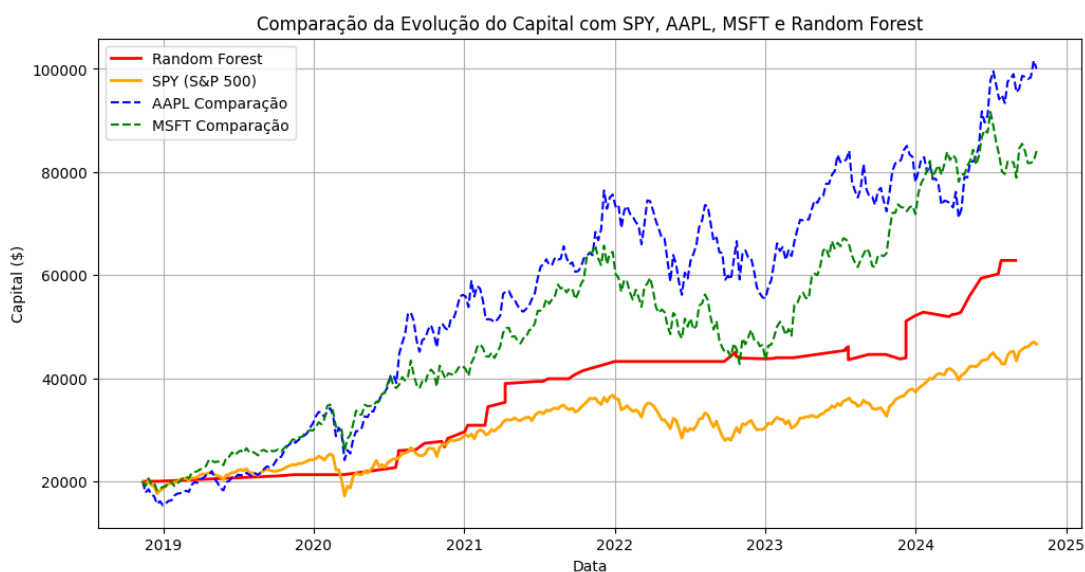
A Tabela 9 e a Figura 9 a seguir representam uma visão geral da performance do modelo, com comparações com o SPY, Apple e Microsoft:

Tabela 9 – Comparação de retorno e *drawdown* máximo entre modelos e ativos

Modelo	Retorno	Drawdown Máximo
RF	214.24%	9.12%
SPY	118.72%	27.22%
AAPL	456.45%	32.35%
MSFT	311.35%	38.11%

Fonte: Elaborada pelo autor

Figura 9 – Gráfico comparando resultados do RF com outros ativos



Fonte: Elaborada pelo autor

A Tabela 9 fornece uma visão clara do desempenho do modelo RF em comparação com o índice de mercado SPY, e as ações da Apple (AAPL) e Microsoft (MSFT). Observa-se que o modelo de RF obteve um retorno expressivo de 214,24%, superando o índice SPY, que apresentou um retorno de 118,72%. Esse resultado indica que a estratégia do modelo foi eficaz em gerar retornos superiores ao índice de referência, destacando seu potencial como uma alternativa aos investimentos tradicionais em ações e fundos de índices.

Embora o modelo tenha ficado atrás das ações da Apple e Microsoft em termos de retorno, que foram de 456,45% e 311,35%, respectivamente, é importante destacar que esses desempenhos são anômalos e improváveis de serem igualados por estratégias de *machine learning* aplicadas de forma genérica. As ações da Apple e Microsoft são conhecidas por suas performances excepcionais nos últimos anos, refletindo o crescimento extraordinário dessas empresas.

Um aspecto crucial em que o modelo de RF se destaca é no controle de risco,

evidenciado por seu *drawdown* máximo de apenas 9,12%, significativamente inferior aos 27,22% do SPY, 32,35% da AAPL e 38,11% da MSFT. Esse *drawdown* reduzido indica que, além de proporcionar retornos atraentes, o modelo oferece uma proteção contra grandes quedas, sendo especialmente útil para investidores que desejam limitar suas perdas durante períodos de volatilidade no mercado. Essa capacidade de minimizar perdas torna o modelo particularmente atrativo, mostrando que ele não só busca maximizar ganhos, mas também prioriza a preservação do capital investido.

