



Flávio Lins da Mota Nascimento

**Análise da técnica de XGBoost para prever com precisão valores máximos e mínimos de ações listadas na B3**



Universidade Federal de Pernambuco

secgrad@cin.ufpe.br

<https://portal.cin.ufpe.br/graduacao/contato/>

Recife

2023

Flávio Lins da Mota Nascimento

**Análise da técnica de XGBoost para prever com precisão valores máximos e mínimos de ações listadas na B3**

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

**Área de Concentração:** *Inteligência Computational*

**Orientador:** *Adriano Lorena Inácio de Oliveira*

Recife

2023

Ficha de identificação da obra elaborada pelo autor,  
através do programa de geração automática do SIB/UFPE

Nascimento, Flávio Lins da Mota.

Análise da técnica de XGBoost para prever com precisão valores máximos e mínimos de ações listadas na B3 / Flávio Lins da Mota Nascimento. - Recife, 2023.

43 p. : il., tab.

Orientador(a): Adriano Lorena Inácio de Oliveira

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Informática, Engenharia da Computação - Bacharelado, 2023.

1. Aprendizagem de máquina. 2. Ciência de dados. 3. Mercado financeiro. 4. Bolsa de valores. 5. Predição. I. Oliveira, Adriano Lorena Inácio de. (Orientação). II. Título.

000 CDD (22.ed.)

## **AGRADECIMENTOS**

Gostaria de agradecer primeiro a toda minha família, como minha irmã, Rafaela e principalmente aos meus pais, Rosemary e José Lins, que sempre me deram suporte, me incentivaram e me apoiaram de várias formas. Agradeço por terem se dedicado para construir uma base que me auxiliou e que me fez capaz de usar meus esforços para o estudo e metas cada vez maiores. Igualmente agradeço a minha companheira Karla que motivou e me deu energia nos momentos mais desanimadores.

Mostrar gratidão a todos meus amigos que foram parceiros de jornada. Meus amigos me proporcionaram momentos essenciais de descontração após dias exaustivos de estudo e trabalho, o que me ajudou na disposição cotidiana.

Por fim, sou muito grato também ao meu Orientador, Professor Adriano Lorena, que me guiou nas melhores decisões para esse trabalho e me possibilitou agregar os conhecimentos necessários para atingir os resultados fundamentais com o tema proposto. De forma semelhante quero deixar registrado meus agradecimentos a todos professores que me acompanharam na trajetória da UFPE, em especial a todos funcionários do Centro de Informática da UFPE, pois me entregaram uma infraestrutura incrível para qualquer problema no dia a dia do estudante ser solucionado, facilitando o caminho do aprendizado dos temas complexos neste curso.

*"Tudo muda quando descobrimos que  
os diferentes, são  
também  
iguais."*

*–Silvio Almeida (2020)*

# ABSTRACT

Forecasting stock values on the stock exchange is one of the most interesting problems for financial professionals, as it can bring almost immediate economic profit. For this reason, academics are constantly working on this purpose. In this work, XGBoost, a prediction regression tool for machine learning, is used. The technique is used with the intention of improving results in a Day Trading system. Data from the last 5 or more days of any stock are used as input to the system to find the accuracy of the predicted maximum and minimum values.

**Keywords:** Machine Learning, XGBoost, Data Science, Stock Market, Prediction, Day Trade.

## RESUMO

Para profissionais da área financeira a previsão dos valores de ações na bolsa de valores é um dos problemas de maior interesse, pois pode trazer o lucro financeiro quase imediato. Por esse motivo, acadêmicos constantemente trabalham nesse propósito. Neste trabalho é utilizado XGBoost, um regressor de predição para aprendizagem de máquina. A técnica é utilizada com propósito de aperfeiçoar resultados em um sistema de Day Trading. São utilizados dados dos últimos 5 ou mais dias anteriores de uma ação como entrada do sistema para encontrar a precisão dos valores máximos e mínimos previstos.

**Palavras-chave:** Aprendizagem de máquina, XGBoost, Ciência de dados, Mercado financeiro, Bolsa de valores, Predição, Day Trade.

## LISTA DE FIGURAS

Figura 1	– Demonstrativo das Bandas de <i>Bollinger</i> . . . . .	19
Figura 2	– Estrutura Árvore de Decisão (DT) . . . . .	21
Figura 3	– Fluxo em Sistemas com <i>Random Forest</i> . . . . .	22
Figura 4	– Fluxo no Sistema <i>XGBoost</i> . . . . .	23
Figura 5	– Forma resumida de agregar DTs em modelos <i>XGBoost</i> . . . . .	24
Figura 6	– Diagrama do Sistema . . . . .	25

## LISTA DE TABELAS

Tabela 1	– Valores MAPE inicial com 15, 25 e 33 variáveis no sistema <i>XGBoost</i> . . . . .	33
Tabela 2	– Porcentagem da importância das variáveis de entrada dos sistemas em <i>XGBoost</i> prevendo valores máximos . . . . .	34
Tabela 3	– Valores MAPE considerando indicadores externos como variáveis adicionais no sistema <i>XGBoost</i> . . . . .	34
Tabela 4	– Porcentagem da importância das variáveis dos sistemas em <i>XGBoost</i> considerando índices externos prevendo valores máximos . . . . .	35
Tabela 5	– Valores MAPE com alteração dos parâmetros de treinamento no sistema <i>XGBoost</i> . . . . .	35
Tabela 6	– Porcentagem da importância das variáveis de entrada dos sistemas em <i>XGBoost</i> com parâmetros iniciais alterados . . . . .	36
Tabela 7	– Porcentagem da importância das variáveis de entrada dos sistemas em <i>XGBoost</i> com parâmetros iniciais alterados e 62 variáveis de entrada . . . . .	38
Tabela 8	– Valores MAPE com alteração dos parâmetros de treinamento no sistema <i>XGBoost</i> e uso de 62 variáveis de entrada . . . . .	39

# LISTA DE ACRÔNIMOS

<b>ANN</b>	<i>Artificial Neural Network</i>
<b>BB</b>	<i>Bandas de Bollinger</i>
<b>DT</b>	<i>Decision Tree</i>
<b>EMA</b>	<i>Exponential Moving Average</i>
<b>MAPE</b>	<i>Mean Absolute Percentage Error</i>
<b>ML</b>	<i>Machine Learning</i>
<b>MMA</b>	<i>Média Móvel Aritimética</i>
<b>OHLC</b>	<i>Open-High-Low-Close</i>
<b>RF</b>	<i>Random Forest</i>
<b>RSI</b>	<i>Relative Strength Index</i>
<b>STA</b>	<i>Sistema Trade Avançado</i>

# SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>12</b>
1.1	Contextualização Financeira . . . . .	12
1.2	Objetivos do Trabalho . . . . .	12
<b>2</b>	<b>Fundamentação Teórica</b>	<b>14</b>
2.1	Previsão de Valores de Ações no Mercado Financeiro . . . . .	14
2.1.1	Entendendo o Mercado Financeiro . . . . .	14
2.1.1.1	Ação . . . . .	15
2.1.1.2	Corretora . . . . .	15
2.2	<i>Day Trading</i> como Estratégia de Negociação . . . . .	15
2.3	Técnicas de Análise . . . . .	15
2.3.1	Médias Móveis . . . . .	16
2.3.1.1	Média Móvel Aritmética (MMA ou MA) . . . . .	16
2.3.1.2	Média Móvel Exponencial (MME ou EMA) . . . . .	16
2.3.2	Erro Médio Percentual Absoluto (EMPA ou MAPE) . . . . .	17
2.3.3	Bandas de <i>Bollinger</i> . . . . .	18
2.3.4	<i>Relative Strength Index</i> . . . . .	19
2.4	<i>Machine Learning</i> e sua Aplicação no Mercado Financeiro . . . . .	20
2.4.1	<i>Features</i> . . . . .	20
2.4.2	<i>Decision Tree</i> . . . . .	21
2.4.2.1	<i>Random Forest</i> (RF) . . . . .	21
2.4.2.2	<i>XGBoost</i> e sua base em <i>Gradient Boosting</i> . . . . .	22
<b>3</b>	<b>Metodologia</b>	<b>25</b>
3.1	Modelos e Estrutura do Trabalho . . . . .	25
3.1.1	Sistema com <i>Machine Learning</i> . . . . .	26
3.1.2	Sistema Trade Avançado . . . . .	26
3.2	Coleta e Preparação de Dados . . . . .	27
3.3	Implementação do <i>XGBoost</i> no Sistema de <i>Machine Learning</i> . . . . .	27
3.3.1	Adaptações para <i>XGBoost</i> . . . . .	27
3.3.2	Criação de <i>Features</i> e Engenharia de <i>Features</i> . . . . .	28
3.3.3	Aperfeiçoamento dos Parâmetros . . . . .	30
3.4	Avaliação da Precisão das Previsões . . . . .	31
3.5	Métricas de Avaliação . . . . .	31
3.5.1	<i>Mean Absolute Percentage Error</i> (MAPE) . . . . .	31

<b>4</b>	<b>Resultados e Discussão</b>	<b>32</b>
4.1	Desempenho do <i>XGBoost</i> nas Previsões de trabalhos anteriores . . . . .	33
4.1.1	Desempenho do <i>XGBoost</i> referente a ações e variáveis em comum . . .	33
4.1.2	Desempenho do <i>XGBoost</i> referente a ações e índices externos . . . . .	34
4.1.3	Resultados e Discussão com <i>XGBoost</i> e seus aprimoramentos . . . . .	35
4.1.3.1	Aprimoramentos baseados em Engenharia de <i>Feature</i> . . . . .	36
<b>5</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>40</b>
5.1	Síntese dos Resultados Obtidos . . . . .	40
5.1.1	Predições . . . . .	40
5.2	Contribuições e Implicações no Campo Financeiro . . . . .	41
5.3	Possíveis Direções Futuras . . . . .	41
5.3.1	Sistema de <i>Day Trading</i> . . . . .	41
	<b>REFERÊNCIAS</b>	<b>43</b>

# 1

## INTRODUÇÃO

### 1.1 CONTEXTUALIZAÇÃO FINANCEIRA

No âmbito da esfera financeira, a previsão dos valores das ações no mercado de capitais representa um desafio crucial e de interesse destacável. A possibilidade de alcançar ganhos financeiros quase instantâneos eleva a importância desse domínio para os profissionais da área. Não surpreende, portanto, que pesquisadores acadêmicos estejam constantemente imersos nessa busca por aprimorar as técnicas de previsão e eficiência nesse cenário altamente dinâmico.

O presente trabalho se concentra na aplicação do *XGBoost*, um poderoso regressor de predição utilizado no aprendizado de máquina, para abordar essa problemática intrincada. A essência do projeto reside em melhorar e otimizar os resultados em um contexto específico: o *Day Trading*. O *Day Trading* é uma estratégia que envolve a realização de diversas transações de compra e venda de ativos dentro do mesmo dia, visando aproveitar as flutuações de preços em curtos períodos.

