

**Efeitos da Latência em Negociações
na Bolsa de Valores B3**

Airton Gomes Finoti

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Efeitos da Latência em Negociações na Bolsa de Valores B3

Airton Gomes Finoti

Versão Revisada

USP - São Carlos
2022

Airton Gomes Finoti

Efeitos da Latência em Negociações na Bolsa de Valores B3

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Willian Dener de Oliveira

Versão Revisada

USP - São Carlos

2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

G633e Gomes Finoti, Airtton
 Efeitos da Latência em Negociações na Bolsa de
Valores B3 / Airtton Gomes Finoti; orientador
Willian Dener De Oliveira. -- São Carlos, 2022.
 44 p.

 Trabalho de conclusão de curso (MBA em
Inteligência Artificial e Big Data) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2022.

 1. Classificador, SVM, Inteligência Artificial.
2. Day-trading, Investimentos, Negociações. 3.
Latência, Atraso. 4. Bolsa de Valores B3, Bovespa.
I. De Oliveira, Willian Dener , orient. II. Título.

*Dedico esse trabalho a todos que arriscam própria
pele. Aproveitar e sofrer dos próprios atos e
decisões, nada mais justo, nada mais belo.*

AGRADECIMENTOS

A minha família pelo constante apoio, incentivo aos estudos e carinho em excesso.

Aos meus amigos de Porto Feliz, sempre companheiros e presentes na minha vida.

Aos meus amigos de São Carlos, cuja paixão pela computação uniu e colocou nessa profissão que amo exercer. Poucos momentos são tão agradáveis e nerds como nossos encontros.

Ao meu orientador, mestre Willian, pelas dezenas de horas de conversas leves e empolgantes sobre o tema e seus paralelos. Sua sensatez me deu o foco para realizar uma pesquisa concreta dentro do prazo.

Ao meu fã número 1, Eud, sempre me motivando a superar meus limites.

66

*tenho um amigo
ele é comum
ele tem nome
tem sentido
tem alma
tem muita constância
ele nem sempre é visível
mas a gente sente
as vezes ele pode ser esquecido
mas nunca deixa de estar contigo
ele é amigo da pedra
é amigo da terra
é amigo do rio
ele às vezes assusta
e assusta mesmo
porque tem muito
muito dele mesmo
é todo ele
real
sem disfarce
o nome dele é:
TEMPO*

99

(Breno Boro)

RESUMO

Gomes Finoti, Airton. **Efeitos da Latência em Negociações na Bolsa de Valores B3**. 2022. 44 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

Com mais investidores entrando no mercado de renda variável, plataformas e corretoras de investimento se atualizam para oferecerem melhores serviços, mas independente da escolha, sempre há um atraso nas ordens de negociações. Em uma primeira etapa, os efeitos das latências no envio das ordens do mini índice da BOVESPA (WIN) são obtidos através de simulações com os mesmos dados que as corretoras provêm aos investidores. Diferenças como sucesso na execução e o preço obtido são comparados evidenciando possíveis prejuízos quando a latência aumenta. Com essa análise dos dados, foram criados alguns critérios mitigando o efeito da latência para estratégia de investimento. Na segunda etapa, um classificador SVM é treinado com o objetivo de escolher qual é a melhor ação a ser realizada dentro de 3 opções (comprar, abster-se e vender) em cada intervalo de 15 minutos. Os resultados demonstram que não se pode ignorar a latência durante a etapa de treinamento e sua inserção na análise de dados permite a criação de estratégias mais lucrativas para cada cenário.

Palavras-chave: IBOVESPA, Aprendizado de máquina, Mercado Financeiro, SVM, *Day-trading*, Latência

ABSTRACT

Gomes Finoti, Airton. Lag effects over day-trading **Efeitos da Latência em Negociações na Bolsa de Valores B3**. 2022. 44 p. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

As more investors enter the equity market, investment platforms and brokers upgrade their services, but there is always a delay in trading orders regardless of choice. As the first step, through simulations with the same data brokerage firms provide to investors, we simulate the effects of latencies in sending orders to BOVESPA mini index (WIN). Differences such as execution success and the price obtained are compared, showing possible losses when latency increases. With this data analysis, we create criteria to mitigate the latency effect of the investment strategy. In the second step, we trained a SVM classifier to choose the best action within three options (buy, abstain and sell) at each 15-minute interval. The results evidenced that latency is relevant during the training stage and the data analysis allows the creation of more profitable strategies for each latency scenario.

Keywords: IBOVESPA, Machine Learning, Financial Market, SVM, Day-trading, Latency

LISTA DE ILUSTRAÇÕES

Figura 1 - Livro de ofertas.....	6
Figura 2 - Modificações do livro de ofertas	7
Figura 3 – Tipos de ordem de em relação ao preço.....	8
Figura 4 - (A) Representação de Barras (B) 5 Barras de 1 Dia Cada.....	10
Figura 5 – Fluxo de processos de aquisição e tratamento de dados	13
Figura 6 - Fluxo de processos de aprendizado e classificação	14
Figura 7 - Calendário de calor com volumes diários.....	16
Figura 8 - Calendário de dias disponíveis	17
Figura 9 - Validação visual dos indicadores.....	20
Figura 10 - Código para criação das ordens em a cada 1 minuto.....	22
Figura 11 - Janela de treinamento.....	24
Figura 12 - Descarte aleatório de amostras	25
Figura 13 - Efeito latência ordens limite	30
Figura 14 - Ordens limite sem entrada por latência.....	31
Figura 15 - Status de ordens tipo limite ao longo do dia.....	32
Figura 16 - Ordens limite em aberto por latência.....	32
Figura 17 - Dispersão dos prejuízos por latência	33
Figura 18 - Classes desbalanceadas.....	34
Figura 19 - Classes balanceadas	34
Figura 20 - Média de acertos por tamanho de janela.....	35
Figura 21 - Média de acertos por janela por latência.....	36
Figura 22 - Taxa de acerto por classificador por latência (28 semanas)	36
Figura 23 - Taxa de acertos SVM com janela de 28 semanas.....	37
Figura 24 - Lucro por latência SVM	37
Figura 25 - Lucro por latência do classificador Aleatório.....	38
Figura 26 - Resultados metodologia com latência ou ingênua.....	39

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Revisão da literatura e trabalhos relacionados	1
1.2 Objetivos	3
1.3 Organização do trabalho	3
2. FUNDAMENTAÇÃO TEÓRICA	5
2.1 Considerações iniciais	5
2.2 Base de dados	5
2.3 Livro de ofertas	5
2.4 Spread e spread prático	6
2.5 Tipos de ordens	7
2.6 Posição	9
2.7 Leilões	9
2.8 Gráfico de barras	10
2.9 Extração de características e indicadores técnicos	10
3. METODOLOGIA	13
3.1 Aquisição das negociações	14
3.2 Extração de velas de 1 minuto	15
3.3 Análise de volume	15
3.4 Concatenação dos arquivos de atributos	18
3.5 Inserção dos indicadores agrupados em 15 Minutos	18
3.6 Relativização dos atributos	20
3.7 Simulação de trades	21
3.8 Concatenação de arquivos de simulação de trades	23
3.9 Análise dos efeitos da latência	23
3.10 Escolha da melhor ação em intervalos de 15 minutos	23
3.11 União de atributos e classes	24
3.12 Divisão por semanas	24
3.13 Amostragem e balanceamento	25
3.14 Aprendizado e classificação	27
4. RESULTADOS	29
4.1 Análise dos trades	29
4.2 Resultados dos classificadores	35
5. CONCLUSÕES E TRABALHOS FUTUROS	41
REFERÊNCIAS	43

Anexo 1 - Exemplo de 5 entradas do classificador.....	44
--	-----------

1. INTRODUÇÃO

A busca por ferramentas e técnicas de análise de séries temporais no mercado financeiro é objeto de estudo de diversas áreas e, com o aumento de investidores iniciantes no mercado de capitais, conhecimento, adaptação e bom uso dessas ferramentas podem ser cruciais para o sucesso financeiro do investidor.

O número de brasileiros investindo na bolsa de valores vem aumentando (CHAGUE e colab., 2020). Uma das modalidades de investimento que exige pouco capital inicial e permite grandes alavancagens é o *day-trading* de índices ou moedas, como o dólar. Operando como *day-trader*, o investidor compra e vende o mesmo ativo financeiro, no mesmo dia e na mesma quantidade, finalizando o dia sem nenhuma posição em sua carteira. Mesmo sendo uma prática comum, pesquisas mostram que os resultados financeiros da grande maioria dos *traders* são negativos ou medíocres em comparação a outras profissões (CHAGUE e colab., 2020).

Enquanto ferramentas de análise gráfica e plataformas de negociação continuam aumentando suas funcionalidades e desempenho, os operadores ainda precisam tomar a decisão final de compra/venda para cada operação arcando com os custos psicológicos de cada decisão. Ferramentas como *Expert Advisors* do Metatrader5® ou “sistemas de apoio à decisão inteligente” [IDSS] detalhado por (DALLAQUA, 2020) facilitam os investimentos automatizando as compras e vendas com base em uma estratégia. Independentemente da plataforma de negociação, sempre haverá um atraso entre o momento da decisão e a realização da operação. Contudo, não foi encontrada na literatura, estudos sobre o feito das latências de negociação na bolsa brasileira.

A estratégia responsável pela tomada de decisões com base em séries temporais já foi vastamente explorada no âmbito de aprendizado de máquina como demonstrado por (ROUF e colab., 2021). Dos diversos métodos de aprendizado de máquina descritos neste trabalho, o SVM é o mais utilizado e será o método adotado aqui.

Esperamos contribuir analisando os efeitos das latências nas negociações e entender como podem afetar a busca por estratégias de negociação quando se utiliza o aprendizado SVM.

1.1 Revisão da literatura e trabalhos relacionados

O uso de aprendizado de máquinas para predição de preços foi vastamente explorado pelo mundo acadêmico como demonstrado por (ROUF e colab., 2021). Mesmo assim, devido aos desafios como a natureza incerta, aleatória e caótica, ainda existe necessidade de atualização e adequação para se trabalhar com volumes cada vez maiores de informação.

Adequação das técnicas para uso no dia a dia do investidor é um problema a ser resolvido pois “a maior parte da literatura sobre predição de preço em tempo real afirma que a metodologia proposta pode ser usada em tempo real, mas somente dentro circunstâncias controladas.” (Rouf et al., 2021, p.17, tradução nossa)

Em sua conclusão, (Rouf et al., 2021) menciona que apesar de técnicas como redes neurais artificiais (ANN) e redes neurais profundas (DNN) serem muito usadas por serem precisas e rápidas, SVM é a técnica mais detalhada nas centenas de artigos encontrados, representando aproximadamente 16% dos trabalhos, seguido por ANN com 14%. Isso é condizente com nossas buscas iniciais por artigos, justificando a escolha do método SVM para análise e adequação nesse estudo.

Comparações entre os diversos métodos de classificação foram feitas em várias bolsas de valores. Ao aplicar técnicas de aprendizado de máquina na bolsa de valores de Kuala Lumpur, (ISMAIL e colab., 2020) conclui ao final da otimização de parâmetros, que o método SVM teve melhor desempenho em todas as categorias, sendo uma delas o uso de indicadores técnicos com entrada.

Indicadores técnicos são usados de forma recorrente em muitos dos trabalhos encontrados como (Aguirre et al., 2020; Felipe Moreira Moura, 2020; Ismail et al., 2020; Teixeira & de Oliveira, 2010; Vinicius et al., 2013).

Investidores que usam indicadores técnicos podem aplicá-los à diversos tipos de agrupamentos de preços, como exemplo, o último preço negociado (*CLOSE*) em um intervalo, mas não somente a eles. (VINICIUS e colab., 2013) utiliza esses valores de agrupamento por intervalos, as barras, como parâmetros de entrada de sua estratégia aplicada à ações da bolsa brasileira.

Com método de aprendizado de máquina definido como SVM, as entradas baseadas em indicadores técnicos e um framework de treinamento e testes, chega-se à contribuição deste estudo trazendo para a prática diária onde compras e vendas deixam de ser uma abstração com execução absoluta em preços determinados e serão tratadas com a possibilidade de execuções em diferentes momentos dependendo da volatilidade, liquidez do mercado e dos atrasos (latência) no envio das ordens. Enquanto a volatilidade pode ser mensurada através de negociações realizadas, a liquidez se refere a diferença (*spread*) entre as ofertas de compra (*bid*) e vendas (*ask*) ainda não realizadas. Durante suas simulações, (SAMPAIO e GUEVARA, 2017) constata “A função estratégia está sujeita ao risco de liquidez expresso pelo *Bid-Ask Spread*. Desta forma, para gerar retorno sobressalentes não basta ter capacidade preditiva, mas a

indicação de movimentação entre compradores e vendedores deve superar a diferença entre a melhor oferta de compra e venda”.

1.2 Objetivos

O objetivo deste trabalho é demonstrar que a melhor operação a ser feita em cada intervalo de tempo do Índice Bovespa (IBOVESPA) no curto prazo pode variar dependendo das condições de latência no momento da negociação. O algoritmo de aprendizado de máquina escolhido para testes foi o SVM por ter sido o mais explorado nessa área segundo por (Rouf et al., 2021). Será demonstrado que a análise dos dados, considerando o atraso nas ordens nas operações, pode evitar prejuízos e melhorar a tomada de decisões do SVM.

1.3 Organização do trabalho

Este trabalho está estruturado em cinco capítulos, sendo os capítulos descritos a seguir:

No capítulo 1 foram apresentadas a introdução, a revisão da literatura, a definição dos objetivos do trabalho e a sua organização.