6 Arquitetura

Além do estudo focado em modelos para otimizar o desempenho dos investimentos, esta seção aborda a arquitetura desenvolvida para executar os modelos e notificar usuários periodicamente, visando analisar ativos e identificar possíveis recomendações de compra ou venda. A escolha da infraestrutura foi orientada pela busca de uma solução financeiramente acessível, escalável e capaz de atender aos requisitos de eficiência e desempenho dos usuários.

A plataforma escolhida para implementar essa solução foi a *Amazon Web Services* (AWS). A AWS oferece uma gama de serviços em nuvem que permitem não só a execução automatizada dos processos, mas também a flexibilidade de escalar conforme a demanda, sem incorrer em altos custos fixos de infraestrutura. Isso garante que o sistema possa atender a um número crescente de usuários, mantendo o foco na eficiência e na agilidade das operações financeiras.

- **EventBridge:** Atua como o orquestrador central de eventos, emitindo eventos periódicos que iniciam o fluxo de trabalho para o processamento dos ativos selecionados. Essa automação garante que o sistema seja acionado de forma regular e sem a necessidade de intervenção manual.
- **Lambda:** Escolhida como a camada de computação serverless para processar os ativos e invocar o modelo hospedado no SageMaker. A Lambda é ideal por escalar automaticamente conforme a demanda e operar no modelo *pay-as-you-go*, eliminando custos com recursos ociosos e garantindo eficiência operacional.
- **SageMaker:** Responsável pela execução dos modelos de machine learning. O SageMaker também fornece APIs nativas para realizar previsões com base nos dados de ativos financeiros e retornar as predições necessárias para tomada de decisões.
- **S3:** Utilizado para o armazenamento de dados, o S3 oferece escalabilidade praticamente ilimitada e alta durabilidade. Ele armazena registros das análises históricas e os resultados gerados pelos modelos de machine learning, além de servir para o armazenamento das predições dos usuários.
- **DynamoDB:** O banco de dados *NoSQL* da AWS foi escolhido para armazenar informações críticas que exigem acesso rápido e eficiente, como os dados de *login* dos usuários e ações relacionadas aos ativos. Ele garante alta disponibilidade e baixa latência, tornando-se adequado para consultas em tempo real.

- **SNS (Simple Notification Service):** Usado para o envio de notificações de compra ou venda aos usuários em tempo real. O SNS facilita a integração com múltiplos canais de comunicação, assegurando que as recomendações de investimento cheguem rapidamente aos investidores, garantindo uma resposta ágil às oportunidades de mercado.
- **API Gateway:** Serve como a camada de *Application Programming Interface* (API) que expõe as funcionalidades do sistema para interações externas, como o acesso às previsões de mercado e aos dados de ativos. O API Gateway permite a criação e o gerenciamento de *APIs RESTful* e *WebSocket*, possibilitando que as aplicações dos usuários consumam os serviços com segurança e confiabilidade.
- **ECS (Elastic Container Service):** Projetado para gerenciar e executar contêineres Docker na infraestrutura da AWS, o ECS permite hospedar aplicações e serviços de forma prática e escalável. Ele oferece controle sobre a rede e segurança, além de gerenciar o ciclo de vida dos contêineres, desde o provisionamento até a execução.
- **EFS (Elastic File System):** O EFS é um sistema de arquivos da AWS que pode ser acessado por várias instâncias ou contêineres ao mesmo tempo. Ele é útil para prover dados compartilhados por várias instâncias concorrentes.