### 1.2 OBJETIVOS DO TRABALHO

A abordagem adotada neste estudo é notável pela sua implementação do *XGBoost*, uma ferramenta que tem demonstrado excelentes resultados em várias tarefas de aprendizado de máquina. A ênfase reside em sua aplicação para aprimorar um sistema de *Day Trading*, buscando identificar oportunidades de negociação com base em previsões de preços de ações.

Esse estudo tem como base duas grandes análises feitas anteriormente com os esforços de Adrion Cavalcanti de Albuquerque Filho (2021) [ALBUQUERQUE FILHO](#), onde pode-se acompanhar com detalhes a criação de um novo sistema de *Day Trading* utilizando Rede Neural Artificial ou *Artificial Neural Network* (ANN) e também esforços de Luiz Antonio de Albuquerque Junior (2022) [ALBUQUERQUE JUNIOR](#), com a produção de um sistema mais aperfeiçoado focando sua análise no indicador drawdown de cada resultado de transação desse sistema de *Day Trading*.

O sistema é concebido considerando uma janela de tempo dos últimos cinco ou mais dias anteriores de uma ação, e os dados são utilizados como entrada para o modelo. Isso permite

avaliar a precisão das previsões dos valores máximos e mínimos projetados. A abordagem deste trabalho foca na melhoria qualitativa das previsões dos valores máximos e mínimos utilizando parâmetros numéricos precisos para sistemas preditores: o Erro Médio Percentual Absoluto.

A pesquisa se insere na conjuntura de constante evolução das estratégias de negociação no mercado financeiro, onde a tecnologia e as inovações no aprendizado de máquina têm demonstrado potencial transformador.

Ao explorar a capacidade do *XGBoost* em melhorar a qualidade das previsões no contexto do *Day Trading*, almejamos contribuir com percepções e visões valiosas para os profissionais da área financeira e acadêmica. Também, com o registro dos testes reproduzidos aqui, almeja-se avançar na compreensão das aplicações da aprendizagem de máquina em um ambiente de mercado tão dinâmico e desafiador.

# 2

## FUNDAMENTAÇÃO TEÓRICA

A Fundamentação Teórica é um pilar essencial para a compreensão das bases conceituais e metodológicas que sustentam este estudo. Neste capítulo são explorados os fundamentos teóricos relacionados à previsão de valores de ações no mercado financeiro, explicar sobre estratégias de *Day Trading*, técnicas de análise, equações necessárias para esses objetivos, e uma introdução sobre a aplicação de *Machine Learning* nesse contexto financeiro. Além disso, serão discutidas as nuances da utilização de múltiplas técnicas de *Machine Learning*, como *Decision Tree*, *Random Forest* e *Gradient Boosting*, para dar contexto suficiente para entendimento da motivação para o uso da técnica de *XGBoost* neste trabalho.

### 2.1 PREVISÃO DE VALORES DE AÇÕES NO MERCADO FINANCEIRO

A previsão de valores de ações no mercado financeiro é uma busca constante para investidores e profissionais da área monetária, em busca de ganhos ou rendimentos. Compreender e antecipar os movimentos de preços das ações é fundamental para a tomada de decisões e conseguir que sejam bem-sucedidas no mercado. Esta seção abordará de forma ampla o conteúdo, as estratégias, os métodos e os modelos utilizados para prever os preços das ações, destacando a importância dessas previsões no contexto financeiro.

#### 2.1.1 Entendendo o Mercado Financeiro

O mercado financeiro é um ambiente complexo e dinâmico onde ocorre a negociação de ativos financeiros, tais como ações, títulos, derivativos, moedas e comódites. Este é um espaço onde investidores, instituições financeiras, empresas e governos interagem para comprar e vender tais ativos. Não é incomum encontrar estrategistas que buscam comprar moedas como Dólar (\$), usando Reais (R\$), por valores baixos para ter a chance de em um futuro fazer a venda dessa moeda obtendo lucro nessa transação. Estratégias como essa são partes vitais da economia global, pois auxiliam na alocação eficiente de recursos e no financiamento das atividades econômicas. Nesse trabalho usaremos alguns parâmetros baseados no próprio Dólar, todavia essa pesquisa foca em ser exclusiva para o mercado brasileiro, então iremos prever e calcular os valores de

---

ações brasileiras usando o Real como padrão.

### 2.1.1.1 Ação

Uma ação representa uma parcela do capital de uma empresa, em suma é seu valor relativo ao mercado. Quando uma empresa decide abrir seu capital e se tornar uma sociedade anônima, ela emite ações que são oferecidas ao público, permitindo que interessados as comprem. Os investidores que adquirem essas ações tornam-se acionistas e, como tal, possuem direitos e participação na empresa proporcional ao número de ações que possuem. As ações são negociadas no mercado financeiro. Seus preços são influenciados por uma série de fatores, incluindo desempenho da empresa, economia, setor de atuação, entre outros. Investir em ações é uma maneira comum de buscar lucro no mercado financeiro, porém é importantíssimo ressaltar que também envolve riscos e não existem formas garantidas de obter lucros.

### 2.1.1.2 Corretora

Uma corretora é uma entidade financeira que atua como intermediária entre os investidores e o mercado financeiro. Essas instituições facilitam a compra e venda das ações, fornecem informações e assessoria, além de oferecerem plataformas de negociação para os investidores. As corretoras desempenham um papel crítico ao conectar os investidores ao mercado, permitindo-lhes participar das transações de forma segura e eficiente. Também são o sistema final para a execução das transações previstas por esse trabalho, apesar de não executar-se transações reais em uma corretora nessa pesquisa, entender o seu funcionamento prático é crucial.

## 2.2 DAY TRADING COMO ESTRATÉGIA DE NEGOCIAÇÃO

O *Day Trading* é uma estratégia de negociação no mercado financeiro em que os traders, em que os operadores (*traders*) compram e vendem ativos no mesmo dia, buscando lucros com as flutuações de curto prazo nos preços das ações. Essa técnica é explorada pela possibilidade rápida da tomada de decisão e por uma chance de rendimento instantâneo, alguns casos chegam a usar aberturas de segundos entre transações, contudo as janelas padrões nessa estratégia são de 5 ou 15 minutos. A ideia principal é o uso desse método já que temos previsões diárias de máximos e mínimos para um dia específico conseguindo certa dinamicidade nas execuções das operações.

## 2.3 TÉCNICAS DE ANÁLISE

No presente trabalho as técnicas de análise desempenham um papel crítico na tomada de decisões. Não diferente, também possuem esse papel no mercado financeiro. Nesta seção, serão discutidas técnicas de análise amplamente desfrutada, incluindo Médias Móveis, Bandas de

*Bollinger* e o *Relative Strength Index* (RSI). Essas métricas possuem uma fundamentação forte de conceitos matemáticos e fornecem uma compreensão aprofundada de como podem auxiliar os investidores a interpretar e prever os movimentos das ações no mercado.

### 2.3.1 Médias Móveis

As Médias Móveis são uma ferramenta fundamental de análise técnica que suaviza os dados de preços das ações, facilitando a identificação de tendências e padrões. Duas das variantes mais comuns são a Média Móvel Aritmética (MMA ou MA em inglês) e a Média Móvel Exponencial (MME ou EMA em inglês), cada uma com suas características distintas. Esta subseção aborda tanto a Média Móvel Aritmética quanto a Média Móvel Exponencial, explicando seus cálculos e aplicabilidades no contexto do mercado financeiro.

#### 2.3.1.1 Média Móvel Aritmética (MMA ou MA)

A Média Móvel Aritimética (MMA) é uma técnica simples e eficaz que calcula a média dos preços de um ativo em um determinado período de tempo. A fórmula para calcular a MMA é:

$$MMA = \frac{P_1 + P_2 + P_3 + \dots + P_n}{n} \quad (2.1)$$

onde:

- $P_1, P_2, P_3, \dots, P_n$  são os preços observados nos  $n$  períodos anteriores;
- $n$  é o número de períodos considerados.

A MMA proporciona uma visão suavizada da tendência de preços, facilitando identificar padrões e comportamentos do mercado. Dependendo da forma que manipula-se esse valor para grandes conjuntos de dados pode-se entender a existência de tendências de crescimento, de decaimento ou de estagnação em uma janela de período longa. Costuma-se usar a descrição de MMA10 para o caso em que  $n$  é 10, MMA20 para quando  $n$  é 20 e assim por diante. Esse é um conceito comumente usado no ambiente de *Day Trading*, pois pode dar os indícios dos valores futuros da ação ao observar os últimos casos e não exige muito poder computacional para seu cálculo de forma rápida.

#### 2.3.1.2 Média Móvel Exponencial (MME ou EMA)

A *Exponential Moving Average* (EMA) é uma variação da MMA, dando mais peso aos preços mais recentes. A fórmula para calcular a EMA é:

$$EMA = \alpha * P_n + (1 - \alpha) * EMA_{n-1} \quad (2.2)$$

onde:

- $n$  é o número de períodos utilizado;
- $\alpha$  é o fator de suavização que se calcula com  $\frac{2}{n+1}$  ;
- $P_n$  é o preço mais recente;
- $EMA_{n-1}$  é o valor da EMA no período anterior.

A EMA reage mais rapidamente às mudanças nos preços, sendo mais sensível às variações recentes. Isso a torna uma ferramenta valiosa para identificar mudanças de tendência em um estágio inicial. Da mesma forma que a MMA, a EMA também é utilizada baseada no período  $n$ , exemplos: EMA\_5, EMA\_14, EMA\_21. Nesse trabalho teremos casos de uso como "EMA\_21\_close", onde o  $n$  é 21 e a coluna utilizada para os cálculos é "close", ou seja, o valor de fechamento da ação durante os últimos 21 dias. Algo a se observar é que para os casos em que o  $n$  não está explícito, o valor padrão utilizado foi  $n$  como 5 dias.

### 2.3.2 Erro Médio Percentual Absoluto (EMPA ou MAPE)

O *Mean Absolute Percentage Error* (MAPE), que significa Erro Médio Percentual Absoluto em português, é uma métrica usada para avaliar a precisão de um modelo de previsão ou uma técnica de previsão, especialmente em contextos de séries temporais ou previsão financeira.

O MAPE é calculado pela média dos valores absolutos das diferenças percentuais entre as previsões ( $V_{prev}$ ) e os valores reais ( $V_{real}$ ), dividido pelo valor real, multiplicado por 100 para expressar o erro em termos percentuais. A fórmula é dada por:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{V_{real} - V_{prev}}{V_{real}} \right| * 100 \quad (2.3)$$

onde:

- $n$  é o número de valores previstos e comparados;
- $V_{real}$  é o valor real no mundo real;
- $V_{prev}$  é o valor previsto pelo sistema;

O MAPE expressa o erro como uma porcentagem da magnitude real do valor previsto, o que é útil para interpretar a precisão das previsões em termos relativos, tornando-o uma métrica popular em análises de previsão. Quanto menor o MAPE, melhor é a precisão das previsões, sendo o valor ideal de MAPE igual a 0 (quando as previsões são perfeitamente precisas). Esse será o nosso principal parâmetro de comparação na pesquisa.

