**CENTRO UNIVERSITÁRIO DO INSTITUTO MAUÁ DE TECNOLOGIA**

**Escola de Engenharia Mauá**

Engenharia de Controle e Automação

BRUNO MARINOTTI BROSTOLINE

IGOR AMARAL CORREA

LEON CHEROTO WEISSBERG

# Inteligência artificial na bolsa de valores

São Caetano do Sul

2019

BRUNO MARINOTTI BROSTOLINE

IGOR AMARAL CORREA

LEON CHEROTO WEISSBERG

# Inteligência artificial na bolsa de valores

Trabalho de Conclusão de Curso apresentado à Escola de Engenharia Mauá do Centro Universitário do Instituto Mauá de Tecnologia como requisito parcial para a obtenção do título de Engenheiro de Controle e Automação.

Orientador: Eduardo Lobo Lustosa Cabral, DSc

Área de concentração: Engenharia de Controle e Automação

São Caetano do Sul

2019

Brostoline, Bruno Marinotti

Inteligência artificial na bolsa de valores / Bruno Marinotti Brostoline, Igor Amaral Correa, Leon Cheroto Weissberg. — São Caetano do Sul: CEUN-IMT, 2019.

61 p.

Trabalho de Conclusão de Curso – Escola de Engenharia Mauá do Centro Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, SP, 2019.

Orientador(a): DSc. Eduardo Lobo Lustosa Cabral

1. Inteligência Artificial. 2. Bolsa de valores. 3. Aprendizado por reforço. 4. Day-trade. 5. Algotrading. I. Amaral, Igor Correa. II. Weissberg, Leon Cheroto. III. Instituto Mauá de Tecnologia. Escola de Engenharia. IV. Título.

BRUNO MARINOTTI BROSTOLINE

IGOR AMARAL CORREA

LEON CHEROTO WEISSBERG

# Inteligência artificial na bolsa de valores

Trabalho de Conclusão de Curso aprovado pela Escola de Engenharia Mauá do Centro Universitário do Instituto Mauá de Tecnologia como requisito parcial para a obtenção do título de Engenheiro de Controle e Automação.

Banca avaliadora:

Eduardo Lobo Lustosa Cabral, DSc.

Nome completo do professor(a) avaliador(a) 1 e título
Avaliador(a) ou Instituição, se externo

Nome completo do professor(a) avaliador(a) 2, se houver, e título
Avaliador(a) ou Instituição, se externo

São Caetano do Sul, data da apresentação de mês de ano.

*Dedicamos este trabalho para*

*nossa família, sem eles nossos sonhos jamais seriam realidade.*

# AGRADECIMENTOS

Ao Professor Dr. Eduardo Lobo Lustosa Cabral pela atenção, paciência e disponibilidade na orientação desta pesquisa e pela amizade que surgiu no decorrer do trabalho.

Ao Instituto Mauá de Tecnologia pela oportunidade de realização do curso de graduação e a todos os professores que nos lecionaram, pois estes serão sempre lembrados por nós.

Aos nossos amigos que sempre nos motivaram e contribuíram direta ou indiretamente no trabalho.

A toda nossa família, principalmente aos nossos pais e avós, que nunca nos deixaram de apoiar e nos amar, mesmo nos momentos mais difíceis e estressantes.

A Deus, que sempre nos conduziu para o melhor caminho.

*“Engenharia mais economia igual a poder. ”*

*Frase de um célebre professor de Economia, que fez nossa visão sobre o mundo mudar.*

# RESUMO

O estudo e aplicação de algoritmos autônomos no mercado financeiro já é uma realidade. Tal tecnologia tem sido profusamente discutida desde o crescimento das pesquisas em inteligência artificial. Em linha com essa ideia, este trabalho tem como objetivo analisar, implementar e discutir os efeitos de "robôs operadores" com aprendizado de máquina na bolsa de valores brasileira. O algoritmo foca em operações de curto prazo, que começam e terminam no mesmo dia, denominadas *day-trade*, no mini contrato de dólar futuro e utiliza a técnica *Deep Q-Learning* com o intuito de aprender as melhores decisões a se tomar (compra, venda ou neutro) para se ter o melhor lucro possível. Por um lado o *day-trade* se mostra uma estratégia arriscada e especulativa de se ganhar dinheiro na bolsa. Em contrapartida, a aplicação de um método robusto de aprendizado por reforço em uma ampla série histórica do ativo escolhido trouxe bons resultados simulados, que podem garantir um lucro positivo ao longo do tempo por meio de um sistema automatizado.