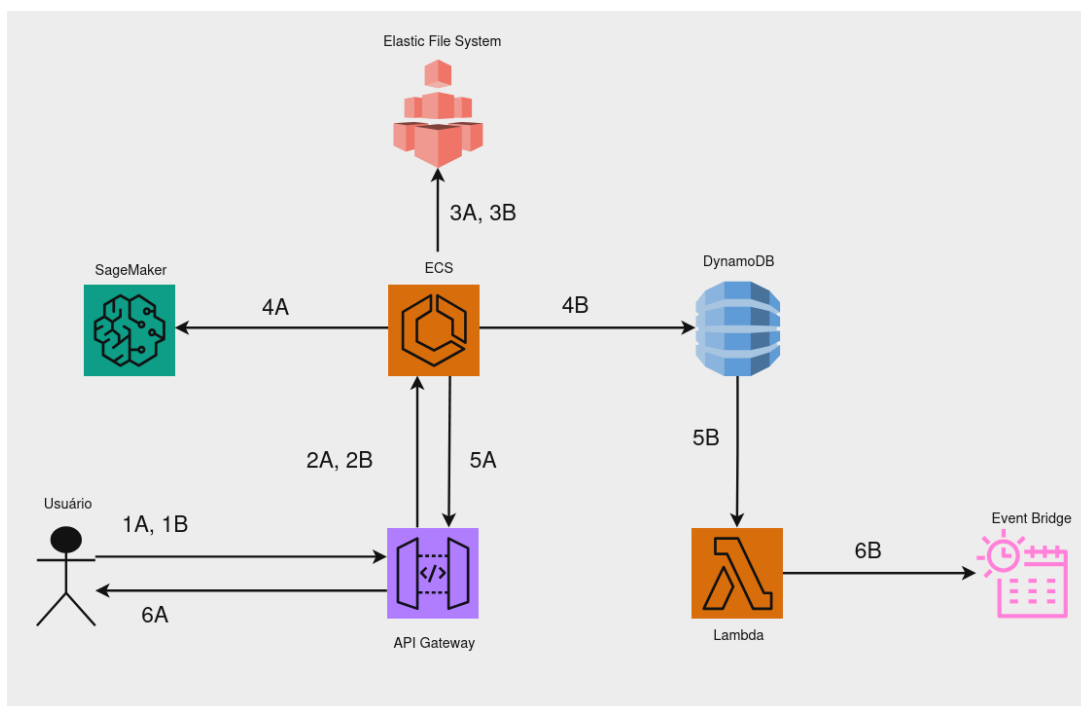
6.1 Arquitetura do Servidor da Aplicação

A camada de servidor da aplicação desempenha três funções principais. A primeira delas consiste em disponibilizar o site de recomendação de ativos para os usuários, proporcionando acesso rápido e intuitivo às informações relevantes sobre os ativos financeiros. Em seguida, essa camada é responsável por executar o modelo de RF para o ativo selecionado pelo usuário, com o objetivo de indicar se há uma recomendação de compra ou venda. Por fim, o servidor realiza o cadastro dos usuários no sistema de notificações, de modo que possam receber alertas sempre que forem detectadas oportunidades de compra.

Ademais, este componente inclui dois *workflows* principais: o primeiro permite ao usuário obter recomendações de compra/venda geradas pelo modelo de RF, enquanto o segundo possibilita a subscrição ao sistema de monitoramento periódico de ativos específicos.

A Figura 10 apresenta o diagrama do funcionamento geral do servidor da aplicação, destacando os dois *workflows* principais: o *Workflow* de Análise de Ativos, identificado pela letra A, e o *Workflow* de Subscrição de Ativos, identificado pela letra B.