### 2.3.3 Bandas de *Bollinger*

As Bandas de *Bollinger* (BB), ou conhecidas como *Bollinger Bands* em inglês, são uma técnica de análise técnica que recorre à volatilidade dos preços para prever futuros movimentos do mercado. Vamos explorar a aplicação e interpretação das Bandas de *Bollinger* na análise de ações, destacando sua função na identificação de pontos de possível maior ou menor volatilidade no futuro da ação, melhorando o entendimento do sistema para previsões mais precisas. É possível extrair 5 valores dos cálculos das BB usando a biblioteca *Pandas* em *Python*, os valores retornados podem nos dizer a banda superior (*BB\_UPPER*), a banda inferior (*BB\_LOWER*), o valor médio entre eles (*BB\_MIDDLE*), a largura da banda (*BB\_WIDTH*) e um valor percentual baseado na largura da banda (*BB\_PERCENT*). As bandas utilizam o valor da MMA como referencial e suas equações são as seguintes:

$$BB_{upper} = MA + (\delta * 2) \quad (2.4)$$

$$BB_{lower} = MA - (\delta * 2) \quad (2.5)$$

onde:

- *MA* é o valor de MMA seguindo a mesma janela da banda;
- $\delta$  é o valor do desvio padrão na janela da banda;

Como demonstrativo visual, temos a Figura a seguir, onde vemos uma faixa acompanhando o valor da ação:

Figura 1: Demonstrativo das Bandas de *Bollinger*

Fonte: *Binomo*

Já os demais três valores usam os resultados dessas duas equações, onde o resultado para *BB\_MIDDLE* é o valor intermediário entre ambas bandas superior e inferior. Esses valores de forma um pouco subjetiva para o entendimento computacional, fundamentado no desvio padrão do período, caso o mesmo possua um valor maior, demonstra que seu valor está variando com maior frequência, seguindo o mesmo raciocínio o contrário também é verdade, quão menor for desvio padrão menos volátil está este valor da ação.

### 2.3.4 *Relative Strength Index*

O RSI é um indicador de *momentum* amplamente desfrutado na análise técnica para avaliar a velocidade das mudanças e a magnitude dessas alterações nos preços das ações. A escala desse índice varia de 0 a 100 em percentual, igualmente ao MAPE. Porém, a ideia desse valor é frequentemente usada no mercado financeiro pelos investidores como uma interpretação direta: quando o RSI ultrapassa 70% , indica-se uma condição de **sobrecompra**; já quando desce abaixo de 30%, indica-se uma condição de **sobrevenida**. A interpretação desses valores percentuais é levada como regra para vários *traders*. Para o cálculo do RSI usa-se a equação:

$$RSI_n = 100 - \frac{100}{1 + RS_n} \quad (2.6)$$

onde  $RS_n$  é encontrado com a seguinte equação:

$$RS_n = \frac{AG_n}{AL_n} \quad (2.7)$$

- $n$  é o período de observação utilizado para o  $RSI_n$ ;
- $AG_n$  é o a média do ganho ou aumento do valor nos últimos  $n$  dias observados;
- $AL_n$  é o a média da perda ou diminuição do valor nos últimos  $n$  dias observados;

Os pensamentos despertados pelas equações são que a condição de **sobrecompra** (com  $1 \ll RSn$ ) é o momento em que a ação foi comprada em excesso, ou seja, seu valor está mais valorizado que o comum e é o momento de se fazer venda dessa ação, como contraponto, a condição de **sobrevenida** (com  $RSn \ll 1$ ) indica uma oportunidade de compra.

## 2.4 MACHINE LEARNING E SUA APLICAÇÃO NO MERCADO FINANCEIRO

O *Machine Learning* (ML) representa uma disciplina para inteligência artificial que se baseia na capacidade dos modelos de aprender e aprimorar suas decisões sem serem explicitamente programados. No contexto do mercado financeiro, o ML se destaca como uma ferramenta poderosa capaz de analisar volumes massivos de dados financeiros e extrair padrões, tendências e relações que podem fornecer insights valiosos para os investidores e traders.

Este campo oferece um vasto conjunto de técnicas que possibilitam a construção de modelos preditivos e analíticos que podem auxiliar na tomada de decisões informadas e na formulação de estratégias de investimento mais eficazes.

Nesse trabalho tiramos vantagem dessa habilidade computacional, justo que seria muito complexo ou até impossível para um humano analisar uma quantidade tão grande de dados para que com eles calcule e chegue em um valor real como as previsões de máximo e mínimo do dia futuro. Vamos citar modelos básicos existentes para executar essa aprendizagem e explicar de forma ampla como isso funciona.

### 2.4.1 Features

No contexto desse trabalho, as *features* ou características, são os atributos preparados como entrada dos modelos de ML. As *features* são as colunas das tabelas de dados e representam informações importantes construídas ou retiradas dos dados originais, que servem como alimento para o modelo aprender de forma matemática qual a relação dessas entradas com a saída desejada. Um exemplo de informação bruta do sistema que temos como *feature* nesse trabalho o valor de abertura da ação por dia. Contudo, é uma possibilidade criar novas *features* usando as demais *features* existentes, o RSI explicado anteriormente é um exemplo disso.

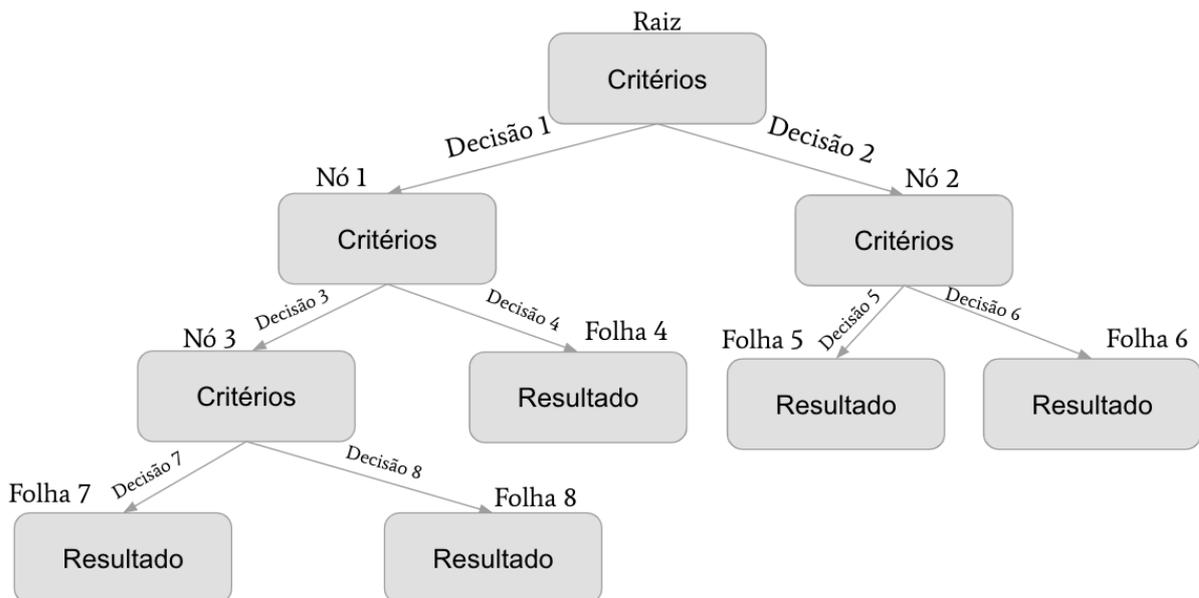
Ainda que criar *features* seja algo crucial, isso não é o suficiente para melhorar a aprendizagem do sistema, por isso existe uma vasta área de estudo aplicando engenharia de *features* (*Feature Engineering*). Os métodos englobam: criação, transformação, seleção, agrupamento e

binarização nas informações usadas como entrada do sistema. Nesse trabalho, tentamos adotar os conceitos que estavam em nosso alcance para esse trabalho.

### 2.4.2 *Decision Tree*

As *Decision Trees* (DTs) são estruturas de árvore que dividem os dados com base em características específicas, geralmente condições matemáticas como maior ou menor. O conceito dessas árvores precisar ser compreendido para dar um panorama de como elas são executadas no contexto de um sistema de ML e na previsão de preços de ações, apresentando suas vantagens e limitações. A Figura 2 pode ilustrar bem como é o funcionamento dessa estrutura e mostrar o porquê recebe o nome de "árvore", já que possui também raiz e folhas.

Figura 2: Estrutura Árvore de Decisão (DT)



Fonte: Produzido pelo próprio Autor

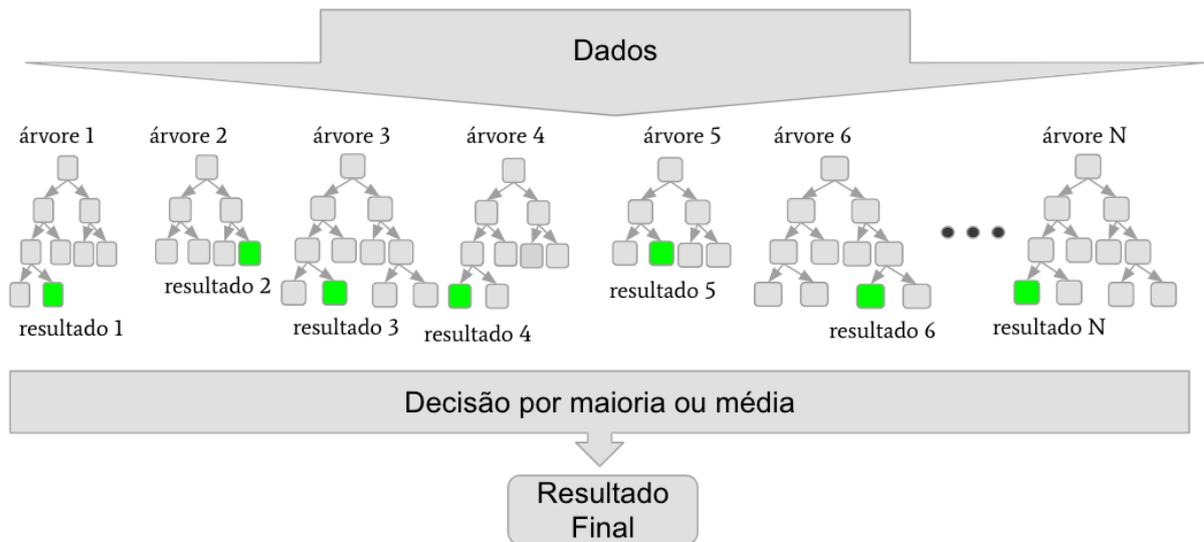
A raiz da árvore recebe um dado ou uma estrutura de dados e substanciado nesses valores, toma uma decisão. Com a decisão adequada para cada critério matemático do nó, chega-se em uma folha que simboliza um resultado final dos vários concebíveis. Assim, podemos entender que, a DT é um modelo de aprendizado supervisionado que toma decisões com base em regras simples de classificação derivadas dos dados. Ela é amplamente usada em finanças para tarefas como categorização de perfis de investidores, identificação de padrões de gastos e classificação de riscos financeiros.