Palavras-chave: Inteligência artificial. Bolsa de valores. Aprendizado por reforço. *Day-trade*. *Deep Q-Learning*.

## *ABSTRACT*

*The research and implementation of autonomous algorithms in the financial markets already is a reality. Such technology has been widely discussed since the recent growth in artificial intelligence research. In the same way, this work aims to analyze, implement and discuss the effects of machine learning "trader robots" in the Brazilian futures exchange. The algorithm has its focus on day-trades on mini US Dollar Futures and uses Deep Q-Learning technique to learn the best decision-making process (buy, sell or neutral) that returns the highest profit. On one hand, day-trade proves to be a risked and speculative strategy of earning money in the exchange. On the other, the application a robust Reinforcement Learning method to a broad historical series of the chosen asset has yielded great simulated results, which may ensure a positive profit over time through an automated system.*



*Keywords: Artificial intelligence. Exchange. Reinforcement Learning. Day-trade. Deep Q-Learning.*

## LISTA DE ABREVIATURAS E SIGLAS

B3 - B3 - Brasil Bolsa Balcão S.A.

Bacen - Banco Central do Brasil

BCB - Banco Central do Brasil

CVM - Comissão de Valores Mobiliários

DQL - *Deep Q-Learning*

DQN - *Deep Q-Networks*

DRL - *Deep Reinforcement Learning*

DT - Day-trade

FGV - Fundação Getúlio Vargas

HFTs - *High-frequency traders*

IA - Inteligência artificial

MSE - Mean squared error

ML - *Machine Learning*

QL - *Q-Learning*

RL - *Reinforcement Learning*

RNA - Redes neurais artificiais

## SUMÁRIO

# 1 INTRODUÇÃO

A modernização das bolsas de valores mundiais e o avanço tecnológico desencadeou o desenvolvimento e a implantação de sistemas automatizados para novos patamares. Atualmente estima-se que os algoritmos sejam responsáveis por quase metade das negociações na bolsa brasileira (FGV, 2019).

Embora sejam comumente mencionados como “robôs de investimento”, sua especialidade é no curto prazo, em operações que duram menos de um dia ou até segundos, sendo o termo “algoritmos especuladores” ou “algoritmos de *trading*” mais adequados a essa prática. Tais tecnologias cada vez mais avançam e se tornam complexas, principalmente com as recentes pesquisas em inteligência artificial e aprendizado de máquina, em que os próprios computadores igualam e até superam a capacidade humana de realizar inúmeras tarefas.

Do outro lado do mercado, encontram-se os *day-traders* ou especuladores, que assim como os algoritmos focam em negociações com a menor duração possível e lucros rápidos. Com a popularização de vídeos e transmissões ao vivo, e da maior facilidade de acesso a bolsa, o número de pessoas físicas aficionadas pelo *day-trade* (ou *day-trading*) aumentou consideravelmente, mais que triplicando entre 2017 e o início de 2019, de 8 mil para mais de 30 mil (B3, 2019).

No entanto, o risco de tais operações torna-se muito mais elevado dado que as variações dos preços no curto prazo são maiores. Segundo uma pesquisa encomendada pela Comissão de Valores Mobiliários à Fundação Getúlio Vargas sobre o uso *day-trade* como único meio de renda (CHAGUE, DE-LOSSO e GIOVANNETTI, 2019), constatou-se que mais de 90% dos especuladores analisados entre 2012 e 2017 tiveram prejuízo, sendo que apenas 0,8% obteve um lucro médio diário maior do que R$ 300,00. Conforme os pesquisadores concluem, “Nós apresentamos fortes evidências de que não faz sentido, ao menos econômico, tentar viver de *day-trading*. Os dados indicam que a chance de obter uma renda significativa é remota para as pessoas que persistem na atividade. Por outro lado, a chance de se obter prejuízo é muito elevada”.