Figura 10 – Arquitetura do servidor principal do sistema

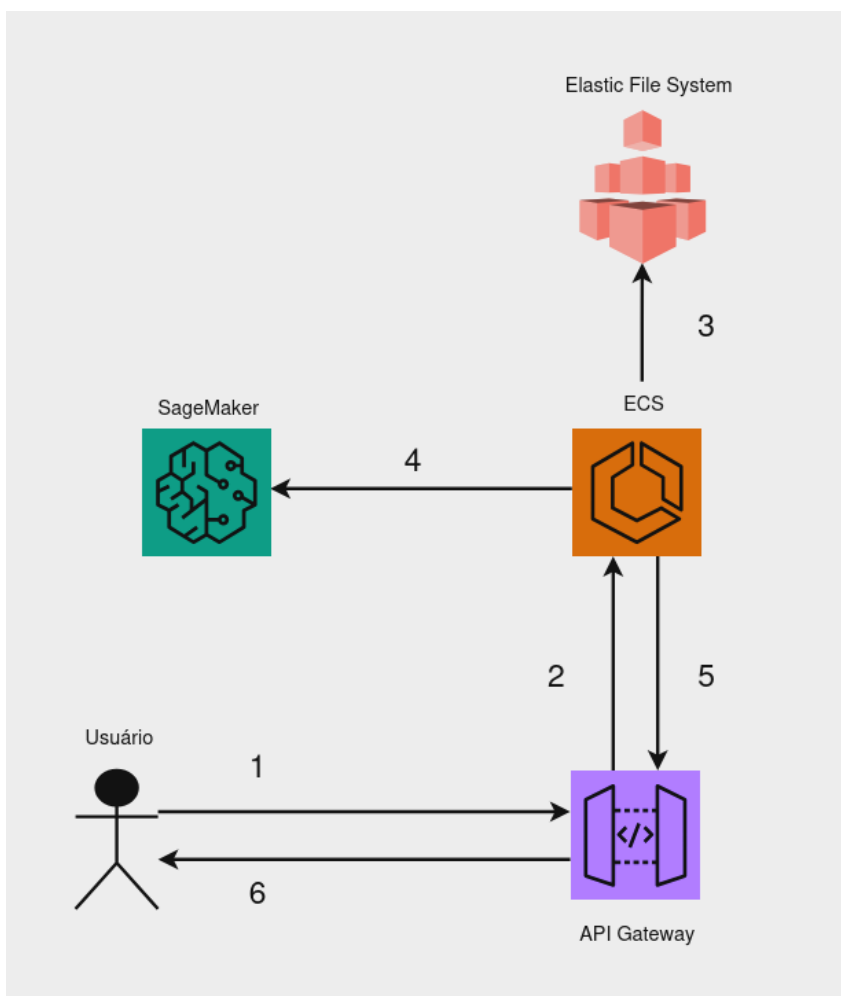


Fonte: Elaborada pelo autor

6.1.1 *Workflow* de Análise de Ativos

Esse *workflow* tem como funcionalidade principal fornecer dados sobre os ativos escolhidos pelo usuário e executar o modelo RF para gerar uma recomendação de compra ou venda, conforme a solicitação do usuário.

A Figura 11 ilustra a arquitetura do *workflow* de análise de ativos, detalhando as etapas desde a escolha do ativo pelo usuário até a geração da recomendação baseada no modelo RF.

Figura 11 – Arquitetura do *workflow* de análise de ativos

Fonte: Elaborada pelo autor

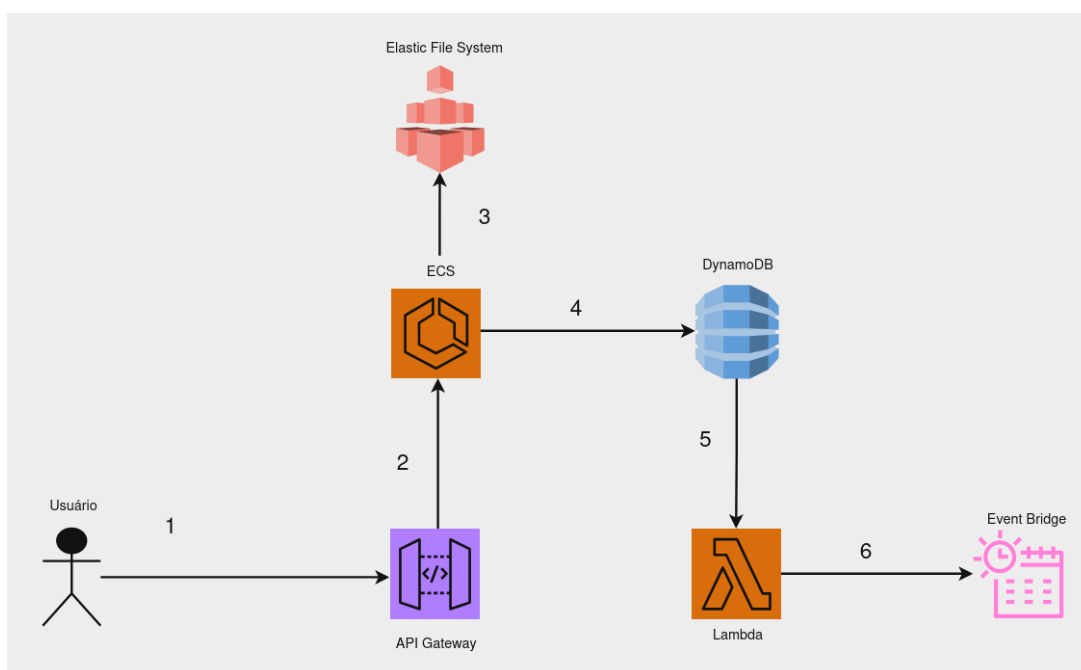
1. O usuário acessa a *endpoint* do site e faz uma requisição para o API Gateway.
2. O API Gateway direciona a requisição para o contêiner ECS.
3. O ECS busca o conteúdo do site para prover para o usuário.
4. O ECS roda os dados do ativo selecionado contra o modelo RF.
5. O ECS retorna o resultado do RF para o API Gateway, podendo ser uma recomendação de compra ou venda, ou uma indicação que não há nada a ser feito no ativo.
6. O API Gateway encaminha a resposta do modelo para o usuário.