No capítulo 2 contém conceitos fundamentais para esse trabalho: aquisição e pré-processamento de dados, extração de características (indicadores técnicos dos índices), tipos de ordens disponíveis e aprendizado de máquina.

No capítulo 3 são apresentadas as metodologias empregadas no desenvolvimento desta pesquisa: o fluxo de trabalho adotado, os detalhes de cada uma das etapas do fluxo.

No capítulo 4 são apresentados os resultados obtidos e os desafios observados ao longo do desenvolvimento deste projeto.

No capítulo 5 são apresentadas as conclusões e as possibilidades de contribuições e trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Considerações iniciais

O presente capítulo tem como enfoque apresentar os conceitos e as terminologias básicas utilizadas para a compreensão do projeto proposto.

2.2 Base de dados

A base de dados é formada por informações de todas as ofertas de compra e venda que resultaram em negociação.

Existem 2 tipos de informação disponíveis:

- Ofertas: ordens enviadas pelos investidores que podem ou não ter fechado negócio.
- Negociações: ofertas que encontraram uma ou várias contrapartes com mesmo preço fechando seu volume total ou parcialmente.

As negociações dispõem dos seguintes dados de interesse: data e horário de execução (precisão de milissegundos), preço e volume.

2.3 Livro de ofertas

Para gerar mais negociações, a bolsa de valores fornece o mecanismo de livro de ofertas onde compradores e vendedores podem observar os volumes e valores disponíveis e demonstrar sua intenção compra e venda nos preços que são convenientes com suas estratégias financeiras.

Ofertas de compra/venda são organizadas por ordem de preço e idade de criação com ordens de compra mais caras e mais antigas no topo no lado esquerdo e ordens de venda mais baratas e mais antigas no lado direito.

Figura 1 - Livro de ofertas

Qtd	Compra	Venda	Qtd
30	104.580	104.585	39
2	104.580	104.585	2
5	104.580	104.585	6
5	104.580	104.585	9
2	104.580	104.585	3
9	104.580	104.590	3
2	104.580	104.590	12
3	104.580	104.590	5
8	104.580	104.590	2
10	104.580	104.590	3
1	104.580	104.590	2
1	104.580	104.590	3
2	104.580	104.590	6
1	104.580	104.590	10
5	104.580	104.590	6
9	104.580	104.590	2
3	104.580	104.590	3
1	104.580	104.590	13
1	104.580	104.590	4
10	104.580	104.590	9
9	104.575	104.590	3
2	104.575	104.590	9
5	104.575	104.590	4

Fonte: Elaborada pelo autor.

Essas ordens ficam em aberto até que uma ordem contrária execute uma negociação ou até tenham seus vencimentos expirados.

2.4 Spread e spread prático

O preço da maior ordem de compra é chamado de *bid* e o preço da menor ordem de venda é chamado de *ask*. A diferença entre o *bid* e o *ask* é chamado de *spread*. A bolsa define valores mínimos para o spread do índice futuro em 5 pontos, porém, não a limite máximo para o spread. Na figura 1 podemos observar um spread de $104585 - 104580 = 5$ pontos. Nessa dissertação, faremos uma diferenciação entre esse spread teórico e o spread real:

- Spread teórico é o observado em qualquer momento livro de ofertas e pressupõe que a última diferença observada é a mesma que será executada.
- Spread prático considera que existe um tempo de atraso entre a observação e a execução.

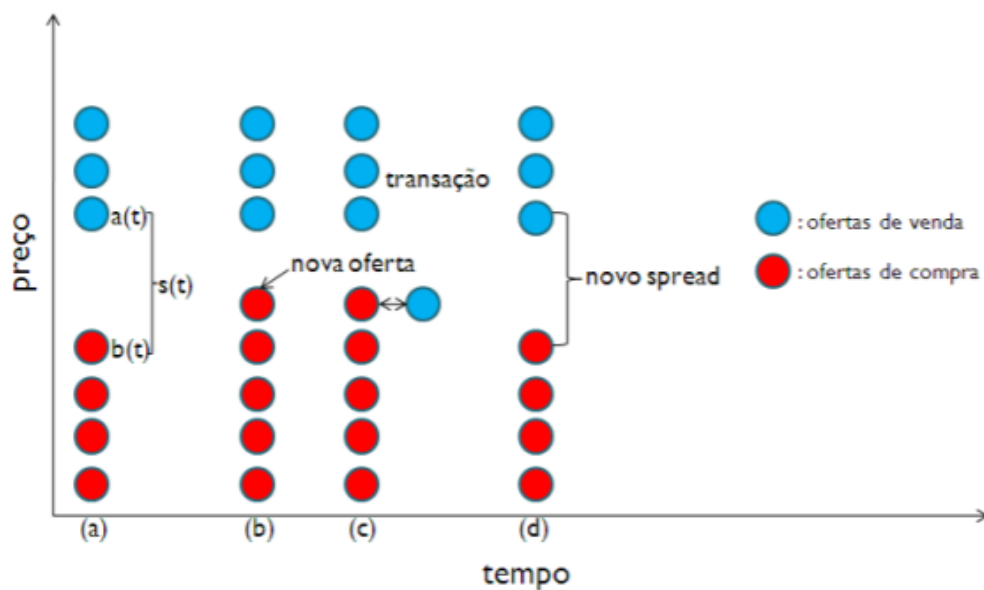
2.5 Tipos de ordens

As ordens disponíveis no livro, que são executadas somente no preço-alvo, são chamadas de ordens do tipo limite ou passivas, no inglês, *limit*. É possível lançar uma ordem de compra com preço maior que a menor ordem de venda, dessa forma, a ordem de compra vai executar a ordem de venda até seu volume seja consumido. Se a ordem de compra tiver volume maior que o disponível no preço de venda, o restante ficará em aberto.

Na figura 1, temos volume de 59 no preço de venda 104590, logo, uma ordem de compra com volume 60, executaria todas as ofertas no preço 104590 e manteria uma ordem de compra com volume 1 tipo limite nesse preço, alterando o livro e subindo o preço do ativo. Porém, como o mercado é dinâmico, outro investidor poderia mandar uma ordem anterior a essa, arrematando o preço 104590 deixando todo o volume de 60 em aberto.

Para garantir a execução, usa-se o tipo ordem a mercado, no inglês, *market*. São ordens ativas que serão executadas independente do preço. No momento que o investidor emite uma ordem de venda *market*, essa ordem será executada com o máximo de volume possível no maior preço de compra e se esse volume não for o suficiente para cobrir toda ordem, será executado com o próximo preço, até que seu volume esteja completo.

Figura 2 - Modificações do livro de ofertas

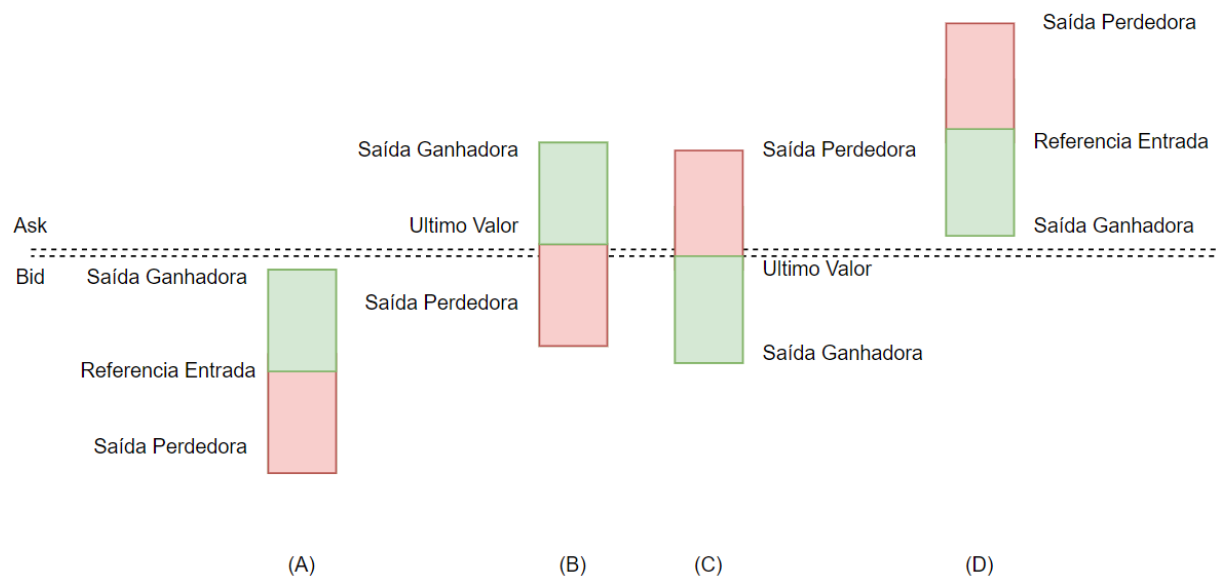


Fonte: (SILVA, 2013)

“Para simplificar, todas as ofertas possuem volume igual a um – os agentes sempre desejam vender/comprar uma unidade de papel. No instante (a), o livro de ofertas possui uma determinada configuração. No instante (b), uma nova oferta de compra é registrada. Porém, como seu preço é inferior a $a(t)$, ela não é executada e é inserida no livro. Ocorre apenas um aumento do *bid* e uma redução do *spread*. Em (c), uma oferta de venda com preço igual a $b(t)$ é registrada, a transação ocorre e ambas desaparecem do livro de ofertas, resultando na configuração retratada em (d).” (SILVA, 2013)

Além das ordens limite e mercado, existe um terceiro tipo de ordem disponível chamadas ordens gatilho ou *stop* que ficam salvas apenas em ambiente local da plataforma de negociação ou na corretora. A partir do momento que o preço do ativo atinge um alvo (gatilho), uma ordem a mercado é lançada. É uma ordem comumente usada para evitar prejuízos maiores quando uma posição começa ficar com resultado negativo dentro de uma estratégia, sendo chamada de *stop-loss*. Pode também ser usada para criar uma posição, sendo chamada de *stop-entry*. Ordens *stop* devem ser analisadas em duas etapas, ativação se o preço atinge o gatilho e execução que é semelhante a uma ordem mercado. O efeito da latência afeta a primeira etapa da mesma forma que uma ordem limite e a segunda como ordem mercado.

Figura 3 – Tipos de ordem de em relação ao preço



Fonte: Elaborada pelo autor.

Na figura 3, temos 4 ordens:

- A) Entrada de compra do tipo limite esperando execução. Será executada apenas quando *ask* diminuir até a entrada.
- B) Entrada de compra do tipo mercado sendo executada. Uma vez enviada é executada no menor *ask* disponível.
- C) Entrada de venda do tipo mercado sendo executada. Uma vez enviada é executada no maior *bid* disponível.
- D) Entrada de venda do tipo limite esperando execução. Será executada apenas quando *bid* aumentar até a entrada.

Uma vez que a ordem de entrada foi executada, as saídas são representadas em verde no caso de lucro ou vermelho em caso de prejuízo. Durante todo esse projeto, as saídas de lucro são do tipo limite e as saídas de prejuízo do tipo *stop-loss*.

2.6 Posição

A partir do momento que uma compra ou venda é executada, independentemente do tipo de ordem usada, o investidor criou uma posição que precisará ser fechada antes do final da sessão, sendo essa a definição de *day-trade*. Caso o investidor tenha iniciado uma posição compradora, que irá lucrar com aumento do preço do ativo, ele precisará em algum momento futuro fechar essa posição com uma venda. Em qualquer momento, a estratégia proposta pode ter posições compradoras e vendedoras simultaneamente. São controladas separadamente as operações, podendo gerar resultados diferentes dependendo do momento de encerramento de cada posição.

2.7 Leilões

Leilões de pré-abertura, pré-fechamento ou extraordinários (ocorrem ao longo do dia) são ferramentas utilizada pela bolsa de valores para equilibrar ofertas de compras e vendas quando o mercado está iniciando ou há um desequilíbrio nas ofertas de compra e venda. Nesses momentos as ordens não são executadas, apenas recebidas, até que o leilão se encerre.

2.8 Gráfico de barras

Como milhões de negociações são realizadas diariamente, analisá-las uma a uma seria uma tarefa humanamente impossível. Investidores modernos usam uma ferramenta chamada gráfico de barras. Todas as negociações dentro de um intervalo de tempo são sumarizadas em 6 valores: horário do primeiro trade, valor da primeira negociação ou abertura (*open*), máxima (*high*), mínima (*low*), valor da última negociação ou fechamento (*close*) e volume de negociações. Esse conjunto de valores é comumente chamado por OHLC e exemplificado na figura 4.

Figura 4 - (A) Representação de Barras (B) 5 Barras de 1 Dia Cada



Fonte: (VINICIUS e colab., 2013)

2.9 Extração de características e indicadores técnicos

A extração de características tem como objetivo encontrar o melhor conjunto de atributos, primitivos ou sintetizados, que forneçam informações complexas com elevado grau de eficácia após a aplicação de operações de transformação nos valores originais. Estas características derivadas são normalmente superiores na representação e agregação da informação em relação às originais.

Esta etapa é normalmente opcional, porém tem grande relevância durante o processo de desenvolvimento de modelos, pois as características extraídas podem apresentar uma elevada qualidade em suas informações, aprimorando significativamente a assertividade do treinamento, e consequentemente o resultado de predição de um modelo.

Em resumo, os dados originais das bases de dados são transformados em características aprimoradas. No presente projeto são extraídos alguns indicadores técnicos a partir das negociações da base de dados considerando sua ordem temporal comum a investidores em todo mundo como mostrado por (TEIXEIRA e DE OLIVEIRA, 2010) .

Uma vez que se usa um gráfico de barras, pode-se calcular médias e outros indicadores em quaisquer dos valores OHLC, mas para essa dissertação sempre se usa o valor de fechamento (*close*) de cada barra.

