

INSTITUTO FEDERAL DE SANTA CATARINA

WELINTON HAAS

APLICANDO MODELOS DE APRENDIZADO DE
MÁQUINA ONLINE PARA TRADING NO MERCADO
FINANCEIRO

Caçador - SC

2022

WELINTON HAAS

APLICANDO MODELOS DE APRENDIZADO DE
MÁQUINA ONLINE PARA TRADING NO MERCADO
FINANCEIRO

Monografia apresentada ao Curso de Sistemas de Informação do Câmpus Caçador do Instituto Federal de Santa Catarina para a obtenção do diploma de Bacharel em Sistemas de Informação.

Orientador: Carlos Henrique Radavelli, Me.

Coorientador: Cristiano Mesquita Garcia, Me.

Caçador - SC

2022

Haas, Welinton

H112a Aplicando modelos de aprendizado de máquina online para trading no mercado financeiro / Welinton Haas ; orientador: Carlos Henrique Radavelli; coorientador: Cristiano Mesquita Garcia. -- Caçador, SC, 2022.

54 f.

Trabalho de Conclusão de Curso (Graduação)-Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, Curso de Sistemas de Informação.

Inclui bibliografias

1. Computação – Matemática. 2. Aprendizado do computador. 3. Sistemas de informação. I. Radavelli, Carlos Henrique. II. Garcia, Cristiano Mesquita. III. Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina. Curso de Sistemas de Informação. IV. Título.

CDD 658.4038

WELINTON HAAS

APLICANDO MODELOS DE APRENDIZADO DE MÁQUINA ONLINE PARA TRADING NO MERCADO FINANCEIRO

Este trabalho foi julgado adequado para obtenção do título de Bacharel em Sistemas de Informação, pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, e aprovado na sua forma final pela comissão avaliadora abaixo indicada.

Caçador - SC, 06 de dezembro de 2021:

Documento assinado digitalmente
 CARLOS HENRIQUE RADAVELLI
Data: 16/12/2022 07:37:46-0300
Verifique em <https://verificador.iti.br>

Carlos Henrique Radavelli, Me.
Orientador

Instituto Federal de Santa Catarina

Cristiano Garcia 
Digitally signed by Cristiano Garcia
DN: C=BR, O=Instituto Federal de Santa Catarina,
CN=Cristiano Garcia, E=cristiano.garcia@
ifsc.edu.br
Reason: Estou aprovando este documento
Location:
Date: 2022.12.16 08:59:44-03'00'
Foxit PDF Reader Version: 12.0.1

Cristiano Mesquita Garcia, Me.
Coorientador

Instituto Federal de Santa Catarina

Documento assinado digitalmente
 CARINE AZEVEDO DANTAS
Data: 19/12/2022 20:47:19-0300
Verifique em <https://verificador.iti.br>

Carine Azevedo Dantas Dra.
Banca Avaliadora

Instituto Federal de Santa Catarina

Paulo Roberto Córdova 
Digitally signed by Paulo Roberto Córdova
DN: C=BR, OU=Coordenadoria de Sistemas de Informação,
O=Instituto Federal de Santa Catarina, CN=Paulo Roberto
Córdova, E=paulo.cordova@ifsc.edu.br
Reason: De acordo com o conteúdo e acurácia do documento.
Location: Caçador
Date: 2022.12.19 21:58:47-03'00'
Foxit PhantomPDF Version: 10.1.0

Paulo Roberto Cordova, Me.

Banca Avaliadora

Instituto Federal de Santa Catarina

RESUMO

O Mercado Financeiro e de ações vêm despertando interesse dos brasileiros. Como indício, o número de contas criadas na Bolsa de valores oficial do Brasil (B3), que teve, em dezembro de 2021, um crescimento de 56% na comparação com dezembro de 2020. O interesse por ferramentas e técnicas para auxiliar nas operações deste mercado são conhecidas de longa data. Tais técnicas auxiliam acionistas, analistas e amadores a tomarem decisões mais qualificadas, auxiliando na gestão de riscos. Entretanto, com a popularização dos robôs *traders*, novas ferramentas como Aprendizado de Máquina estão tomando conta do cenário. A utilização de uma base histórica dos ativos permite o treinamento de um ou mais modelos de Aprendizado de Máquina para operar de forma “inteligente” e rápida, com operações na escala de milissegundos, além que operar ininterruptamente, aumentando assim as possibilidades de lucros. Dessa forma, espera-se, como resultado deste trabalho, o desenvolvimento de um robô capaz de realizar operações de forma automática utilizando dois modelos incrementais e, com base nas métricas, ser comparável ou superior a outros trabalhos da literatura de modo a obter de lucro nas operações propostas.

Palavras-chave: Aprendizado de Máquina, Aprendizado *Online*, Mercado Financeiro, Ações, *Trading*

ABSTRACT

The financial and stock markets have been attracting the interest of Brazilians. As an indication, the number of accounts created at B3, which had, in December 2021, a growth of 56% compared to December 2020. The interest in tools and techniques to assist in the operations of this market is known long time. Such techniques help shareholders, analysts and amateurs to make more qualified decisions, assisting in risk management. However, with the popularization of robots *traders*, new tools such as Machine Learning are taking over the scene. The use of a historical base of the assets allows the training of one or more Machine Learning models to operate in an “smart” and fast way, with operations in the millisecond scale, in addition to operating uninterruptedly, thus increasing the possibilities of profits . In this way, it is expected, as a result of this work, the development of a robot capable of performing operations automatically using two incremental models and, based on the metrics, being comparable or superior to other works in the literature in order to obtain profit in the proposed operations.

Keywords: Machine Learning, Online Learning, Financial Market, Stocks, Trading

LISTA DE ILUSTRAÇÕES

Figura 1 – Dinâmica das possíveis operações a serem realizadas pelo agente. . . .	17
Figura 2 – Gráfico utilizando Médias Móveis, indicador representado pela linha azul.	19
Figura 3 – Gráfico utilizando Bandas de Bollinger, indicador representado pelas linhas azuis.	19
Figura 4 – Gráfico utilizando Média Móvel Convergente e Divergente (MACD), Linha MACD (LM), representada pela linha de cor azul. Linha Sinal (LS), representada pela linha de cor preta.	20
Figura 5 – Fluxo da pesquisa	25
Figura 6 – Representação das possíveis classificações dos registros preditivos do modelo	29
Figura 7 – Modelo Naive Bayes - Visualização completa dos dados	34
Figura 8 – Modelo Naive Bayes - Detalhe de aproximadamente 500 registros . . .	34
Figura 9 – Modelo KNNRegressor - Detalhe de aproximadamente 500 registros . .	35
Figura 10 – Modelo KNNRegressor - Visualização completa dos dados	36
Figura 11 – KNNRegressor - Detalhe de aproximadamente 500 registros	36
Figura 12 – Modelo SoftmaxRegression - Visualização completa dos dados	37
Figura 13 – Modelo SoftmaxRegression - Detalhe de aproximadamente 500 registros	37
Figura 14 – Diagrama Funcionamento Robô Trader	39

LISTA DE TABELAS

Tabela 1	– Resultados da predição utilizando o modelo Naive Bayes - Fonte: Autor	34
Tabela 2	– Resultados da predição utilizando o modelo KNNRegressor - Fonte: Autor	35
Tabela 3	– Resultados da predição utilizando o modelo SoftmaxRegression - Fonte: Autor	36
Tabela 4	– Resultados dos experimentos dos modelos de classificação - Fonte: Autor	40
Tabela 5	– 3 melhores resultados dos testes em <i>loop</i> utilizando os papeis da IVVB11 - Fonte: Autor	41
Tabela 6	– 3 melhores resultados dos testes em <i>loop</i> utilizando os papeis da VALE3 - Fonte: Autor	41
Tabela 7	– 3 melhores resultados dos testes em <i>loop</i> utilizando os papeis da ABEV3 - Fonte: Autor	41
Tabela 8	– Melhores resultados dos experimentos objetivando maior lucratividade utilizando os papeis da IVVB11 e o modelo Classificador HoelfdingTree - Fonte: Autor	42
Tabela 9	– Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da IVVB11 e modelo Classificador HoelfdingTree - Fonte: Autor	43
Tabela 10	– Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papeis da IVVB11 e o modelo Classificador HoelfdingTree	43
Tabela 11	– Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da IVVB11 e modelo Classificador HoelfdingTree - Fonte: Autor	44
Tabela 12	– Melhores resultados dos experimentos objetivando maior lucratividade utilizando os papeis da ABEV3 e o modelo Classificador HoelfdingTree - Fonte: Autor	44
Tabela 13	– Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da ABEV3 e modelo Classificador HoelfdingTree - Fonte: Autor	45
Tabela 14	– Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papeis da ABEV3 e o modelo Classificador HoelfdingTree - Fonte: Autor	45
Tabela 15	– Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da ABEV3 e modelo Classificador HoelfdingTree - Fonte: Autor	45
Tabela 16	– Melhores resultados dos experimentos objetivando maior lucratividade utilizando os papeis da VALE3 e o modelo Classificador HoelfdingTree - Fonte: Autor	46
Tabela 17	– Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da VALE3 e modelo Classificador HoelfdingTree - Fonte: Autor	46

Tabela 18 – Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papeis da VALE3 e o modelo Classificador HoelffdingTree	
- Fonte: Autor	46
Tabela 19 – Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papeis da VALE3 e o modelo Classificador HoelffdingTree	
- Fonte: Autor	47

LISTA DE ABREVIATURAS

B3 Bolsa de valores oficial do Brasil

GPS Global Positioning System

IA Inteligência Artificial

LSTM Long Short-Term Memory

HFT High Frequency Trading

MD Mineração de Dados

API Application Programming Interface

VALE3 VALE S.A

ABEV3 Ambev S.A

IVVB11 iShares S&P 500 Fundo de Investimento

ANN Artificial Neural Network

NB Naive-Bayes

RF Random Forest

SVM Support Vector Machine

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Problema de Pesquisa	13
1.2	Hipótese de Pesquisa	13
1.3	Objetivos	13
1.3.1	Objetivo Geral	13
1.3.2	Objetivos Específicos	13
1.4	Justificativa	14
1.5	Organização do texto	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Mercado Financeiro	16
2.1.1	Ações	16
2.1.2	<i>Day Trade</i>	17
2.1.3	Formas de Análise	18
2.1.4	Indicadores	18
2.2	Aprendizado de Máquina	20
2.2.1	Tipos de Aprendizado	21
2.2.2	Algoritmos <i>Online</i>	21
2.2.3	Processamento de Fluxo	22
3	TRABALHOS RELACIONADOS	24
4	METODOLOGIA	25
4.1	Coleta de Dados	26
4.2	Processamento	26
4.2.1	Python	26
4.2.2	Pandas	27
4.2.3	River	27
4.3	Métricas	28

4.3.1	Precisão	29
4.3.2	Revocação	29
4.3.3	Acurácia	30
4.3.4	F1-Score	30
4.3.5	RMSE	30
4.3.6	R2 Score	31
4.4	Operação no Mercado Financeiro	31
5	RESULTADOS E DISCUSSÕES	32
5.1	Considerações Iniciais	32
5.2	Dados	32
5.3	Variáveis	32
5.4	Codificação	33
5.5	Experimentos	33
5.5.1	Treinamento dos Modelos Regressores utilizando o Dataset IVVB11	33
5.5.1.1	Naive Bayes	33
5.5.1.2	KNNRegressor	35
5.5.1.3	SoftmaxRegression	35
5.5.2	Robô Trader	37
5.5.2.1	Considerações Iniciais	37
5.5.2.2	Escolha dos modelos de classificação	40
5.5.2.3	Experimentos Iniciais	40
5.5.2.4	Experimentos com papéis da IVVB11	42
5.5.2.5	Experimentos com papéis da ABEV3	44
5.5.2.6	Experimentos com papéis da VALE3	45
6	CONCLUSÕES	48
	REFERÊNCIAS	50

1 INTRODUÇÃO

O dinheiro influencia diretamente em nossas escolhas profissionais e pessoais. Dessa forma, novos estudos na área de psicologia são realizados constantemente para avaliar a relação humana com dinheiro. Estudos apresentados por [Vohs, Mead e Goode \(2008\)](#) evidenciam como o conceito de dinheiro provoca uma mudança nas nossas relações interpessoais. Em uma das pesquisas do estudo de [Vohs, Mead e Goode \(2006\)](#), foi concedido aos participantes 2 dólares. Em seguida, eles tiveram a oportunidade de doar para um fundo de estudos da universidade. Foi identificado uma porcentagem de doações menor, cerca de 39% oriunda dos participantes lembrados da dificuldade e esforços para adquirir dinheiro, entretanto, 67% dos participantes que não tiveram esta lembrança exposta efetuaram a doação, mostrando assim o efeito que o dinheiro traz na motivação das pessoas.