6.1.2 Workflow de Subscrição de Ativos

Esse *workflow* tem como principal responsabilidade reunir as informações do usuário e a lista de ativos que ele deseja monitorar, registrando-o no sistema de eventos para que receba notificações sobre oportunidades de compra e venda.

A Figura 12 apresenta a arquitetura do *workflow* de subscrição de ativos, destacando como as informações do usuário são processadas e como o sistema é configurado para enviar notificações com base em eventos futuros.

Figura 12 – Arquitetura do *workflow* de subscrição de ativos



Fonte: Elaborada pelo autor

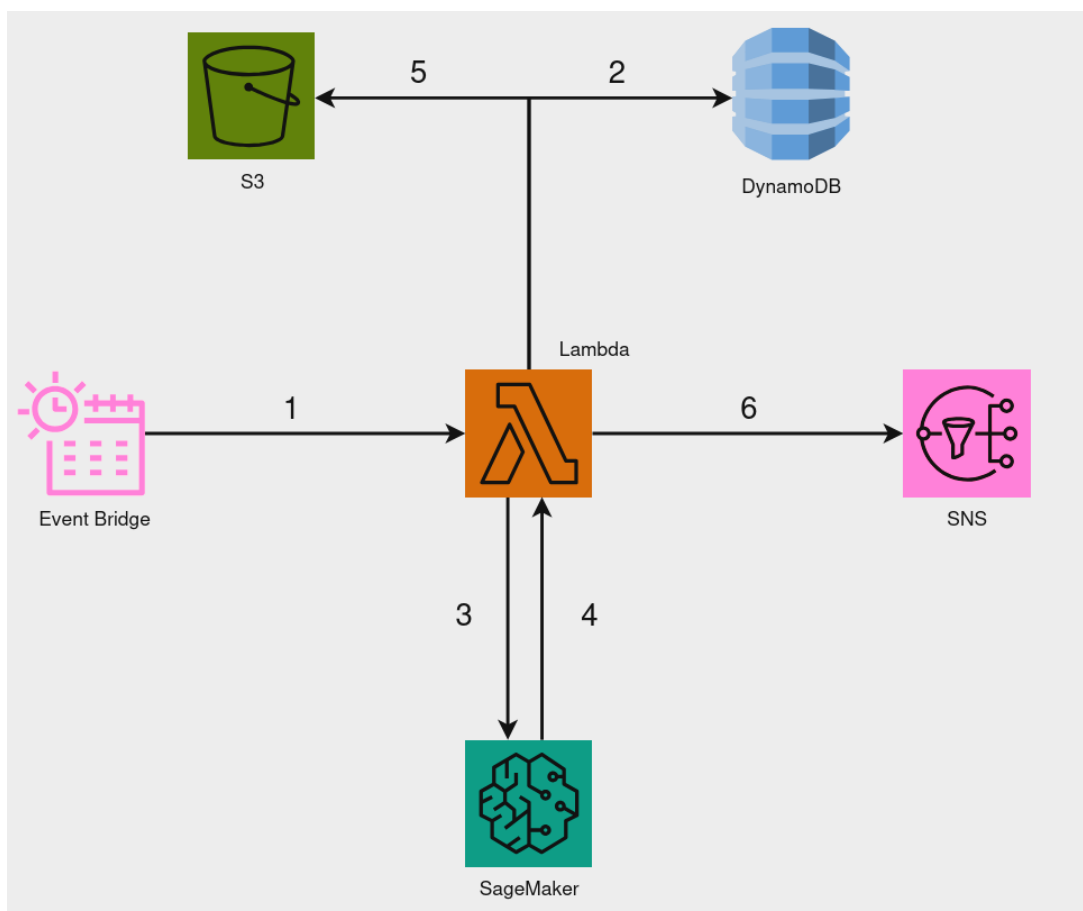
1. O usuário acessa a *endpoint* do site e faz uma requisição para cadastrar seus dados para receber notificações de entrada/saída em ativos previamente selecionados.
2. O API Gateway redireciona o request para uma instância ECS.
3. O ECS busca o conteúdo do site para prover para o usuário.
4. O ECS salva os dados do usuário e seus ativos selecionados no DynamoDB.
5. O DynamoDB envia um evento de nova entrada no banco de dados e dispara uma Lambda.
6. A Lambda lê os dados salvos no DynamoDB e configura o EventBridge para disparar eventos no período escolhido pelo usuário.

