
CADEIAS DE MARKOV EM ESTRATÉGIA DE NEGOCIAÇÃO DE AÇÕES EM ALTA FREQUÊNCIA

Cesar Almiñana

Leandro Augusto da Silva

Universidade Presbiteriana Mackenzie (UPM)

Resumo

O problema de predição do preço ou valor de uma determinada ação tem sido objeto de estudo, aprofundamento e evolução nas últimas décadas, buscando alcançar o objetivo básico de realizações financeiras positivas com o menor risco possível. Uma estratégia conhecida para redução de riscos é a operação em um regime de *high frequency trading*, que se beneficia de pequenas variações de preço e permite realizar ganhos múltiplas vezes ao dia. Partindo da hipótese e da interpretação de que a variação dos preços de ações equivale a um passeio aleatório (*random walks*), é possível caracterizar um processo (estocástico) de Markov com estados e probabilidades de transição conhecidos, descrevendo as variações de preço de uma ação específica ao longo do tempo. Desse modo, este estudo busca estimar o retorno financeiro potencial a partir do uso de cadeias de Markov para tomada de decisões automatizada em uma estratégia de compra e venda de ações em alta frequência, comparando os resultados obtidos com a valorização (ou desvalorização) do papel no mesmo período.

Palavras-chave: Ações. Random walks. Cadeias de Markov. *High frequency trading*.

1 INTRODUÇÃO

Prever com certo nível de precisão o preço ou o valor de um determinado ativo – como um índice, uma ação, uma moeda ou um fundo – tem sido objeto de estudo, aprofundamento e evolução nas últimas décadas, principalmente com o objetivo básico de realizações financeiras positivas e com o menor risco atrelado possível. Analistas técnicos (ou gráficos), por exemplo, dedicam-se a estudar e estabelecer projeções sobre o comportamento de uma ação com base em indicadores, desempenho observado, parâmetros de oferta e procura, evolução de cotações e volumes transacionados, buscando formas de obter estimativas de desempenho futuro. Paralelamente, autores dedicam-se a estudar técnicas e algoritmos de *machine learning* e regressão, principalmente atrelados a redes neurais artificiais (*artificial neural networks* – ANN) e profundas, para prever com exatidão o valor futuro que um determinado ativo pode adquirir.

Abordagens como as de Henrique, Sobreiro e Kimura (2018), Almasarweh e Wadi (2018), Chou e Nguyen (2018) e Sadorsky (2021) buscam realizar a predição dos valores exatos no curto prazo; já Ballini *et al.* (2010), Parmar *et al.* (2018), Pandey e Bajpai (2019), Yadav, Jha e Sharan (2020) e Vijh *et al.* (2020) agregam complexidade com os conceitos de ANN; Lachiheb e Gouider (2018), Nikou, Mansourfar e Bagherzadeh (2019) e Jiang (2020) utilizam redes neurais profundas ainda com o objetivo de prever valores futuros com maior grau de precisão, adaptabilidade dos modelos e refinamento; Nikou, Mansourfar e Bagherzadeh (2019), por exemplo, conseguiram reduzir consideravelmente métricas de erro médio absoluto, erro médio quadrático e a raiz do erro médio quadrático a partir da abordagem de *deep learning*.

Em outras palavras, conseguir prever com exatidão o preço de uma ação pode ser sinônimo de realizar operações de compra e venda nos momentos certos, permitindo maximizar lucros, reduzir riscos e otimizar estratégias de investimento.

Contudo, conforme Conegundes e Pereira (2020) observam, um sistema de operação só é completamente automatizado quando dois subsistemas se combinam: de previsão e de operação. Paradoxalmente, os objetivos de cada um desses sistemas diferem: um busca reduzir o erro de previsão; e o outro, maximizar o retorno acumulado. Mas, como as regras de operação não são normalmente incorporadas no processo de aprendizado e predição, diversas variáveis do mundo real acabam sendo desconsideradas, tais como liquidez, latência ou custos de transação.

Em contraponto ao problema de predição, este estudo aborda o desafio de operar ações gerando lucro sob a perspectiva do sistema operação, e não previsão conforme

comumente é realizado. Métodos mais ativos de operação (*high frequency trading* – HFT), por exemplo, beneficiam-se de pequenas variações de preço, realizando retornos múltiplas vezes ao dia. Em tese, essa estratégia permite redução de riscos, uma vez que realiza e acumula lucros em função de alta volumetria de operações, focando o modelo de operação, muito mais que o modelo de previsão.

Sendo assim, o objetivo geral deste trabalho é desenvolver, implementar, testar e avaliar um sistema automático de compra e venda de ações, dispensando a necessidade de realizar predições exatas dos valores de um ativo em um instante no tempo, mas levando em consideração apenas a direção para qual o mercado aponta. Ainda, esse algoritmo deverá ser capaz de tomar decisões automatizadas, instantâneas e realizadas em tempo real, de tal modo que possibilite e suporte o regime de HFT.

Os resultados obtidos a partir de um período de operações serão comparados com a variação percentual do ativo transacionado no mesmo intervalo de tempo. Desse modo, será possível avaliar a premissa básica de eliminação – parcial ou total – dos riscos envolvidos quando se negociam ações utilizando a estratégia de alta frequência.

A seção 2 apresenta a metodologia proposta e os procedimentos de execução do experimento, além de uma breve revisão da literatura sobre o tema; a seção 3 aborda todo o referencial teórico, descrevendo as especificidades do mercado de compra e vendas de ações no Brasil e os conceitos de estatística sobre variáveis aleatórias, processos estocásticos, *random walks* e cadeias de Markov; a seção 4 contém, detalha e discute todos os resultados experimentais observados; e por fim, na seção 5, há as conclusões deste trabalho, bem como observações para trabalhos futuros e considerações finais.

2 METODOLOGIA

Toda vez que o mercado opera na direção contrária à tendência prevalecente e então retorna a ela, um ciclo se encerra, e um novo se inicia. Esse padrão, que ocorre constantemente em todo e qualquer mercado financeiro, pode ser chamado de *swing*. Quando se operam nesse tipo de movimento, segundo Crane (2003), os chamados *swing trades*, é possível beneficiar-se de movimentos cíclicos dentro de uma tendência existente. Além desses ciclos, Abell (1999) cita o método mais ativo de operação, chamado de *scalping*, que busca beneficiar-se de pequenos movimentos ou microvariações de preço, de modo a buscar pequenos retornos múltiplas vezes ao dia.

Essas pequenas variações de preço podem ser descritas como consequência de sucessivos “impulsos” que ocorrem aleatoriamente ao longo do tempo. Rundo (2019) e Briola *et al.* (2021) introduzem o conceito de alta frequência de operações (HFT), aproximando a estratégia mais ativa de execução de operações, *scalping*, citada por Abell (1999).

Esses impulsos, sob o ponto de vista estatístico, podem ser caracterizados como um passeio aleatório (*random walk*), isto é, um processo estocástico que descreve uma sequência de variáveis independentes e aleatórias que assumem valores e se distanciam entre si por meio de intervalos regulares. Um processo de Markov (ou um processo markoviano) pode representar a generalização mais simples de um processo independente, isto é, um processo que, em qualquer instante do tempo, dependa apenas do resultado imediatamente anterior e de nenhum outro. Os trabalhos de Chitenderu, Mareza e Sibanda (2014) e Vigg e Arora (2018) sobre a aleatoriedade foram capazes de mostrar a aderência dos mercados indiano e sul-africano de ações, respectivamente, à hipótese de que os mercados se comportam como *random walks*.