2.9.1 Índice da barra durante o dia

Olhando-se dentro de cada sessão (dia de operação da bolsa), cada sessão pode ser analisada usando barras e o horário das barras depende do intervalo de análise. Se o intervalo for 10 minutos e a sessão começar às 9:00, a barra que agrupa os trades entre 9:00 e 9:09:59.999 será 1. O próximo intervalo 2 até o final da sessão. O número de barras dependerá do intervalo em análise. O indicador de índice da barra é iniciado na primeira barra independente do horário da primeira barra. Dias de meio expediente podem ter sua primeira barra depois das 12:00.

2.9.2 Média Móvel Simples (SMA)

Média Móvel Simples é o valor médio dos preços de fechamento em um certo número de barras (Metatrader, 2022).

$$SMA(N) = \text{SUM}(\text{CLOSE}, N) / N$$

Tal que:

SUM (CLOSE, N) – a soma dos últimos N valores de fechamento (CLOSE);

N – número de períodos para o cálculo.

2.9.3 Média móvel exponencial (EMA)

A média móvel exponencial é o valor médio dos preços de fechamento em um certo número de barras onde os últimos preços de fechamento são de maior peso. ("Metatrader", 2022). A média móvel exponencial será igual a:

$$EMA = (\text{CLOSE} * P) + (EMA(-1) * (1 - P))$$

Tal que:

CLOSE – preço de fechamento;

EMA (-1) – valor da média móvel do período anterior;

P – porcentagem de uso do valor do preço.

2.9.4 Convergência e divergência de médias móveis (MACD)

Convergência e Divergência de Médias Móveis (MACD) é um indicador seguidor de tendências. Ele indica a correlação entre duas Médias Móveis de um preço. O MACD é a diferença entre as Médias Móveis Exponenciais (EMA) 26 e 12 períodos (Metatrader, 2022).

$$\text{MACD}(12, 26) = \text{EMA}(12) - \text{EMA}(26)$$

2.9.5 Índice de força relativa (RSI)

O Índice de Força Relativa (IFR) é um indicador de análise técnica utilizado na medição entre as forças vendedoras e as compradoras de um ativo durante um intervalo de tempo. O IFR pode ser considerado um oscilador, pois move-se dentro de um intervalo de valores entre 0 e 100 (“Metatrader”, 2022). A fórmula desse indicador é dada por:

$$\text{RSI}(N) = 100 - 100 / (1 + \text{AVG_GAIN} / \text{AVG_LOSS})$$

Tal que:

AVG_GAIN – soma da altura das barras de alta em N períodos / N

AVG_LOSS – soma da altura das barras de baixa em N períodos / N

2.10 Support vector machine (SVM)

SVM foi desenvolvido na teoria do aprendizado estatístico buscando regular a generalização e encontrar um ponto ótimo entre complexidade estrutural e risco empírico. SVM classifica pontos categorizando-os em uma de duas classes distintas. A principal ideia do SVM é construir um hiperplano cuja superfície de separação entre as classes é maximizada (Huang & Liu, 2019).

3. METODOLOGIA

O trabalho é dividido em 17 processos e agrupados em 2: (1) Aquisição e tratamento de dados (ilustrado na figura 5) e (2) Aprendizado e Classificação (ilustrado na figura 6).

Essa separação favoreceu o processo de desenvolvimento. A primeira etapa permite uma validação mais detalhada dos dados obtidos e processados, podendo ser averiguados em plataformas de negociação ou planilhas de controle. A segunda etapa é exploratória e os resultados são validados com medidas estatísticas.

Figura 5 – Fluxo de processos de aquisição e tratamento de dados

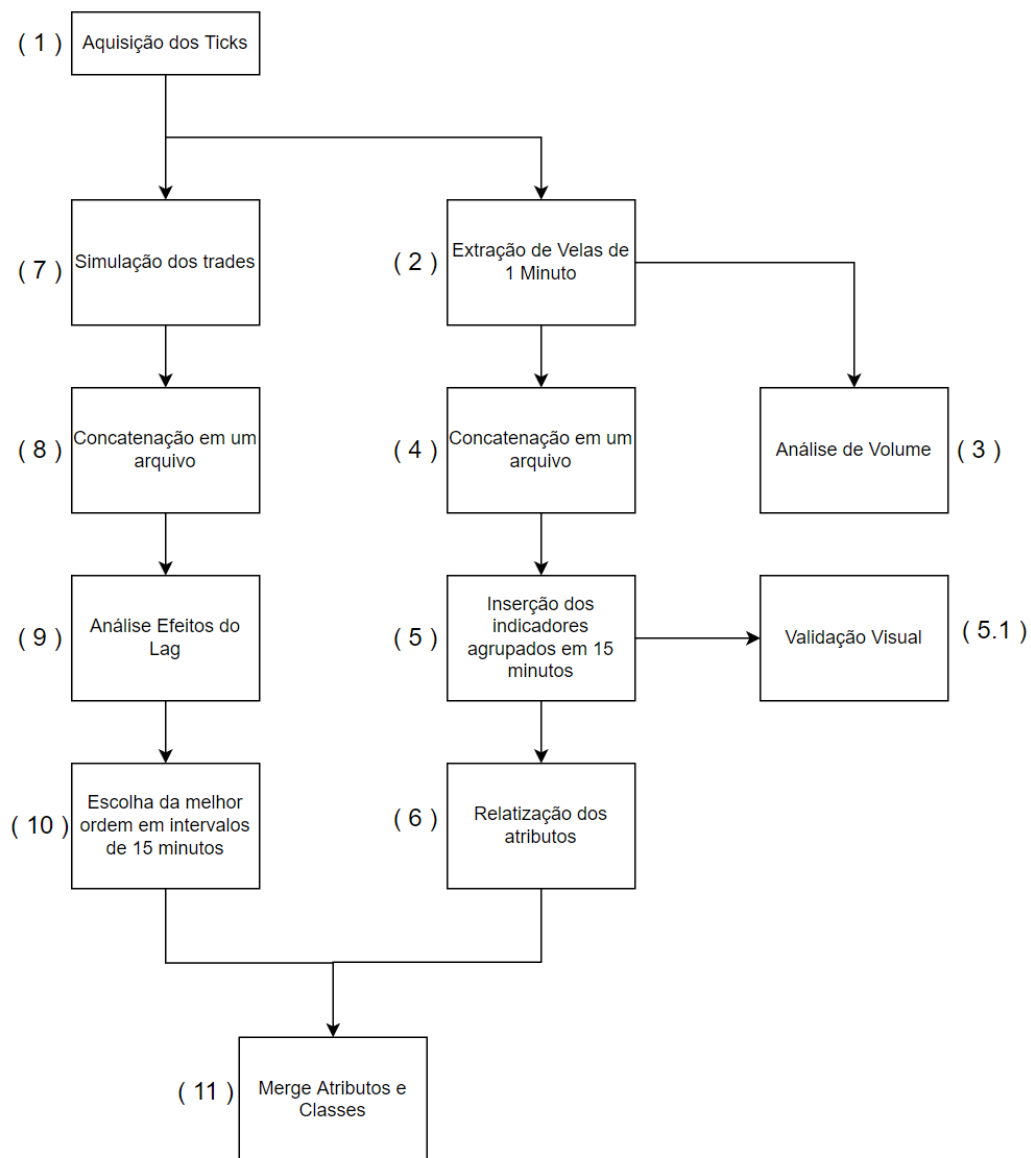
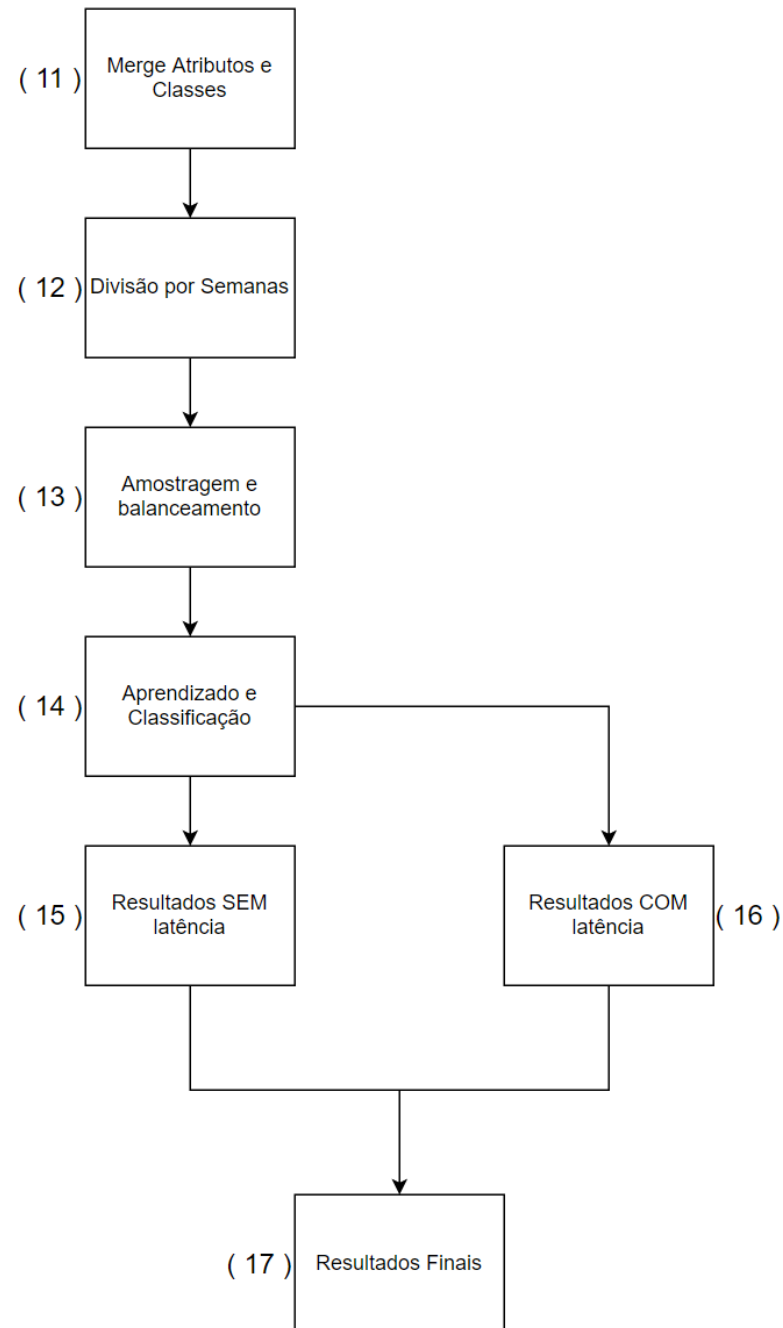


Figura 6 - Fluxo de processos de aprendizado e classificação



3.1 Aquisição das negociações

Este processo foca na obtenção dos dados de negociações da B3. Os dados foram obtidos através de API python do Metatrader 5 ("Metatrader", 2022). Através das corretoras ModalMais® e XP Investimentos®. Foram agrupados por dia sendo salvos no disco em

arquivos CSV. As negociações extraídas foram da série contínua do mini índice da B3 (símbolo WIN@N) sem correção monetária, ou seja, exatamente os valores negociados.

Ticks que não geraram negociações foram removidos, como exemplo, os que ocorreram durante os leilões e podemos afirmar neste e nos próximos passos que cada *tick* resulta em uma negociação.

O arquivo foi organizado de forma que cada linha representa uma ordem de compra ou venda que teve execução. Para este trabalho, é adotado o período de aquisição de 01/08/2018 a 30/04/2022, sendo a informação completa de negociações disponível para o ativo de mini índice.

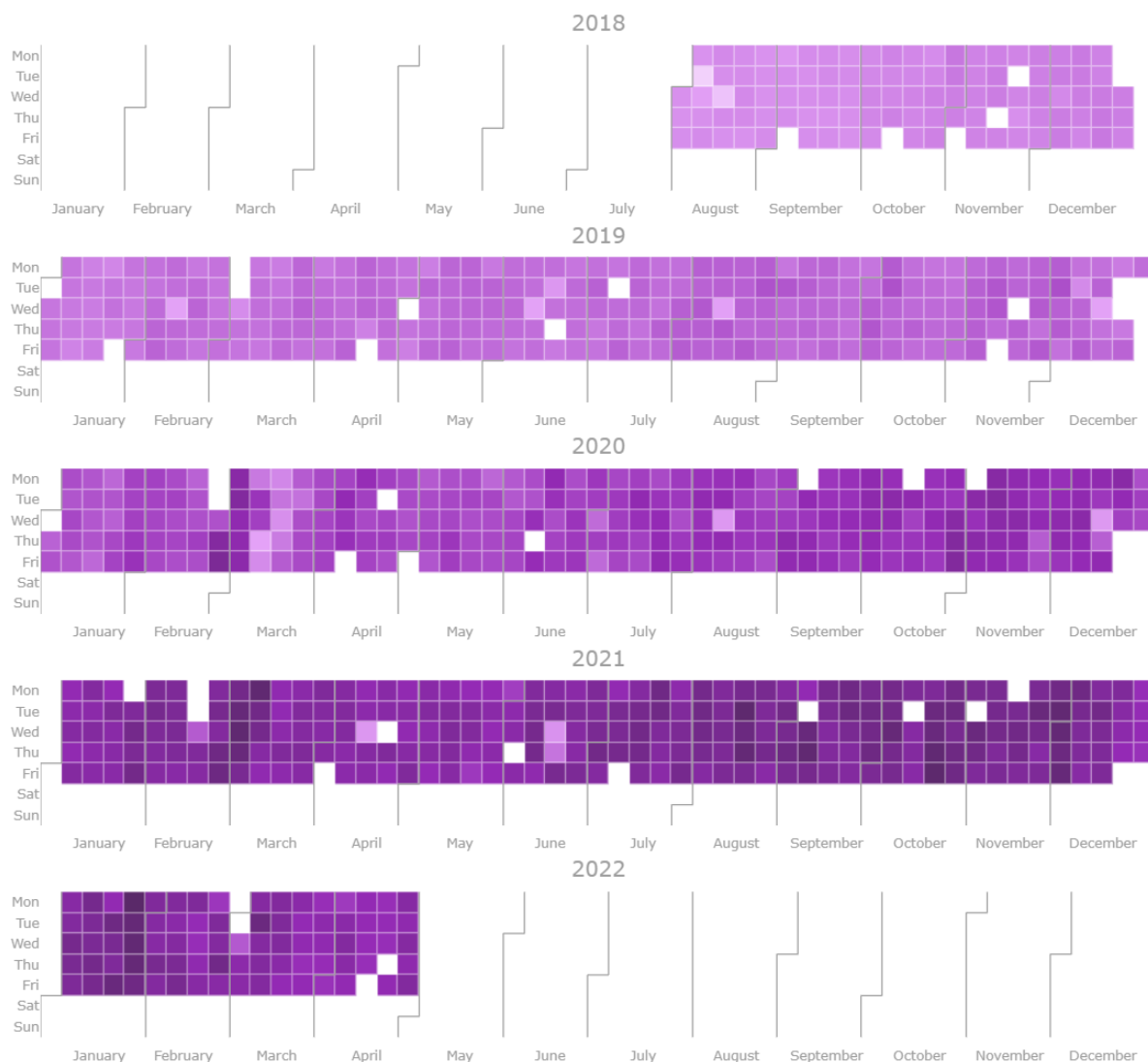
3.2 Extração de velas de 1 minuto

Nesse processo, cada arquivo extraído é processado e velas de 1 minuto sumarizam a negociações no período, ou seja, agrupa-se cada negociação (*tick*) no minuto que ele pertence. Como resultado, esse processo produz arquivos por dia em que cada linha é uma vela de 1 minuto.