Não há como citar dinheiro e não mencionar a bolsa de valores, que possibilita a qualquer pessoa negociar ações de corporações de capital aberto. A prática de compra e venda durante as oscilações de ações diante da possibilidade de ganhos financeiros é cada vez mais comum. Segundo [Laval \(2022\)](#), a B3 atingiu a marca de 5 milhões de contas de investidores pessoa física no início de 2022, o crescimento foi de 56% na comparação com dezembro de 2020. Negociar praticamente qualquer valor a curto, médio e longo prazo cria um ambiente propício para o crescimento de investidores e entusiastas em busca de novas fontes de renda, enriquecimento ou investimento. Conseqüentemente, a corrida por tecnologias que automatizem e aumentem a taxa de ganhos cresce diariamente.

No mercado financeiro, existem diversas modalidades de operações. Entre elas, estão o *Buy and Holding* que utiliza de uma análise fundamentalista onde os investidores se veem como sócios/donos da organização, visto que suas ações são de longo prazo ([SILVA, 2022](#)). Outra modalidade é o *Day Trade* onde os investidores buscam retornos rápidos interagindo com o mercado em transações

diárias que iniciam e terminam em um único dia (SPRITZER, 2017). Além de diversas outras técnicas, ainda é possível citar o *Swing Trade*, voltada para operações em curto prazo, exigindo maior precisão nas operações, uma vez que o mercado tem uma oscilação maior que a médio e longo prazo (ANDRADE, 2018).

Na era da informação, milhares de dados são gerados e armazenados a cada segundo, sejam eles por sensores, dispositivos *Smart*, *Global Positioning System* (GPS), redes sociais, sistemas de vigilância, mercados financeiros ou qualquer outro sistema gerador de dados (SEMELER, 2019). Tal volume de dados inviabiliza qualquer esforço humano para análise ou compreensão, enaltecendo uma tecnologia que têm avançando exponencialmente na última década: o Aprendizado de Máquina. Sua ascensão, além de outros fatores, está atrelada ao aumento do poder computacional e a possibilidade de coletar, armazenar e transmitir grandes volumes de dados (NETO; BONACELLI; PACHECO, 2021).

Segundo Monard e Baranauskas (2003), “Aprendizado de Máquina é uma área de Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática”. O Aprendizado de Máquina atualmente é alvo de pesquisas nos mais diversos âmbitos da ciência; Boeing et al. (2019) testaram a abordagem de Aprendizado de Máquina para ensinar um robô a julgar no âmbito judiciário. Já Amorim, Barone e Mansur (2008) conseguiu 91% de acurácia na previsão de evasão escolar; na área da medicina Takakura et al. (2018), apresentou resultados na casa dos 90% para identificar patologias em diagnósticos médicos.

Dois paradigmas coexistem na abordagem de Aprendizado de Máquina, sendo, em lote ou *online*. No paradigma em lote, todo o conjunto de dados é utilizado para ajustar o modelo. Uma vez ajustado, é impossível atualizar o modelo a menos que seja executado um novo treinamento. Já no paradigma *online*, os dados chegam de forma contínua, e o modelo é atualizado constantemente. Este último paradigma é utilizado quando o modelo deve se adaptar às novas condições que os dados representam (FONTENLA-ROMERO et al.,

2013).

A utilização de uma forma inteligente de operação no mercado financeiro faz-se necessária quando o objetivo é a maior lucratividade. Com o advento do Aprendizado de Máquina, busca-se análises extremamente rápidas e precisas, reduzindo os riscos envolvidos nas operações e dessa forma, maximizar a possibilidade de lucros.

1.1 Problema de Pesquisa

Como realizar operações financeiras de forma automática utilizando modelos de aprendizado de máquina para obter ganhos acima da taxa básica de juros SELIC, consumindo pouco tempo para tomada de decisão e pouco tempo para a atualização do modelo em relação a novos padrões?

1.2 Hipótese de Pesquisa

É possível realizar operações financeiras de forma automática utilizando modelos de aprendizado de máquina incremental (online), o qual tem tempo de tomada de decisão similar a modelos tradicionais, mas que conseguem se adaptar a novas informações sem um retreino completo do modelo. Dessa forma acredita-se que os ganhos monetários superam a taxa básica de juros SELIC.

1.3 Objetivos

1.3.1 Objetivo Geral

Com a presente pesquisa, pretende-se verificar a possibilidade de aplicar algoritmos de Aprendizado de Máquina para realizar operações financeiras de forma automática e inteligente a fim de maximizar ganhos no mercado financeiro.

1.3.2 Objetivos Específicos

- Realizar um levantamento bibliográfico;

- Identificar filosofias (ou métodos) mais frequentes de *trading* de ações;
- Identificar os principais algoritmos de Aprendizado de Máquina para avaliar o mercado financeiro;
- Montar o *dataset* para ser utilizado no treinamento de Aprendizado de Máquina;
- Implementar os algoritmos e testar abordagens e parâmetros;
- Interpretar os resultados obtidos e comparar com a Literatura, onde possível.

1.4 Justificativa

Mesmo analistas experientes do mercado financeiro enfrentam um grande risco ao tentar prever o mercado, que está cada dia mais flutuante devido ao crescimento da economia e informatização dos processos (SUTTER, 2018) dificultando as operações de *Day Trade* e consequentemente, a obtenção de lucros. O volume de dados gerado pelo mercado financeiro inviabiliza a análise em um curto espaço de tempo por humanos. Segundo Santana (1997), os sinais utilizados para prever variações positivas ou negativas do mercado necessitam sempre de uma confirmação que, em geral, não acontece em tempo hábil para os analistas realizarem as operações.

Neste sentido, técnicas de Aprendizado de Máquina apresentam resultados excelentes quando comparados a análises tradicionais (ATSALAKIS; VALAVANIS, 2010). Técnicas de Aprendizado de Máquina têm a capacidade de analisar um volume de dados expressivos em um curto espaço de tempo e encontrar padrões ou tendências, que sugestivamente pode ser uma vantagem ou ainda um benefício para corretoras e acionistas. Tais técnicas de Aprendizado de Máquina possibilitam a predição de oscilações futuras e os investidores podem obter retorno financeiro das ações escolhidas (SUTTER, 2018).

Esta pesquisa destaca-se por diferir os trabalhos que utilizam Aprendizado de Máquina no contexto do Mercado Financeiro e de Ações, objetivando prever

tendências ou operar no mercado, dado que nesta pesquisa ambos aspectos serão implementados de forma integrada mas com papéis diferentes, um para predição de valores futuros e outro para determinação da ação a ser tomada (comprar, manter, vender). Também pretende-se utilizar aprendizado *online*, técnica desenvolvida recentemente que permite atualização constante do modelo de forma rápida e com baixo consumo de recursos.

1.5 Organização do texto

A continuação do trabalho está escrito da seguinte forma: no Capítulo 2 são abordados os assuntos, Mercado Financeiro, Ações, *Day Trade*, Indicadores e Aprendizado de Máquina; no Capítulo 3, é apresentada a metodologia a ser utilizada neste trabalho. Já o Capítulo 4 apresenta o cronograma, o Capítulo 5 são descritos os Resultados e por fim, a conclusão da pesquisa é apresentada no capítulo 6.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordadas noções do Mercado Financeiro e Aprendizado de Máquina para ambientar o leitor sobre o assunto pesquisado.

2.1 Mercado Financeiro

Segundo [Kiyosaki e Lechter \(2001\)](#) é possível classificar bens de consumo de uma pessoa ou instituição em duas categorias: ativos e passivos. Ativo pode ser adquirido e tem potencial de lucros ou dividendos, como exemplo uma conta poupança, cotas de fundo de investimento ou um aluguel de imóvel. Já passivos, não geram renda ou tem valorização com o passar do tempo, como exemplo destes são móveis, eletrodomésticos ou roupas quando não utilizados para locação.

A avaliação coletiva dos investidores reflete o preço das ações de uma empresa. O preço das ações tende a subir quando as perspectivas atuais e futuras da empresa são boas, facilitando levantar capital de investimento para a empresa [LIRA e ALMEIDA \(2020\)](#). Dessa forma, [LIRA e ALMEIDA \(2020\)](#) recorre a [Bodie, Kane e Marcus \(2014\)](#) afirmando que o preço das ações são determinantes na alocação de capital, visto que os investidores canalizam o uso mais eficiente do dinheiro, podendo atuar inversamente como efeito da volatilidade sistêmica. Durante um período, empresas ou setores podem ser altamente lucrativas, atraindo muito capital investidor, mas em um período muito curto falhar, causando prejuízos irreversíveis.

2.1.1 Ações

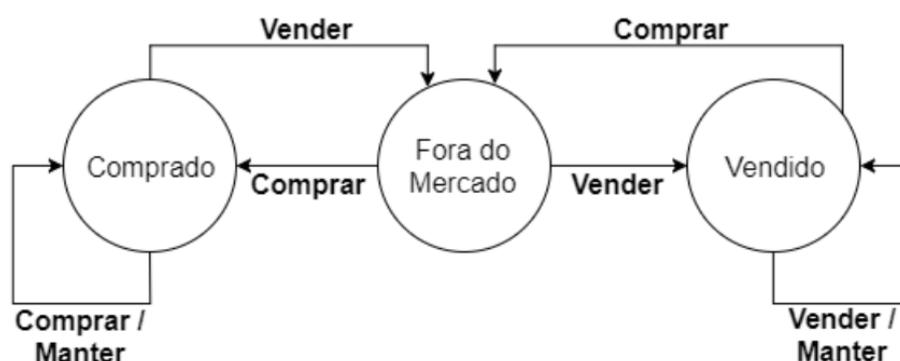
Ação é uma pequena parcela do capital de uma empresa com potencial financeiro para investidores. Comprar uma ação torna o investidor coproprietário da empresa podendo gerar uma remuneração dependendo da capacidade da empresa de gerar lucro. O retorno financeiro pode variar a médio e longo prazo

através dos dividendos da empresa ou na venda da ação por um montante maior que do momento da aquisição (SUTTER, 2018).

A presente pesquisa irá utilizar as seguintes decisões para efetuar a interação com ações, são elas:

- Comprar: adquire uma porção pré determinada de ações.
- Manter: continua de posse da ação.
- Vender: vende uma quantidade pré determinada de ações.

Figura 1 – Dinâmica das possíveis operações a serem realizadas pelo agente.



Fonte: Dantas (2017)

Na Figura 1, vemos que caso o agente ainda não tenha realizado nenhuma operação de compra ou venda, ele está na posição “Fora do Mercado”, caso o agente realize a decisão de compra de determinada ação, ele estará na nova posição “Comprado”, ao tentar realizar uma nova operação de compra, ele irá realizar a decisão de manter. O mesmo ocorre na posição “Vendido”, onde o agente também irá manter a posição caso identificar que a posição deverá ser vendida. Ao vender ou comprar uma ação com sucesso, ele retorna para a posição “Fora do Mercado”.

2.1.2 Day Trade

Segundo Sasaki (2021) são chamadas operações de *Day Trade* ações realizadas em um curto período, visto que o encerramento da operação deve

ocorrer no mesmo Pregão Eletrônico¹ em que foi iniciado. Em outras palavras, operações para serem consideradas *Day Trade* envolvem compra e venda das mesmas ações em um único dia, realizadas pelo mesmo CPF/CNPJ.