Toda cadeia é descrita a partir de um número finito de estados de transição, que descrevem como um sistema se encontra em determinado instante de tempo, como ligado ou desligado, aberto ou fechado, alegre, apático, triste ou bravo, e ainda sob a perspectiva de ações, valorizando, neutro ou desvalorizando.

Baseando-se e inspirando-se na simplicidade técnica do trabalho de Huang (2018) e na abordagem com cadeias de Markov de Cruz e Gonzalez (2018), este estudo dispensa a necessidade de realizar predições exatas e foca a direção que o mercado está apontando em dado instante do tempo. Ainda, é possível usar os processos de Markov para a criação da camada lógica de operação, que se converte na tomada de decisões no mercado, uma vez que é possível calcular a probabilidade de transição de estados. Na prática, traduz-se em quão provável uma variação aleatória, positiva ou negativa, pode ocorrer, sabendo o estado atual em que tal sistema se encontra.

Um sistema completamente automatizado possui a vantagem de reduzir erros humanos ou atrasos. Para certos sistemas de alta frequência – tal como descrito anteriormente –, um sistema automatizado é indispensável, considerando que qualquer intervenção humana irá causar atraso suficiente, capaz de impactar a *performance* das operações (CHAN, 2009).

No desenvolvimento do sistema de operações, utilizou-se a linguagem de programação Python 3.8 que permite a modelagem apropriada do problema e a conexão direta com a plataforma de operação em tempo real MetaTrader 5, em ambiente *on-line*, de modo a observar os dados reais de qualquer ação negociada e trabalhar com eles. Este estudo utilizou dados que foram atualizados em tempo real, conforme a evolução do mercado nos dias de operação, além da extração de histórico dos últimos 12 meses para análises exploratórias, definição dos estados de transição da cadeia de Markov

e análise das probabilidades de transição entre estados (em detalhe na seção 4). Além disso, esses dados foram utilizados para o cálculo da variação percentual do ativo transacionado no intervalo de tempo (*benchmark*), a fim de comparação dos resultados obtidos pelo sistema automático, e a valorização (ou desvalorização) do ativo em questão.

Todos os testes foram realizados em uma máquina virtual com o sistema Windows Server, hospedada nos servidores do Google Cloud Platform (GCP), na região “southamerica-east1-b” (fisicamente em São Paulo, Brasil), com quatro núcleos de processamento e seis *gigabytes* (GB) de memória RAM. Realizaram-se essas escolhas por dois motivos: 1. o *software* MetaTrader5 pode ser executado apenas em Windows (e não em uma arquitetura ARM) e 2. Há redução da latência de recebimento e envio de informações.

3 FUNDAMENTAÇÃO TEÓRICA

Conforme observa Grami (2020, p. xiii, tradução nossa), “aleatoriedade e incerteza, que sempre caminham lado a lado, existem virtualmente em todo aspecto da vida”. Uma variável aleatória, conforme descrito por Papoulis e Pillai (2002), é um valor atribuído a todo resultado de um experimento. Pode ser o ganho em um jogo de azar, a voltagem de uma fonte aleatória de energia ou qualquer grandeza numérica de interesse quando se realiza um experimento.

Por sua vez, um processo aleatório é a extensão do conceito de variável aleatória, diferenciando-se por ser computado em função do tempo. Sendo assim, em qualquer instante de tempo escolhido para uma observação, o processo aleatório definirá uma variável aleatória (GRAMI, 2020). A noção de tempo agregada a variáveis aleatórias caracteriza um processo estocástico. *Random walks* (RW) são apenas um caso particular de processo, que descreve uma sequência de variáveis independentes e aleatórias que assumem valores que se distanciam entre si por meio de intervalos regulares (PAPOULIS; PILLAI, 2002). A variação no preço de uma ação em particular, uma vez que varia em consequência de sucessivos “impulsos” que ocorrem aleatoriamente ao longo do tempo, pode ser caracterizada como um RW (VISHWANATH; KRISHNAMURTI, 2009).

Ainda, um processo estocástico será a família de funções de tempo t dependendo do resultado ζ ou igualmente uma função de t e ζ . Assim, o domínio de ζ será o conjunto de todos os resultados experimentais, e o domínio de t é o conjunto de números reais \mathbb{R} não negativos (PAPOULIS; PILLAI, 2002).

Um processo de Markov representa a generalização mais simples de um processo independente, de modo que o resultado obtido, em qualquer instante do tempo, depende apenas do resultado imediatamente anterior e de nenhum outro. Sendo assim, em um processo de Markov, o passado não terá qualquer interferência no futuro se o presente for especificado (PAPOULIS; PILLAI, 2002).

Toda a cadeia de Markov pode ser descrita a partir de uma quantidade enumerável (ou, em alguns casos, infinita) de estados de transição. Um estado descreve o estado em que um sistema se encontra em determinado instante t do tempo. Portanto, a probabilidade de esse processo mover-se para o próximo estado dependerá apenas do estado presente, e não dos estados anteriores.

$$P[\mathbf{x}(t_n) = \mathbf{x}_n \mid \mathbf{x}(t_{n-1}), \dots, \mathbf{x}(t_1)] = P[\mathbf{x}(t_n) = \mathbf{x}_n \mid \mathbf{x}(t_{n-1})] \quad (1)$$

Para $n \geq m \geq 0$, a sequência $\mathbf{x}_{m+1} \rightarrow \mathbf{x}_m \rightarrow \dots \rightarrow \mathbf{x}_n, \dots$ representa a evolução do sistema. Assim, $p_i(m) = P[\mathbf{x}_m = e_i]$ representa a probabilidade de que, no instante $t = t_m$, o sistema ocupará o estado e_i , $p_{ij}(m, n) = P[\mathbf{x}_n = e_j \mid \mathbf{x}_m = e_i]$; além disso, há a probabilidade de que o sistema, no instante $t = t_n$, ocupe o estado e_j , dado que estava no estado e_i no instante $t = t_m$.

A partir da ideia de transição entre estados, é possível descrever a probabilidade $p_{ij}(m, n)$ na forma de uma matriz de probabilidades, conforme Equação 2.

$$P(m, n) = \begin{pmatrix} p_{11}(m, n) & p_{12}(m, n) & \cdots & p_{1j}(m, n) & \cdots \\ p_{21}(m, n) & p_{22}(m, n) & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{i1}(m, n) & \cdots & \cdots & p_{ij}(m, n) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \quad (2)$$

Vale ressaltar que a matriz $P(m, n)$ tem, nas linhas e colunas, apenas valores positivos, e a soma dos elementos de cada linha é igual a 1.

É possível representar uma cadeia de Markov por meio de um grafo, conforme ilustrado na Figura 1, e cada nó representa um estado, e, por sua vez, cada aresta representa a transição entre dois estados (nós). O peso atribuído a essa aresta representa a probabilidade de transição entre dois estados.

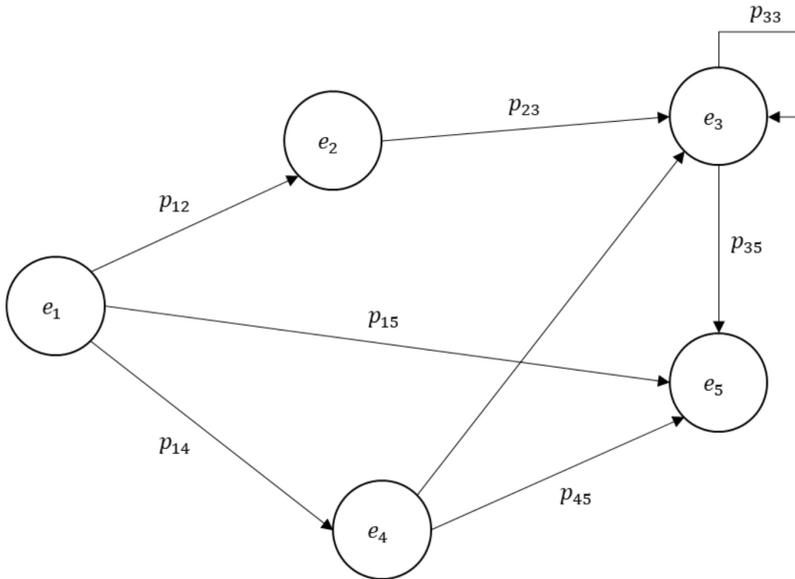


Figura 1 Representação de uma cadeia de Markov, com estados e probabilidades de transição

Fonte: Elaborada pelos autores.