6.2 Arquitetura do Componente de Eventos

Esse componente será responsável por periodicamente analisar dados de ativos dos usuários cadastrados e notificá-los caso haja oportunidade de investimento.

A Figura 13 apresenta a arquitetura do componente de eventos, detalhando o fluxo de processamento de dados e envio de notificações, destacando como os dados são analisados para identificar oportunidades de investimento.

Figura 13 – Arquitetura do componente de eventos



Fonte: Elaborada pelo autor

Com isso, o fluxo de trabalho final segue os seguintes passos:

1. O EventBridge dispara eventos periodicamente, iniciando o *workflow* de avaliação dos ativos.
2. A Lambda acessa o DynamoDB para obter os clientes e as ações que devem ser avaliadas no ciclo corrente.

3. Em seguida, a Lambda faz uma chamada à API do SageMaker para executar o modelo de *machine learning* que prevê possíveis oportunidades de entrada no mercado.
4. Após a execução do modelo, o SageMaker retorna o resultado da execução para a função Lambda.
5. A Lambda armazena os resultados da análise no S3 para fins de auditoria e análise futura.
6. Caso a previsão indique uma oportunidade positiva, uma recomendação de entrada é enviada ao usuário final por meio de uma notificação do SNS.

6.3 Tecnologias Utilizadas

O servidor da aplicação foi desenvolvido utilizando o *framework* Flask¹, configurado para ouvir requisições na porta 5000. A estrutura do servidor conta com diferentes *endpoints*, organizados para atender às principais funcionalidades da aplicação, como o acesso à página inicial, a execução do modelo de RF para gerar recomendações de compra e venda, e o cadastro de usuários no sistema de subscrição para notificações de oportunidades. Essa organização modular permite uma fácil manutenção e escalabilidade da aplicação, garantindo a separação das responsabilidades de cada funcionalidade.

A interface do usuário foi implementada com HTML, CSS e JavaScript, tecnologias escolhidas para otimizar o desenvolvimento. Essa abordagem proporciona uma interface simples e eficiente, permitindo que os usuários naveguem pela aplicação com rapidez e acessem as informações de forma clara. O uso dessas tecnologias também favorece o carregamento ágil das páginas, fundamental em uma aplicação que se propõe a oferecer recomendações de mercado em tempo hábil.

Para o funcionamento da arquitetura de eventos, foi utilizado Python com o Boto3², o *Software Development Kit* (SDK) oficial da AWS para essa linguagem. Um SDK é um conjunto de ferramentas e bibliotecas que permite aos desenvolvedores interagir e integrar aplicações com plataformas específicas. O Boto3, sendo o SDK da AWS para Python, oferece uma interface programática para acessar serviços da AWS, como DynamoDB, Lambda, SNS e S3, simplificando o desenvolvimento e automação de tarefas.

¹ Disponível em: <<https://flask.palletsprojects.com/>>. Acesso em: 30 out. 2024.

² Disponível em: <<https://boto3.amazonaws.com/v1/documentation/api/latest/index.html>>. Acesso em: 30 out. 2024.

A segurança da aplicação foi reforçada com medidas como o uso de uma Virtual Private Cloud (VPC), que limita o acesso externo e controla o tráfego de rede dentro da infraestrutura. Para o armazenamento de dados no S3, foi configurado o *Public Access Blocking*, restringindo o acesso apenas aos componentes autorizados. Além disso, foram implementadas chaves de criptografia usando o *Key Management Service* (KMS) tanto para o DynamoDB quanto para o S3, assegurando a proteção dos dados sensíveis.

Por fim, o sistema utiliza o protocolo HTTP para a comunicação entre o usuário e a aplicação. Esse protocolo, por ser amplamente adotado mundialmente, facilita a transferência de dados e garante compatibilidade com a maioria dos navegadores e dispositivos, aumentando o alcance da aplicação.

6.4 Produto Final

Esta seção apresenta telas da aplicação e exemplo de uma recomendação do modelo RF, além das notificações enviadas pelo *workflow* de subscrição de ativos.

A Figura 14 mostra a tela principal de busca de ativos, que permite ao usuário pesquisar ativos financeiros e visualizar recomendações de compra ou venda.


Figura 14 – Tela principal de busca de ativos



Fonte: Elaborada pelo autor

A Figura 15 apresenta a tela de subscrição a eventos de recomendação, onde o usuário pode registrar ativos para receber notificações de oportunidades futuras.