São chamadas High Frequency Trading (**HFT**) operações realizadas por robôs *trades*, devido ao alto volume de negociações em um mesmo dia. Verificou-se que a participação de HFT na bolsa de valores brasileira, o volume médio de operações cresce diariamente, ao pé que em maio de 2019, representavam 34,5% (**ANDRADE, 2019**).

2.1.3 Formas de Análise

Duas formas de análises existem para auxiliar o investidor para a tomada de decisão, são elas: Análise Técnica e Análise Fundamentalista. A Análise Técnica é normalmente utilizada para auxiliar no momento de entrada ou saída do mercado, já a Análise Fundamentalista é muito utilizada para definir em qual ativo investir (**DANTAS, 2017**).

2.1.4 Indicadores

Alguns indicadores são utilizados para análises técnicas ou estatísticas através de gráficos, fornecendo informações para criar e sustentar hipóteses de padrões e previsões do Mercado Financeiro. Como exemplos, pode-se citar Médias Móveis, Bandas de Bollinger e Média Móvel Convergente e Divergente (MACD).

- **Médias Móveis (MM)** - Um dos mais utilizados e mais flexíveis dos indicadores da análise técnica, a sua construção permite ser facilmente quantificada e testada, é a base para diversas ferramentas de investimentos (**MURPHY, 1999**).

A média móvel é basicamente uma ferramenta que apresenta claramente tendências do mercado, objetivando identificar onde iniciou uma antiga tendência e sua reversão (**SPRITZER, 2017**).

¹ Pregão Eletrônico - Procedimento ágil e prático, destinado apenas à aquisição de bens e serviços considerados comuns, independente do valor (<https://e-dou.com.br/qual-a-diferenca-entre-pregao-eletronico-e-presencial>)

Figura 2 – Gráfico utilizando Médias Móveis, indicador representado pela linha azul.



Fonte: Autor

- **Bandas de Bollinger (BB)** - John Bollinger verificou que em gráficos de ativos, uma trégua ocorria tanto em partes compradoras e vencedoras. Além disso, observou que raramente havia um distanciamento de uma determinada região, invariavelmente aproximado para uma região de equilíbrio, indiferentemente da volatilidade [Correia \(2008\)](#). As BB são basicamente três linhas, sendo uma central uma superior e uma inferior ([SPRITZER, 2017](#)).

Figura 3 – Gráfico utilizando Bandas de Bollinger, indicador representado pelas linhas azuis.



Fonte: Autor

- **Média Móvel Convergente e Divergente (MACD)** - Utiliza indicadores de tendência para indicar o momento. Possui duas linhas: Linha MACD (LM) formada a partir da diferença das médias móveis exponenciais de 12 e 26 períodos, e a Linha Sinal (LS), formada pela média móvel exponencial de 9 períodos da LM ([DANTAS, 2017](#)).

Figura 4 – Gráfico utilizando Média Móvel Convergente e Divergente (MACD), Linha MACD (LM), representada pela linha de cor azul. Linha Sinal (LS), representada pela linha de cor preta.



Fonte: Autor

2.2 Aprendizado de Máquina

O processo de aprendizado está presente na vida dos seres humanos desde o nascimento. Um único contato com alguma ação dificilmente pode fazer o observador replicar o que foi apresentado. Normalmente, há necessidade de um processo repetitivo que comprove o padrão observado. Os computadores atuais conseguem acelerar aprendizados de determinado assunto significativamente. Aprendizado de Máquina pode resumir-se a habilidade dos computadores de aprender sem ser explicitamente programados ([SUTTER, 2018](#)).

A ideia por trás de construir conhecimento de forma automática baseia-se em aprender com dados, ou seja, criar modelos analíticos a partir de experiências acumuladas de solução bem-sucedida de problemas anteriores. Essencialmente, o Aprendizado de Máquina utiliza-se de padrões encontrados nos dados passados para gerar um modelo que possibilita a tomada de decisões em novas previsões ([ANDRADE; SOUZA et al., 2019](#)).

2.2.1 Tipos de Aprendizado

- **Aprendizado supervisionado:** No paradigma mais estudado atualmente segundo [Dantas \(2017\)](#), o conjunto de dados de Aprendizado de Máquina possui previamente a informação verdadeira dos dados. O treinamento baseia-se no comparativo do resultado obtido pelo modelo e a informação verídica. Repetindo até a obtenção do menor erro ([PAIXÃO et al., 2022](#)).
- **Aprendizado não supervisionado:** Ao contrário do aprendizado supervisionado, no aprendizado não supervisionado, o modelo utiliza padrões dos dados e constrói uma representação própria dos dados sem o conhecimento prévio da classificação de cada dado, ou seja, de forma especulatória. Isto pode representar em alguns casos, uma vantagem, uma vez que os padrões pré-determinados não são considerados ([PAIXÃO et al., 2022](#)).
- **Aprendizado semi-supervisionado:** No Aprendizado Semi-Supervisionado, os dados utilizados no treinamento são parcialmente classificados, enquanto outra consiste em dados não classificados. Este paradigma pode ser útil quando os dados não possuem informação suficiente para impor uma regra geral, utilizando-se o conjunto de teste como fonte extra de informação para resolução da problemática ([SILVA, 2008](#)).
- **Aprendizado por reforço:** No paradigma de aprendizado por reforço o conjunto de dados não possui nenhuma classificação ou rótulo. Para ocorrer o aprendizado, as decisões assertivas geram recompensa para o modelo. O objetivo do modelo é conseguir o máximo de recompensas positivas possíveis, não importando a maior recompensa, mas sim a soma das mesmas a longo prazo ([TEIXEIRA, 2016](#)).

2.2.2 Algoritmos *Online*

Em base de dados volumosas é comum atualizações e adição de novos dados. Dessa forma, algoritmos clássicos de Mineração de Dados ([MD](#)) precisam ser reprocessados para manter a consistência e precisão nos padrões encontrados. A Mineração de Dados de forma *online* tem por objetivo a manutenção dos

padrões aprendidos, por meio da inclusão de novos dados. Ou seja, os padrões já encontrados são atualizados para se alinharem as novas realidades evitando reprocessamento (JUNIOR, 2013).

Reis (2017) reflete sobre algoritmos *online* da seguinte forma: “Os algoritmos estudados retêm informação adaptando ou expandindo um único modelo através do tempo, ao invés de reconstruir modelos a partir do zero e treiná-los em novos dados”. Algumas vantagens dos algoritmos *online* são:

- Adapta-se as mudanças ou novos padrões;
- Não necessita de retreinamento;
- Tem a capacidade de aprender do zero;
- Consome menos recursos (memória, processamento e armazenamento).

Neste tipo de algoritmo, o sistema é alimentado sequencialmente em pequenos grupos ou individualmente. Cada etapa do aprendizado é leve e performática do ponto de vista computacional, ideal para sistemas de fluxo contínuo (por exemplo, preço de ações) ou que precisam se adaptar as mudanças de forma rápida e autônoma (GÉRON, 2019).

2.2.3 Processamento de Fluxo

Os fluxos de dados podem ser vistos como processos aleatórios nos quais os eventos ocorrem de forma contínua e independente uns dos outros. Consultar fluxos de dados é bem diferente de consultar no modelo relacional convencional. A ideia chave é que no processamento de fluxo os dados são transitórios, já nos modelos convencionais, os dados são armazenados (GAMA, 2010). Algumas diferenças relevantes incluem:

1. Os elementos de dados no fluxo chegam constantemente;
2. O sistema não tem controle sobre a ordem em que os dados chegam, seja dentro de um fluxo de dados ou entre fluxos de dados;
3. Os fluxos de dados são potencialmente ilimitados em tamanho;

4. Depois que um elemento de um fluxo de dados é processado, ele é descartado ou arquivado. Ele não pode ser recuperado facilmente a menos que seja explicitamente armazenado na memória, que é pequena em relação ao tamanho dos fluxos de dados.

3 TRABALHOS RELACIONADOS

No contexto de Aprendizado de Máquina e do Mercado Financeiro, existem diversos trabalhos na Literatura. Em [Fernandes \(2019\)](#) o autor conseguiu alcançar aproximadamente 76,34% de precisão máxima utilizando Regressão Logística com 119 dias de dados referente aos papéis da PETR4 para treino, dessa forma, concluiu que “a tomada de decisão apoiada por algoritmos de inteligência artificial pode ser muito mais eficaz do que uma simples análise técnica tradicional”.

[Jr et al. \(2021\)](#) utilizou a rede neural Long Short-Term Memory ([LSTM](#)) em ativos financeiros e concluiu ser possível predizer valores de fechamento dos ativos, dado que sua ferramenta conseguiu obter um lucro médio em 0,15% tanto para os dados reais quanto para a predição da LSTM, para o ativo da PETR4.

[Rocha et al. \(2019\)](#) obteve resultados superiores tanto aos *baselines* quanto em termos financeiros utilizando redes neurais LSTM, onde em alguns casos ao se comparar a Acurácia dos modelos, a rede neural obteve valores até 15% maiores em relação ao classificador aleatório, entretanto, mesmo obtendo taxas de Acurácia elevadas e retornos financeiros positivos, houve vários casos onde o custo de operação se sobrepôs ao lucro bruto obtido pela estratégia.

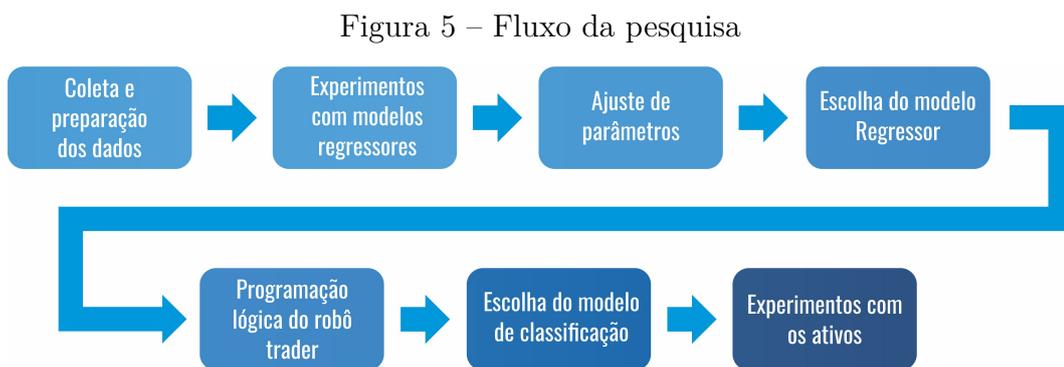
[Henrique \(2018\)](#) realizou a predição da direção do dia seguinte sob um limiar de variação nos preços, concluiu que todos os modelos utilizados (Artificial Neural Network ([ANN](#)), Support Vector Machine ([SVM](#)), Random Forest ([RF](#)) ou Naive-Bayes ([NB](#))) apresentam acurácia próxima a de um teórico classificador aleatório, com 50% de taxa de acerto.

Na tentativa de prever o comportamento diário das ações, ([ANDRADE; SOUZA et al., 2019](#)), obteve resultados com baixa assertividade (menos de 50%), entretanto ao prever janelas de curto prazo (entre 25 e 30 dias), os resultados alcançaram quase 80% de acurácia.

4 METODOLOGIA

Neste capítulo serão abordados a coleta de dados, incluindo as fontes e os tipos de papéis utilizados, a forma de processamento dos dados coletados, bem como as métricas que serão utilizadas: classificação Precisão, Revocação, Acurácia e F1-Score, e as métricas de regressão RMSE e R2. Por fim e não menos importante é apresentado o fluxo da pesquisa e considerações sobre o Robô Trader.

A fim de organizar e delimitar o fluxo da pesquisa, foi definido o andamento da mesma conforme representado pela Figura 5.



Fonte: Autor

Inicialmente será realizada a coleta e preparação dos dados fornecidos pela plataforma MetaTrader 5¹. Em seguida, se iniciará os experimentos com os modelos regressores selecionados e será realizado os ajustes dos parâmetros disponíveis, por fim, com base nas melhores métricas será delimitado o modelo regressor para andamento da pesquisa.