Ainda sobre as aplicações da teoria da probabilidade, Pappoulis e Pillai (2002) ainda questionam sobre quais conclusões podem ser tiradas sobre a probabilidade P de um evento em uma próxima e única ocorrência. Ou seja, por mais que a probabilidade de ocorrência de determinado evento – para este estudo, o valor ou preço de um ativo – seja conhecida, ainda é necessário tomar decisões a fim de conquistar o objetivo básico e principal de realizações financeiras.

Um processo decisório de Markov (*Markov decision process* – MDP) pode ser considerado apenas extensões estocásticas de autômatos finitos ou, ainda, um processo de Markov aumentando com ações e recompensas.

Quando se aplica a ação $a \in A$ no estado $e \in E$, o sistema realiza a transição do estado e para um novo estado $e' \in E$, baseado em uma distribuição de probabilidade em relação ao conjunto de possíveis transições.

Uma função recompensa indica a recompensa por estar em um estado ou realizar uma ação a partir de um estado, portanto $R: E \times A \times E \rightarrow \mathbb{R}$ como a função recompensa obtida para uma transição particular entre estados. A recompensa por adotar a ação a , resultando na transição do estado e para o estado e' , pode ser denotada por $R(e, a, e')$. Embora a conotação básica de recompensa seja algo positivo, cabe

ressaltar também que isso pode ser interpretado como penalização (WIERING; OTTERLO, 2012).

Conforme ilustrado pela Figura 2, na representação gráfica do conceito de ação em um processo markoviano de decisão, ações são ilustradas a partir de nós intermediários que, a partir da sua execução, permitem a transição entre estados. Em alguns sistemas, nem todas as ações podem ser aplicadas em todos os estados.

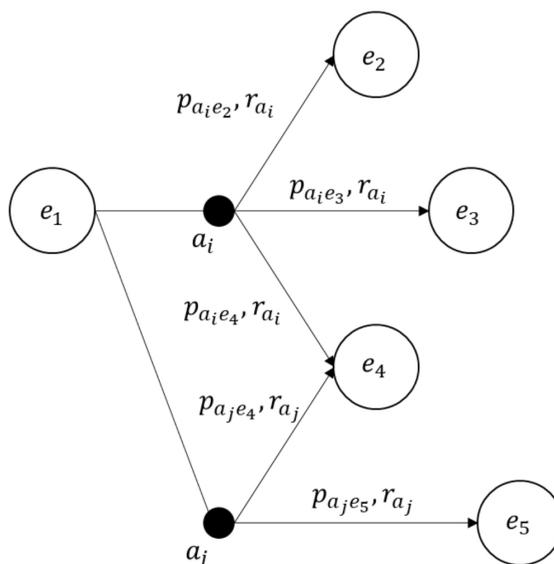


Figura 2 Representação de um processo decisório de Markov, com estados, ações, probabilidades de transição e recompensas

Fonte: Elaborada pelos autores.

Para os efeitos deste estudo, uma ação pode ser representada pelas possíveis decisões tomadas em um momento específico para uma determinada ação negociada, ou seja, compra, venda ou *hold* (não realizar nenhuma operação).

Alguns trabalhos que também buscaram criar um sistema de tomada de decisões para operação em mercados financeiros podem ser referenciados. Wu *et al.* (2019), Shin, Ra e Choi (2019) e Rundo (2019) utilizaram como princípio básico modelos baseados em *deep learning* e implementação de agentes de operação em um *framework* de aprendizado por reforço. Os dois primeiros trabalhos utilizaram o mercado de ações (chinês e sul-coreano, respectivamente) como objeto de estudo; já Rundo (2019)

baseou-se no mercado cambial (*foreign exchange* – Forex) e no conceito de arbitragem financeira.

Todos os trabalhos diferem entre si em relação à quantidade de diferentes papéis utilizados – isto é, operados em testes e simulações – e ao período da série histórica. Rundo (2019), por exemplo, estudou a operação no mercado Forex para 12 pares de moedas, entre os períodos de 2012 e 2017, levando em conta os valores de abertura, fechamento, mínimo e máximo a cada 15 minutos.

O trabalho de Sattarov *et al.* (2020) diferenciou-se pela utilização de dados de criptomoedas e pelo curto período de tempo analisado para o processo de aprendizado dos modelos.

Yang *et al.* (2020) utilizou a estratégia de um *ensemble* de modelos para o estudo de 30 ações norte-americanas no intervalo de 2009 a 2020. O retorno acumulado para o período de teste foi avaliado em mais de 70%, equiparando-se com outros os estudos anteriormente abordados.

Uma tônica comum entre todos os trabalhos supracitados é estratégia de aprendizado *off-line*, aliada a 1. diferentes volumes de dados históricos (meses ou anos), 2. diferentes variáveis para treinamento dos agentes de operação (valores de abertura, fechamento, mínimos e máximos) e 3. diferentes frequências de operação (por hora, por dia ou outras janelas de tempo).

Este estudo limita-se ao uso dos conceitos de cadeia de Markov e MDP supracitados, não realizando o uso qualquer algoritmo de *machine learning*.

Sobre o processo de compra e venda, conforme descrito por Vishwanath e Krishnamurti (2009), o melhor preço para o comprador equivale ao menor valor oferecido por um vendedor. Em termos técnicos: o comprador, conhecido por *buyer*, estabelece a melhor (menor) oferta de compra de determinado ativo, conhecido por *bid*. Por sua vez, o vendedor, conhecido por *seller*, estabelece a melhor (maior) oferta de venda de determinado ativo, conhecido por *ask* ou *offer*. A diferença entre *bid* e *ask* é conhecida por *spread* e, conforme descrito por Rampell e Kupor (2021), equivale ao custo implícito de ser capaz de imediatamente realizar uma operação de compra ou venda no mercado.

O ambiente nacional de operações possui três importantes aspectos básicos a serem considerados: 1. no mercado de ações brasileiro, a fração monetária mínima é de R\$ 0,01; 2. o lote mínimo de operação equivale a múltiplos de 100 papéis; e, assim como em outros mercados, 3. os valores de abertura e fechamento de cada papel são atualizados minuto a minuto, e, em tempo real (*real time*), o valor corrente (para compra e venda) também é atualizado.

Além disso, vários tipos de ordens de compra ou de venda podem ser executados. Para efeito deste estudo, serão descritos apenas dois tipos: ordem a mercado e ordem limitada. Ordens a mercado acontecem ou são realizadas quando não existe um limi-

tante quanto ao preço de negociação, ou seja, acontecem pelo melhor preço possível identificado no momento de sua execução; ordens limitadas, por sua vez, acontecem atreladas a um limite de preço de negociação, considerando geralmente um preço mínimo para a venda e outro máximo para a compra. Vale ressaltar uma vez mais que todas as negociações descritas são processadas por um sistema eletrônico de negociação, sem a necessidade de presença física de qualquer representante nos pregões.