3.3 Análise de volume

Esse processo poderia ter sido realizado com as negociações, não haveria diferença de resultados, mas arquivos de negociações tem ao mínimo 80 MB, enquanto arquivos de velas em média 34 KB. Essa etapa visa avaliar e eliminar problemas na aquisição de dados analisando o volume de negociação. Alguns dos dias da análise não tem todas as negociações possivelmente por falta de informações na corretora ou problemas na API python do Metatrader5. A figura 7 mostra um calendário de calor onde cores escuras representam maiores volumes de negociação e permite uma análise visual sobre os dias de operações.

Figura 7 - Calendário de calor com volumes diários

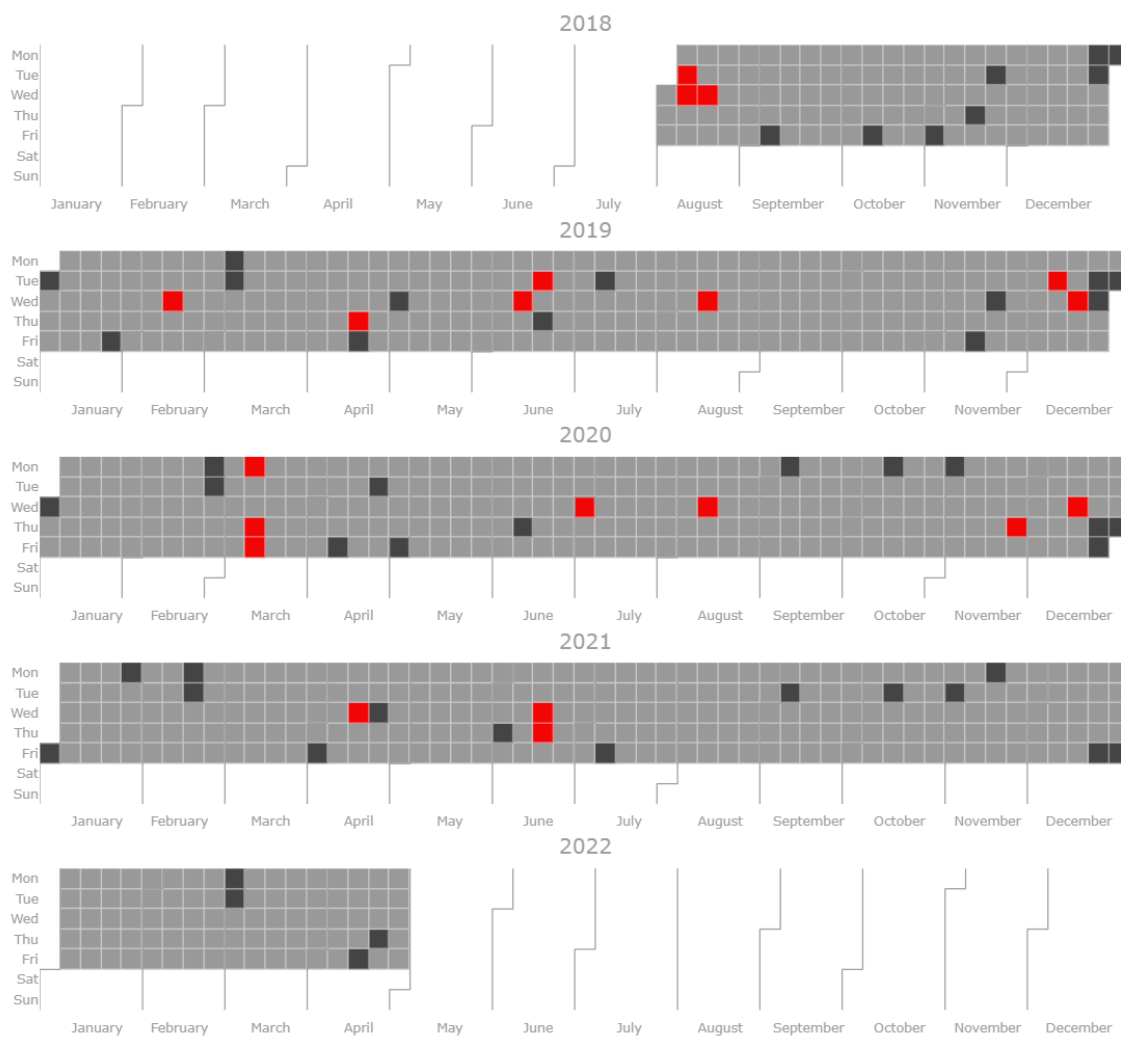


Fonte: Elaborada pelo Autor

Esse processo gerou uma heurística que analisa o volume de operações de cada dia comparando com a média dos últimos 5 dias e validando datas de feriados e meio-expedientes através do site da B3 ("B3", [S.d.]):

- Feriados são marcados para remoção
- Dias de meio expediente são marcados como normais
- Dias de expediente inteiros cujo volume é menor de 30 % são marcados para remoção

Figura 8 - Calendário de dias disponíveis



Fonte: Elaborada pelo Autor

Na Figura 8, dias em vermelho são dias com volumes abaixo da média. Dias em preto são feriados, em cinza normais e em branco, finais de semana. Os dias com volume abaixo da média não são removidos nessa etapa, apenas marcados para remoção. A tabela 1 contabiliza a remoção de dias do período aquisição inicial.

Tabela 1 - Remoção de dias de baixo volume

Total de Dias do Estudo	=	1369
Finais de Semana		-391
Feriados		-52
Total de arquivos obtidos	=	926
Volume 30% abaixo da média		-20
Dias disponíveis para análise	=	906

A heurística de 30% não foi criada em apenas uma etapa, mas em vários ciclos de desenvolvimento do projeto todo. Pequenos detalhes em cada processo evidenciaram erros na aquisição de dados que eram marcados para remoção manualmente e após vários ciclos, foi criada a heurística.

A marcação para remoção ao invés da remoção direta permitiu testes contínuos e paralelizar desenvolvimento do código com aquisição de dados. Ela consiste em um arquivo com todos os dias marcados como disponíveis ou não.

3.4 Concatenação dos arquivos de atributos

Todos os arquivos, inclusive com baixo volume, são lidos e ordenados por data, formando um único arquivo sendo cada linha composta por: time, O, H, L, C, volume. O tempo é com intervalos de 1 minuto provenientes da etapa de extração de velas. O arquivo resultando com aproximadamente 500k linhas. As técnicas posteriores consideram que o equipamento de processamento contém memória RAM e disco suficientes, mas, como trabalho futuro, pode-se adotar técnicas de leituras e escritas parciais para ampliar a possibilidade de uso da metodologia desenvolvida neste trabalho.

3.5 Inserção dos indicadores agrupados em 15 Minutos

O arquivo gerado na etapa 3.4 foi agrupada em intervalos de 15 minutos. O valor de fechamento (*close*) foi usado no cálculo dos indicadores analisados na etapa 2.10. Ao final do processo, é salvo um arquivo CSV como todas as colunas descritas na tabela 2

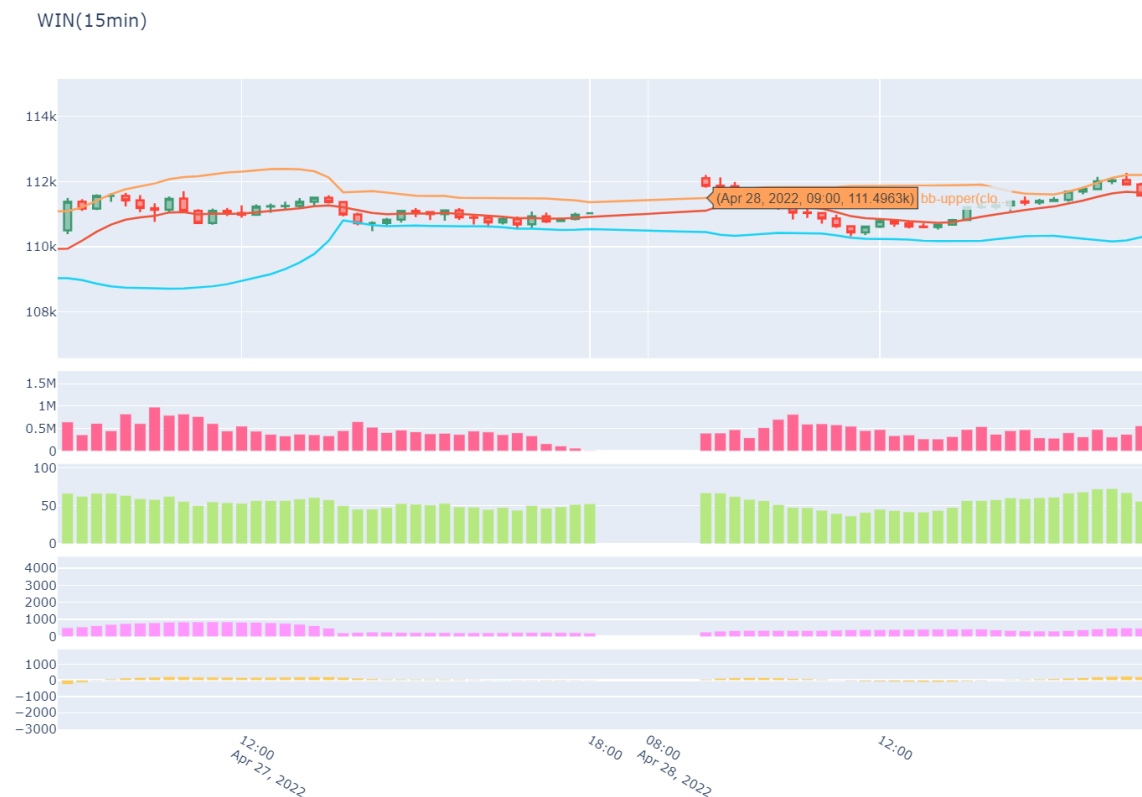
Tabela 2 - Características

Características	Descrição
<i>time</i>	Início do intervalo de 15 minutos
<i>open</i>	Primeira negociação
<i>high</i>	Maior negociação
<i>low</i>	Menor negociação
<i>close</i>	Última negociação
<i>volume</i>	Total volume negociado
<i>ema(close,9)</i>	Média móvel exponencial de 9 períodos no <i>close</i>
<i>rsi(close,13)</i>	Índice de força relativa de 13 períodos no <i>close</i>
<i>sma(close,20)</i>	Média móvel simples de 20 períodos no <i>close</i>
<i>day_index</i>	Índice da vela no dia
<i>MACD(close,12,26)</i>	Distância entre médias 12 e 26 no <i>close</i>

Validação visual

Foi construído um gráfico para validação visual de toda série temporal com os indicadores. Através dessa validação pode-se averiguar a corretude dos indicadores. A Figura 9 é um exemplo das velas e indicadores nos dias 27/04/2022 e 28/04/2022.

Figura 9 - Validação visual dos indicadores



A visualização busca por erros nos artefatos gerados, não sendo usada em qualquer outro processo do projeto.

3.6 Relativização dos atributos

Os valores calculados na etapa anterior são absolutos dificultando o reconhecimento de padrões. Para mitigar esse problema, os dados são relativizados utilizando as distâncias entre cada atributo e o ponto anterior na série temporal da seguinte forma:

A Tabela 3 mostra o conjunto de 12 características de entrada para o classificador. Todas essas características são calculadas em cada intervalo de 15 minutos.