#### 2.4.2.1 *Random Forest (RF)*

O *Random Forest* (RF) é um modelo de aprendizado de máquina baseado em DTs, conhecido por sua capacidade de lidar com *overfitting* e melhorar a precisão das previsões. No

contexto financeiro, o RF é amplamente adotado, tanto em previsão de preços de ações, quanto em detecção de fraudes, análise de risco de crédito e diversificação de portfólio. Atualmente, as vastas utilidades e os bons resultados trouxeram maior atenção a esse tema e conseqüentemente surge uma nova forma de pensar em resolver problemas matemáticos. Podemos observar a ideia na seguinte Figura:

Figura 3: Fluxo em Sistemas com *Random Forest*

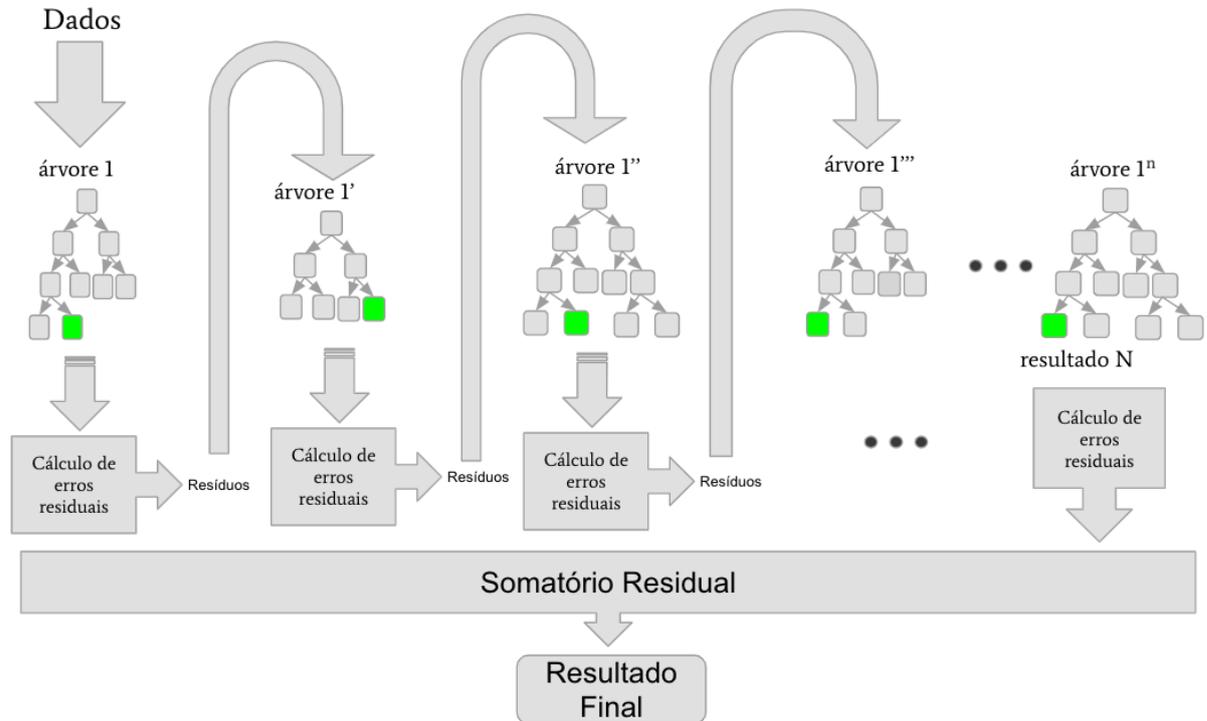


Fonte: Produzido pelo próprio Autor

De forma elegante, os modelos em RF desfrutam de inúmeras DTs, organizando uma floresta de árvores de decisão. As decisões são aderidas independentemente nessa construção, entretanto a quantidade de decisões pode trazer uma precisão melhor ao eleger um resultado final. Como contraponto, não temos interação entre as próprias decisões, já que as DTs não se interferem.

#### 2.4.2.2 XGBoost e sua base em Gradient Boosting

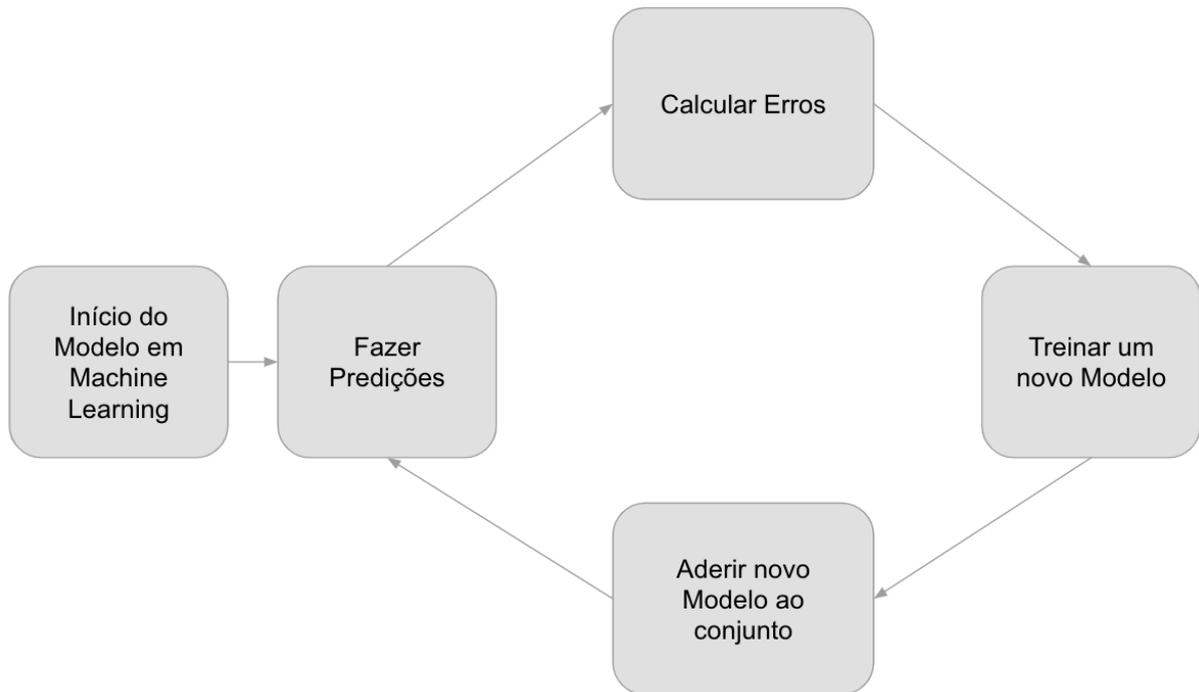
O *Gradient Boosting* é uma técnica que combina vários modelos mais fracos para criar um modelo robusto e preciso. É especialmente útil para previsões de séries temporais financeiras e análise de risco. O *Gradient Boosting* pode ser empregado para aperfeiçoar as previsões de preços das ações e identificar padrões ocultos nos dados financeiros.

Figura 4: Fluxo no Sistema *XGBoost*

Fonte: Produzido pelo próprio Autor

O conceito de reuso das DTs e seus erros residuais pode ser observada na Figura 4. Esses algoritmos de ML são importantes para a construção de modelos preditivos no mercado financeiro, permitindo a análise de padrões complexos e fornecendo informações valiosas que podem orientar as estratégias de investimento e tomada de decisões no ambiente financeiro. A seleção criteriosa do algoritmo adequado pode impactar significativamente a precisão e a confiabilidade das previsões, sendo essencial para o sucesso das operações financeiras.

Além dos algoritmos mencionados, o *Extreme Gradient Boosting (XGBoost)*, que pode-se traduzir como Tunagem Extrema de Parâmetros, merece destaque especial no contexto do mercado financeiro. Pois, atualmente ele representa uma extensão poderosa da técnica de *Gradient Boosting*, vastamente recomendada e válida devido à sua eficácia e desempenho excepcional, comparado aos modelos atuais. Em resumo, temos os processos do *XGBoost* na Figura a seguir:

Figura 5: Forma resumida de agregar DTs em modelos *XGBoost*

Fonte: Produzido pelo próprio Autor

O *XGBoost* compartilha os princípios básicos do *Gradient Boosting*, sendo uma técnica que gera grupos combinando múltiplos modelos menos complexos, geralmente as próprias árvores de decisão (mencionadas anteriormente como DTs), para formar um modelo robusto e preciso. Por outro lado, o *XGBoost* cria melhorias significativas, como regularização para evitar *overfitting*, uma consequência indesejada para estudiosos que desferem ML em seus sistemas. Com essa tunagem interna dos parâmetros, ainda torna capaz o tratamento eficaz de valores ausentes e uma abordagem otimizada para o treinamento de modelos.

No contexto financeiro, o *XGBoost* é comumente adotado por sua habilidade de prever os preços das ações com alta precisão e por sua viabilidade em lidar com conjuntos de dados complexos, por exemplo, séries temporais financeiras. Ele oferece uma ferramenta valiosa para aprimorar as previsões e identificar padrões ocultos nos dados financeiros. Em sua aplicação é facilitada o entendimento da importância de cada *feature* usada como entrada no sistema.

Assim, a seleção criteriosa de entradas do sistema que recorre ao *Gradient Boosting*, mais precisamente em um sistema com *XGBoost*, é crucial, pois ela pode impactar de forma substancial na precisão e na confiabilidade das previsões no mercado financeiro. Cada técnica possui suas próprias vantagens e cenários de usabilidade, e com essa compreensão aprofundada de suas características foi essencial a escolha adequada para a otimização dos resultados.

# 3

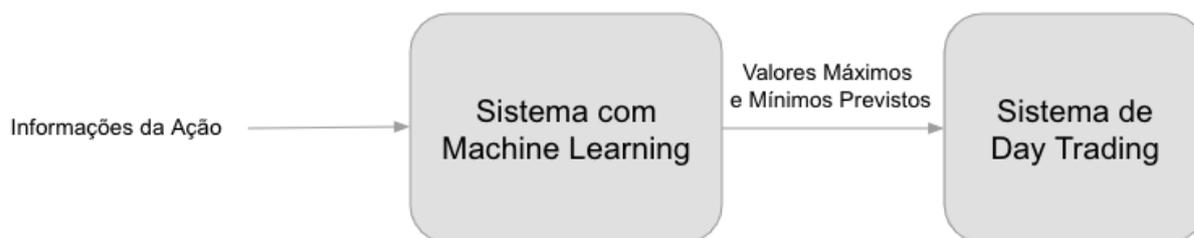
## METODOLOGIA

Na sessão presente irá ser descrita com detalhes a ideação e os processos utilizados em que temos como base nessa pesquisa. Além disso, usando como referência os trabalhos anteriores, iremos enfatizar as modificações necessárias feitas para os devidos resultados de métricas de avaliações desse trabalho.