Toda vez que uma ordem atinge o valor desejado para obtenção de lucro – conforme ilustrado pela Figura 3 –, seja a partir de uma ordem limitada ou de uma ordem a mercado, diz-se que foi uma ordem de tomada de lucro (*take profit*); por sua vez, quando uma ordem atinge o valor máximo permitido para minimização de perdas, seja a partir de uma ordem limitada ou de uma ordem a mercado, diz-se que foi uma ordem de encerramento de perdas (*stop loss*).

Existem ainda, em relação às características de uma operação, duas formas de negociação: comprando e depois vendendo – o que tende a ser a ordem mais intuitiva quando se fala sobre compra e venda de ações – ou vendendo (ou pegando emprestado) e depois comprando (ou devolvendo o empréstimo).

Na operação de compra e posterior venda, a negociação se caracteriza de forma bastante simples: emite-se uma primeira ordem (limitada ou a mercado) para compra de uma determinada ação por um valor específico, respeitando-se o lote mínimo de operação; estabelece-se um preço-alvo, seja este com objetivo de realização de lucro ou para limitar um máximo valor de perda; uma segunda ordem oposta, nesse caso de venda, é realizada com o objetivo de encerrar a posição aberta, realizando-se o lucro (ou o prejuízo). Ou seja, essa operação de compra/venda é composta por duas ordens: uma primeira ordem para compra e uma segunda ordem para venda.

Um exemplo dessa prática seria comprar ações da empresa ABCD3 por R\$ 10,00, por papel. Considerando-se o lote mínimo de 100 ações, investem-se R\$ 1.000,00 em uma ordem de compra. Quando se emite a ordem de venda para as mesmas 100 ações pelo valor unitário de R\$ 11,00, realiza-se o lucro de R\$ 1,00 por ação, o que totaliza R\$ 100,00.

Contudo, é possível realizar a operação imediatamente contrária, ou seja, a operação de venda e posterior compra. Conforme explicado por Assaf Neto (2014), essa prática também é conhecida por negociação a descoberto e consiste na venda de títulos que o investidor não possui.

Na prática, para realizar esse tipo de operação, a corretora negocia o empréstimo desses papéis com outro investidor que esteja disposto a “alugar” suas ações, o que possibilita a posterior venda no mercado à vista. Decorrido o prazo do empréstimo, que pode ser de um ou vários dias, realiza-se a devolução das ações emprestadas, liquidando assim a posição por meio de uma ordem de compra. O resultado dessa operação, conforme observado por Assaf Neto (2014), é calculado por meio da diferença entre

o preço da ação à vista e o seu preço no final do empréstimo. Essa estratégia de operação, embora estranha, pode ser utilizada para apurar ganhos em movimentos de baixa de uma ação ou do mercado como um todo.

Um exemplo dessa prática seria vender (ou seja, tomar emprestado e então vender) ações da empresa ABCD3 por R\$ 10,00, por papel. Considerando-se o lote mínimo de 100 ações, investem-se R\$ 1.000,00 em uma ordem de venda. Quando se emite a ordem de compra (e então se devolvem as ações que foram tomadas por empréstimo) para as mesmas 100 ações pelo valor unitário de R\$ 9,00, realiza-se o lucro de R\$ 1,00 por ação, o que totaliza R\$ 100,00.

Para toda e qualquer ordem emitida, devem-se ainda considerar, nos ganhos ou nas perdas, os valores relativos a impostos e taxas de corretagem.

Como descrito por Takar (2019), o investidor que negocia ações pode ter que pagar imposto de renda na venda desses ativos, em função do prazo em que determinada posição foi liquidada e com o lucro obtido.

No caso de operações de *day trade*, o tributo de 20% deve ser considerado sob qualquer lucro obtido, e, dessa forma, a declaração deve ser feita pelo investidor mensalmente. Retido diretamente pela corretora e repassado à Receita Federal brasileira, há incidência de 1% sobre o lucro, “na fonte”. Prejuízos podem ser abatidos e considerados no montante total (CORRAL, 2021).

De acordo com a B3 (2021), outros impostos também devem ser considerados de acordo com as tabelas de tarifas, sempre com base nos valores máximos da taxa de negociação de 0,005% e da taxa de liquidação de 0,018%, ambos sobre o valor financeiro da operação, e desde que caracterizadas como *day trade*. Outros impostos devem incidir de acordo com as regras de cada estado ou região do Brasil.

Cabe observar ainda que um dos fatores mais importantes a ser considerado são os valores referentes a taxas de corretagem e custódia. Ou seja, um valor fixo cobrado pelas corretoras a cada ordem (de compra ou de venda) realizada e um valor cobrado apenas para registrar os investimentos, respectivamente. Cada corretora aplica suas respectivas taxas com base nas diferentes naturezas de operação e regras internas, muito embora algumas até pratiquem atualmente taxas reduzidas em função do volume operado ou até, conforme observado em Clear (2021) e Rico (2021), isenção para operações com ações.

Tipo de ordem	Características de negociação	Resultado da operação
<ul style="list-style-type: none"> • À mercado: acontece pelo melhor preço possível identificado no momento de sua execução 		<p>Compra-venda $R = [(P_{t'} - P_t) \cdot vol] - tributos$</p> <p>Venda-compra (ordem à descoberto) $R = [(P_t - P_{t'}) \cdot vol] - tributos$</p>
<ul style="list-style-type: none"> • Limitada: acontece atreladas a um limite de preço de negociação 		<p>Compra-venda $R = [(P_n - P_A) \cdot vol] - tributos$</p> <p>Venda-compra (ordem à descoberto) $R = [(P_A - P_B) \cdot vol] - tributos$</p>

Figura 3 A relação entre os tipos de ordem, as características de negociação e os resultados observados a cada operação

Fonte: Elaborada pelos autores.

4 RESULTADOS E DISCUSSÃO

A partir dos conceitos teóricos até aqui apresentados, é possível desenvolver, implementar, testar e avaliar um sistema automático de compra e venda de ações, dispensando a necessidade de realizar predições exatas dos valores de um ativo em um instante no tempo, mas levando em consideração apenas a direção para qual o mercado aponta. Esse algoritmo será capaz de tomar decisões automatizadas, instantâneas e realizadas em tempo real, de tal modo que possibilite e suporte o regime de HFT.

A plataforma de *trading* MetaTrader 5 permite a utilização de lógica de programação para operação automática (*algotrading*), além de realizar a coleta de dados históricos de valores de abertura, fechamento, mínimo e máximo de qualquer ação (na bolsa brasileira ou internacional), com granularidade mínima de um minuto. Com esses dados, será possível 1. definir os estados de transição e 2. as probabilidades de transição entre estados. Este estudo utiliza a ação CCRO3 (ações da empresa brasileira de concessões de infraestrutura, CCR SA) como objeto inicial de estudo.

A Tabela 1 trata de um exemplo de extração de dados históricos realizado para a ação CCRO3 por meio do *software* MetaTrader 5. Essa extração possui os valores de abertura (*open*), fechamento (*close*), mínimo (*low*) e máximo (*high*) de cada minuto (*timestamp*), entre junho e dezembro de 2020.

Com esses dados, foi possível validar a premissa básica da estratégia de operação, que consiste na realização de pequenos ganhos em um intervalo de tempo curto.

Para que isso seja possível:

- 1) o preço dessa ação deve variar dentro do mesmo minuto, ou seja, a diferença absoluta entre o preço mínimo e o preço máximo deve ser diferente de R\$ 0,00;
- 2) o preço dessa ação deve variar entre um minuto e outro.