Tabela 3 - Relativização dos Características

Características	Fórmula
open[0]	df.open.shift(0) - df.close.shift(1)
high[0]	df.high.shift(0) - df.close.shift(1)
low[0]	df.low.shift(0) - df.close.shift(1)
close[0]	df.close.shift(0) - df.close.shift(1)
close[1]	df.close.shift(1) - df.close.shift(2)
close[2]	df.close.shift(2) - df.close.shift(3)
volume[0]	df.volume.shift(0) - df.volume.shift(1)
ema(close,9)[0]	df.ema_close_9_.shift(0) - df.close.shift(1)
sma(close,9)[0]	df.sma_close_21_.shift(0) - df.close.shift(1)
rsi(close,13)[0]	df.rsi_close_13_.shift(0)
MACD(close,12,26)[0]	df.MACD_close_12_26_.shift(0)
day_index(close,20)[0]	df.day_index.shift(0)

3.7 Simulação de trades

O trade é composto por 1 ordem de entrada e 2 possíveis ordens de saída. Se a ordem de entrada é de compra, ambas ordens de saída serão de venda, porém, uma será 250 pontos acima do valor da entrada do tipo limite e uma será 250 pontos abaixo do tipo *stop*. Ambas são verificadas e a ordem que foi executada primeiro será a ordem considerada como fechamento para a abertura e a outra será ignorada. Assim, podemos ter expectativa de ganho de 250 pontos (“gain”) ou perda de 250 pontos (“loss”). Para ordens de vendas, é o mesmo procedimento, porém as ordens de saídas são compras.

As ordens de saída com ganho são sempre do tipo limite e as ordens de saída com perda são do tipo *stop*, mas as ordens de entrada podem ser de dois tipos: limite ou mercado.

Todas as ordens calculadas sofrem de alguma forma com a latência. Ordens do tipo limite podem não ser executadas devido ao atraso. Ordens do tipo mercado, podem ser executadas em preços distintos dependendo do atraso. Pela própria natureza da ordem tipo *stop*, essa sofre o efeito 2 vezes, a primeira para gatilho como a ordem limite e segunda para execução como ordem mercado.

O processo de simulação dos efeitos da latência foi feito para diversos tipos de entradas (compra/venda e limite/mercado) e latências de 0, 250, 500, 1000, 1500, 2000 e 3000ms foram aplicados a todas as ordens, tanto de entrada como saídas. A Figura 10 ilustra o código com essa variedade de análise.

Figura 10 - Código para criação das ordens em a cada 1 minuto

```
for lag in [0, 250, 500, 1000, 1500, 2000, 3000]:
    for direction in ["buy", "sell"]:
        for mode in ["limit", "market"]:
            order_sim.create_order(
```

Foram testados 7 cenários com latências diferentes para cada tipo de ordem. Como resultado há 28 ordens de entrada e 56 ordens de saída para cada intervalo de tempo.

A simulação consiste em verificar:

- Se ordens do tipo limite são executadas ou não. A latência pode causar a “perda” da negociação que executaria a ordem, deixando-a aberta.
- Quais são valores que ordens do tipo mercado são executados usando o próximo valor de *ask* para compra e próximo valor de *bid* para vendas considerando a latência.
- O tempo exato de cada execução.

Para cada ordem de entrada, temos as seguintes possibilidades:

1. Entrada do tipo limite ficam em aberto. Ou seja, não teve entrada. Status “sem entrada” (*no_entry*).
2. Entrada (limite ou mercado) é executada, mas as saídas ficam em aberto. Status “aberto” (*open*).
3. Entrada é executada e apenas uma ordem de saída é executada definindo se é de ganho ou perda. Status ganhou (*gain*) ou perdeu (*loss*).
4. Entrada é executada e ambas saídas são executadas, é definido se é ganho ou perda dependendo qual foi executada primeiro. Status ganhou (*gain*) ou perdeu (*loss*)

Como essa etapa foi realizada em paralelo com a etapa de indicadores, não se sabia ainda, qual intervalo de tempo seria utilizado, assim, decidiu-se testar para cada 1 minuto de intervalo. Em outras palavras, são 28 cenários aplicados a cada 1 minuto.

Ao final dessa etapa, temos para cada 1 minuto da série histórica os resultados das 28 opções [7 latências] [compra/venda] [entrada mercado/entrada limite]. O resultado consiste em:

- Status [*no_entry*, *open*, *gain*, *loss*],
- Resultado financeiro em pontos positivo para *gain*, negativo para *loss* ou 0 para *no_entry* e *open*.

- Horário do fechamento da entrada em caso de *gain* ou *loss*

3.8 Concatenação de arquivos de simulação de trades

Semelhante a etapa 3.4, porém para os arquivos de simulação de negociações.

3.9 Análise dos efeitos da latência

Nessa etapa foi analisado como as classes ficaram divididas no banco de dados e quais os efeitos da latência sobre as negociações independentemente do classificador ou da estratégia de *trading*. Os resultados dessa etapa estão descritos na análise de resultados (capítulo 4).

3.10 Escolha da melhor ação em intervalos de 15 minutos

Para cada latência tem-se os resultados de para os 4 tipos de ordens e abstinência:

- Compra limite
- Compra mercado
- Venda limite
- Venda mercado
- Abster-se (que não precisa de simulação)

O primeiro critério usado foi lucro na ação escolhida e o segundo foi tempo de encerramento, quanto mais cedo, melhor. Na etapa 3.9, percebeu-se que esse critério de seleção das classes, gerou 5 classes desbalanceadas e, ao invés de tentar criar formas de balancear durante o treinamento, preferiu-se substituir 2 classes:

- Compra mercado => Compra limite
- Venda mercado => Venda limite

O classificador não tomará as melhores decisões com essas trocas que representam menos de 5%, mas, espera-se que se acertar a direção, provavelmente, o resultado será positivo.

Fechamento de ordens em aberto

Para *day-trading*, não se pode manter as posições em aberto até o dia seguinte. Para isso, é obrigatório fechar em antes do final da sessão. O momento escolhido foi as 17:00. O status ‘em aberto’ é alterado para ‘fechado por tempo’ e o resultado depende do preço as 17:00. E a melhor ação é ‘abster-se’.

Além das ordens em abertos, todas ordens que encerraram depois das 17:00 são fechados as 17:00 usando o mesmo procedimento das ordens em aberto. No dia a dia, não teria como saber se essas ordens fechariam ou não, por isso o fechamento forçado.

A substituição de duas classes e o fechamento forçado é o segundo critério considerado para análises finais. Assim terminamos o processo com apenas 3 classes (compra limite, venda limite e abster-se) em intervalos de 15 minutos.

3.11 União de atributos e classes

É feita a união dos arquivos de simulação de trade e de indicadores. Ao final dessa etapa de união tem-se o conjunto final da base de dados. Atributos e classes em intervalos de 15 minutos. Cada classe é a melhor ação dependendo da latência.

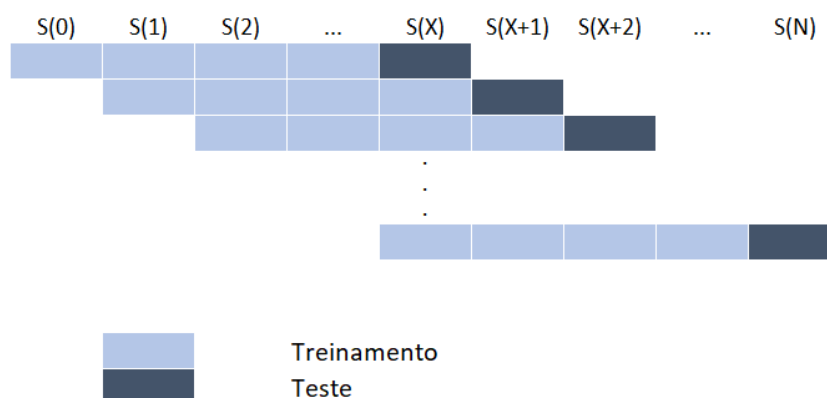
3.12 Divisão por semanas

O treinamento foi realizado separando os dias em semanas. A motivação é ter um usuário humano que treina um classificador com os dados durante o final de semana e deixa o classificador operar ao longo da semana sem outras interferências. Ao organizar os dados em semanas, temos 200 semanas para análise.

Janela de treinamento

Foi escolhido um período de (X) semanas para treinamento e 1 semana para teste, progredindo como uma janela de (X+1) semanas. Dessa forma, as primeiras X semanas não são testadas, apenas usadas para treinamento. A figura 11 mostra como a janela fixa se desloca com o ciclo de treinamento.

Figura 11 - Janela de treinamento



3.13 Amostragem e balanceamento

Para evitar uma escolha de datas e amostras que fossem enviesadas para o treinamento, 1/4 da amostra é aleatoriamente descartada durante o treinamento. Como várias rodadas de treinamentos são realizados com o mesmo banco de dados, esse procedimento tenta diminuir a dependência da qualidade do classificador ao período de testes que foi escolhido. A figura 12 mostra que a variável “*random_state*” permite ter amostras diferentes para um mesmo período.

Figura 12 - Descarte aleatório de amostras

```
# randomly discarding training data
discard_total = int(len( training_data)/4 )
to_drop = training_data.sample(n=discard_total, random_state=random_state)
training_data = training_data.drop(to_drop.index)
```

Balanceamento de classes

Dentro do intervalo de (X) semanas de treinamento, existe um desbalanceamento das classes. Períodos de alta tendem a ter mais classes de compra, por exemplo. A amostragem na etapa anterior, por ser aleatória, não promove o balanceamento e pode até invertê-lo. Foi feito um descarte das amostras da classe majoritária até que se permaneçam menos de 5% desbalanceadas. Nenhum balanceamento é realizado nas amostras de testes, mas pode-se ver pela Tabela 4 que as médias e desvio padrão são próximas entre compras e vendas nas amostras de testes.

Tabela 4 - Exemplo de estatísticas em rodada de treinamento

	week_index	training_buy	training_sell	testing_buy	testing_sell
count	182.000000	182.000000	182.000000	182.000000	182.000000
mean	106.934066	720.483516	724.472527	68.093407	70.912088
std	53.304360	69.199426	72.608439	18.812452	18.676749
min	15.000000	525.000000	528.000000	5.000000	16.000000
25%	61.250000	689.500000	693.250000	57.000000	60.000000
50%	106.500000	735.500000	746.000000	69.000000	74.500000
75%	152.750000	770.000000	775.000000	77.750000	85.000000
max	199.000000	842.000000	843.000000	133.000000	109.000000

Percebe-se que as amostras de compra e vendas do treinamento diferem menos de 1% na média e nos desvios padrões.

3.14 Aprendizado e classificação

Uma vez validado todo o sistema desde aquisição até balanceamento, foram testadas:

- as janelas de treinamento: [12, 16, 20, 24, 28, 32] semanas
- descartes e balanceamentos: 5 descartes aleatórios
- latências: [0, 250, 500, 1000, 1500, 2000, 3000] ms
- diferentes classificadores:
 - "Linear SVM": SVC(kernel="linear", C=0.025)
 - "RBF SVM": SVC(gamma=2, C=1),
 - "Decision Tree": DecisionTreeClassifier(max_depth=3),
 - "Random Forest": RandomForestClassifier(max_depth=5, n_estimators=10, max_features=1)
 - "Neural Net": MLPClassifier(alpha=1, max_iter=1000,)
 - "Naive Bayes": GaussianNB(),

Esse conjunto de testes teve dois objetivos:

- buscar a melhor janela de treinamento
- averiguar a que SVM tinha melhor desempenho como observado na revisão bibliográfica

O score calculado é a taxa de acerto entre compra limitada, venda limitada e abster-se.

4. RESULTADOS

Os resultados foram divididos em duas partes, primeira, análise dos artefatos produzidos na etapa de extração de dados e criação do banco de dados, segunda, análise dos modelos treinados.

4.1 Análise dos trades

O processo de ‘Análise dos Efeitos da Latência’ (3.9) foi realizado sobre o artefato gerado pelos processos ‘Simulação dos Trades’ (3.7) com intervalos de 1 minutos e ‘Concatenação dos Arquivos’ (3.8).

4.1.1 Decisões e pré-requisitos da simulação

Efeitos encontrados da latência na simulação de negociações partem de alguns pré-requisitos e algumas decisões de negócio.

Pré-requisitos:

- A latência é sempre a mesma. No ambiente real, a latência pode variar ao longo do dia.
- A latência é única. Não existem outras latências como tempo de processamento das informações ou pode-se considerar que outros atrasos estão inclusos dentro dessa latência simulada.

Decisões:

- Alvos de ganho de 250 pontos com ordens limite
- Alvos de perda de 250 pontos com ordens stop
- Operações começam na segunda barra de cada dia
- 5 Ações estão disponíveis em cada barra (compra mercado, compra limite, abster-se, venda limite e venda mercado).
- A ação é realizada no final da barra, uma vez que todas as negociações da barra foram processadas, agrupadas e os indicadores calculados.