Realizada a escolha do modelo regressor iniciará a programação lógica do Robô Trader. Posteriormente será realizada a seleção do modelo através das métricas disponíveis, a modo de consolidar o melhor cenário possível. Finalmente será realizado uma bateria de testes com todos os ativos selecionados a fim de encontrar o melhor desempenho possível do Robô Trader.

¹ MetaTrader 5 - Plataforma que fornece terminais de negociação

4.1 Coleta de Dados

A coleta de dados será através de uma *Application Programming Interface (API)* em *Python* fornecida gratuitamente por uma comunidade denominada MetaTrader 5, onde os dados podem ser consultados de forma contínua, alimentando assim o modelo de Aprendizado de Máquina. Os papéis que serão utilizados para o treinamento serão as ações da Ibovespa VALE S.A ([VALE3](#)) e Ambev S.A ([ABEV3](#)), uma vez que sua participação teórica em porcentagem de mercado são representativas de acordo com o Índice Bovespa ([B3, 2022](#)), sendo ABEV3 com 3,113% e VALE3 com 14,966% na carteira da data de 10 de maio de 2022. Também será utilizado o iShares S&P 500 Fundo de Investimento ([IVVB11](#)), sendo um fundo de índice que reflete a performance do índice americano S&P 500, que representa as 500 maiores empresas listadas em bolsa nos Estados Unidos da América (EUA). Para esta pesquisa serão desconsiderados os dividendos.

4.2 Processamento

Será utilizada a linguagem de programação *Python* que é amplamente difundida para a criação de modelos de Aprendizado de Máquina. Diversas bibliotecas são desenvolvidas para a linguagem, entre elas a biblioteca chamada *Pandas*² que permite a análise dos dados de forma fácil e rápida. Não menos importante, para a construção do modelo de Aprendizado de Máquina será utilizado a biblioteca chamada *River*³, que trabalha com processamento de fluxo.

4.2.1 Python

Criada por Guido van Rossum (1991), *Python* é uma linguagem de programação que possui uma cultura muito forte na produtividade e legibilidade. Para isto, poucos caracteres especiais são utilizados ([BORGES, 2014](#)). *Python*

² Pandas - Biblioteca em *Python* para manipulação e análise de dados

³ River - Biblioteca em *Python* utilizada para treinamento de modelos de Aprendizado de Máquina e aprendizado *online*

conta com um gerenciador de memória automático que possui a característica de descartar dados armazenados quando não possuem mais nenhuma referência.

Devido a vasta biblioteca de códigos produzidos pela comunidade e por ser multiplataforma, a linguagem é largamente utilizada para desenvolvimento das mais diversas categorias de aplicações, por exemplo: aplicações WEB e *desktop*, *scripts* de automação, API's, Inteligência Artificial, Mineração de Dados, Aprendizado de Máquina entre outras.

Conforme [McKinney \(2019\)](#), para análise de dados, processamento interativo e visualização de dados, *Python* inevitavelmente atrairá comparações com outras linguagens de programação comerciais e de código aberto, e com ferramentas amplamente utilizadas, como *R*, *MATLAB*, *SAS*, *Stata* e outras. Nos últimos anos, o suporte melhorado de *Python* para bibliotecas (como o *Pandas* e o *Scikit-Learn*) o transformou em uma opção popular para tarefas de análise de dados. Em conjunto com a robustez de *Python* para uma engenharia de software de propósito geral, é uma excelente opção como uma linguagem principal para a construção de aplicações de dados.

4.2.2 Pandas

O *Pandas* oferece estruturas de dados de alto nível e funções, projetadas para trabalhar com dados de forma rápida, fácil e expressiva. Desde o seu surgimento em 2010, o *Pandas* ajudou o *Python* a tornar-se um ambiente eficaz e produtivo para análise de dados. A biblioteca combina as ideias de processamento de alto desempenho de *arrays* da biblioteca *NumPy*⁴ com os recursos flexíveis de manipulação de dados. Ele disponibiliza uma funcionalidade sofisticada de indexação para facilitar a reformatação, a manipulação e a limpeza dos dados ([MCKINNEY, 2019](#)). A biblioteca *Pandas* auxiliará no tratamento e preparação dos dados para posterior treinamento utilizando a biblioteca chamada *River*.

4.2.3 River

River é uma biblioteca que permite trabalhar com fluxos de dados dinâmicos e aprendizado *online*. A biblioteca dispõe de diversos métodos e funcionalidades utilizados no aprendizado de fluxo por ser uma fusão de dois

⁴ NumPy é uma biblioteca em *Python* que permite o processamento de grandes matrizes utilizando funções matemáticas.

projetos em *Python: Creme* e *Scikit-Multiflow*, resultando em uma biblioteca de última geração que traz consigo ganhos no que diz respeito a tempo de processamento, mais métricas para classificação, regressão e agrupamento e métodos de agrupamento extras. Além disso, o fato de se tratar de uma biblioteca de código aberto, possui o suporte de uma grande comunidade de pesquisadores e profissionais (MONTIEL et al., 2020). Será utilizado métodos regressores *online* disponíveis na biblioteca, como *Hoeffding Tree Regressor*, *KNN Regressor* e *MLP Regressor*. Para o segundo modelo, será utilizado métodos classificadores *online* como *Hoeffding Trees*, *Naive Bayes* e *KNN*, estando aberto a outras possibilidades durante a execução do trabalho.

4.3 Métricas

Para avaliar o desempenho dos sistemas de Aprendizado de Máquina são utilizadas algumas métricas específicas de modo a subsidiar as discussões e resultados obtidos. Neste trabalho serão empregados dois modelos de aprendizado de máquina sendo um para previsão de valores futuros e outro para classificação e regressão. Para cada uma dessas tarefas, existem diferentes métricas, dessa forma, serão utilizadas as métricas de classificação Precisão, Revocação, *Acurácia* e *F1-Score*, e as métricas de regressão *RMSE* e *R2*.

Para realização dos cálculos, primeiramente é necessário compreender em quais classes se enquadram os registros após o modelo realizar a previsão. Para toda tarefa de classificação são gerados acertos em relação a uma classe específica (verdadeiros positivos e verdadeiros negativos) e erros (falsos positivos e falsos negativos).

- Verdadeiro Positivo (TP) são registros que o modelo classificou corretamente;
- Falso Positivo (FP) são registros que o modelo classificou incorretamente como verdadeiro;
- Verdadeiro Negativo (TN) são registros que o modelo classificou corretamente como falso;

- Falso Negativo (FN) são registros que o modelo classificou incorretamente como falso.

Figura 6 – Representação das possíveis classificações dos registros preditivos do modelo

		Valores Reais	
		POSITIVO	NEGATIVO
Valores Preditos	POSITIVO	TP (True Positive) Verdadeiro Positivo	FP (False Positive) Falso Positivo
	NEGATIVO	FN (False Negative) Falso Negativo	TN (True Negative) Verdadeiro Negativo

Fonte: Autor

4.3.1 Precisão

É uma métrica utilizada que representa a proporção de registros classificados como positivos que estavam realmente corretos. Para esta métrica, quanto maior o resultado, melhor. Considerando os Verdadeiros Positivos (TP), Falsos Positivos (FP), Verdadeiros Negativos (TN) e Falsos Negativos (FN), o cálculo é feito pela Equação 4.1.

$$Precisão = \frac{TP}{TP + FP} \quad (4.1)$$

4.3.2 Revocação

Conforme Géron (2019), revocação é a taxa de instâncias positivas que são corretamente detectadas pelo classificador. Para esta métrica, quanto maior o valor de Revocação, melhor. Considerando os Verdadeiros Positivos (TP), Falsos Positivos (FP), Verdadeiros Negativos (TN) e Falsos Negativos (FN), o cálculo é feito pela Equação 4.2.

$$Revocação = \frac{TP}{TP + FN} \quad (4.2)$$

4.3.3 Acurácia

Acurácia é razão de acertos, resultando no percentual de acerto do modelo utilizado. Uma vez que utiliza a taxa de acertos, quanto maior o valor da acurácia, melhor a performance do modelo. Considerando os Verdadeiros Positivos (TP), Falsos Positivos (FP), Verdadeiros Negativos (TN) e Falsos Negativos (FN), o cálculo é feito pela Equação 4.3.

$$Acurácia = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.3)$$

4.3.4 F1-Score

O *F1-Score* utiliza uma combinação da precisão e da revocação, auxiliando na comparação entre classificadores. A pontuação F1 é uma média harmônica da precisão e revocação representada pela Equação abaixo (GÉRON, 2019).

$$F1 = \frac{2}{\frac{1}{precisão} + \frac{1}{revocação}} = 2 \times \frac{precisão \times revocação}{precisão + revocação} = \frac{2TP}{TP + FN + FP} \quad (4.4)$$

Ainda segundo Géron (2019) a média harmônica dá muito mais peso aos valores mais baixos, diferentemente da média regular que trata todos os valores com pesos iguais. Como resultado, o classificador só obterá uma pontuação F1 alta se a revocação e a precisão forem altas simultaneamente.

4.3.5 RMSE

Raiz do Erro Quadrático Médio ou (sigla RMSE, em inglês) é uma medida de desempenho que dá ideia da quantidade de erros gerados após realizadas as previsões, com peso maior para grandes erros (GÉRON, 2019). Ou seja, quanto menor o valor de RMSE obtido, melhor.

RMSE permite medir os erros gerados pelas previsões é representada pela Equação abaixo.

$$RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2} \quad (4.5)$$

Onde m representa o número de registros, $h(x^{(i)})$ a previsão do regressor e $y^{(i)}$ a saída real esperada.

4.3.6 R2 Score

O R2 tem por objetivo medir a performance do modelo utilizando a média simples dos valores. $R2\ Score = 1$ indica que as previsões do modelo foram perfeitas, $R2\ Score = 0$ indica que o modelo é uma média de todos os valores, enquanto $R2\ Score$ abaixo de 0 significa que o modelo não está tendo resultados bons (PENA; MENA-CHALCO, 2021). Esta unidade de medida descreve a qualidade do ajuste realizado no modelo, ou seja, quanto o modelo conseguiu explicar dos dados utilizados. Para o cálculo utiliza-se a soma dos quadrados da regressão (SQR) e a soma dos quadrados total (SQT) (SALMAZO, 2021).

A Equação 4.6 é utilizada pelo $R2\ Score$ para indicar a performance do modelo. Para esta métrica, busca sempre o valor mais próximo a 1, ou seja, desempenho perfeito.

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQR}{SQT} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (4.6)$$

4.4 Operação no Mercado Financeiro

Alguns parâmetros foram estabelecidos para delimitação desta pesquisa. Quanto à operação do mercado, optou-se por utilizar *Day Trade*, que consiste em iniciar e fechar uma operação de compra ou venda em um único dia. Quanto a linguagem de programação, será utilizada *Python* e suas bibliotecas que permitem a coleta, tratamento, treinamento e análise dos dados. Os dados serão processados de forma contínua utilizando algoritmos incrementais, capazes de aprender de forma *online*. Os resultados serão fornecidos para outro modelo que irá realizar uma classificação, mas dessa vez para operar o mercado de ações.

5 RESULTADOS E DISCUSSÕES

Neste capítulo serão detalhados os resultados obtidos durante os experimentos realizados. Os resultados obtidos serão apresentados em forma de tabelas e gráficos.

5.1 Considerações Iniciais

A proposta do trabalho foi utilizar dois modelos de Aprendizado de Máquina para operar no mercado financeiro, sendo um para predição de uma posição futura e outros dois para juntamente com essa informação determinar o momento certo para operar no mercado financeiro, sendo estes modelos de classificação. Para todos os experimentos, foi utilizado o a seguinte configuração de máquina: Processador AMD Ryzen 5600G de 3.9GHz, 2 memórias DDR4 de 8GB, clock de 3000MHz, armazenamento M.2 PNY CS3030 500GB de leitura sequencial máxima 3,500 MB/s e gravação sequencial máxima de 3,000 MB/s. A versão do Python utilizada foi 3.10.7 e a biblioteca River versão 0.13.0.