Figura 15 – Tela de subscrição a eventos de recomendação



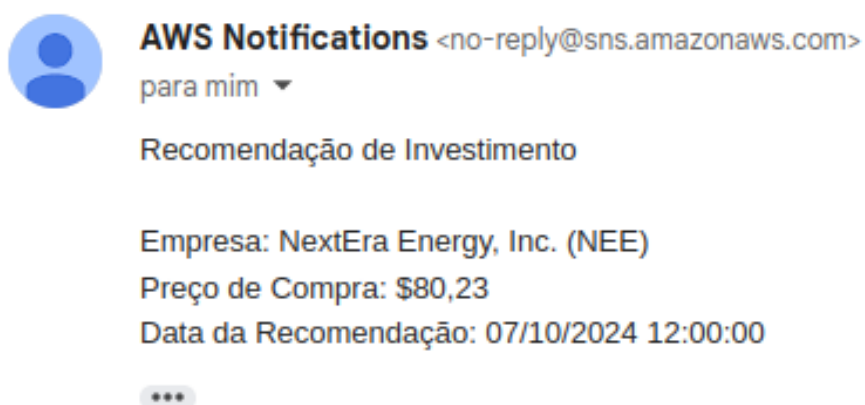
O formulário, intitulado "Cadastro de Notificações", contém os seguintes campos e elementos:

- Um cabeçalho com o título "Cadastro de Notificações" e um toggle para "Modo Claro" (Modo Claro ☺).
- Um campo "Email:" com o placeholder "Digite seu email".
- Um campo "SMS:" com o placeholder "Digite seu número de celular".
- Um campo "Ações para Monitorar:" com o placeholder "Ex: AAPL, MSFT, etc.".
- Um campo "Tempo de Notificação:" com um menu suspenso atualmente configurado para "1 Dia".
- Um botão azul "Cadastrar" no rodapé.

Fonte: Elaborada pelo autor

A Figura 16 ilustra uma notificação de compra enviada ao usuário via email, detalhando o ativo recomendado e as razões para a recomendação.

Figura 16 – Notificação de compra enviada via email



Fonte: Elaborada pelo autor

Por fim, a Figura 17 apresenta um exemplo de notificação de compra enviada ao usuário via SMS, permitindo acesso rápido às informações pelo celular.

Figura 17 – Notificação de compra enviada via SMS



Fonte: Elaborada pelo autor

7 Conclusão

Nesta monografia, foi explorada uma aplicação do modelo RF para a análise e previsão de sinais de compra e venda de ativos financeiros, abordando tanto os aspectos teóricos quanto os práticos da sua implementação. A RF, caracterizada por ser um modelo de *ensemble*, demonstrou bons resultados ao lidar com diferentes indicadores e ativos financeiros para a tomada de decisões de investimento. Com a aplicação de técnicas de *bagging* e da divisão aleatória dos dados em conjuntos de treino, teste e validação, o modelo apresentou uma taxa de acerto satisfatória, validando sua eficácia no cenário de previsões financeiras.

A escolha da RF se justificou pela sua capacidade de reduzir o *overfitting* e por garantir robustez contra dados ruidosos, uma característica essencial em séries temporais financeiras voláteis. O uso de métricas de avaliação, como o SR e a acurácia serviu para evidenciar que a RF pode ser uma alternativa eficiente aos métodos tradicionais de regressão linear, ou aos métodos de *deep learning* que são comumente adotados no *forecasting* de mercados financeiros.

Para trabalhos futuros, uma sugestão é a implementação de modelos de *machine learning* voltados para análise e gestão financeira. A inclusão desses modelos pode agregar valor à estratégia de investimento a longo prazo, proporcionando uma visão mais completa do risco e da rentabilidade dos ativos. Assim, a aplicação de algoritmos para análise financeira além da previsão de sinais permite ao investidor uma abordagem mais integrada e sólida, favorecendo melhores retornos e controle dos riscos associados.

Ademais, a arquitetura proposta neste estudo pode ser expandida para incluir funcionalidades adicionais, como a análise de ativos em tempo real. Outra melhoria possível seria a integração da arquitetura de eventos com API's de corretoras para automatizar completamente o processo de compra/venda de ativos.