### 3.1 MODELOS E ESTRUTURA DO TRABALHO

Ao ter em consideração que a pesquisa feita por [ALBUQUERQUE FILHO](#) fez uma avaliação inicial do método proposto por [MARTINEZ](#), onde o método consiste em prever os valores máximo e mínimo de uma ação para tomar decisão de compra e venda em busca do maior lucrar em momentos precisos, com a ajuda de métodos de ML nestas predições. E pode-se notar que [ALBUQUERQUE FILHO](#) executa uma ideação com todos os passos fundamentais para execução do sistema completo, com a obtenção dos dados, aplicação do modelo com ML e também o sistema de operações com transações de compra e venda que seriam executadas em uma corretora. O modelo visual pode ser entendido com a seguinte imagem:

Figura 6: Diagrama do Sistema



Fonte: Produzido pelo próprio Autor

Na Figura 6, mostra-se que o sistema com o processamento de dados em ML recebe de entrada as informações da ação do mercado. Essas informações para nosso cenário são valores da mesma em um período entre o dia 01/08/2017 e o dia 01/06/2021, seguindo o padrão usado

no trabalho proposto por [ALBUQUERQUE JUNIOR](#) com o sistema de ML baseado em RF. Porém, falaremos com mais detalhes sobre esses dados em seguida.

O trabalho em uma visão abstrata é segmentado em dois sistemas como pode ser observado na Figura 6. Um sistema representa a análise, estudo e preparação das informações iniciais, também como a implementação, avaliação e ajustes da predição dos valores máximos e mínimos usando a inteligência do processo de ML. Em outra parte, tem-se o sistema com foco em receber os parâmetros preditos para simular as operações de compra e venda na bolsa de valores, ou seja, o processo de execução do *Day Trading*. Esse sistema é predefinido com regras e estratégias que indicam os momentos de compra e venda, baseado nas informações recebidas previamente nas predições.

### 3.1.1 Sistema com *Machine Learning*

No presente trabalho, os esforços em aperfeiçoamento foram direcionados primordialmente aos detalhes do sistema de processamento de dados e previsão, com o modelo de aprendizagem conhecido como *XGBoost*. Entraremos em algumas especificidades no desenvolvimento atingido por [ALBUQUERQUE JUNIOR](#) no bloco do sistema de *Day Trading* que já é eficiente o suficiente para as regras que desejamos. Por esses motivos, as explicações práticas gerais desse trabalho serão mais destinadas aos processos envolvendo o primeiro bloco da Figura 6.

O bloco em peculiar com a inteligência da máquina possui a implementação do tratamento e da preparação dos dados, a criação de *feature* ou novas colunas na tabela de dados, melhoria de parâmetros para o modelo de treinamento e avaliação em métricas percentuais das predições em valores de teste. Então, as análises, resultados, discussões e conclusões são relacionadas aos avanços atingidos nesse bloco do sistema em comparação aos trabalhos desenvolvidos pelos outros pesquisadores já citados.

### 3.1.2 Sistema Trade Avançado

O sistema de *Day Trading* foi iniciado no trabalho de [ALBUQUERQUE FILHO](#) e aperfeiçoado por [ALBUQUERQUE JUNIOR](#) com novas regras e melhorias necessárias tomando como fundamentação as informações colhidas em seus estudos. O novo sistema aperfeiçoado não muda a estrutura da Figura 6, pois suas novas regras e melhorias são executadas internamente ao bloco do sistema mencionado. As melhorias englobaram refinamento das regras que não eram contempladas genuinamente. O sistema foi nomeado como Sistema *Trade Avançado* (STA).

No STA as principais regras definidas são: Transações baseadas no preço máximo e mínimo, justo que as predições são necessárias do sistema com ML para uma execução precisa da operação prevendo o maior lucro baseado em cada *candle* de 15 minutos nessa tomada de decisão; Transações realizadas conforme a tendência da ação, onde o sistema se prepara para a lógica de compra e venda tentando entender a variação do período seguinte e se adapta melhor

a dias atípicos nas variações da ação analisada, diminuindo a tendência de falhas; e Prejuízo limitado unicamente pelo *Stop Loss*, já que essa regra na prática é uma obrigação a ser seguida em momentos precisos e havia sido observado um *delay* nessa tomada de decisão, o que estava causando um prejuízo maior que o desejado, que havia sido previsto pelo sistema.

## 3.2 COLETA E PREPARAÇÃO DE DADOS

Para este estudo, os dados foram obtidos de uma fonte confiável de informações financeiras, que inclui históricos de preços de ações das empresas listadas na bolsa de valores. Os dados brutos incluíam informações diárias sobre o preço de abertura, fechamento, máxima e mínima das ações, bem como volumes de negociação.

Antes de usar os dados, realizamos um processo de pré processamento que envolveu limpeza dos dados, ou seja, a identificação e tratamento de outliers ou inconsistências e remoção de pontos de dados ausentes. Entretanto, em sua maioria os dados já estavam sanitizados, visto que a fonte de coleta dos dados é realmente confiável. O volume de negociação não foi uma informação utilizada nesse projeto, por isso foi indispensável a reestruturação dos dados em novas colunas do que seria útil ao sistema.

Os dados utilizados estão em um conjunto do período entre dia 01/08/2017 e 01/06/2021 das ações PETR4, VALE3, ABEV3, BBDC4, ITUB4, EGIE3 e JBSS3, que em suma foram obtidos do site ([GOOGLE/FINANCE](#)). Com as seguintes informações: data; valor de abertura; valor de fechamento; valor máximo; e valor mínimo, somos capazes de construir todas as informações extras fundamentais num processo de engenharia de *Features*.

## 3.3 IMPLEMENTAÇÃO DO *XGBOOST* NO SISTEMA DE *MACHINE LEARNING*

Nesta seção do trabalho será descrito os caminhos cruciais executados no percorrer da pesquisa para atingir os resultados observados. Entrando de forma objetiva e subjetiva nas alterações fundamentais, para que trabalhos futuros tenham em mãos a capacidade de construir o mesmo sistema para testes mais abrangentes caso essa seja a aspiração.

### 3.3.1 Adaptações para *XGBoost*

O funcionamento do sistema do modelo com *XGBoost* é semelhante ao do sistema com RF. Primordialmente, usando a linguagem de programação *Python*, ambos modelos seguem os mesmos passos: importação do código da biblioteca do modelo; criação do modelo; ajustes dos parâmetros antes do início do treinamento; execução do treinamento com os dados preparados; e por fim, a realização das predições.

Por outro lado, é necessário ter em mente que o formato do treinamento dos modelos que usam RFs sofrem mais com os efeitos colaterais nas alterações dos seus parâmetros de treinamento iniciais. De forma resumida, RF usa inúmeras DTs e elas usam os mesmos parâmetros, para por fim o resultado da decisão ser o resultado mais frequente entre todas "mini" decisões de cada DT. Sabendo disso, não houve a possibilidade de replicar os esforços feitos no trabalho de [ALBUQUERQUE JUNIOR](#) para esse projeto com *XGBoost*. Então para propósito de comparação, será difícil levar o ponto de aperfeiçoamento de parâmetros como algo relevante para o impacto nos resultados.

### 3.3.2 Criação de *Features* e Engenharia de *Features*

Para alimentar o modelo *XGBoost*, foram criadas várias *features* com base nas informações disponíveis iniciais *Open-High-Low-Close* (OHLC). Isso incluiu a construção de métodos exclusivos para cálculo de métricas específicas para cada dia. Um dos métodos necessários foi para produzir MMA e EMA para diferentes janelas de tempo, como 5, 9 e 21 dias. Um padrão do mercado é utilizar a coluna de fechamento (*close*), como base para o cálculo dos valores de EMA, entretanto para nosso caso, fomos mais profundamente e no método em que geramos a coluna com os valores EMAs recebemos a coluna que deseja-se usar de base e o  $n$  para o período do cálculo. Além disso, foram gerados métodos de cálculo das características técnicas, como o RSI e as BBs.

A engenharia de *features* também envolveu a criação de variáveis temporárias para reestruturar a tabela completamente com apenas as colunas desejadas. Para melhorar o desempenho do modelo, aplicamos técnicas de engenharia de *features*, como a normalização de algumas dessas variáveis numéricas e a remoção de variáveis que não se mostraram relevantes durante nossos primeiros testes ou que não acrescentavam informações significativas.

Por padrão, inicialmente utilizamos a janela de 5 dias para os principais cálculos, onde de começo temos as 4 colunas da tabela original OHLC em 5 dias, resultando em 20 informações ou *features* diferentes. Para esta o padrão usado foi: `d-dia_coluna`.

Outra categoria de *features* criada foi com a equação 2.2, usada para calcularmos valores de EMAs. Como já mencionado anteriormente, os EMAs são baseados em um período  $n$ , que nesse caso são o número de dias que desejamos passar como parâmetro. Para as *features* em que na sua nomenclatura não possua o número do período explícito, significa que o período usado é 5 dias e que o valor dessa coluna não está normalizado. Já as *features* de base está explícita no nome de todas as colunas que usaram a equação 2.2. Esse cálculo está representado pelo seguinte padrão nos nossos resultados: `EMA_coluna`.

Seguindo o mesmo exemplo anterior com a equação 2.2, as colunas com o período  $n$  explícitos em seus nomes para calcular valores de EMA. Outro detalhe é que esses valores são normalizados em seus métodos de construção dos conteúdos das colunas. Estas *features* têm a seguinte estrutura: `EMA_n_coluna`. Nos resultados teremos exemplos com o período  $n$  com 5, 9

e 21 dias construindo *features* nomeadas como: *EMA\_21\_close*.

Para o tema BB que já foi destrinchado, temos os valores inicialmente não normalizados (equações 2.4 e 2.5) utilizando uma janela de 5 dias em colunas que seguem o padrão para banda superior: *UB\_coluna*; e para banda inferior: *LB\_coluna*. A proximidade, entre as bandas superior e inferior resultam na largura da banda, o que nos deu a decisão de não utilizar como *feature* do sistema o valor de *BB\_WIDTH* e seu percentual *BB\_PERCENT*. Para nossos testes com valores normalizados, as colunas criadas no sistema foram apenas as bandas superior, inferior e média. Com o padrão a seguir: *BBANDS\_n\_coluna\_posição*, onde a "posição" pode ser os 3 casos: *LOWER* para banda inferior; *MIDDLE* para média; e *UPPER* para superior.

Outro valor calculado baseado em parâmetros externos de outras pesquisas foi o oscilador estocástico, sendo mencionado no trabalho feito por KHAIDEM *et al.* (3) em 2016. O nosso exemplo usa a seguinte equação para esse cálculo:

$$O_{estocastico} = 100 * \frac{F_{atual} - Min_5}{Max_5 - Min_5} \quad (3.1)$$

onde:

- *F<sub>atual</sub>* é o último valor mais atual de fechamento da análise;
- *Min<sub>5</sub>* é o mínimo valor em uma janela de 5 dias;
- *Max<sub>5</sub>* é o máximo valor em uma janela de 5 dias;

O retorno é sempre um valor entre 0 e 100, pois é um valor percentual e seu valor segue a velocidade da oscilação da ação, possivelmente entregando indícios sobre os valores máximos e mínimos futuros. Nesse trabalho a nomenclatura entregue para essa *feature* é *OSCILADOR%K\_5\_close*.