5.2 Dados

Para realizar os experimentos, foram utilizados datasets contendo exatos 100 mil registros referentes à 16/06/2021 à 23/05/2022. Os registros são referentes a cotação dos ativos a cada minuto (M1), contendo as seguintes informações: abertura, fechamento, menor valor, maior valor, volume de tick¹ e volume real, dos ativos selecionados para esta pesquisa (IVVB11, VALE3 e ABEV3).

5.3 Variáveis

A fim de melhorar os resultados obtidos, foram utilizadas variáveis que auxiliam no treinamento do modelo e os momentos de encerrar ou manter uma

¹ Tickl - tick é a mínima variação de preço possível para um ativo.

operação no mercado. As variáveis utilizadas no treinamento são:

- `janela`: Quantidade de momentos passados que são utilizados para treinar os modelos.
- `diferencaPredicaoEntradaComprando`: Diferença mínima entre a primeira posição da janela e a posição atual, utilizada para treinar o modelo classificador de entrada referente ao momento mais adequado para entrar no mercado comprando.
- `diferencaPredicaoEntradaVendendo`: Diferença mínima entre a primeira posição da janela e a posição atual, utilizada para treinar o modelo classificador de saída referente ao momento mais adequado para entrar no mercado venda.
- `diferencaMercadoSaida`: Variável utilizada para definir a diferença mínima necessária para sair do mercado e obter lucro.
- `perdaMaxima`: Variável utilizada para definir a perda máxima aceitável.

5.4 Codificação

Para predição do valor do ativo foi realizado experimentos com alguns modelos de regressão disponíveis na biblioteca do River, sendo eles: `KNNRegressor`, `Naive Bayes` e `SoftmaxRegression`. Para operação no mercado financeiro foi utilizados os modelos de classificação: `KNNClassifier`, `PAClassifier` e `Perceptron`, também presentes na biblioteca.

5.5 Experimentos

5.5.1 Treinamento dos Modelos Regressores utilizando o Dataset IVVB11

5.5.1.1 Naive Bayes

A Tabela 1 apresenta os resultados obtidos utilizando as métricas de RMSE e R2 durante o teste do modelo regressor Naive Bayes. É possível notar

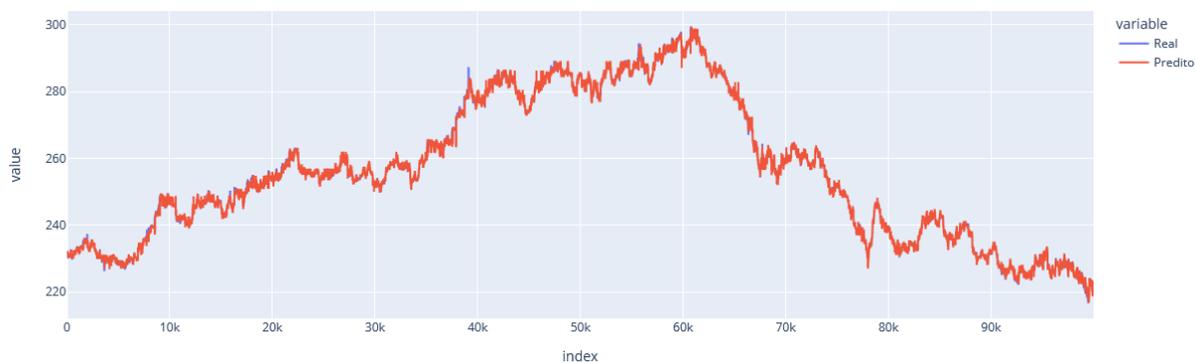
que os valores das métricas são bons. Entretanto, o tempo de execução do modelo é extremamente moroso.

Naive Bayes			
Iteração	RMSE	R2	Tempo (s)
1	0.5895	0.99921	20885

Tabela 1 – Resultados da predição utilizando o modelo Naive Bayes - Fonte: Autor

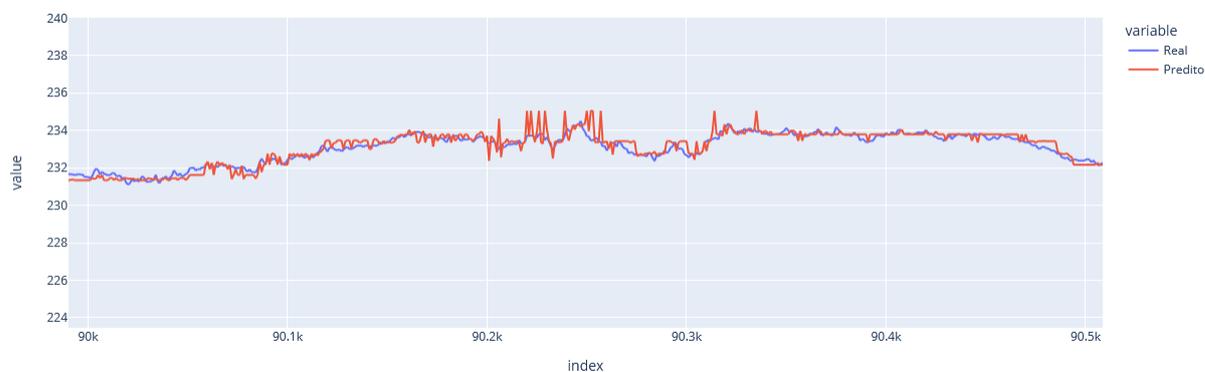
A Figura 7 apresenta o comportamento do ativo utilizando o modelo Naive Bayes, que em um primeiro momento acredita-se que esteja acompanhando o mercado de forma precisa. Entretanto ao visualizar a Figura 8 que representa uma ampliação a fim de exibir aproximadamente 500 registros, o equivalente a aproximadamente um dia. Verificamos que o modelo não mantém uma predição estável, gerando excesso de ruídos indesejados.

Figura 7 – Modelo Naive Bayes - Visualização completa dos dados



Fonte: Autor

Figura 8 – Modelo Naive Bayes - Detalhe de aproximadamente 500 registros



Fonte: Autor

5.5.1.2 KNNRegressor

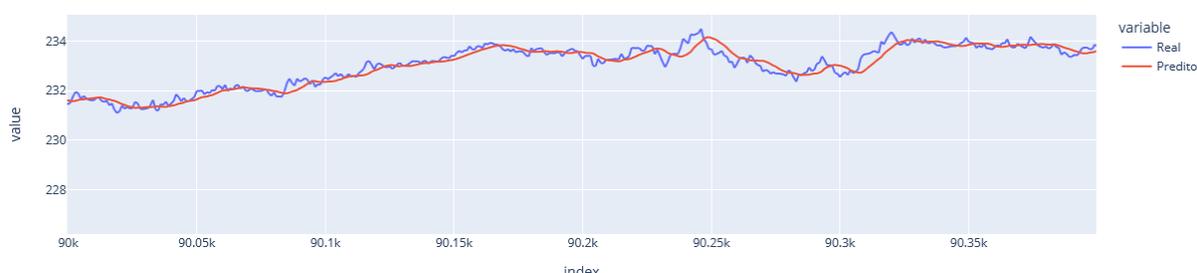
A Tabela 2 apresenta os resultados obtidos utilizando o modelo regressor KNNRegressor, o modelo permite definir o parâmetro `window_size` (tamanho da janela). Este parâmetro define o tamanho máximo da janela dos últimos registros.

KNNRegressor				
Iteração	RMSE	R2	Tempo (s)	window_size
1	0.2854	0.99981	10	10
2	0.3814	0.99967	10	20
3	0.5289	0.99936	10	40
4	0.7456	0.99874	11	80

Tabela 2 – Resultados da predição utilizando o modelo KNNRegressor - Fonte: Autor

Verifica-se que quanto maior o tamanho da janela, maior é o RMSE, indicando uma menor taxa de acerto do modelo. Já valores menores superajusta o modelo (**Overfitting**), replicando os dados da posição anterior ($n-1$), conforme apresentado na Figura 9 (`window_size = 10`).

Figura 9 – Modelo KNNRegressor - Detalhe de aproximadamente 500 registros



Fonte: Autor

Na Figura 10 (`window_size = 80`) é possível verificar que o modelo também consegue acompanhar a evolução do mercado, entretanto, ao analisar aproximadamente 500 registros (Figura 11), é possível identificar que o modelo prediz a média do comportamento do mercado.

5.5.1.3 SoftmaxRegression

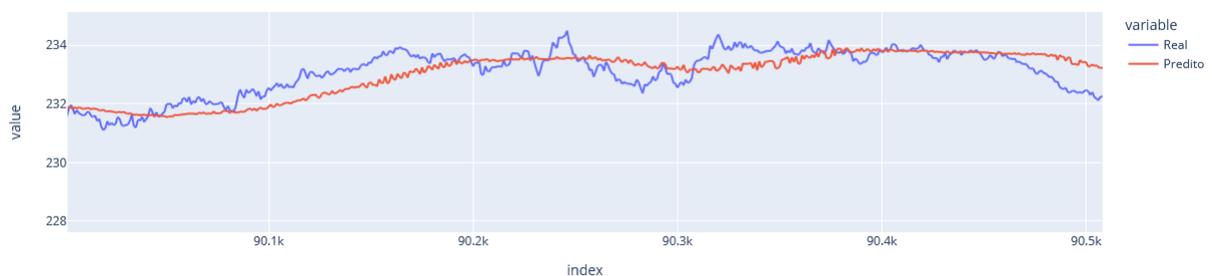
A Tabela 3 apresenta os resultados obtidos utilizando o modelo regressor SoftmaxRegression, os valores de RMSE e R2 demonstram bons resultados

Figura 10 – Modelo KNNRegressor - Visualização completa dos dados



Fonte: Autor

Figura 11 – KNNRegressor - Detalhe de aproximadamente 500 registros



Fonte: Autor

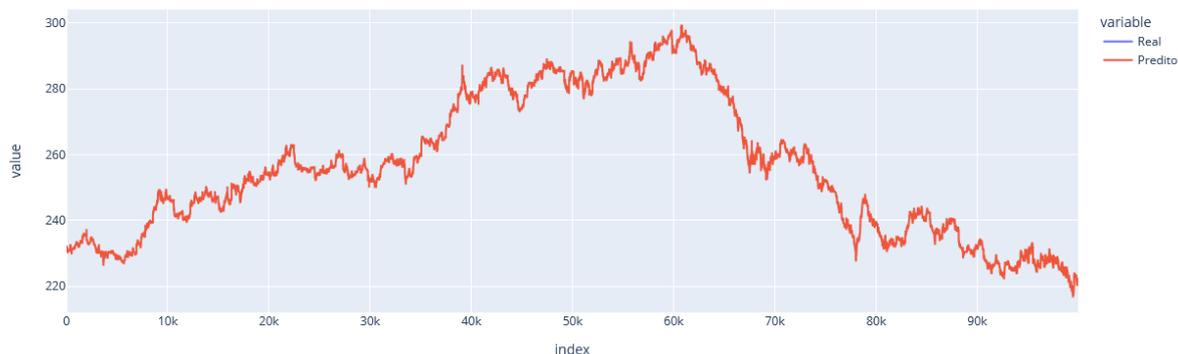
para o modelo em questão. O único ponto negativo deste modelo em relação ao KNN é o tempo de execução.

SoftmaxRegression				
Iteração	RMSE	R2	Tempo (s)	l2
1	0.5358	0.93531	5566	0
2	0.2054	0.99990	5072	20
3	0.1989	0.99991	4695	30
4	0.1820	0.99992	4735	50
5	0.1673	0.99993	5840	100

Tabela 3 – Resultados da predição utilizando o modelo SoftmaxRegression - Fonte: Autor

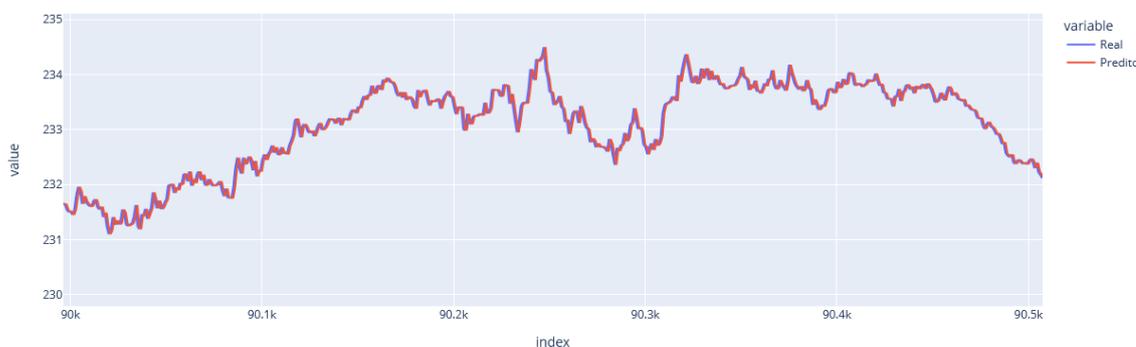
O modelo SoftmaxRegression permite definir o parâmetro l2 (parâmetro do modelo). Este parâmetro é utilizado para regularizar pesos e melhorar a performance do modelo. Verifica-se que quanto maior o valor de l2, maior a precisão do modelo, entretanto, mais super ajustado fica o modelo, replicando os dados da posição anterior (n-1) (Figura 13).