TABELA 1

Valores de abertura, fechamento, mínimo e máximo de cada minuto da ação CCRO3, entre junho e dezembro de 2020

Timestamp	Open	High	Low	Close
1592509260	14,51	14,51	14,48	14,49
1592509320	14,50	14,54	14,49	14,50
1592509380	14,50	14,54	14,50	14,53
1592509440	14,52	14,52	14,51	14,51
1592509500	14,51	14,52	14,49	14,51
1592509560	14,52	14,54	14,50	14,52
1592509620	14,52	14,56	14,52	14,54
1592509680	14,54	14,56	14,54	14,55
1592509740	14,55	14,56	14,51	14,51

Fonte: Elaborada pelos autores.

Em uma análise exploratória, conforme ilustrado pela Figura 4, foi possível observar a distribuição de frequências da diferença absoluta (em reais) entre os valores mínimos e máximos da ação no intervalo de um minuto. Aproximadamente 90% das diferenças estão compreendidas entre R\$ 0,00 e \pm R\$ 0,07. Potencialmente, há possibilidade teórica de realizar uma operação que resulte em ganhos de pelo menos R\$ 0,01 a cada minuto.

Uma vez com os dados extraídos, a operação de *shift* da série temporal permitiu relacionar a série consigo mesma considerando o incremento de tempo de um minuto, conforme ilustra a Figura 4. Em outras palavras, o preço de abertura do minuto m_{i+1} será equivalente ao preço de fechamento do minuto m_i . Em contraponto ao trabalho de Huang (2018), este estudo utiliza o preço de abertura e não o de fechamento de cada minuto, uma vez que a tomada de decisão de compra e venda será feita no início de cada minuto. As diferenças entre valores de abertura do último minuto do dia d_{i+1} versus o primeiro minuto do dia d_{i+1} foram desconsideradas.

Na Figura 4, pode-se observar a distribuição de frequências da diferença absoluta (em reais) entre os preços de abertura no intervalo de um minuto. Aproximadamente 95% das diferenças estão compreendidas entre R\$ 0,00 e ± R\$ 0,03.

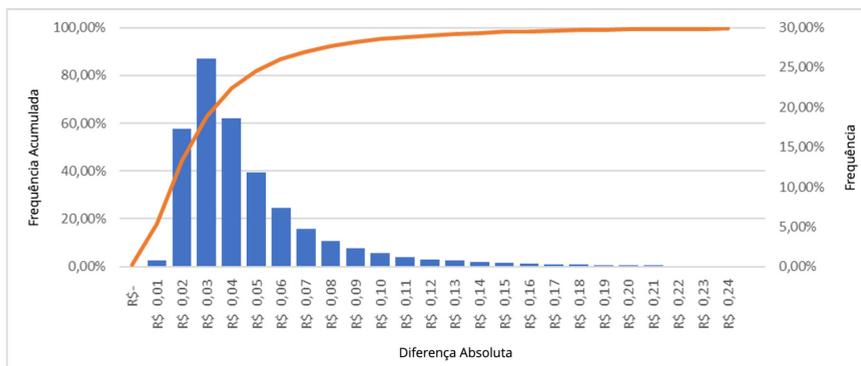


Figura 4 Distribuição de frequências da diferença entre preços mínimos e máximos dentro do intervalo de um minuto

Fonte: Elaborada pelos autores.

Início		Fim	
timestamp	open		
1592509260	14,51	1592509320	14,50
1592509320	14,50	1592509380	14,50
1592509380	14,50	1592509440	14,52
1592509440	14,52	1592509500	14,51
1592509500	14,51	1592509560	14,52
1592509560	14,52	1592509620	14,52
1592509620	14,52	1592509680	14,54
1592509680	14,54	1592509740	14,55
1592509740	14,55		

Figura 5 Exemplo de operação de shift, considerando o preço de abertura

Fonte: Elaborada pelos autores.

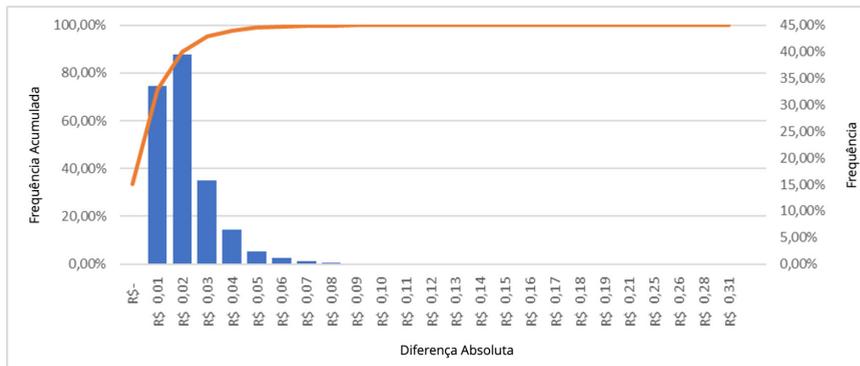


Figura 6 Distribuição de frequências da diferença entre preços de abertura dentro de intervalo de um minuto

Fonte: Elaborada pelos autores.

Utilizar a variação nominal (isto é, em unidade monetária) não permitiria descrever com o devido detalhe as mudanças de estado do sistema. Por exemplo: a mesma variação de R\$ 0,01 pode ser observada entre os preços R\$ 10,00 e R\$ 10,01, e também entre R\$ 100,00 e R\$ 100,01, contudo, trata-se de movimentos substancialmente diferentes em função da ordem de grandeza do valor absoluto em determinado momento. Portanto, os estados do processo de Markov não serão definidos a partir dos valores absolutos de preço ou da diferença de preço absoluta entre início e fim de cada intervalo de tempo, mas sim a partir da respectiva variação percentual.

A operação de *shift*, conforme mostrado anteriormente, tornou possível calcular a variação percentual (o que foi chamado de “delta”) de um minuto para o outro. Os resultados da operação seguem exemplificados na Tabela 2.

TABELA 2

Cálculo do delta percentual entre valores de abertura no intervalo de um minuto

Início	Fim	Preço inicial	Preço final	Delta
1592509260	1592509320	14,51	14,50	-0,00068918
1592509320	1592509380	14,50	14,50	0
1592509380	1592509440	14,50	14,52	0,00137931
1592509440	1592509500	14,52	14,51	-0,00068870
1592509500	1592509560	14,51	14,52	0,00068918
1592509560	1592509620	14,52	14,52	0
1592509620	1592509680	14,52	14,54	0,00137741
1592509680	1592509740	14,54	14,55	0,000687758

Fonte: Elaborada pelos autores.

O critério básico para a escolha do tamanho dos intervalos regulares foi aquele que pudesse representar as menores variações percentuais do valor de uma ação (R\$ 0,01), independentemente do valor da ação. Nesse caso, o cálculo do delta percentual entre início e fim de cada intervalo de um minuto revelou mais de 4.100 valores únicos. Contudo, trabalhar com tantos estados diferentes, além de pouco prático, demonstrou-se pouco relevante, uma vez que nominalmente representam as mesmas variações (por exemplo: -0,068918% e -0,068870% representam a mesma variação de R\$ 0,01 em intervalos de valores absolutos muito próximos – R\$ 14,51-14,50 e R\$ 14,54-14,52, respectivamente – diferentemente do exemplo com valores muito diferentes apresentados antes). Portanto, a criação de intervalos regulares de variação foi realizada para que a variação mínima de R\$ 0,01 fosse levada em consideração, mas também que o número de estados fosse reduzido a fim de que as mesmas variações ficassem dentro do mesmo intervalo (por exemplo: $e_x = [-0,019\%, 0,118\%]$, $e_y = [0,709\%, 0,808\%]$). Além disso, esse processo permitiu eliminar estados considerados praticamente “inalcançáveis”, isto é, eventos extremamente raros com probabilidade de ocorrência é muito baixa (perto de 0,000%).