Referências

- APPEL, Gerald. *Technical Analysis: Power Tools for Active Investors*. New Jersey: Financial Times Prentice Hall, 2005. ISBN 9780131479029.
- BOLLINGER, John. *Bollinger on Bollinger Bands*. New York: McGraw-Hill, 2002.
- BOX, George E. P.; JENKINS, Gwilym M. *Time Series Analysis: Forecasting and Control*. 2. ed. San Francisco: Holden-Day, 1976. ISBN 9780816211046.
- BREIMAN, Leo. Random forests. *Machine Learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Disponível em: <<https://link.springer.com/article/10.1023/A:1010933404324>> . Acesso em: 27 de outubro de 2024.
- BROCKWELL, Peter J.; DAVIS, Richard A. *Introduction to Time Series and Forecasting*. 3. ed. Cham: Springer, 2016. ISBN 9783319298536.
- BUFFETT, Warren; CUNNINGHAM, Lawrence A. *The Essays of Warren Buffett: Lessons for Corporate America*. 3rd. ed. Durham: Cunningham Group, 2008.
- CHAN, Ernest P. *Algorithmic Trading: Winning Strategies and Their Rationale*. [S.l.]: John Wiley & Sons, 2013.
- HAMILTON, James D. *Time Series Analysis*. Princeton: Princeton University Press, 1994. ISBN 9780691042893.
- HASAN, Farhan; SANTOS, Thiago dos; KUMAR, Ravi. A performance comparison of machine learning models for stock market prediction with novel investment strategy. *PLOS ONE*, v. 18, n. 7, p. e0286362, 2023. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0286362>> . Acesso em: 27 de outubro de 2024. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0286362>>.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. [S.l.]: Springer, 2009.
- JONES, Mark; SMITH, Carla; THOMPSON, David. Evaluating risk adjusted performance in hedge funds. *Financial Review*, v. 39, n. 1, p. 12–25, 2018. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0031320315000989>> . Acesso em: 27 de outubro de 2024.
- KAHNEMAN, Daniel; TVERSKY, Amos. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, JSTOR, v. 47, n. 2, p. 263–291, 1979. Disponível em: <<https://www.jstor.org/stable/1914185>> . Acesso em: 25 de outubro de 2024.
- LANE, George C. *Lane's Stochastics*. Lakewood, Colorado: Investment Educators, Inc., 1984. Discusses the Stochastic Oscillator and its application in technical analysis.
- LEVINE, Paul; ROSS, Michael. *Advanced Portfolio Management Techniques*. [S.l.]: Springer, 2021.

- LYNCH, Peter; ROTHCHILD, John. *One Up On Wall Street: How to Use What You Already Know to Make Money in the Market*. New York: Simon & Schuster, 1989.
- MAIMON, Oded; ROKACH, Lior. *Data Mining and Knowledge Discovery Handbook*. Boston: Springer, 2014.
- MISHKIN, Frederic S. *The Economics of Money, Banking, and Financial Markets*. 13. ed. United States: Pearson, 2021.
- MUNSCH, Kurt. Sharpe ratio analysis in investment strategies. *Journal of Portfolio Management*, v. 45, n. 2, p. 34–46, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0031320315000989>> . Acesso em: 27 de outubro de 2024.
- MURPHY, John J. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York: New York Institute of Finance, 1999.
- PRING, Martin J. *Technical Analysis Explained: The Successful Investor's Guide to Spotting Investment Trends and Turning Points*. 5. ed. New York: McGraw-Hill Education, 2014.
- SCHWAGER, Jack D. *Market Wizards: Interviews with Top Traders*. New York: HarperBusiness, 1989.
- SHILLER, Robert J. Phishing for phools: The economics of manipulation and deception. *The American Economic Review*, American Economic Association, v. 105, n. 5, p. 328–353, 2015. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/aer.20140924>> . Acesso em: 28 de março de 2024.
- TSAY, Ruey S. *Analysis of Financial Time Series*. 3. ed. Hoboken: Wiley, 2010. ISBN 9780470414354.
- WILDER, J. Welles. *New Concepts in Technical Trading Systems*. Greensboro: Trend Research, 1978.
- WONG, Tze-Haw; YEH, Andy; LIN, Kai-Min. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition Letters*, Elsevier, v. 20, n. 1, p. 1–10, 2015. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0031320315000989>> . Acesso em: 27 de outubro de 2024.
- WOOLDRIDGE, Jeffrey M. *Introductory Econometrics: A Modern Approach*. 6. ed. Boston, MA: Cengage Learning, 2016.