Figura 12 – Modelo SoftmaxRegression - Visualização completa dos dados



Fonte: Autor

Figura 13 – Modelo SoftmaxRegression - Detalhe de aproximadamente 500 registros



Fonte: Autor

5.5.2 Robô Trader

5.5.2.1 Considerações Iniciais

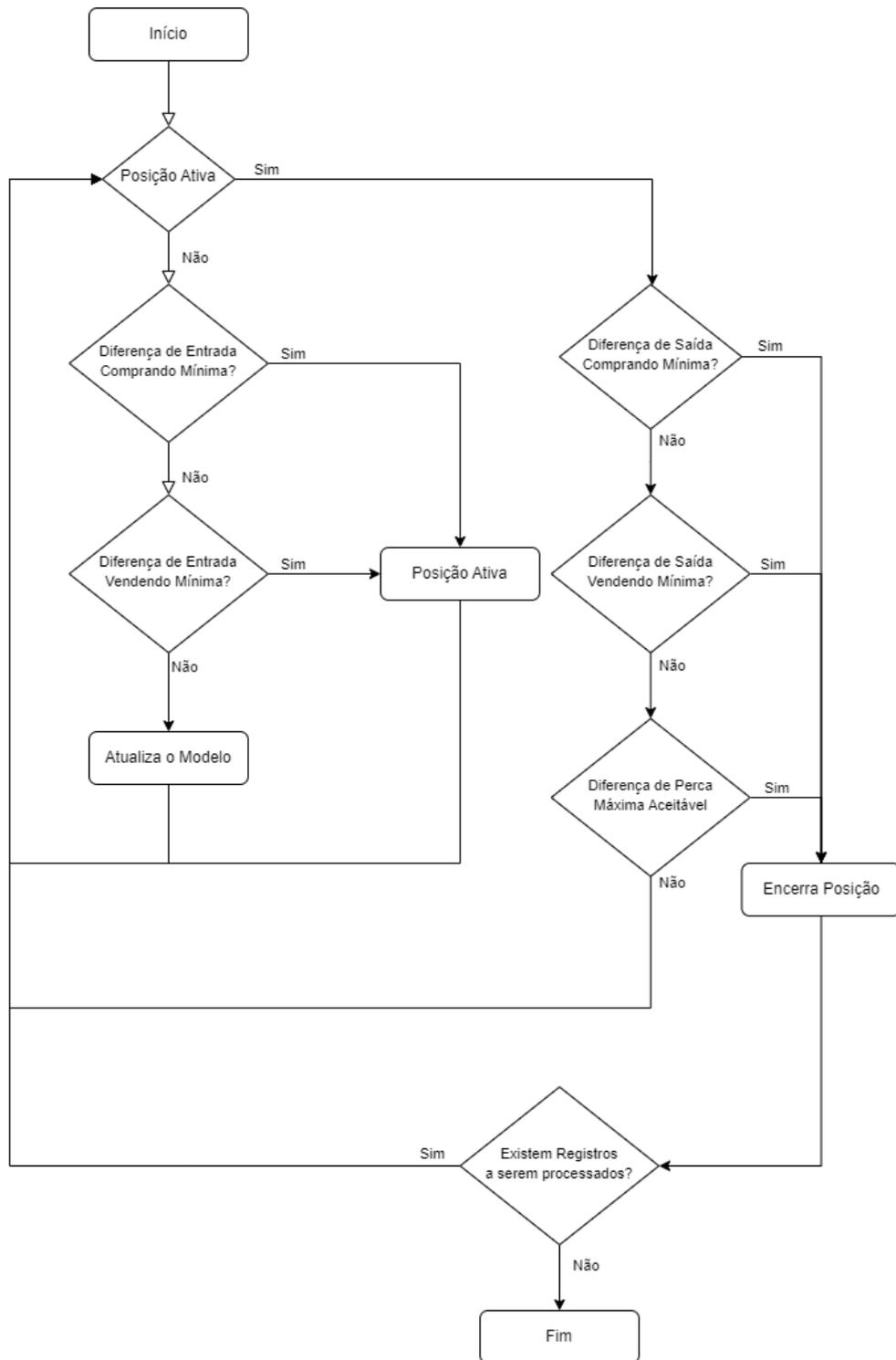
Após os experimentos dos modelos de regressão, optou-se pela utilização do modelo `KNNRegressor` devido a sua precisão e tempo de execução. Utilizou-se o valor de 20 para o parâmetro `window_size`, visto que este valor apresentou melhor relação entre a réplica dos dados da posição anterior ($n-1$) e média do mercado. A variável “janela” é utilizada para prever o mercado em “x” posições a frente. A partir deste preposto, iniciam-se as decisões do Robô Trader.

A lógica utilizada para desenvolver o Robô Trader é apresentada no diagrama da Figura 14, onde inicialmente, é verificado se existe uma posição ativa no mercado. Em caso negativo, é feita verificação se a posição atual

possui uma diferença “x” positiva, esta, por sua vez é definida na variável “diferencaPredicaoEntradaComprando”. A mesma verificação é feita para entrar no mercado vendendo, entretanto com uma diferença “x” negativa, definida na variável “diferencaPredicaoEntradaVendendo”. Caso não exista alguma das diferenças requeridas para entrar no mercado, o modelo é atualizado retornando a verificação de posição ativa.

Quando existe uma posição ativa, é verificado se o mercado movimentou-se “x” pontos, definido na variável “diferencaMercadoSaida”, caso a condição seja satisfeita, encerra-se a posição com lucro. A mesma verificação é feita quando a entrada foi realizada com posição de venda. A fim de minimizar as perdas e evitar que a posição fique ativa por dias, descaracterizando o *Day Trade*, foi definido um valor de perda máxima aceitável, sendo este calculado entre a diferença da posição de entrada e a posição atual. Sendo esta a última verificação antes de encerrar a posição ou verificar o próximo registro. Ao processar todos os registros do *dataset*, o programa é encerrado.

Figura 14 – Diagrama Funcionamento Robô Trader



Fonte: Autor

Ao todo foram utilizados 3 modelos de *Machine Learning*, um modelo regressor e 2 modelos de classificação, sendo um modelo para predição de entrada comprando e um modelo para predição de entrada vendido.

Modelo	Acurácia Compra	Acurácia Venda	F1 Compra	F1 Venda	Lucro (R\$)	Acertos (%)
HoeffdingTree	0.77	0.77	0.98	0.98	423.8	59.07
Naive Bayes	0.75	0.75	0.95	0.95	593.4	60.0
KNNClassifier	0.74	0.74	0.94	0.95	608.5	59.17
PAClassifier	0.75	0.75	0.94	0.94	951.8	61.89
Perceptron	0.65	0.65	0.84	0.84	513.6	58.02

Tabela 4 – Resultados dos experimentos dos modelos de classificação - Fonte: Autor

5.5.2.2 Escolha dos modelos de classificação

Após a codificação referente ao funcionamento do Robô Trader, realizou-se alguns experimentos com os modelos disponíveis na biblioteca River. Nestes experimentos apresentados na Tabela 4, utilizou-se o ativo IVVB11 e empiricamente os valores para as variáveis: “janela” = 5, “diferencaPredicaoEntradaComprando” = 0.1, “diferencaPredicaoEntradaVendendo” = 0.5, “diferencaMercadoSaida” = 2 e “perdaMaxima” = 10. Para todos os testes realizados na pesquisa utilizou-se a quantia de 10 contratos.

Com os resultados apresentados na Tabela 4, foi possível verificar que o modelo de classificação HoeffdingTree apresentou os melhores resultados de acurácia e F1 score. Entretanto este modelo apresentou a pior lucratividade entre os modelos testados mas manteve a taxa de acerto semelhante. O modelo que apresentou maior lucratividade foi o PAClassifier, o qual manteve a média de acurácia e F1 Score. Ambos modelos serão utilizados nos experimentos a fim de verificar os resultados.

5.5.2.3 Experimentos Iniciais

Para fins de experimentos, utilizou-se 10 contratos em todos ativos. O R2 Score foi condicionado a valores acima de 0.9 para o modelo regressor a fim de maximizar a precisão nas operações. As taxas de corretagem, emolumentos, custódia e imposto de renda foram desconsideradas nesta pesquisa.

Utilizou-se o desvio padrão com intuito de mensurar o coeficiente de variação dos ativos. Ao realizar os experimentos, o ativo IVVB11 apresentou o desvio padrão de R\$0,95, já o ativo VALE3 apresentou o valor de R\$0,53 e o ativo ABEV3 resultou em R\$0,08. Dessa forma é possível identificar que o ativo IVVB11 possui uma variabilidade maior dos dados se comparado a VALE3

e principalmente a ABEV3.

A fim de encontrar os melhores valores possíveis para as variáveis definidas, foi utilizado um *loop* para cada variável. Os valores utilizados para a variável “janela” foram 5, 10, 20, 30 e 60, para a variável “diferençaPredicaoEntradaComprando” (DEC) utilizou-se os valores de 0.1, 0.5, 1 e 2, para a variável “diferençaPredicaoEntradaVendendo” (DEV) também utilizou-se os valores 0.1, 0.5, 1 e 2, já na variável “diferençaMercadoSaida” (DS) os valores são 1, 2, 4, 6, 8 e 10 e a variável “perdaMaxima” (PM) utilizou-se 1, 2, 4, 6, 8 e 10.

A Tabela 5, Tabela 6 e Tabela 7 apresentam os resultados obtidos realizando o *loop* dos parâmetros. Nestes testes optou-se por utilizar o modelo de classificação PAClassifier visto que este modelo apresentou a maior lucratividade.

Janela	DEC	DEV	DS	PM	Acertos (%)	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.1	0.5	2	6	63.71	36.29	938.7	40.39
5	0.1	0.5	2	8	61.85	38.15	939.4	40.42
5	0.1	0.5	2	10	61.89	38.11	951.8	40.96

Tabela 5 – 3 melhores resultados dos testes em *loop* utilizando os papeis da IVVB11 - Fonte: Autor

Janela	DEC	DEV	DS	PM	Acertos (%)	Erros (%)	Lucro (R\$)	Lucro (%)
10	0.1	0.5	4	1	44.98	55.02	236.6	10.18
10	0.1	0.5	6	1	46.45	53.55	258.6	11.13
20	0.1	0.1	1	1	50.35	49.65	353.1	15.19

Tabela 6 – 3 melhores resultados dos testes em *loop* utilizando os papeis da VALE3 - Fonte: Autor

Janela	DEC	DEV	DS	PM	Acertos (%)	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.1	0.1	1	1	51.96	48.04	5.9	0.25
5	0.1	1	1	1	51.96	48.04	11.5	0.49
5	0.1	0.1	2	1	53.07	46.93	19.2	0.83

Tabela 7 – 3 melhores resultados dos testes em *loop* utilizando os papeis da ABEV3 - Fonte: Autor

Dados os experimentos, verificou-se a baixa lucratividade dos ativos ABEV3 e VALE3. Ao analisar os registros, é perceptível que os valores setados

para as variáveis de “diferencaPredicaoEntradaComprando”, “diferencaPredicaoEntradaVendendo” e “diferencaMercadoSaida” não são compatíveis com a variação do ativo. Outro ponto observado foi que a janela acima de 30 não apresentaram resultados úteis para pesquisa. A fim de otimizar os resultados, fez-se os ajustes necessários e uma nova bateria de testes.