4.1.2 Efeito da latência na taxa de acerto

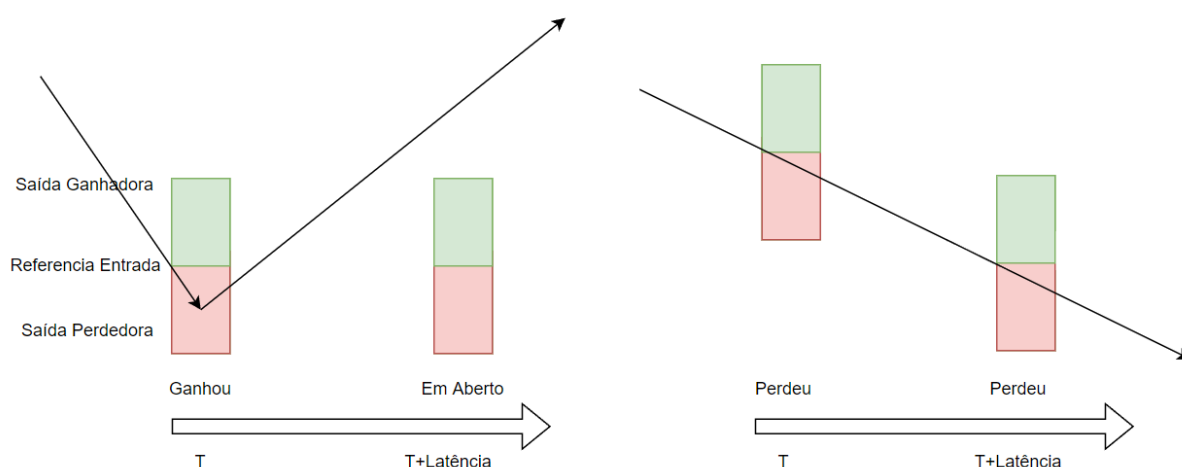
A Tabela 5 mostra que a taxa de ‘perdeu’ não se altera com a latência, a taxa de ‘ganhou’ diminui e taxa de ‘sem entrada’ aumenta. Verificou-se que o comportamento ocorre devido as ordens de entrada do tipo limite. À medida que aumentou a latência, aumento a quantidade dessas ordens não serem tomadas.

Tabela 5: Status por latência

Latência em ms	Ganhou	% Ganhou	Perdeu	% Perdeu	Sem Entrada	% Sem Entrada	Em Aberto	% Em Aberto	Total
0.0	825970	44.19	867300	46.40	4364	0.23	171406	9.17	1869040
250.0	824626	44.12	864434	46.25	5091	0.27	174889	9.36	1869040
500.0	823230	44.05	865229	46.29	5660	0.30	174921	9.36	1869040
1000.0	821405	43.95	866087	46.34	6583	0.35	174965	9.36	1869040
1500.0	820039	43.87	866775	46.38	7243	0.39	174983	9.36	1869040
2000.0	818919	43.81	867322	46.40	7777	0.42	175022	9.36	1869040
3000.0	817233	43.72	868155	46.45	8662	0.46	174990	9.36	1869040

Ordens a limite perdedoras precisam que o preço se desloque apenas em uma direção, enquanto ordens ganhadoras precisam ser tomadas e depois que o preço siga em direção oposta. Essa dinâmica está exemplificada na Figura 13, quando uma ordem compradora ganhadora fica em aberto se sofresse atraso e uma ordem compradora perdedora continua perdendo mesmo com preço de entrada inferior. O aumento da latência em mercados direcionais dificulta o ganho de ordens a limite.

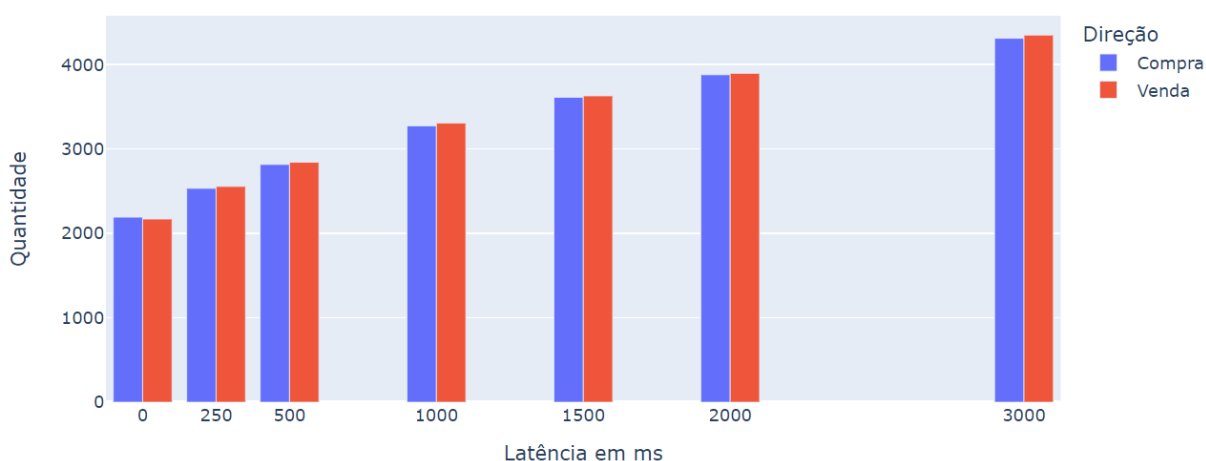
Figura 13 - Efeito latência ordens limite



4.1.3 Efeito da latência nas entradas tipo limite

Na Figura 14 pode-se averiguar o efeito acima filtrando apenas as ordens do tipo limite que ficaram sem entrada.

Figura 14 - Ordens limite sem entrada por latência



A taxa de ordens em aberto dobra quando é testado com 3000 ms de latência. Essa foi a pergunta base para esse estudo e pode-se perceber que a hipótese é correta, mas a ordem de grandeza encontrada foi menor que se esperava. Como mostrado anteriormente na Tabela 5, isso afeta menos que 0,5% a taxa total de acerto na maior latência analisada.

4.1.4 Ordens em aberto do tipo limite ao longo do dia

As ordens do tipo limite são distribuídas ao longo do dia como mostra a Figura 15. Uma forma de evitar o efeito da latência em ordens do tipo limite é abrir operações até as 16:00 de cada dia independente da latência como mostrado na Figura 16.

Figura 15 - Status de ordens tipo limite ao longo do dia

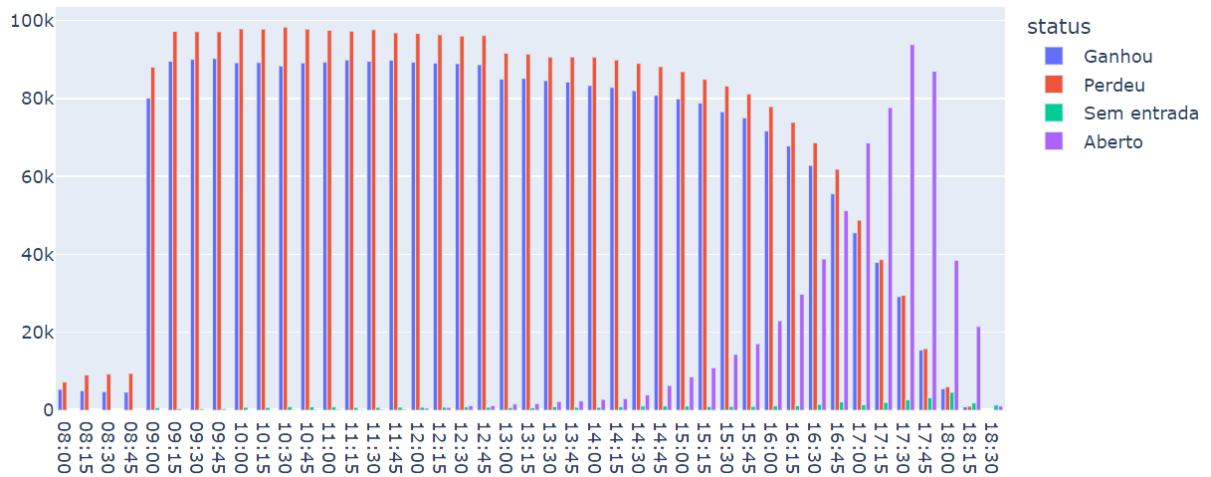
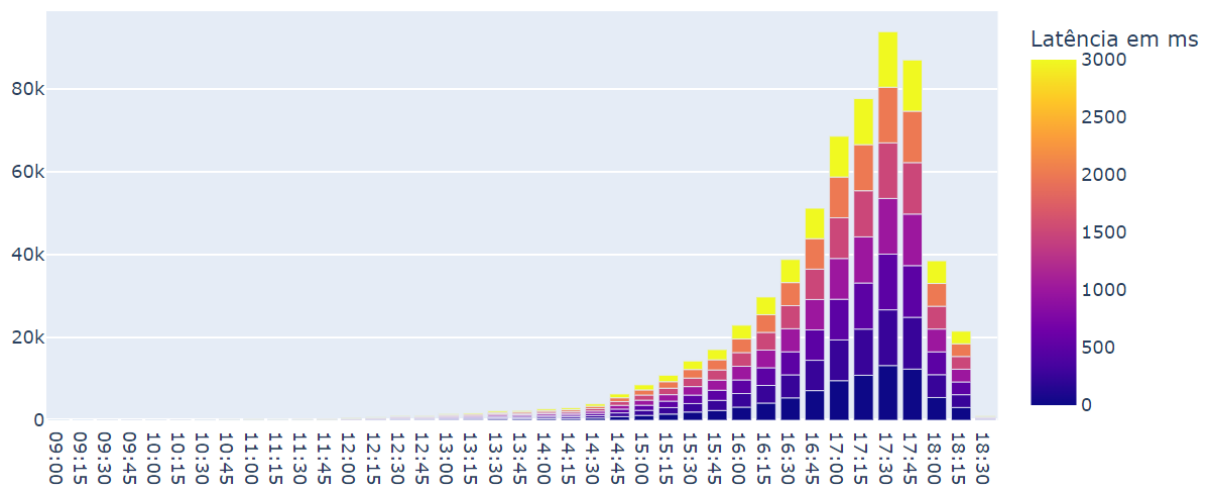


Figura 16 - Ordens limite em aberto por latência

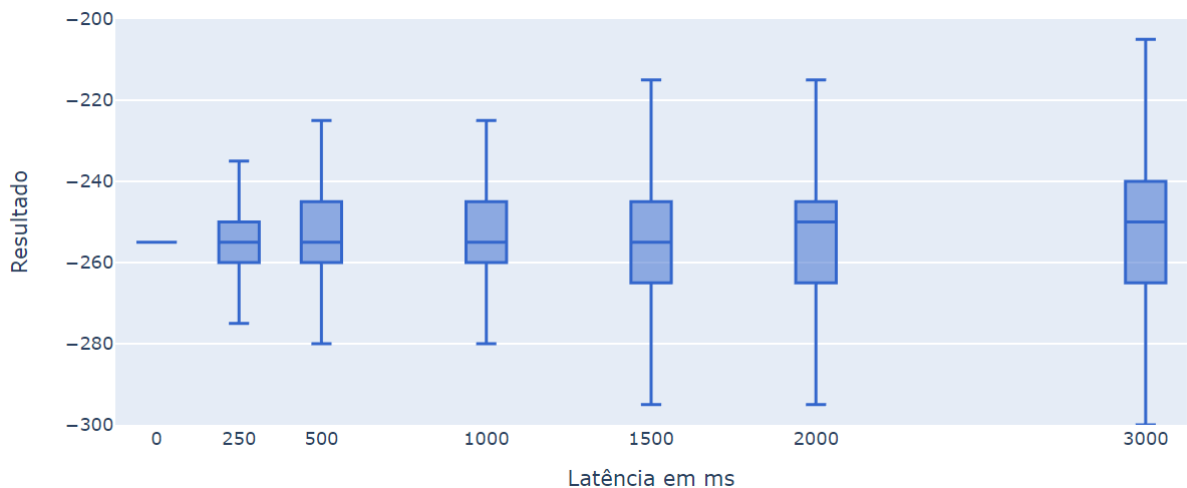


4.1.5 Efeito da latência nos prejuízos

O prejuízo das perdas é decorrente de uma ordem do tipo *stop*. Ordens do tipo *stop* dependem do preço no momento de sua execução. O resultado esperado é uma curva normal com média em 250 pontos. A intuição inicial era que o aumento da latência aumentaria o prejuízo em cada *stop*, porém, ao invés disso, os dados mostram uma diminuição no prejuízo

médio e um alargamento na curva normal analisando-se o desvio padrão. Com diferença menor que 1 % na média, não foi considerada necessidade de aprofundar o assunto. A Figura 17 mostra o ligeiro aumento das médias (diminuindo o prejuízo) e o aumento do desvio. *Outliers* foram cortados porque não havia diferença visual por serem valores discretos e semelhantes em todas as latências.

Figura 17 - Dispersão dos prejuízos por latência



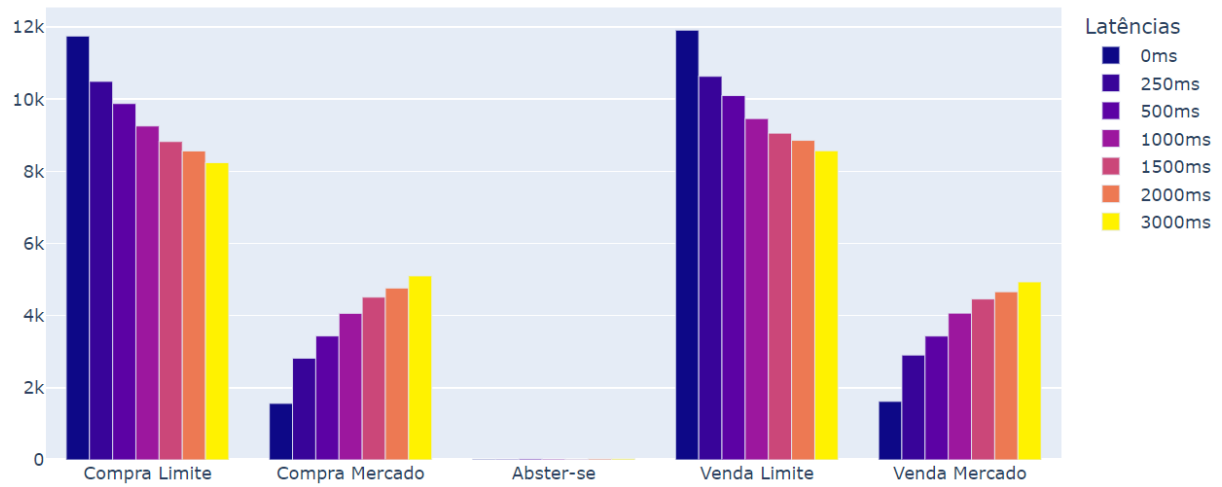
4.1.6 Fechamento de ordens em aberto

O resultado das posições que estavam em aberto e foram forçadas a fechar as 17:00, como descrito em 3.10, é uma curva normal com média -1 pontos e variância 85 pontos para todas as latências.