Recorrendo ao RSI explicado previamente, construímos uma coluna com o seguinte padrão *RSI\_n\_coluna*, sendo *n* o período de dias observados para comparação do aumento e diminuição. Esse cálculo foi feito também usando a biblioteca *Pandas* mencionada anteriormente.

Por fim, precisa-se ressaltar algumas informações sobre as colunas simples do conjunto OHLC, que representam os valores originais, onde temos o valor de abertura do dia exato (*open*), o valor máximo, mínimo e de fechamento desse mesmo dia atingido pela ação, respectivamente *high*, *low* e *close*. É necessário enfatizarmos aqui que não podemos utilizar os valores máximo, mínimo e de fechamento como *features* do nosso sistema, justo que máximo e mínimo são os valores exclusivamente desconhecidos para a predição do dia desejado. Além disso, o valor de fechamento também é desconhecido, pois, só acontece no fim do dia analisado. O valor de fechamento não foi aprofundado nessa pesquisa, contudo é analisado em predições do valor de abertura do dia seguinte de outros trabalhos como de MARTINEZ. Então, em síntese a única *feature* que será vista do conjunto OHLC será a abertura ou *open*.

De início criamos um conjunto de *features* específicas para entender os seus desempenhos dos treinamentos em importâncias para cada *feature* e MAPE dos conjuntos. Um conjunto possui 15 colunas, seguindo os padrões mencionados anteriormente, a lista dessas *features* segue abaixo: *open*; *EMA\_high*; *EMA\_low*; *UB\_close*; *LB\_close*; *d-1\_high*; *d-1\_low*; *d-2\_high*; *d-2\_low*; *d-3\_high*; *d-3\_low*; *d-4\_high*; *d-4\_low*; *d-5\_high*; e *d-5\_low*.

Outro conjunto usado para comparação nos testes possui 25 *features* *open*; *EMA\_high*; *EMA\_low*; *UB\_close*; e *LB\_close*; *d-1\_open*; *d-1\_high*; *d-1\_low*; *d-1\_close*; *d-2\_open*; *d-2\_high*; *d-2\_low*; *d-2\_close*; *d-3\_open*; *d-3\_high*; *d-3\_low*; *d-3\_close*; *d-4\_open*; *d-4\_high*; *d-4\_low*; *d-4\_close*; *d-5\_open*; *d-5\_high*; *d-5\_low*; e *d-5\_close*.

Já o terceiro conjunto possui 33, sendo as seguintes: *open*; *EMA\_open*; *EMA\_high*; *EMA\_low*; *EMA\_close*; *UB\_open*; *LB\_open*; *UB\_high*; *LB\_high*; *UB\_low*; *LB\_low*; *UB\_close*; *LB\_close*; *d-1\_open*; *d-1\_high*; *d-1\_low*; *d-1\_close*; *d-2\_open*; *d-2\_high*; *d-2\_low*; *d-2\_close*; *d-3\_open*; *d-3\_high*; *d-3\_low*; *d-3\_close*; *d-4\_open*; *d-4\_high*; *d-4\_low*; *d-4\_close*; *d-5\_open*; *d-5\_high*; *d-5\_low*; e *d-5\_close*.

Foi observada uma melhora nos MAPE com o aumento do número de *features* nesse cenário. Por isso, seguiu-se com a tentativa de adicionar novas *features* que pudessem trazer ainda melhores resultados no sistema. Contudo, muitas tentativas, que adicionam índices exteriores aos valores relacionados à própria ação, prejudicam o desempenho do MAPE. Entendemos que isso ocorreu pela adição de conteúdos não polidos o suficiente para mostrar relação direta com as previsões necessárias. Ao fazer a remoção de colunas lidas como menos relevantes para o sistema no seu treinamento, o resultado dos MAPes volta ao esperado.

### 3.3.3 Aperfeiçoamento dos Parâmetros

Os parâmetros do modelo *XGBoost* foram ajustados usando um processo de otimização. Realizamos uma pesquisa em grade para encontrar a combinação ideal de hiperparâmetros, incluindo os principais analisados: (**eta**; **booster**; **base\_score**; e **n\_estimators**). Onde o **eta** é uma taxa de aprendizado que afeta os pesos de cada passo da aprendizagem, o método **booster** que representa o formato do reforço que seria aplicado às árvores, o formato da pontuação base do viés global chamado de **base\_score** que tem mais efeito nas primeiras interações de aprendizagem do modelo e o critério **n\_estimators** que é o número de aplicações da tunagem das DT.

Utilizando como referência a ação PETR4 foram feitas várias execuções obtendo valores MAPE para cada caso. Com os parâmetros inalterados mantendo alguns padrões iniciais que são: **booster** como "*gbtree*"; **base\_score** com "0.5"; **n\_estimators** com "500"; e **eta** com "0.3", os resultados de teste foram observados. Em seguida, tentou-se todas as combinações possíveis com quatro possibilidades de **eta**: 0, 0.3, 0.1 e 0.5; com duas possibilidades de **boosters**: *gblinear* e *gbtree*; com quatro possibilidades de **base\_score**: 0.5, 0.4, 0.3 e 0.2; e com quatro possibilidades de **n\_estimators**: 100, 500, 1000 e 2000. Totalizando 128 combinações plausíveis.

O método encarregado de executar todas essas possibilidades retornou a lista dos melhores parâmetros baseados nos melhores MAPEs em ambos casos de valores máximos e mínimos. Os valores finais nesse processo foram utilizados em diante para o restante das execuções de treinamento do modelo e apresentaremos eles nos resultados da pesquisa. Essa otimização foi realizada com base no desempenho do modelo no conjunto de teste que por padrão utilizamos do dia 01/08/17 à 31/05/20.

### 3.4 AVALIAÇÃO DA PRECISÃO DAS PREVISÕES

Para avaliar a precisão das previsões geradas pelo modelo *XGBoost*, dividimos nossos dados em um conjunto de treinamento e um conjunto de teste. O conjunto de teste está definido entre o dia 01/08/17 e o dia 31/05/20, já o de teste está variando do dia 01/06/20 ao dia 01/06/21, ou seja, um período de 34 meses de treinamento (equivalente a 74% do conjunto de dados) e 12 meses de teste (resultando em 26% do conjunto de dados). O modelo foi treinado no conjunto de treinamento e, em seguida, suas previsões foram avaliadas no conjunto de teste utilizando como referência a métrica de erro que falaremos em seguida.

### 3.5 MÉTRICAS DE AVALIAÇÃO

#### 3.5.1 *Mean Absolute Percentage Error (MAPE)*

Utilizamos o MAPE como a métrica principal para avaliar o desempenho do modelo, já que o MAPE é calculado como a média das porcentagens de erro absoluto entre as previsões do modelo e os valores reais. Um MAPE mais baixo indica um melhor desempenho do modelo. Decidimos desfrutar desse cálculo pela facilidade de observação e por ter muitos valores de referências nos trabalhos citados. Nossos resultados mostraram que o MAPE médio para o valor máximo para o melhor sistema foi da mesma magnitude do valor mínimo. Essas métricas foram comparadas com os resultados anteriores obtidos com o regressor RF desenvolvidos por [ALBUQUERQUE FILHO](#) e [ALBUQUERQUE JUNIOR](#), demonstrando uma melhoria significativa no desempenho do sistema da inteligência em ML.

# 4

## RESULTADOS E DISCUSSÃO

Esta seção abrange os resultados obtidos por meio da implementação do modelo *XGBoost* no sistema de Day Trading, bem como a discussão detalhada dos achados. A análise apresentada aqui tem como objetivo fornecer insights sobre o desempenho do modelo, sua eficácia em prever valores máximos e mínimos diários de ações e suas capacidades de superar abordagens anteriores. Além disso, serão examinadas as previsões específicas de valores máximos e mínimos das ações, relacionadas com os parâmetros e *features* utilizadas em cada cenário. Assim, permitindo uma estimativa e uma avaliação mais completa das capacidades do modelo em um contexto de negociação de curto prazo.

O desempenho do modelo é avaliado com base em métricas como MAPE e na análise das previsões geradas para diferentes ativos financeiros. A comparação com as abordagens anteriores, onde uma delas empregou o processo de compra e venda no período de um dia com ANN no método desenvolvido por [MARTINEZ](#), em que [ALBUQUERQUE FILHO](#) fez a reprodução e o comparou com experimentos baseados em RF, mostrando as vantagens e desvantagens de cada modelo. Utilizando esse exemplo de modelo de RF, pode-se observar uma relação direta da diminuição do MAPE com o resultado melhor do sistema de preditor. O que permite avaliar como possibilidade de improvisação do sistema de *Day Trading* em melhoras nos resultados anteriores, caso a previsão baseada no MAPE obtenha uma melhora significativa também.

Na Tabela 7 do trabalho desenvolvido por [ALBUQUERQUE FILHO](#) pode-se acompanhar a evolução do MAPE para os cenários experimentados em que o menor MAPE foi para a ação PETR4 com 33 variáveis de entrada, onde 1,01% foi o MAPE do valor máximo e 1,00% foi o MAPE do valor mínimo. Já para a ação VALE3 o menor MAPE foi com apenas 3 variáveis e seus valores para o valor máximo e mínimos, foram os respectivos, 1,17% e 1,16%.

Um estudo subsequente realizado por [ALBUQUERQUE JUNIOR](#) investigou o desempenho do regressor em RF nos resultados apresentados. Na Tabela 4.2 desse estudo, é possível observar que a ação EGIE3 alcançou o melhor desempenho ao utilizar 33 variáveis, resultando em um MAPE de 0,75% para o valor máximo e 0,68% para o valor mínimo. Por outro lado, a ação ABEV3, também com 33 variáveis, apresentou um MAPE mais elevado, atingindo 1,3% para o valor máximo e 2,33% para o valor mínimo.

Pode-se observar que as melhorias levando em consideração o MAPE não são tão

expressivas entre esses dois trabalhos referenciados anteriormente. Com essa visão, o foco desse trabalho manteve-se em gerar um MAPE consideravelmente baixo e também estável para o modelo *XGBoost*. Além disso, a análise das previsões de valores máximos e mínimos dos trabalhos anteriores ofereceram insights valiosos sobre a capacidade de um modelo regressor baseado em DT, como RF, então foi considerado de grande valor experimentar um modelo novo da área que tenha uma base em decisões de DT. Juntos, esses tópicos fornecerão uma visão abrangente do impacto do modelo *XGBoost* nas operações de Day Trading e sua relevância para o ambiente financeiro atual.

## 4.1 DESEMPENHO DO *XGBOOST* NAS PREVISÕES DE TRABALHOS ANTERIORES

Neste tópico, apresentaremos uma análise detalhada do desempenho do modelo *XGBoost* nas previsões de valores máximos e mínimos diários das ações. Os resultados obtidos serão apresentados e discutidos em termos de precisão e confiabilidade das previsões. Serão destacados aspectos como acurácia média, desvio padrão e outras métricas relevantes que refletem o quão próximo o modelo foi capaz de prever os valores reais das ações. Além disso, serão fornecidos exemplos concretos de previsões corretas e incorretas, a fim de ilustrar as capacidades e limitações do modelo *XGBoost*.