Nesta nova bateria de testes, o tempo de execução do *loop* em todos ativos foi de aproximadamente 8 horas, os valores de maior lucratividade e taxa de acerto foram coletados toda vez que um melhor resultado era encontrado. Os resultados foram obtidos no mesmo *loop* de cada ativo.

5.5.2.4 Experimentos com papeis da IVVB11

A Tabela 8 apresenta os melhores resultados do *loop* considerando os maiores valores de lucro obtidos para o ativo IVVB11, neste experimento os valores utilizados para a variável “janela” foram 5, 10, 20 e 30, para a variável “diferencaPredicaoEntradaComprando” (DEC) foram 0.1, 0.5, 1 e 2, para a variável “diferencaPredicaoEntradaVendendo” (DEV) também utilizou-se os valores 0.1, 0.5, 1 e 2, já na variável “diferencaMercadoSaida” (DS) os possíveis valores são 0.1, 0.5, 1, 2 e a variável “perdaMaxima” (PM) utilizou-se 1, 2, 4, 6, 8 e 10.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.1	0.1	0.1	2	1872	88.97	232	11.03	426.3	18.34
5	0.1	0.5	0.5	4	494	82.61	104	17.39	465.0	20.01
5	0.1	0.5	2	4	156	59.77	105	40.23	552.5	23.77
10	0.1	0.5	2	2	166	53.72	143	46.28	613.4	26.39
10	0.1	1	2	2	166	53.72	143	46.28	624.6	26.88
10	0.1	0.5	0.5	4	461	83.67	90	16.33	638.1	27.46
10	0.1	1	0.5	4	458	83.73	89	16.27	659.0	28.36
20	0.1	0.5	1	2	312	67.53	150	32.47	677.0	29.13
20	0.1	0.5	0.5	4	424	84.46	78	15.54	723.3	31.12

Tabela 8 – Melhores resultados dos experimentos objetivando maior lucratividade utilizando os papeis da IVVB11 e o modelo Classificador HoelfdingTree - Fonte: Autor

A fim de validar os experimentos, utilizou-se um novo *dataset* com 3 meses posterior ao *dataset* utilizado para o treinamento. Neste caso, os registros são referentes a data de 01/07/2022 à 05/10/2022 contendo 27.910 registros.

Dessa forma vislumbra-se validar se os lucros e taxas de acerto encontrados com os melhores valores. Para isso, utilizou-se os 3 melhores valores de variáveis encontrados na Tabela 8, que indicaram maior lucratividade.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
10	0.1	1	0.5	4	161	80.1	40	19.9	132.9	5.72
20	0.1	0.5	1	2	106	61.99	65	68.01	139.2	5.99
20	0.1	0.5	0.5	4	162	80.2	40	19.8	137.8	5.93

Tabela 9 – Validação dos resultados utilizando 3 meses posteriores ao experimento, papéis da IVVB11 e modelo Classificador HoelffdingTree - Fonte: Autor

Ao analisar a Tabela 9, é perceptível a baixa lucratividade, mesmo com uma pequena redução na taxa de acerto. Entretanto ao levar em consideração que o *dataset* utilizado é referente a 3 meses, logo, ao realizar o cálculo de juros compostos representado pela equação $M = C (1+i)^t$, obtém-se os valores de 24,91%, 26,20% e 25,91% de lucro. Sendo este um resultado satisfatório para o experimento.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.1	0.1	0.1	2	1872	88.97	232	11.03	426.3	18.34
5	0.1	0.5	0.1	2	1862	89.01	230	10.99	365.4	15.72
5	0.1	1	0.1	2	1868	89.08	229	10.92	404.4	17.40
5	0.1	0.5	0.1	4	1199	92.23	101	7.77	380.5	16.37
5	0.1	0.5	0.1	6	1041	92.45	85	7.55	411.0	17.69
10	0.1	0.1	0.1	4	1121	92.80	87	7.20	518.6	22.31
10	0.1	1	0.1	4	1105	92.86	85	7.14	569.0	24.48
10	0.1	0.1	0.1	10	930	92.91	71	7.09	631.7	27.18
10	0.1	1	0.1	10	933	92.93	71	7.07	628.2	27.03
20	0.1	0.1	0.1	4	1013	93.19	74	6.81	640.9	27.58
20	0.1	0.1	0.1	6	859	93.27	62	6.73	687.9	29.6
20	0.1	0.1	0.1	8	855	93.34	61	6.66	709.2	30.52
20	0.1	0.1	0.1	10	855	93.44	60	6.56	701.1	30.17

Tabela 10 – Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papéis da IVVB11 e o modelo Classificador HoelffdingTree

Vislumbrando melhorar a consistência de lucro ao longo do tempo, mesmos que este não seja o melhor resultado, experimentou-se encontrar as maiores taxas de acerto. Verificou-se que os resultados foram satisfatórios, com taxas de acerto em geral acima de 90%, e lucratividade acima dos 30%, conforme apresentado na Tabela 10.

Na Tabela 11, buscou-se novamente validar a lucratividade apresentada nos resultados da Tabela 10. Observa-se que a taxa de acerto e o lucro apre-

sentaram resultados levemente melhores que os experimentos que almejavam a maior lucratividade, apresentados na Tabela 9. Entretanto, desta vez os resultados foram ainda mais satisfatórios ao considerar este lucro ao longo de um ano, com 29,08%, 26,62% e 26,62%, indicando que a maior taxa de acerto ao longo do tempo pode representar maior lucro ao longo do tempo.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
20	0.1	0.1	0.1	6	387	91.92	34	8.08	153.1	6.59
20	0.1	0.1	0.1	8	375	91.69	34	8.31	141.2	6.08
20	0.1	0.1	0.1	10	375	91.69	34	8.31	141.2	6.08

Tabela 11 – Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da IVVB11 e modelo Classificador HoelfdingTree - Fonte: Autor

5.5.2.5 Experimentos com papeis da ABEV3

Os mesmos testes foram realizados nos demais ativos, a Tabela 12 apresenta os resultados objetivando maior taxa de acerto para os papeis da ABEV3, os possíveis valores utilizados na variável “janela” foram 5, 10 e 20, para a variável “diferencaPredicaoEntradaComprando” (DEC) utilizou-se os valores de 0.02, 0.05, 0.1, 0.5 e 1, para a variável “diferencaPredicaoEntradaVendendo” (DEV) também utilizou-se os valores 0.02, 0.05, 0.1, 0.5 e 1, já na variável “diferencaMercadoSaida” (DS) os possíveis valores são 0.1, 0.5, 1, 2 e a variável “perdaMaxima” (PM) utilizou-se 0.1,0.5,1,2.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.02	0.05	0.1	748	62.44	450	37.56	23.4	12.65
5	0.02	0.05	0.5	0.1	121	40.07	181	59.93	48.5	26.22
5	0.02	0.1	0.5	0.1	120	39.8	181	60.13	49.8	26.92
10	0.02	0.05	0.5	0.1	89	39.38	137	60.62	51.8	28.0

Tabela 12 – Melhores resultados dos experimentos objetivando maior lucratividade utilizando os papeis da ABEV3 e o modelo Classificador HoelfdingTree - Fonte: Autor

Analisando a Tabela 12 é perceptível que o número de combinações possíveis para obter-se lucro é pequena, bem como os lucros obtidos pelo Robô Trader, observa-se também que a taxa de acertos é baixa em relação aos valores encontrados quando comparado com o ativo IVVB11.

Quando realizado os testes para validar os melhores parâmetros encontrados conforme apresentado na Tabela 13, identificou-se uma perda significativa de

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.05	0.5	0.1	40	30.76	90	69.23	-27.3	-147.56
5	0.02	0.1	0.5	0.1	41	31.06	91	68.94	-36.8	-198.91
10	0.02	0.05	0.5	0.1	39	31.70	84	68.30	-21.7	-117.29

Tabela 13 – Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da ABEV3 e modelo Classificador HoelffdingTree - Fonte: Autor

lucro, possivelmente relacionado ao baixo número de combinações encontradas pelo Robô Trader para obter maior lucratividade.

Buscando minimizar os lucros negativos, executou-se também os experimentos visando encontrar os melhores parâmetros para maior taxa de acerto no ativo ABEV3. Os resultados obtidos são apresentados na Tabela 14.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.02	0.05	0.1	748	62.44	450	37.56	23.4	12.65
5	0.02	0.02	0.02	0.5	550	83.59	108	16.41	8.8	4.76
20	0.02	0.02	0.02	0.5	172	85.57	29	14.43	21.3	11.51

Tabela 14 – Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papeis da ABEV3 e o modelo Classificador HoelffdingTree - Fonte: Autor

Ao analisar a Tabela 14 é possível verificar que houve um aumento significativo na taxa de acertos, entretanto o lucro não acompanhou os resultados positivos.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.05	0.5	0.1	40	30.76	90	69.24	-27.3	-147.56
5	0.02	0.02	0.02	0.5	195	82.27	42	17.73	-18.4	-99.45
20	0.02	0.02	0.02	0.5	174	82.08	38	17.92	-21.0	-113.51

Tabela 15 – Validação dos resultados utilizando 3 meses posteriores ao experimento, papeis da ABEV3 e modelo Classificador HoelffdingTree - Fonte: Autor

Após o experimento visando validar os melhores parâmetros encontrados, apresentados na Tabela 15, é possível concluir que o Robô Trader não comporta-se bem no ativo ABEV3, uma vez que as taxas de lucros continuaram negativas apresentando apenas uma pequena melhora no lucro.

5.5.2.6 Experimentos com papeis da VALE3

A Tabela 16 apresenta os resultados objetivando maior taxa de acerto para os papeis da VALE3, os possíveis valores utilizados nas variáveis foram os

mesmos que para o ativo ABEV3 visto a semelhança na variação de pontos dos ativos.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.02	0.02	0.1	7427	57.8	5423	42.2	28.5	2.9
5	0.02	0.05	0.02	0.1	7446	58.09	5372	41.91	43.4	4.41
5	0.02	0.02	0.05	0.1	5647	51.29	5364	48.71	126.7	12.88
5	0.02	0.1	0.05	0.1	5690	52.03	5245	47.97	199.8	20.3
5	0.02	0.1	0.1	0.1	3893	43.67	5022	53.33	215.3	21.88
5	0.02	0.05	0.5	0.1	740	20.43	2882	79.57	237.2	24.11
5	0.02	0.1	0.5	0.1	748	20.75	2856	79.25	462.6	47.01

Tabela 16 – Melhores resultados dos experimentos objetivando maior lucratividade utilizando os papéis da VALE3 e o modelo Classificador HoelfdingTree - Fonte: Autor

Na Tabela 16 é possível perceber que foi encontrado um parâmetro com altas taxas de lucro, entretanto com baixa taxa de acerto.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.1	0.1	0.1	911	42.50	1233	57.50	-140.4	-173.17
5	0.02	0.05	0.5	0.1	183	21.53	667	78.47	35.8	3.64
5	0.02	0.1	0.5	0.1	174	20.14	690	79.86	-21.6	-21.95

Tabela 17 – Validação dos resultados utilizando 3 meses posteriores ao experimento, papéis da VALE3 e modelo Classificador HoelfdingTree - Fonte: Autor

Ao validar os melhores parâmetros encontrados utilizando o dataset dos 3 meses posteriores, apresentados na Tabela 17 percebe-se que houve uma divergência na porcentagem de lucro, indicando uma possível aleatoriedade.

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.02	0.02	0.1	7427	57.8	5423	42.2	28.5	2.9
5	0.02	0.05	0.02	0.1	7446	58.09	5372	41.91	43.4	4.41
5	0.02	0.5	0.02	0.5	3891	83.03	795	16.97	72.5	7.37
5	0.02	1	0.05	2	1443	90.93	144	9.07	17.6	1.79

Tabela 18 – Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papéis da VALE3 e o modelo Classificador HoelfdingTree - Fonte: Autor

Ainda buscando a maior taxa de acerto do ativo VALE3, a Tabela 18 apresenta os melhores valores para os parâmetros. É perceptível que houve uma elevação na taxa de acertos, entretanto o lucro não apresentou valores consideráveis.