Dessa forma, pôde-se consolidar o processo de Markov com apenas 15 estados de transição, conforme é possível observar na representação da matriz de probabilidade de transições ilustrada pelo mapa de calor (*heatmap*) da Figura 7. Na Figura, o vermelho representa uma probabilidade que se aproxima de 0%, enquanto o verde indica uma probabilidade que se aproxima de 100%.

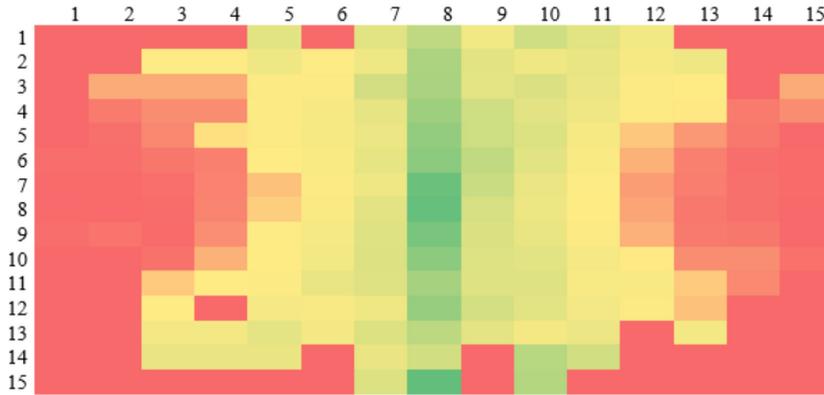


Figura 7 Heatmap representando a probabilidade de transição entre estados da cadeia de Markov

Fonte: Elaborada pelos autores.

Por sua vez, os estados podem ser descritos pelos limites inferior e superior da variação (delta) percentual entre valores de abertura, conforme mostra a Tabela 3.

TABELA 3

Os intervalos de variação para definição dos estados de transição da cadeia de Markov

Estado	Inferior	Superior
1	-0,671%	-0,572%
2	-0,572%	-0,474%
3	-0,474%	-0,375%
4	-0,375%	-0,277%
5	-0,277%	-0,178%
6	-0,178%	-0,079%
7	-0,079%	0,019%
8	0,019%	0,118%
9	0,118%	0,216%
10	0,216%	0,315%
11	0,315%	0,413%
12	0,413%	0,512%
13	0,512%	0,610%
14	0,610%	0,709%
15	0,709%	0,808%

Fonte: Elaborada pelos autores.

Após determinar os estados que definem a cadeia de Markov e o cálculo das respectivas probabilidades de transição entre eles, é necessário estabelecer a lógica que irá guiar o processo de tomada de decisão de compra e venda, a partir de um MDP. De forma genérica, a cada estado inicial será possível tomar uma ação a partir do conjunto de ações possíveis que, posteriormente, irá permitir a transição de estados para o estado final. Esse MDP pode ser ilustrado conforme a Figura 8.

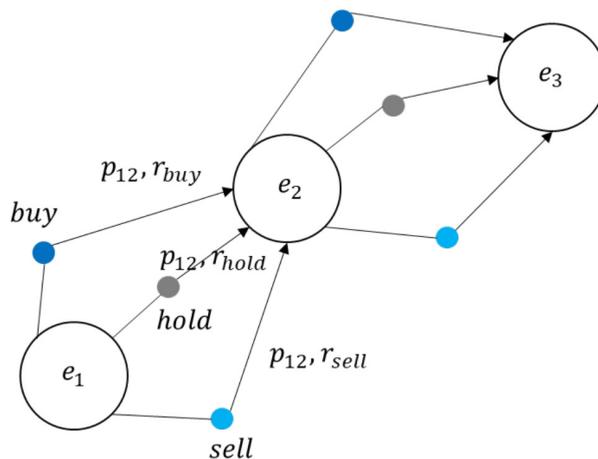


Figura 8 Exemplo do MDP

Fonte: Elaborada pelos autores.

Uma ordem de compra (*buy*) será realizada sempre que o somatório das probabilidades que indicam transição para um estado de valorização for maior que a probabilidade de queda ou a probabilidade de estabilização; uma ordem de venda (*sell*) será realizada sempre que o somatório das probabilidades que indicam transição para um estado de desvalorização for maior que a probabilidade de valorização ou de estabilização do valor; nenhuma ação será tomada (*hold*) caso a probabilidade de estabilização seja a maior entre o somatório que representa a probabilidade de valorização e o somatório que representa a probabilidade de desvalorização.

Portanto, o sistema realizará (1) uma ordem de compra, venda ou *hold* de uma ação no início de cada minuto, ao longo de todo horário comercial de operação, (2) buscando o ganho máximo equivalente à fração monetária mínima de R\$ 0,01 por papel transacionado – ou seja, sempre que esse ganho estimado for alcançado, a operação será finalizada a fim da realização do lucro – e respeitando o lote mínimo de operação; (3) para limitação de perdas, um valor máximo será atribuído para encerrar

compulsória e automaticamente uma posição (de compra ou venda) caso tal prejuízo seja observado. Esse valor poderá ser estabelecido no início de cada rodada de operações, e quanto menor for esse valor, menor será a permissividade a variações de preço dentro de um intervalo de tempo. Para os efeitos deste estudo, o valor será de R\$ 3,00 por ordem (representando 95% das variações de preço dentro de um minuto, conforme é possível observar na Figura 4).

O primeiro minuto de operação funciona apenas como ponto de partida para estabelecer os valores iniciais de preço p . No segundo minuto, atualiza-se o preço atual com o último valor de abertura. Calcula-se o delta percentual ente os valores conhecidos, e, a partir desse valor, será possível determinar o estado atual na cadeia de Markov.

Então, calcula-se o somatório σ^+ das probabilidades de transição entre o estado atual σ^- e todos os estados superiores (isto é, estados que representem um aumento no preço atual da ação), e das probabilidades de transição ente o estado atual e todos os estados inferiores, excluindo a probabilidade de permanência (estabilização de valor) no estado atual de ambas as somas.

Com base nos estados descritos na Tabela 3, um exemplo hipotético seria considerar que o estado atual é o 5. As transições desse estado inicial para os estados de 1 a 6 representam desvalorizações, portanto o somatório dessas probabilidades será σ^- . Por sua vez, as transições desse estado inicial para os estados de 8 a 15 representam valorizações, portanto o somatório dessas probabilidades será σ^+ .

Intuitivamente, é possível estabelecer que, dessa forma, qualquer somatório que seja maior que 33% (caso a distribuição de probabilidade seja uniforme entre valorização, estabilização e desvalorização) será dado como a ação vencedora. Contudo, não há garantia de que essa seja a probabilidade que melhor direciona decisões de negociação. Portanto, o algoritmo também leva em consideração um valor limite (*threshold*) para que outros valores de probabilidade mínima (por exemplo, 40%, 50% etc.) possam ser estabelecidos.

Sendo assim, um ponto de decisão é definido, e, se $\sigma^+ \geq \text{threshold}$, cria-se uma ordem de compra; se $\sigma^- \geq \text{threshold}$, envia-se uma ordem de venda; caso contrário, nenhuma ordem será enviada. Para os efeitos deste estudo, o valor do limite inicia-se em 0,33 (ou 33%) e poderá ser modificado e ajustado, caso se observe que o valor faça com que operações em excesso sejam executadas.