4.1.7 Melhores ações em cada momento

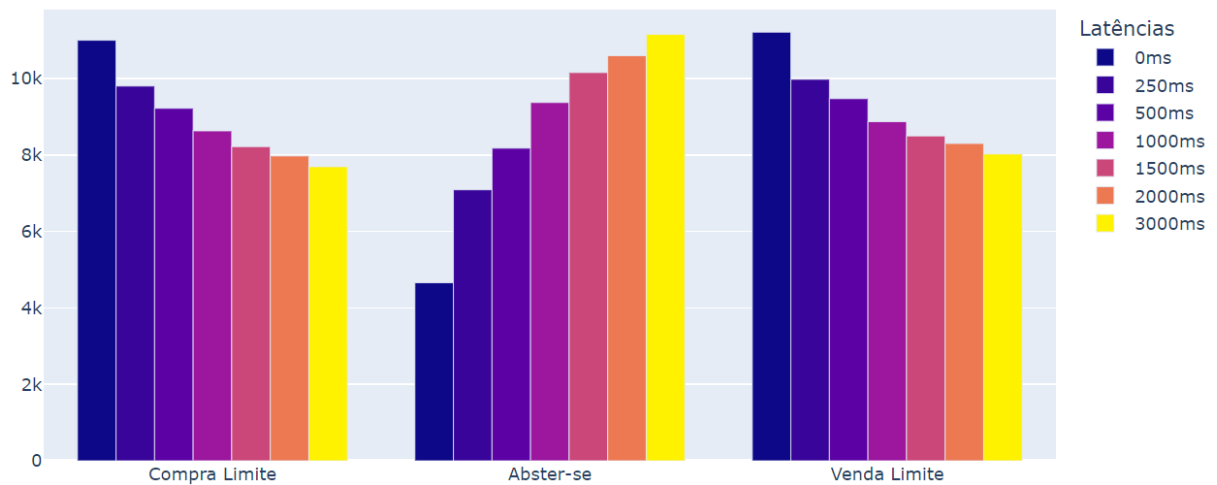
Ao final do processamento dos trades, temos a escolha da melhor ação a ser realizada em intervalos de 15 minutos. A melhor ação é a classe que será usada no classificador. No primeiro momento, demonstrado na Figura 18, obteve-se classes desbalanceadas para o treinamento.

Figura 18 - Classes desbalanceadas



O segundo critério, discutido em 3.10, gerou 3 classes mais balanceadas, compra limite, venda limite e abster-se mostrado na Figura 19.

Figura 19 - Classes balanceadas



À medida que aumenta a latências, mais ordens ficam em aberto depois das 17:00 e, portanto, usando o segundo critério, prefere-se abster-se dessa posição.

4.2 Resultados dos classificadores

Neste capítulo verifica-se como os classificadores desempenharam na separação das classes compra limite, venda limite e abster-se considerando as latências e janelas de treinamento. Todo processo foi executado 5 vezes descartando aleatoriamente 25% das amostras.

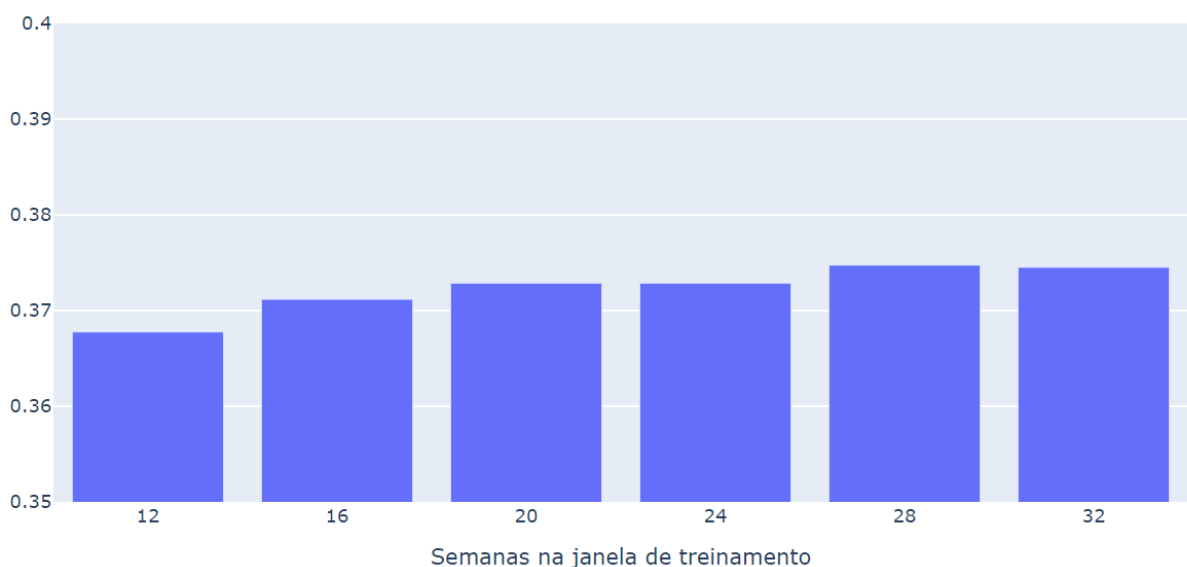
4.2.1 Entrada do classificador

O merge entre os indicadores e as classes pode ser observada no Anexo 1. Uma mesma entrada de indicadores pode ter classes diferentes em latências diferentes como mostrada na linha 5, onde a classe melhor ação para latência 0ms ('best_action_0') possui a classe compra limite ('buy_limit') e a classe de melhor ação para a latência 250 ms ('best_action_250') possui a classe venda limite ('sell_limit').

4.2.2 Resultados por janela de treinamento

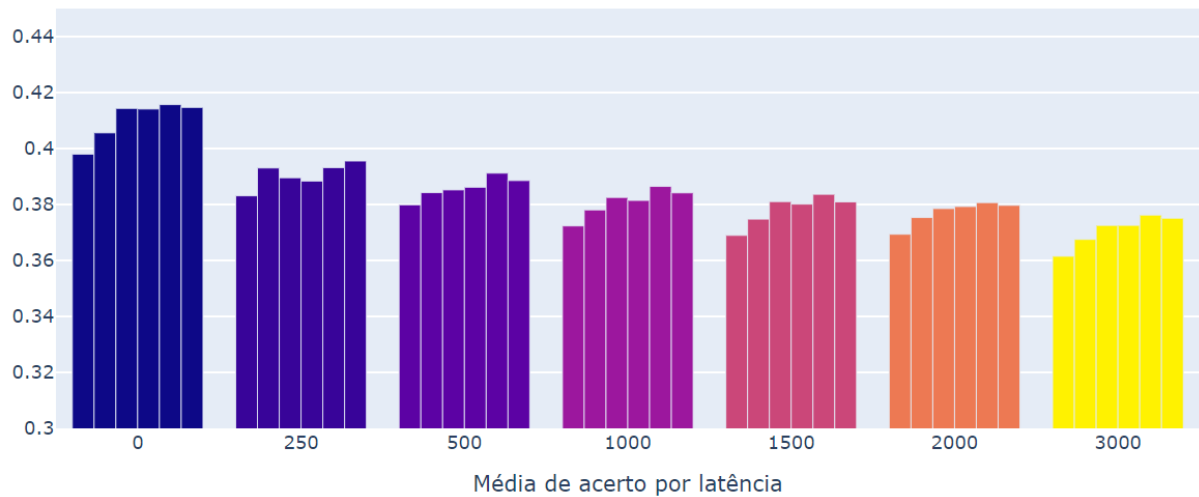
O efeito das janelas de treinamento pode ser observado na Figura 20, onde a taxa de acerto aumento quanto aumenta a janela, até 28 semanas e depois estabiliza. Escolheu-se 28 semanas para análises mais detalhadas.

Figura 20 - Média de acertos por tamanho de janela



Porém, se analisar as janelas de treinamento separadas pela latência, percebe-se que o efeito da latência é mais importante que a janela de treinamento.

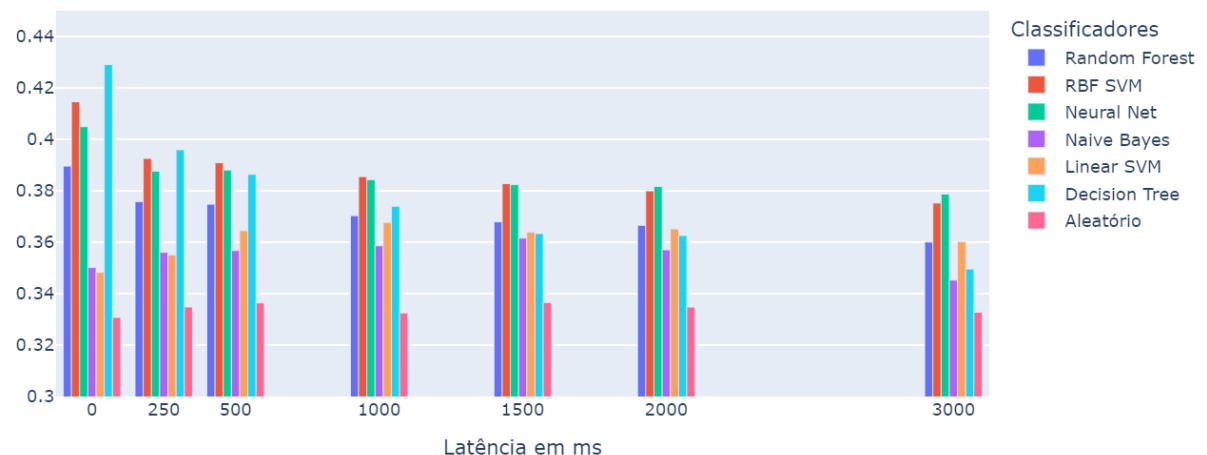
Figura 21 - Média de acertos por janela por latência



4.2.3 Resultados dos classificadores por latência

Ao escolher 28 semanas como janela de treinamento pode-se verificar na Figura 22 que a taxa de acertos para cada classificador testado. O classificador Aleatório (*DummyClassifier* do *Sklearn*) é uma boa referência para mostrar que as classes estavam balanceadas em 1/3 cada pois a taxa de acerto mantém-se em aproximadamente 33%.

Figura 22 - Taxa de acerto por classificador por latência (28 semanas)



4.2.4 Resultados somente RBF SVM

Na Figura 23, outros classificadores foram removidos e observa-se com mais detalhes que o treinamento do RBV SVM com 28 semanas tem uma queda na taxa de acertos à medida que a latência aumenta.