Ao realizar a verificação dos parâmetros, percebe-se que não houveram resultados positivos, indicando que o Robô Trader não comporta-se bem no

Janela	DEC	DEV	DS	PM	Acertos	Acertos (%)	Erros	Erros (%)	Lucro (R\$)	Lucro (%)
5	0.02	0.05	0.02	0.1	8094	58.25	5800	41.75	-12.5	-12.70
5	0.02	0.5	0.02	0.5	959	82.31	206	17.69	-184.8	-187.80
5	0.02	1	0.05	2	310	89.33	37	10.66	-80.9	-82.21

Tabela 19 – Melhores resultados dos experimentos objetivando maior taxa de acerto utilizando os papéis da VALE3 e o modelo Classificador HoelfdingTree - Fonte: Autor

ativo VALE3, uma vez que as taxas de lucros não apresentaram valores positivos ou significativos após os testes em meses posteriores.

6 CONCLUSÕES

Foi possível utilizar os modelos de aprendizado contínuo para predição do mercado financeiro, analisar os resultados e criar um protótipo de Robô Trader, e com este realizar os experimentos propostos nesta pesquisa.

Esta pesquisa visou responder o seguinte questionamento: Como realizar operações financeiras de forma automática utilizando modelos de aprendizado de máquina para obter ganhos acima da taxa básica de juros SELIC, consumindo pouco tempo para tomada de decisão e pouco tempo para a atualização do modelo em relação a novos padrões?

Para responder o questionamento, o estudo foi direcionado a utilização de algoritmos de aprendizado de máquina contínuo, por sua característica de rápida tomada de decisão e aprendizado *on-the-fly*. Para tanto, realizou-se o desenvolvimento de um robô com dois modelos contínuos em sequência: um para realizar a predição dos próximos valores do ativo e outro para decidir o melhor momento para comprar ou vender, de acordo com o valor obtido da predição do modelo antecessor.

Para os experimentos foram utilizados os modelos de regressão Naive Bayes, KNNRegressor e SoftmaxRegression. Dentre estes, verificou-se que o modelo KNNRegressor apresentou melhores resultados devido a sua precisão satisfatória e baixo tempo de execução. Já para o modelo de classificação, os experimentos foram realizados utilizando HoeffdingTreeClassifier, Naive BayesClassifier, KNNClassifier, PAClassifier e Perceptron Classifier, dentre estes, o modelo HoeffdingTreeClassifier apresentou a melhor acurácia e precisão (F1 Score), sendo este selecionado para execução dos experimentos no decorrer do estudo.

Observando os resultados obtidos, conclui-se que os modelos de aprendizado de máquina contínuo conseguem realizar operações de forma automática e apresentar ganhos acima da taxa básica de juros SELIC nos papeis da IVVB11

e desconsiderando as taxas de corretagem, emolumentos, custódia e imposto de renda. Entretanto, nos papeis da ABEV3 e VALE3 o robô não atingiu os resultados esperados. Ao verificar os resultados obtidos com o desvio padrão dos ativos apresentados, observa-se que quando as variações são mais claras (maiores), ou seja, ativos mais voláteis, o robô obtém resultados melhores. Outro ponto a ser observado é que a alta variação pode favorecer o robô. Uma alta variação pode significar potenciais ganhos maiores, acompanhados de riscos maiores. Ao que tudo indica, o sucesso do robô está diretamente relacionado a escolha adequada do ativo.

Sugestões para trabalhos futuros:

- Utilizar outros modelos disponíveis na biblioteca do River;
- Realizar experimentos com outros ativos como por exemplo Mini Dólar;
- Realizar experimentos utilizando indicadores do mercado financeiro para auxiliar o Robô Trader a encontrar padrões;
- Otimizar os parâmetros disponíveis nos modelos utilizados.

REFERÊNCIAS

- AMORIM, M. J.; BARONE, D.; MANSUR, A. U. Técnicas de aprendizado de máquina aplicadas na previsão de evasão acadêmica. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2008. v. 1, n. 1, p. 666–674. Citado na página 12.
- ANDRADE, A. D. d.; SOUZA, Y. K. R. d. et al. Inteligência artificial como auxílio ao investidor na tomada de decisão na bolsa de valores. Florianópolis, SC, 2019. Citado 2 vezes nas páginas 20 e 24.
- ANDRADE, G. G. J. *Os robôs do mercado financeiro: entre oportunidades e riscos*. 2019. Url<https://arte.estadao.com.br/focas/estadaoqr/materia/os-robos-do-mercado-financeiro-entre-oportunidades-e-riscos>. Citado na página 18.
- ANDRADE, M. V. Z. d. Operando ativos com base em análise técnica: aplicação do método da agulhada. Universidade Federal de Campina Grande, 2018. Citado na página 12.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques-part i: Conventional methods. *Journal of Computational Optimization in Economics and Finance*, v. 2, n. 1, p. 45–92, 2010. Citado na página 14.
- B3. *Índice Bovespa (Ibovespa B3)*. 2022. Urlhttps://www.b3.com.br/pt_br/market_-_data_-_e_-_indices/indices/indices_-_amplos/indice_-_ibovespa_-_ibovespa_-_composicao_-_da_-_carteira.htm. Citado na página 26.
- BODIE, Z.; KANE, A.; MARCUS, A. *EBOOK: Investments-Global edition*. [S.l.]: McGraw Hill, 2014. Citado na página 16.
- BOEING, D. H. A. et al. Ensinando um robô a julgar: pragmática, discricionariedade e vieses no uso de aprendizado de máquina no judiciário. Florianópolis, SC, 2019. Citado na página 12.
- BORGES, L. E. *Python para desenvolvedores: aborda Python 3.3*. [S.l.]: Novatec Editora, 2014. Citado na página 26.
- CORREIA, J. S. *Operando na bolsa de valores utilizando análise técnica*. [S.l.]: Novatec Editora, 2008. Citado na página 19.

DANTAS, S. G. Utilização de aprendizado por reforço para operações em bolsa de valores. 2017. Citado 4 vezes nas páginas 17, 18, 20 e 21.

FERNANDES, L. S. de B. *TÉCNICAS DE APRENDIZADO DE MÁQUINA APLICADAS A ALGOTRADING NO MERCADO DE AÇÕES*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2019. Citado na página 24.

FONTENLA-ROMERO, Ó. et al. Online machine learning. In: *Efficiency and Scalability Methods for Computational Intellect*. [S.l.]: IGI Global, 2013. p. 27–54. Citado na página 13.

GAMA, J. *Knowledge discovery from data streams*. [S.l.]: CRC Press, 2010. Citado na página 22.

GÉRON, A. *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow*. [S.l.]: Alta Books, 2019. Citado 3 vezes nas páginas 22, 29 e 30.

HENRIQUE, B. M. Predição da direção dos preços de ativos do mercado financeiro usando aprendizagem de máquina. 2018. Citado na página 24.

JR, L. A. et al. Predição de comportamento de ações no mercado financeiro usando redes neurais recorrentes lstm. Pontifícia Universidade Católica de Goiás, 2021. Citado na página 24.

JUNIOR, C. R. S. Algoritmo para a extração incremental de sequências relevantes com janelamento e pós-processamento aplicado a dados hidrográficos. Universidade Federal de São Carlos, 2013. Citado na página 22.

KIYOSAKI, R. T.; LECHTER, S. L. *Rich Dad Poor Dad: What the Rich Teach Their Kids About Money-That the Poor and the Middle Class Do Not!* [S.l.]: Business Plus, 2001. Citado na página 16.

LAVAL, L. *B3 atinge 5 milhões de contas de investidores pessoa física em renda variável*. 2022. Url <https://www.cnnbrasil.com.br/business/b3-atinge-5-milhoes-de-contas-de-investidores-pessoa-fisica-em-renda-variavel/>. Citado na página 11.

LIRA, M. C.; ALMEIDA, S. A. de. A volatilidade no mercado financeiro em tempos da pandemia do (novo) coronavírus e da covid-19: Impactos e projeções. *Facit Business and Technology Journal*, v. 1, n. 19, 2020. Citado na página 16.

MCKINNEY, W. *Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython*. [S.l.]: Novatec Editora, 2019. Citado na página 27.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, Manole, v. 1, n. 1, p. 32, 2003. Citado na página 12.

MONTIEL, J. et al. *River: machine learning for streaming data in Python*. 2020. Citado na página 28.

MURPHY, J. J. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. [S.l.]: Penguin, 1999. Citado na página 18.

NETO, V. J. d. S.; BONACELLI, M. B. M.; PACHECO, C. A. O sistema tecnológico digital: inteligência artificial, computação em nuvem e big data. *Revista Brasileira de Inovação*, SciELO Brasil, v. 19, 2021. Citado na página 12.

PAIXÃO, G. M. de M. et al. Machine learning na medicina: Revisão e aplicabilidade. *Arquivos Brasileiros de Cardiologia*, Sociedade Brasileira de Cardiologia, v. 118, n. 1, p. 95, 2022. Citado na página 21.

PENA, M. de S.; MENA-CHALCO, J. P. Análise exploratória e preditiva de textos de divulgação científica. 2021. Citado na página 31.

REIS, A. M. d. Incremental learning applied to streaming environments. 2017. Citado na página 22.

ROCHA, C. M. H. S. da et al. Ciência de dados e aprendizado de máquina para predição em séries temporais financeiras. Universidade Federal de Minas Gerais, 2019. Citado na página 24.

SALMAZO, G. C. Validação de um modelo dinâmico preditivo de crescimento da cocultura de streptococcus thermophilus e lactobacillus bulgaricus para a fermentação de iogurte em escala laboratorial e industrial. 2021. Citado na página 31.

SANTANA, G. F. *O poder de previsão da análise técnica: uma aplicação para o mercado futuro de IBOVESPA*. Tese (Doutorado), 1997. Citado na página 14.

SASSAKI, B. H. Projeto de robô trader assistido com inteligência artificial. Universidade Estadual Paulista (Unesp), 2021. Citado na página 17.

SEMELER, M. W. A. A arte e a ciência de dados na era digital. *Revista de Artes Visuais*, v. 24, p. 2–3, 2019. Citado na página 12.

- SILVA, M. M. Uma abordagem evolucionária para aprendizado semi-supervisionado em máquinas de vetores de suporte. Universidade Federal de Minas Gerais, 2008. Citado na página 21.
- SILVA, V. H. M. N. *Renda passiva: a busca pela independência financeira com a estratégia Buy and Hold*. Dissertação (B.S. thesis) — Universidade Federal do Rio Grande do Norte, 2022. Citado na página 11.
- SPRITZER, F. A. Análise técnica para day trade: rentabilidade de indicadores no longo prazo. 2017. Citado 3 vezes nas páginas 12, 18 e 19.
- SUTTER, L. F. M. *Aplicação de técnicas de aprendizado de máquina na predição de tendência das ações nas bolsas de valores*. Tese (Doutorado), 2018. Citado 3 vezes nas páginas 14, 17 e 20.
- TAKAKURA, A. M. et al. Uso do aprendizado de máquina no diagnóstico médico de patologias. In: *Colloquium Exactarum*. ISSN: 2178-8332. [S.l.: s.n.], 2018. v. 10, n. 1, p. 78–89. Citado na página 12.
- TEIXEIRA, L. de A. Métodos de regressão para aprendizado por reforço. 2016. Citado na página 21.
- VOHS, K. D.; MEAD, N. L.; GOODE, M. R. The psychological consequences of money. *science*, American Association for the Advancement of Science, v. 314, n. 5802, p. 1154–1156, 2006. Citado na página 11.
- VOHS, K. D.; MEAD, N. L.; GOODE, M. R. Merely activating the concept of money changes personal and interpersonal behavior. *Current Directions in Psychological Science*, SAGE Publications Sage CA: Los Angeles, CA, v. 17, n. 3, p. 208–212, 2008. Citado na página 11.