O fluxograma contido na Figura 9 ilustra o processo de decisão do algoritmo com base no MDP criado e nas regras de operação.

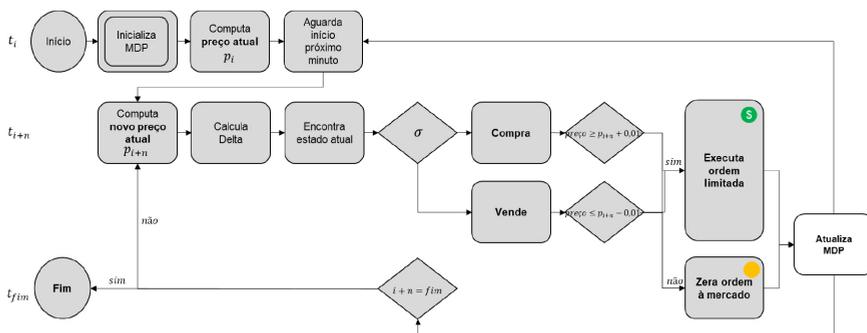


Figura 9 Fluxograma do algoritmo de tomada de decisão

Fonte: Elaborada pelos autores.

O algoritmo foi testado em um ambiente *on-line* de demonstração, isto é, um ambiente que é abastecido exatamente pelas mesmas informações e tende a funcionar de forma exatamente análoga ao funcionamento do ambiente para operação real. Sendo assim, os resultados apresentados fazem referência aos preços reais observados a cada instante de tempo, no mercado brasileiro, sem qualquer outra modificação ou ajuste.

Essa análise, portanto, consistiu em operar, ao longo de 12 dias, o total de 3.984 operações, entre 29 de dezembro de 2020 e 15 de janeiro de 2021, para não só testar a lógica estabelecida anteriormente, mas também para realizar a validação do hiperparâmetro *threshold*, mantido em 33%, uma vez que não tenha sido observado descontrole na quantidade de ordens executadas em função do valor estabelecido inicialmente.

No período, o ganho bruto relativo acumulado foi de 61,8% operando apenas o papel CCRO3, conforme cadeia caracterizada pela Tabela 3 e *heatmap* da Figura 5. As operações resultaram em retorno bruto diário médio de 5,2% (em termos absolutos, o equivalente a, aproximadamente, R\$ 68,00 de lucro bruto por dia) em relação ao montante investido inicial de, aproximadamente, R\$ 1.300,00. É importante lembrar que cada operação positiva resultou em R\$ 1,00 de lucro.

Por sua vez, cada operação negativa resultou, em média, em R\$ 1,70 de prejuízo – em função do valor de R\$ 3,00 estabelecido como limite de perdas por operação.

A variação percentual da ação no período foi de -4,9% (*benchmark*). Além disso, conforme ilustrado pela Tabela 4, é possível visualizar que, entre os 12 dias de operação, apenas dois apresentaram retorno bruto negativo, em média, de -0,7%. Picos de ganho foram observados também em duas ocasiões quando o retorno percentual médio ultrapassou o valor de 10%, enquanto a taxa de acerto aproximou-se dos 80%.

Contudo, quando se consideram os impostos de negociação, liquidação e emolumentos somados, aproximadamente de 0,02% do montante total operado, e o imposto de renda de 20% sobre o lucro – o que totaliza aproximadamente R\$ 1.970,00 –, os resultados líquidos são negativos. É importante lembrar que esses impostos são aplicados automaticamente pela Bolsa e pelo governo brasileiro. Sendo assim, o ganho líquido relativo acumulado nesse período foi de aproximadamente de -151%. A corretora utilizada para a realização dessas operações não aplica nenhum valor de corretagem ou custódia, e os prejuízos foram abatidos do imposto de renda.

Também foi possível observar taxa de acerto média de 75%, sendo, em média, 63% em operações resultando ganho e outros 11% em operações neutras (com zero lucro e zero prejuízo). A taxa de acerto (*win rate*) e a taxa neutra (*neutral rate*) foram calculadas por meio das equações (3) e (4), sendo w' e t' , respectivamente, a quantidade de operações que resultaram em lucro e a quantidade de operações neutras. A taxa de ganho total equivale à soma $T = WR + NR$.

$$WR = \frac{w'}{h} \quad (3)$$

$$NR = \frac{t'}{h} \quad (4)$$

Relacionando com o conceito/estratégia de HFT, observou-se média diária de 332 operações, equivalente a 0,8 operações por minuto (considerando período de operação das 10 às 17 horas).

TABELA 4

Resumo das taxas de acerto e ganhos absolutos e percentuais do período de testes

Data	Ganho absoluto	Ganho relativo	Win rate	Neutral rate	Total
29/12/2020	R\$ 49,00	3,6%	56%	15%	71%
30/12/2020	R\$ 82,00	6,0%	56%	18%	74%
04/01/2021	R\$ 59,00	4,4%	61%	12%	73%
05/01/2021	R\$ 130,00	10,0%	67%	12%	79%
06/01/2021	R\$ 84,00	6,5%	66%	10%	76%
07/01/2021	-R\$ 12,00	-0,8%	51%	14%	65%
08/01/2021	R\$ 65,00	5,0%	85%	2%	87%
11/01/2021	R\$ 33,00	2,5%	65%	10%	75%
12/01/2021	-R\$ 7,00	-0,5%	50%	16%	66%
13/01/2021	R\$ 106,00	8,1%	66%	12%	78%
14/01/2021	R\$ 148,00	10,9%	74%	5%	79%
15/01/2021	R\$ 80,00	6,1%	64%	11%	75%

Fonte: Elaborada pelos autores.

5 CONSIDERAÇÕES FINAIS

Foi possível utilizar os conceitos de MDP e HFT para elaboração de um sistema automático de compra e venda de ações, focando o processo de tomada de decisões e não especificamente o processo de predição de valores específicos para dado intervalo de tempo. Os resultados financeiros brutos demonstraram-se positivos quando comparado ao *benchmark* (-4,9% de desvalorização pela estratégia de *buy-and-hold versus* ~60% de valorização pela operação automatizada).

Por sua vez, a taxa de acerto precisa melhorar consideravelmente para que compense os impostos que incidem sobre o volume total operado e somados aos prejuízos observados (intrínsecos aos erros de operação), impedindo que os resultados se tornem muito negativos, o que tornaria a operação inviável na prática.

Além disso, ainda se faz necessário ampliar o escopo de testes para maior período de operação assistida, em um ambiente *on-line* real e não simulado como o utilizado ao longo deste estudo. Ademais, a diferença entre *buy* e *ask* (*spread*) não foi levada em

consideração nesses testes e precisará ser incorporada nesse processo de transição (possivelmente, ajustando o ganho mínimo de R\$ 0,01 para R\$ 0,02). É importante ainda ampliar os testes para outras ações mais relevantes e representativas nos movimentos da Bolsa brasileira.

Existe uma oportunidade clara para avançar com este estudo adicionando uma camada de aprendizado no algoritmo de tomada de decisão, como o uso efetivo de um algoritmo de *machine learning*. A lógica de aprendizado por reforço pode ser incorporada a fim de tornar o algoritmo de tomada de decisão inteligente com base na interação com o ambiente de operação, as ações tomadas e as recompensas recebidas.