Figura 23 - Taxa de acertos SVM com janela de 28 semanas

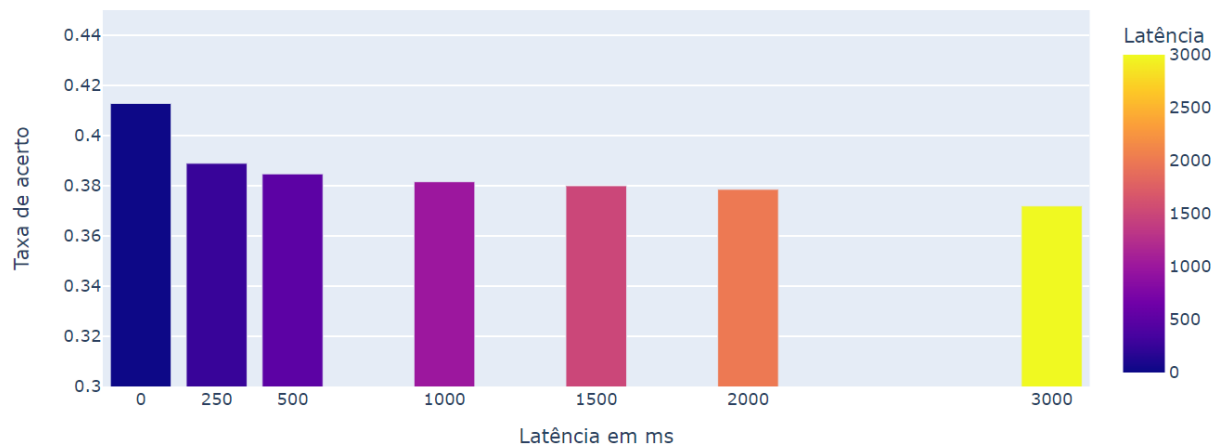
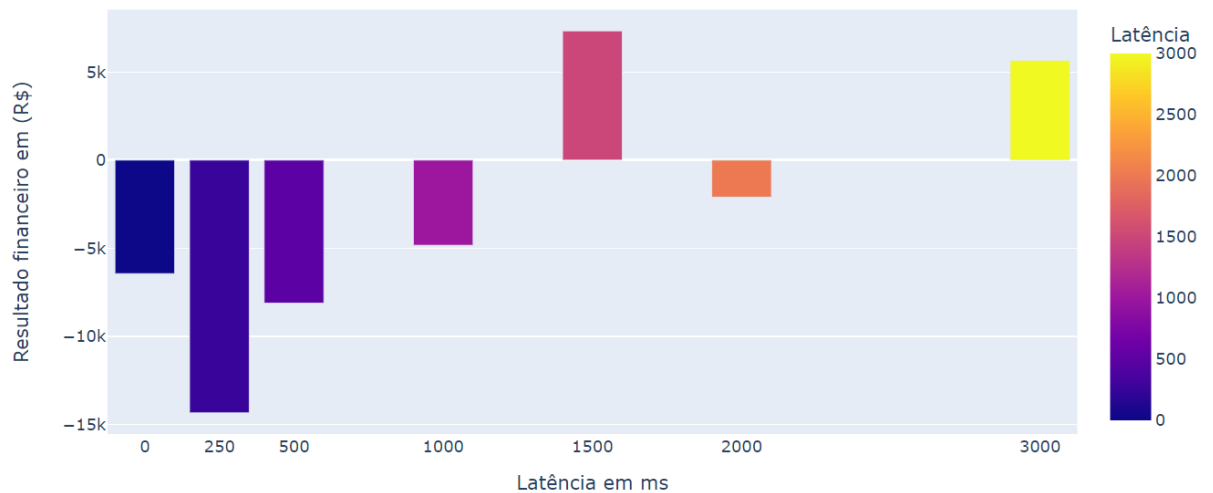
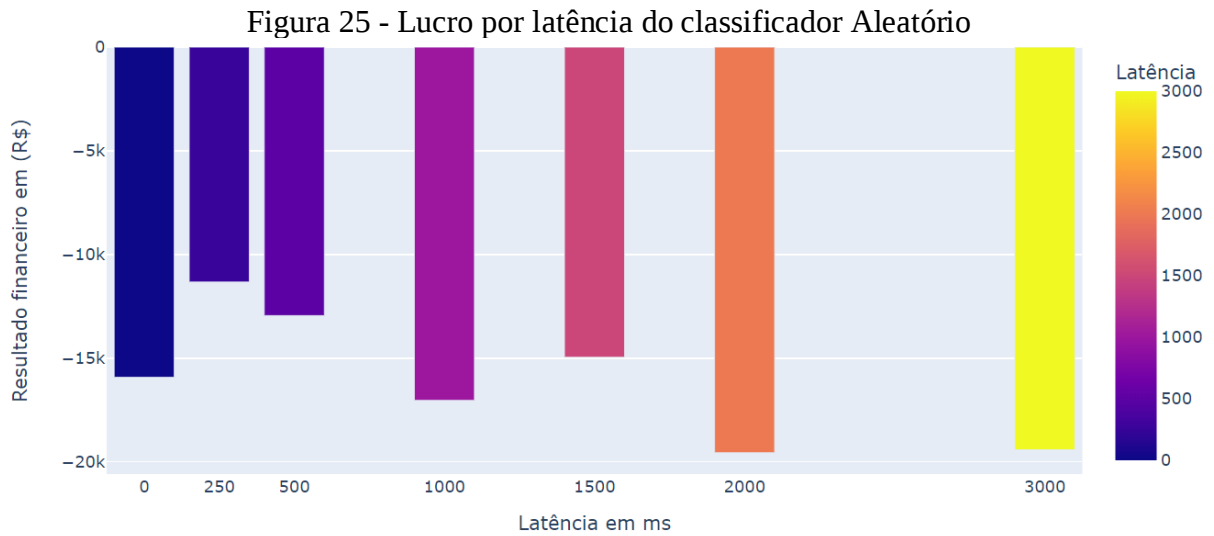


Figura 24 mostra um aumento no lucro líquido por latência para o SVM. Para comparar e validar o aprendizado, a Figura 25 mostra o resultado do classificador Aleatório para as mesmas condições.

Figura 24 - Lucro por latência SVM



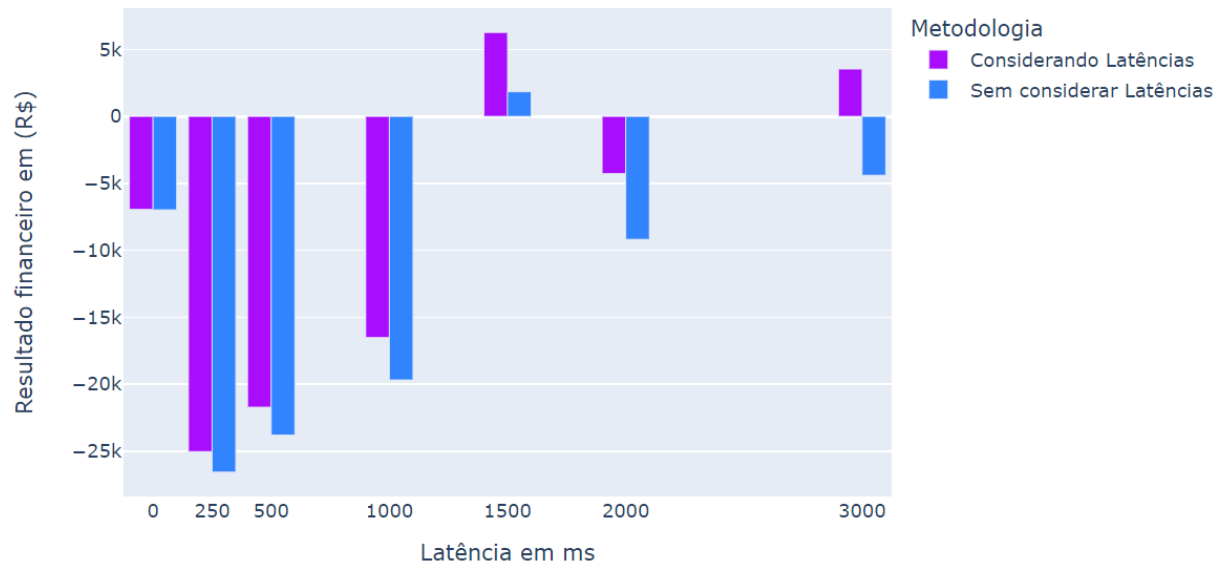


O classificador RBF SVM conseguiu utilizar melhor a classe abster-se. Quanto mais latência, mais abstenções, evitando entradas negativas. Não se deve atribuir a maior lucratividade ao aumento da latência em si, mas ao segundo critério usado para o balanceamento. Se o critério de fechamento fosse criado de forma diferente para cada latência, haveria resultados diferentes.

4.2.5 Avaliando Metodologia Proposta

Se o classificador RBF SVM tivesse sido treinado apenas com latência 0ms (sem atrasos) e fosse aplicado um mercado real (simulação considerando os atrasos), que possui latências, os resultados seriam diferentes dos demonstrados até o momento. Essa diferença entre aplicar um classificador treinado apenas com 0ms (sem considerar latências) e o classificador treinado exatamente para a latência para o qual será usado é demonstrada na Figura 26.

Figura 26 - Resultados metodologia com latência ou ingênua



Com exceção da latência 0 ms, que é o mesmo resultado, o classificador desempenhou melhor usando as latências no treinamento do que desconsiderando as latências durante o treinamento.

5. CONCLUSÕES E TRABALHOS FUTUROS

Esse trabalho apresentou diversos efeitos da latência em negociações da bolsa de valores:

Primeiro, ignorando as estratégias de escolha de compra/venda, foi observado que ocorre:

- um aumento nas ordens do tipo limite que ficam em aberto
- um aumento na dispersão dos prejuízos das ordens *stop*
- uma alteração na melhor ordem a ser feita em cada momento

Segundo efeito, os classificadores podem ser treinados considerando as melhores ordens para cada latência. Observados os efeitos que as latências causam em cada ordem, foi possível criar critérios para escolher a melhor ordem para aquela latência ressaltando a importância das análises dos dados antes da escolha do classificador.

Terceiro, ignorar as latências e aplicar um classificador ingênuo (sem considerar as latências) no mercado pode levar a resultados piores do que esperado.

Diversas possibilidades ficam em aberto para trabalhos futuros. Ao longo do desenvolvimento do projeto, muitas variáveis foram fixadas para avaliar somente o efeito da latência, como exemplo, o tamanho dos alvos de ganho e de perda e outros preços de entrada para ordens limite. Algumas foram brevemente avaliadas, mas a avaliação completa aumentaria o escopo para a restrição de tempo do projeto. Como trabalho futuro, avaliar diferentes tamanhos nos alvos de ganho e perda de cada operação.

As características de entrada foram escolhidas com base na pesquisa inicial e ainda assim, usou-se apenas 12 das diversas possibilidades presente na literatura. Diferentes estratégias podem ser extraídas usando-se indicadores, como exemplo, bandas de Bollinger poderiam ser analisadas para uma estratégia de retorno a média.

No processo de treinamento, apenas configurações iniciais (*default*) foram testadas, não foi feita uma busca por melhores parâmetros. Outros classificadores tiveram desempenhos similares ao SVM, como redes neurais e árvores de decisão e seus resultados financeiros não foram explorados.

A melhor janela de treinamento encontrada, 28 semanas, foi pior que janelas menores (4, 8 e 12) em alguns várias situações que algumas características de entrada do classificador foram removidas. As janelas menores aparentemente se adaptaram mais rápido com menos

características. Uma pesquisa para usar janelas dinâmicas poderia trazer um parâmetro de qual a melhor em cada momento.

Não foram analisadas estatísticas do mundo financeiro, como retorno por investimentos, capital inicial, perda máximas que são fundamentais para todo investidor. Ao buscar uma estratégia de investimento, todas essas variáveis precisam ser avaliadas e a latência precisa estar entre elas. Uma variável importante para alavancagens é o *drawdown* (maior prejuízo consecutivo). Nesse projeto não fizemos análise de indicadores financeiros além do lucro.

REFERÊNCIAS

- AGUIRRE, Alberto Antonio Agudelo e MEDINA, Ricardo Alfredo Rojas e MÉNDEZ, Néstor Darío Duque. **“Machine learning applied in the stock market through the Moving Average Convergence Divergence (MACD) indicator”**. Investment Management and Financial Innovations, v. 17, n. 4, p. 44–60, 2020.
- B3. Disponível em: <https://www.b3.com.br/pt_br/>. Acesso em: 13 ago 2022.
- CHAGUE, Fernando e DE-LOSSO, Rodrigo e GIOVANNETTI, Bruno. **Day trading for a living? ***. [S.l: s.n.], 2020. Disponível em: <<https://www.investopedia.com/articles/trading/09/how-to-trade-for-a-living.asp>>.
- DALLAQUA, MARCEL FERNANDES. **Sistema de Apoio à Decisão Inteligente para Day Trading: contornando a Racionalidade Limitada e tomando boas decisões consistentemente com o uso da Inteligência Artificial**. [S.l: s.n.], 2020.
- HUANG, Chin-Sheng e LIU, Yi-Sheng. **International Journal of Economics and Financial Issues Machine Learning on Stock Price Movement Forecast: The Sample of the Taiwan Stock Exchange**. International Journal of Economics and Financial Issues |, v. 9, n. 2, p. 189–201, 2019.
- ISMAIL, Mohd Sabri e colab. **Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange**. Applied Soft Computing Journal, v. 93, 1 Ago 2020.
- Metatrader. Disponível em: <www.metatrader.com>. Acesso em: 16 ago 2022.
- MOURA, Felipe Moreira. **Previsão de tendência de curto prazo do IBOVESPA usando dados históricos e aprendizado profundo**. 2021. 2021.
- ROUF, Nusrat e colab. **Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions**. Electronics (Switzerland). [S.l.]: MDPI. , 1 Nov 2021
- SAMPAIO, Yuri e GUEVARA, Cira. **Hawkes Processes: A Modeling of the Offer Book in the Brazilian Stock Market**. Selecciones Matemáticas, v. 4, n. 1, p. 38–50, 30 Jun 2017.
- SILVA, Michel Alexandre Da. **Livro de ofertas e dinâmica de preços: evidências a partir de dados da BOVESPA**. 2013. São Paulo, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/100/100132/tde-04122013-114154/>>.
- TEIXEIRA, Lamartine Almeida e DE OLIVEIRA, Adriano Lorena Inácio. **A method for automatic stock trading combining technical analysis and nearest neighbor classification**. Expert Systems with Applications, v. 37, n. 10, p. 6885–6890, 2010.
- VINICIUS, Marcos e colab. **LNCS 8056 - A Similarity-Based Approach for Financial Time Series Analysis and Forecasting**. [S.l: s.n.], 2013.

Anexo 1 - Exemplo de 5 entradas do classificador

		open[0]	high[0]	low[0]	close[0]	close[1]	close[2]	volume[0]	ema(close,9) [0]	sma(close,9) [0]	rsi(close,13) [0]	day_index [0]	MACD(close,12,26,9) [0]
1	2018-08-01 14:00:00-03:00	0.0	160.0	-55.0	75.0	105.0	-15.0	23.0	2.169312	-62.75	61.570299	20	19.838260
2	2018-08-01 14:15:00-03:00	0.0	60.0	-95.0	40.0	75.0	105.0	-20979.0	-50.054477	-114.25	63.123978	21	25.626378
3	2018-08-01 14:30:00-03:00	0.0	35.0	-105.0	-60.0	40.0	75.0	-766.0	-83.998900	-134.50	59.254173	22	25.777677
4	2018-08-01 14:45:00-03:00	0.0	60.0	-130.0	-80.0	-60.0	40.0	-1998.0	-35.265627	-59.25	54.460182	23	20.154380
5	2018-08-01 15:00:00-03:00	0.0	70.0	-50.0	55.0	-80.0	-60.0	-5512.0	46.797240	45.00	57.033904	24	19.330083

		best_action_0	best_action_250	best_action_500	best_action_1000	best_action_1500	best_action_2000	best_action_3000
1	2018-08-01 14:00:00-03:00	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit
2	2018-08-01 14:15:00-03:00	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit
3	2018-08-01 14:30:00-03:00	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit
4	2018-08-01 14:45:00-03:00	buy_limit	buy_limit	buy_limit	buy_limit	buy_limit	buy_limit	buy_limit
5	2018-08-01 15:00:00-03:00	buy_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit	sell_limit