### 4.1.1 Desempenho do *XGBoost* referente a ações e variáveis em comum

Para os primeiros experimentos foram utilizadas as mesmas ações de referência em [ALBUQUERQUE FILHO](#), que utilizou a ação Petrobrás de código PETR4 para os principais testes. Primeiramente, experimentamos utilizar 15, 25 e 33 variáveis de entrada seguindo o padrão adotado por [ALBUQUERQUE JUNIOR](#). E os resultados podem ser acompanhados na Tabela 1 a seguir.

Tabela 1: Valores MAPE inicial com 15, 25 e 33 variáveis no sistema *XGBoost*

<b>PETR4</b>	<b>MAPE (Máximo)</b>	<b>MAPE (Mínimo)</b>
15 variáveis	1,44%	1,46%
25 variáveis	1,44%	1,41%
33 variáveis	1,43%	1,36%

Nesse caso os parâmetros padrões usados para o modelo a ser treinado do *XGBoost* foram utilizados como base para a configuração deste estudo. Esses parâmetros são amplamente documentados na biblioteca oficial do *XGBoost*, disponível em ([READTHEDOCS](#)). Pode-se notar que pouco se alterou no MAPE ao aumentar o número de variáveis, porém existe uma leve tendência a diminuição do MAPE para as 33 variáveis.

Sobre a importância de cada variável para cada modelo testado, utilizando o conjunto de 33 variáveis de entrada, diferente dos modelos em RF em que a variável mais importante foi o valor de abertura do dia da previsão (coluna *Open*) chegando a representar 91,88% no caso da PETR4. Isso mudou para o sistema XGBoost em cada caso de ação a variável mais importante foi diferente. A Tabela 2 a seguir pode representar isso.

Tabela 2: Porcentagem da importância das variáveis de entrada dos sistemas em *XGBoost* prevendo valores máximos

	<b>Primeira</b>	<b>Segunda</b>	<b>Terceira</b>
PETR4	open (com: 82,20%)	LB_open (com: 9,24%)	EMA_low (com: 4,10%)
VALE3	open (com: 82,34%)	EMA_close (com: 4,70%)	UB_open (com: 3,87%)
ABEV3	d-1_close (com: 39,22%)	EMA_close (com: 31,02%)	open (com: 26,14%)
BBDC4	d-1_close (com: 57,39%)	open (com: 20,00%)	EMA_close (com: 17,47%)
ITUB4	open (com: 69,67%)	d-1_close (com: 15,69%)	EMA_close (com: 4,38%)

#### 4.1.2 Desempenho do *XGBoost* referente a ações e índices externos

Em outro procedimento, foram executados testes com a adição de novas variáveis tomando como base os estudos anteriores [ALBUQUERQUE JUNIOR](#) e onde tenta-se utilizar indicadores externos ao contexto da ação em si, foram usados testes com quatro indicadores externos: valor do Dólar em Reais, Índice Bovespa da B3, *Volatility Index da Chicago Board Options Exchange* (VIX) e valor da criptomoeda Bitcoin (BTC-USD) em Reais. Porém, os resultados não foram satisfatórios para o sistema com *XGBoost*, igualmente como aconteceu no treinamento para RF nos resultados "Por fim, aplicamos a mesma metodologia para indicadores externos, Dólar, Bitcoin, IBovespa e VIX, porém, em nenhuma das ações essas *features* tiveram importância suficiente para aparecerem entre as 10 *features* mais importantes.[...]" [ALBUQUERQUE JUNIOR](#).

Para os resultados do sistema padrão com *XGBoost* utilizando os indicadores externos como variáveis adicionais, além das 33 variáveis iniciais, temos a Tabela 3 a seguir.

Tabela 3: Valores MAPE considerando indicadores externos como variáveis adicionais no sistema *XGBoost*

<b>Base PETR4</b>	<b>MAPE (Máximo)</b>	<b>MAPE (Mínimo)</b>
Ao adicionar Dólar	1,48%	1,53%
Ao adicionar VIX	1,61%	1,72%
Ao adicionar BitCoin	2,03%	1,69%
Ao adicionar BM&FBovespa	2,05%	1,83%

Para nível de comparação, mostrar apenas o resultado da PETR4 é o suficiente por enquanto, pois serve de referência em contraste aos outros estudos que também utilizaram essa

ação. A primeira impressão que esses resultados nos entregam é que independentemente de mudarmos o modelo de RF para XGBoost, os resultados do MAPE não irão variar tanto. Porém, ainda assim, chegamos a testar esses mesmos ambientes com outras ações, como por exemplo a VALE3, entretando o cenário não é esperançoso com relação ao MAPE. Então iremos evitar prolongar esse tema com mais resultados de outras ações. Para averiguação, outros resultados podem ser observados no Código Fonte desse trabalho.

Abordando as importâncias dos ambientes onde foram adicionados índices externos, o número de variáveis é duplicado, pois seguindo o padrão das 33 variáveis para ação, construímos o mesmo cálculo para as 33 variáveis baseadas no índice adicional. A Tabela 4 mostra o resultado das principais importâncias entre todas as 33 variáveis padrão e as 33 adicionais com base no índice desejado.

Tabela 4: Porcentagem da importância das variáveis dos sistemas em *XGBoost* considerando índices externos prevendo valores máximos

Base PETR4	Primeira	Segunda	Terceira
Dólar	open (com: 73,37%)	LB_open (com: 13,91%)	EMA_low (com: 6,28%)
VIX	open (com: 73,42%)	d-1_low (com: 7,26%)	LB_open (com: 7,13%)
BitCoin	open (com: 76,47%)	LB_open (com: 10,89%)	EMA_low (com: 6,17%)
BM&FBovespa	open (com: 72,91%)	LB_open (com: 11,18%)	EMA_low (com: 5,96%)

### 4.1.3 Resultados e Discussão com *XGBoost* e seus aprimoramentos

Após o algoritmo de procura dos melhores parâmetros para a biblioteca usada do *XGBoost*, os parâmetros ideais resultantes, para a maioria das ações, foram: (**eta: 0.5; booster: gblinear; base\_score: 0.4; n\_estimators: 2000**). Com essa nova configuração do modelo de treinamento e mantendo as 33 variáveis de entrada do sistema, os MAPE para os valores máximos e mínimos chegaram a ter resultados melhores. Como pode-se acompanhar na Tabela 5 seguinte.

Tabela 5: Valores MAPE com alteração dos parâmetros de treinamento no sistema *XGBoost*

	MAPE (Máximo)	MAPE (Mínimo)
PETR4	0,96%	0,99%
VALE3	0,79%	0,82%
ABEV3	1,03%	1,02%
BBDC4	1,00%	0,91%
ITUB4	0,99%	0,92%

Utilizando os novos parâmetros para o modelo, observamos a melhora no MAPE, já os resultados das variáveis mais importantes acabou seguindo um padrão indesejado com 'open' na primeira posição em todas as execuções, por outro lado, o valor da porcentagem não foi tão destoante como na execução com indicadores externos, pode-se ver isso na seguinte Tabela 6.

Tabela 6: Porcentagem da importância das variáveis de entrada dos sistemas em *XGBoost* com parâmetros iniciais alterados

	<b>Primeira</b>	<b>Segunda</b>	<b>Terceira</b>
PETR4 Máximo	open (31,56%)	d-2_open (16,53%)	EMA_high (6,09%)
PETR4 Mínimo	open (29,64%)	d-2_open (16,28%)	EMA_low (6,15%)
VALE3 Máximo	open (30,73%)	d-2_open (17,09%)	EMA_high (5,38%)
VALE3 Mínimo	open (30,51%)	d-2_open (17,09%)	EMA_low (5,79%)
ABEV3 Máximo	open (27,42%)	d-2_open (19,64%)	EMA_high (2,82%)
ABEV3 Mínimo	open (27,49%)	d-2_open (19,46%)	EMA_low (3,10%)
BBDC4 Máximo	open (28,46%)	d-2_open (17,89%)	EMA_high (4,53%)
BBDC4 Mínimo	open (27,36%)	d-2_open (16,72%)	EMA_low (4,36%)
ITUB4 Máximo	open (28,24%)	d-2_open (17,70%)	EMA_high (4,12%)
ITUB4 Mínimo	open (27,60%)	d-2_open (16,64%)	EMA_low (4,06%)
EGIE3 Máximo	open (26,20%)	d-2_open (18,75%)	EMA_high (4,05%)
EGIE3 Mínimo	open (26,54%)	d-2_open (18,36%)	EMA_low (3,58%)
JBSS3 Máximo	open (15,07%)	EMA_high (11,83%)	d-2_open (10,42%)
JBSS3 Mínimo	open (15,79%)	EMA_low (10,65%)	d-2_open (10,00%)

Interessante observar que para as previsões de valores 'high' (Máximo) a variável EMA\_high está presente entre as três mais relevantes para essa previsão mais precisa, da mesma forma com a coluna 'low' (Mínimo) acontece a mesma equivalência de importância das variáveis dessa vez para o valor EMA\_low. A quarta posição em diante são ocupadas por variáveis diferentes para cada ação sem seguir um padrão. Apesar disso, as variáveis baseadas no valor EMA tendem a receber importâncias maiores.

#### 4.1.3.1 Aprimoramentos baseados em Engenharia de Feature

Em um momento posterior à definição dos parâmetros de treinamento do modelo, os esforços foram direcionados em encontrar as variáveis que de alguma forma ajudassem ainda mais da diminuição do MAPE. A construção dessas variáveis seguiu um padrão adotado por [ALBUQUERQUE FILHO](#). Onde se explica as variáveis com estrutura dX\_high, dX\_low, dX\_open, dX\_close, EMA\_high, EMA\_low, EMA\_open, EMA\_close, UB\_high, LB\_high, UB\_low, LB\_low, UB\_open, LB\_open, UB\_close e LB\_close. Da mesma forma, as variáveis que utilizam RSI\_X\_, EMA\_X\_ e OSCILADOR%K\_5\_close são baseados no estudo

de [ALBUQUERQUE JUNIOR](#), as quais já explicamos detalhadamente cada uma delas na seção de Metodologia. Já as variáveis com o padrão `BBANDS_X_close`, `BBANDS_X_high`, `BBANDS_X_low` e `BBANDS_X_open` são explicadas no setor de Fundamentação Teórica desse trabalho. Essa lista final contém 62 variáveis de entrada para o sistema, são elas:

1. *open*;
2. *BBANDS\_5\_close-BB\_LOWER*;
3. *BBANDS\_5\_close-BB\_MIDDLE*;
4. *BBANDS\_5\_close-BB\_UPPER*;
5. *BBANDS\_5\_high-BB\_LOWER*;
6. *BBANDS\_5\_high-BB\_MIDDLE*;
7. *BBANDS\_5\_high-BB\_UPPER*;
8. *BBANDS\_5\_low-BB\_LOWER*;
9. *BBANDS\_5\_low-BB\_MIDDLE*;
10. *BBANDS\_5\_low-BB\_UPPER*;
11. *BBANDS\_5\_open-BB\_LOWER*;
12. *BBANDS\_5\_open-BB\_MIDDLE*;
13. *BBANDS\_5\_open-BB\_UPPER*;
14. *d-1\_close*;
15. *d-1\_high*;
16. *d-1\_low*;
17. *d-1\_open*;
18. *d-2\_close*;
19. *d-2\_high*;
20. *d-2\_low*;
21. *d-2\_open*;
22. *d-3\_close*;
23. *d-3\_high*;
24. *d-3\_low*;
25. *d-3\_open*;
26. *d-4\_close*;
27. *d-4\_high*;
28. *d-4\_low*;
29. *d-4\_open*;
30. *d-5\_close*;
31. *d-5\_high*;
32. *d-5\_low*;
33. *d-5\_open*;
34. *EMA\_21\_close*;
35. *EMA\_21\_high*;

36. *EMA\_21\_low*;
37. *EMA\_21\_open*;
38. *EMA\_5\_close*;
39. *EMA\_5\_high*;
40. *EMA\_5\_low*;
41. *EMA\_5\_open*;
42. *EMA\_9\_close*;
43. *EMA\_9\_high*;
44. *EMA\_9\_low*;
45. *EMA\_9\_open*;
46. *EMA\_close*;
47. *EMA\_high*;
48. *EMA\_low*;
49. *EMA\_open*;
50. *LB\_close*;
51. *LB\_high*;
52. *LB\_low*;
53. *LB\_open*;
54. *OSCILADOR%K\_5\_close*;
55. *RSI\_9\_close*;
56. *RSI\_9\_high*;
57. *RSI\_9\_low*;
58. *RSI\_9\_open*;
59. *UB\_close*;
60. *UB\_high*;
61. *UB\_low*;
62. *UB\_open*.

Com essa lista de variáveis de entrada para o treinamento do modelo em *XGBoost* o mais interessante de se acompanhar primeiramente nos resultados são as importâncias das variáveis já que dessa vez são sessenta e duas e nos casos anteriores a variável 'open' estava se destacando. Por isso mostraremos de imediato as importâncias dos modelos treinados na Tabela 7.

Tabela 7: Porcentagem da importância das variáveis de entrada dos sistemas em *XGBoost* com parâmetros iniciais alterados e 62 variáveis de entrada

	<b>Primeira</b>	<b>Segunda</b>	<b>Terceira</b>
PETR4 Máximo	open (27,92%)	d-5_low (17,99%)	EMA_high (8,63%)
PETR4 Mínimo	open (25,46%)	d-5_low (18,76%)	EMA_high (6,70%)

Apesar desse teste ter sido executado com sete ações (PETR4, VALE3, ABEV3, BBDC4,

ITUB4, EGIE3 e JBSS3), os resultados da PETR4 já conseguem ilustrar muito bem o padrão de comportamento das importâncias dessa listagem. Em todos os casos a variável mais importante continuou por ser a 'open' e os segundo e terceiro lugares foram ocupados respectivamente pelas mesmas variáveis que a Tabela 7, a única diferença foi com a execução da ABEV3 em que o terceiro lugar da previsão do valor Mínimo foi a variável 'EMA\_low'. Outro detalhe a ser observado é que das sete ações treinadas nesse sistema, a variável 'open' que recebeu menor importância teve o valor 25,46% (com a PETR4 prevendo Mínimo) e a variável 'open' que recebeu maior importância teve o valor 34,58% (com a VALE3 prevendo Máximo).

Contudo, os valores MAPE para essas execuções conseguem um avanço significativo. Os erros não atingem a faixa do 1,00% como pode-se observar na Tabela 8.

Tabela 8: Valores MAPE com alteração dos parâmetros de treinamento no sistema *XGBoost* e uso de 62 variáveis de entrada

	MAPE (Máximo)	MAPE (Mínimo)
PETR4	0,68%	0,63%
VALE3	0,57%	0,56%
ABEV3	0,78%	0,72%
BBDC4	0,66%	0,56%
ITUB4	0,76%	0,54%
EGIE3	0,48%	0,45%
JBSS3	0,50%	0,55%

Nesse contexto notou-se que independente da ação usada para treinamento do modelo, pode-se prever máximos e mínimos das outras seis ações ao passar as 62 variáveis e ainda ter um MAPE abaixo de 2%. Possível até de prever indicadores externos, como o Dólar ou *BitCoin*, com essa mesma precisão. Apenas existindo o esforço de calcular e construir a formatação das 62 variáveis de entrada do sistema.

Consideramos que com esse modelo de previsão o sistema de *Day Trading* conseguiria ter um resultado mais satisfatório anualmente por atingir precisão maior no MAPE, notando que o melhor resultado com RF foi com a ação EGIE3 onde o MAPE para prever o Máximo foi 0,75% e para prever o Mínimo 0,68%, enquanto nesse experimento temos 0,48% e 0,45%, respectivamente.

# 5

## CONCLUSÃO E TRABALHOS FUTUROS

### 5.1 SÍNTESE DOS RESULTADOS OBTIDOS

Em suma, os resultados do trabalho desenvolvido indicam que o uso do algoritmo *XGBoost* com parâmetros otimizados e o uso de técnicas do *Feature Engineering* pode resultar em previsões altamente precisas dos valores máximos e mínimos diários das ações. Não significando melhor desempenho no Sistema de *Day Trading*, pois seria necessário executar testes do sistema completo com os valores preditos pelo Sistema com *Machine Learning* em *XGBoost*.

Os resultados indicaram que, independentemente da ação usada para treinar o modelo, o MAPE ao prever os valores máximos e mínimos diários das ações ficou consistentemente abaixo de 2%. Isso sugere que o modelo *XGBoost*, com os parâmetros otimizados e as 62 variáveis de entrada, é capaz de prever com alta precisão os valores máximos e mínimos de ações, não apenas das estudadas, indicando ser um modelo generalista.

#### 5.1.1 Predições

A pesquisa identificou os melhores parâmetros para o algoritmo *XGBoost*, que incluem um valor de aprendizado 'eta' de 0.5, o uso do *booster* 'gblinear', um valor 'base\_score' de 0.4 e 2000 em 'n\_estimators'. Esses parâmetros otimizados resultaram em melhorias nos erros de previsão, medidos pelo MAPE, para os valores máximos e mínimos das ações analisadas.

Esse trabalho igualmente investigou as importâncias das variáveis no modelo *XGBoost* treinado. Em todas as execuções, a variável 'open' foi a mais importante, em seguida temos como de grande valor 'EMA\_low' e 'EMA\_high'. Essa consistência na importância das variáveis sugere que os cálculos para EMA desempenham um papel fundamental na previsão dos valores máximos e mínimos das ações.

## 5.2 CONTRIBUIÇÕES E IMPLICAÇÕES NO CAMPO FINANCEIRO

O atual estudo pode trazer benefícios para operações de *Day Trading* automatizadas ou até mesmo *Swing Trade*, onde a compra e venda de ações acontece em prazos de um dia ou mais, chegando até mesmo algumas semanas de diferença. Da mesma forma, esses baixos valores MAPEs em previsões de longo prazo podem trazer vantagens empresariais em uma análise financeira sobre o futuro de ações brasileiras, como as mencionadas durante este documento.

## 5.3 POSSÍVEIS DIREÇÕES FUTURAS

Uma direção promissora para trabalhos futuros é a expansão da janela de análise para o sistema de *Day Trading* automatizado. Até o momento, o nosso trabalho e os passados utilizaram um padrão de janelas de 5 dias para prever os valores máximos e mínimos das ações. No entanto, é interessante investigar se o aumento dessa janela, por exemplo, para 7, 10, 14 ou mais dias, pode levar a previsões com MAPEs ainda menores.

Isso pode permitir uma visão de médio a longo prazo das tendências do mercado e ajudar a melhorar a tomada de decisões de compra e venda. Por outro lado, aumentaria o número de variáveis do sistema, tornando-o mais complexo e com maior tempo de treinamento, requisitando um ambiente de treinamento robusto com maior espaço de armazenamento. Além de poder, causar *Overfitting*, onde o sistema não consegue ser generalista e analisar novos dados, pois está treinado ou condicionado apenas a uma entrada específica.

Outro caminho aberto para exploração seria considerar o modelo de aprendizagem para previsões a longo prazo. O cerne dessa ideia seria prever mais de um dia a frente, tentando encontrar valores máximos e mínimos de dois ou mais dias a frente. Para implementação com *XGBoost* existem dois caminhos viáveis para teste. O primeiro seria utilizar um modelo diferente para cada dia, onde as colunas 'high' e 'low' (máximo e mínimo) seriam substituídas pelas colunas dos dias subsequentes desejados. Caso o modelo deseje prever o segundo dia, o modelo seria treinado com os dados de 'd+1\_high' e 'd+1\_low', ou seja, os valores 'high' e 'low' do próximo dia, e assim por diante para as outras previsões. O segundo caminho possível, seria alterar a tabela de dados para inserir esses valores desejados, de acordo com o intervalo escolhido. Com essa alteração, tornaria viável a criação de um modelo que conseguiria prever uma lista de valores.

### 5.3.1 Sistema de *Day Trading*

Finalmente, à medida que se avança na automação do *Day Trading* e na previsão de longo prazo, é importante considerar as implicações éticas e regulatórias. Garantir que o sistema opere de maneira ética e esteja conforme as regulamentações financeiras é essencial. Para a aplicação em tempo real, onde se prevê dados não existentes de datas futuras, é fundamental desenvolver um sistema bem estruturado de avaliação de desempenho que permita monitorar

continuamente o lucro ou prejuízo do sistema automatizado. Isso ajudará a ajustar o sistema à medida que as condições do mercado mudam também levando em consideração as regras de operações na Bolsa de Valores. Consideramos possível até mesmo a previsão futura de moedas, como o Dólar ou criptomoedas como o *Bitcoin*, obtendo resultados MAPE com mesma faixa de variação dos valores atingidos ao prever ações brasileiras.

## REFERÊNCIAS

- [1] ALBUQUERQUE FILHO, A. C. d. (2021). Avaliação do método de random forest para indicar o melhor momento de compra e venda de ações no mercado financeiro. *UFPE, Centro de Informática*.
- [2] ALBUQUERQUE JUNIOR, L. A. d. (2022). Sistema automático para operações day trade baseado em previsão de valores máximos e mínimos com random forest. *UFPE, Centro de Informática*.
- [GOOGLE/FINANCE] GOOGLE/FINANCE. *Disponível em:*  
*<https://www.google.com/finance/>*, Acesso em: jan. 2023.
- [3] KHAIDEM, L., SAHA, S., & DEY, S. R. (2016). Predicting the direction of stock market prices using random forest. *Applied Mathematical Finance*.
- [4] MARTINEZ, L. C. (2009). From an artificial neural network to a stock market day-trading system: A case study on the bm&f bovespa. *Atlanta: IEEE Computer Society*.
- [READTHEDOCS] READTHEDOCS. Dmlc xgboost, xgboost parameters. *Disponível em:*  
*<https://xgboost.readthedocs.io/en/stable/parameter.html#xgboost-parameters>*, Acesso em: fev. 2023.