MARKOV CHAINS FOR HIGH FREQUENCY STOCK TRADING STRATEGY

Abstract

The problem of predicting the price or value of a given share has been studied, deepened, and evolved in recent decades, seeking to achieve the primary objective of positive financial achievements with the lowest possible risk. A well-known risk-reduction strategy is high-frequency trading, which takes advantage of small price variations and allows one to make gains multiple times a day. Starting from the hypothesis and interpretation that the variation of stock prices is equivalent to a random walk, it is possible to characterize a (stochastic) Markov process with known states and transition probabilities, describing the price variations of a stock specific over time. Thus, this study seeks to estimate the potential financial return from using Markov chains for automated decision-making in a strategy for buying and selling shares at high frequency, comparing the results obtained with the valuation (or devaluation) of the stock in the same period.

Keywords: Stocks. Random walks. Markov Chains. High frequency trading.

REFERÊNCIAS

- ABELL, H. *Digital day trading: moving from one winning stock position to the next*. [S. l.]: Dearborn Trade Pub, 1999.
- ALMASARWEH, M.; WADI, S. ARIMA model in predicting banking stock market data. *Modern Applied Science*, v. 12, n. 11, p. 309-312, 2018. Disponível em: <https://www.ccsenet.org/journal/index.php/mas/article/view/0/37282>. Acesso em: 7 dez. 2022.
- ASSAF NETO, A. *Mercado Financeiro*. 12. ed. São Paulo: Atlas, 2014.
- B3. *Ações*. 2021. Disponível em: http://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/acoes.htm. Acesso em: 12 jun. 2021.
- BALLINI, R. *et al.* A comparative analysis of neurofuzzy, ANN and ARIMA models for Brazilian stock index forecasting. *In: INTERNATIONAL CONFERENCE ON COMPUTING IN ECONOMICS AND FINANCE – CEF*, 16., Londres, Inglaterra, 2010.
- BRIOLA, A. *et al.* Deep reinforcement learning for active high frequency trading. *Arxiv*, Cornell University, Feb. 2021. Disponível em: <https://arxiv.org/abs/2101.07107>. Acesso em: 7 dez. 2022.
- CHITENDERU, T. T.; MAREDZA, A.; SIBANDA, K. The random walk theory and stock prices: evidence from Johannesburg stock exchange. *International Business & Economics Research Journal*, v. 13, n. 6, p. 1241-1250, Dec. 2014.
- CHAN, E. P. *Quantitative trading: how to build your own algorithmic trading business*. New Jersey: John Wiley & Sons, 2009.
- CHOU, J.; NGUYEN, T. Forward forecast of stock price using sliding-window metaheuristic-optimized machine learning regression. *IEEE Transactions on Industrial Informatics*, v. 14, n. 17, July 2018.
- CLEAR. Custos. 2021. Disponível em: <https://www.clear.com.br/site/custos>. Acesso em: 15 jun. 2021.
- CONEGUNDES, L.; PEREIRA, A. C. M. Beating the stock market with a deep reinforcement learning day trading system. *In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN)*, Belo Horizonte, 2020, p. 1-8
- CORRAL, H. *Day trade no imposto de renda: como declarar?* Expert XP, 8 fev. 2021. Disponível em: <https://conteudos.xpi.com.br/aprenda-a-investir/relatorios/day-trade-no-imposto-de-renda/>. Acesso em: 15 jun. 2021.
- CRANE, J. *Advanced swing trading: strategies to predict, identify, and trade future markets swings*. New York: John Wiley & Sons, 2003.
- CRUZ, R. R.; GONZALEZ, A. D. D. Investment portfolio trading based on Markov chain and fuzzy logic. *In: IEEE LATIN AMERICAN CONFERENCE ON COMPUTATIONAL INTELLIGENCE*, Guadalajara, 2018. p. 1-6. DOI 10.1109/LA-CCI.2018.8625246.

GRAMI, A. *Probability, random variables, statistics and random processes: fundamentals & applications*. New York: John Wiley & Sons, 2020.

HENRIQUE, B. M.; SOBREIRO, V. A.; KIMURA, H. Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, v. 4, n. 3, p. 183-201, Sept. 2018.

HUANG, C. Financial trading as a game: a deep reinforcement learning approach. *Arxiv*, Cornell University, July 2018.

JIANG, W. Applications of deep learning in stock market prediction: recent progress. *Arxiv*, Cornell University, Feb. 2020.

LACHIHEB, O.; GOUIDER, M. S. A hierarchical deep neural network design for stock returns prediction. *Procedia Computer Science*, v. 126, p. 264-272, 2018.

NIKOU, M.; MANSOURFAR, G.; BAGHERZADEH, J. Stock price prediction using deep learning algorithms and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management: An International Journal*, v. 26, p. 164-174, Sept. 2019.

PANDEY, V. S.; BAJPAI, A. Predictive efficiency of ARIMA and ANN models: a case analysis of Nifty Fifty in Indian stock market. *International Journal of Applied Engineering Research*, v. 14, n. 2, p. 232-244, 2019.

PAPOULIS, A.; PILLAI, S. U. *Probability, random variables, and stochastic processes*. 4th. ed. New York: McGraw-Hill, 2002.

PARMAR, I. *et al.* Stock market prediction using machine learning. *In: INTERNATIONAL CONFERENCE ON SECURE CYBER COMPUTING AND COMMUNICATION*, 1., Punjab, India: Dr B R Ambedkar National Institute of Technology, 2018.

RAMPELL, A.; KUPOR, S. Breaking down the payment for order flow debate. A16Z, 2021. Disponível em: <https://a16z.com/2021/02/17/payment-for-order-flow/>. Acesso em: 13 jun. 2021.

RICO. Custos. 2021. Disponível em: <https://www.rico.com.vc/custos>. Acesso em: 15 jun. 2021.

RUNDO, F. Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems. *Applied Sciences*, 2019.

SADORSKY, P. A Random Forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management*, v. 14, n. 2, p. 1-20, 2021.

SATTAROV, O. *et al.* Recommending cryptocurrency trading points with deep reinforcement learning approach. *Applied Sciences*, v. 10, n. 4, 1506, Feb. 2020.

SHIN, H.; RA, I.; CHOI, Y. A deep multimodal reinforcement learning system combined with CNN and LSTM for stock trading. *In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY CONVERGENCE (ICTC)*, Keju, Coreia do Sul, 2019.

TAKAR, T. Investe e ações? Pode ter de pagar IR todo mês, com muita cara se esquecer. UOL, 28 maio 2019. Disponível em: <https://economia.uol.com.br/financas-pessoais/noticias/>

redacao/2019/05/28/como-calculiar-imposto-de-renda-sobre-aco.es.htm. Acesso em: 15 jun. 2021.

VIGG, S.; ARORA, D. A study of randomness in the national stock exchange. *Prestige International Journal of Management & IT-Sanchayan*, v. 7, p. 112-125, 2018.

VIJH, M. et al. Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, v. 167, p. 599-606, 2020.

VISHWANATH, S. R.; KRISHNAMURTI C. (Ed.). *Investment management: a modern guide to security analysis and stock selection*. [S. l.]: Springer, 2009.

WIERING, M.; OTTERLO M. *Reinforcement learning state-of-the-art*. [S. l.]: Springer, 2012.

WU J. et al. Quantitative trading on stock market based on deep reinforcement learning. *In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS*, Budapeste, Hungria, 2019.

YADAV, A.; JHA, C. K.; SHARAN, A. Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*, v. 167, p. 2091-2100, 2020.

YANG, H. et al. Deep reinforcement learning for automated stock trading: an ensemble strategy. *SSRN*, Sept. 2020. Disponível em: <http://dx.doi.org/10.2139/ssrn.3690996>. Acesso em: 8 dez. 2022.

Contato

Leandro Augusto da Silva
leandroaugusto.silva@mackenzie.br

Tramitação

Recebido em novembro de 2021.
Aprovado em maio de 2022.