Com este estudo em mente, pressupõe-se a ideia de que uma pessoa utilizando seu próprio dinheiro não consegue obter resultados positivos por muito tempo no *day-trade*. No entanto, levando em consideração os avanços contemporâneos da inteligência artificial mencionados inicialmente, esse trabalho demonstra que um algoritmo com técnicas de aprendizado de máquina pode obter resultados positivos e significativos em operações que começam e terminam no mesmo dia. Além disso proporciona um estudo prático sobre o aprendizado por reforço, um método relativamente recente, mas que tem atraído uma atenção especial no ramo científico, especialmente em um ambiente com comportamentos quase que estocásticos como a bolsa de valores.

No capítulo 2 é apresentada uma revisão de toda a literatura relacionada ao *day-trade* na bolsa brasileira, aos algoritmos de *trading* e às técnicas de inteligência artificial utilizadas nesse trabalho. No capítulo seguinte é explicado o conceito por trás do modelo escolhido e desenvolvido e no capítulo 4 são explicados os processos e a metodologia utilizada para sua elaboração e consequente implementação. No capítulo 5 discute-se os resultados e os próximos passos que a pesquisa pode ter. Por fim, são apresentadas as conclusões no capítulo derradeiro.

## 1.1 JUSTIFICATIVA

Há diversos incentivos para desenvolver um algoritmo que opere *day-trade* na bolsa. Um deles, obviamente, é a possibilidade de se obter lucros por um sistema confiável e automatizado.

O principal, no entanto, é demonstrar se a inteligência artificial supera um ser humano no mercado financeiro e tem um retorno positivo dado que, segundo pesquisa da Fundação Getúlio Vargas (CHAGUE, DE-LOSSO e GIOVANNETTI, 2019), no Brasil, apenas menos de 10% das pessoas que operam *day-trade* não tem prejuízo ao longo do tempo.

Outra motivação é o desenvolvimento de um algoritmo com aprendizado por reforço em uma rede neural, duas técnicas recentes e muito estudadas no meio acadêmico, embora ainda haja dificuldades na análise dos efeitos que cada parâmetro gera e consenso na melhor maneira de dispor os dados para o aprendizado. Além disso, também torna-se interessante a

implementação de tais métodos em um ambiente totalmente caótico, em que há fatores externos e aleatórios influenciando seu comportamento simultaneamente.

## 1.2 OBJETIVOS

O principal objetivo deste trabalho consiste no desenvolvimento de um algoritmo para especular na bolsa de valores por meio de redes neurais artificiais (RNAs) e o método de aprendizado por reforço. Para que o objetivo seja atingido, os seguintes objetivos específicos devem ser realizados:

a) desenvolver um algoritmo de aprendizado por reforço utilizando RNAs;
b) validar a implementação desenvolvida através de testes em uma base de dados dos negócios intradiários do mercado de dólar futuro da B3;
c) verificar o retorno obtido durante todo o período testado.

## 1.3 DEFINIÇÃO DO PROBLEMA

Desde sua criação até hoje, a bolsa de valores chama a atenção devido às oportunidades de investimentos ou de ganhar rapidamente uma quantidade expressiva de dinheiro. Este último caso, conhecido como especulação, visa operações no curto prazo a partir de análises imediatistas e fascina milhares de pessoas na busca por métodos que consigam prever o valor futuro dos títulos. Muitos especuladores profissionais, com experiência e resultados consistentes, optam pelo *day-trade*, operações que começam e terminam no mesmo dia, e pela análise dos preços recentes no mercado para obterem lucro. Com a evolução do processamento computacional e a inteligência artificial desbancando muitos humanos em várias tarefas, existe a discussão de até que ponto um algoritmo consegue superar pessoas no mercado financeiro. Este trabalho propõe comparar o resultado de ambos e apresentar se a revolução tecnológica conseguiu superar um profissional nos *day-trades*, por meio de algoritmos de aprendizado de máquina e redes neurais artificiais.

## 1.4 QUESTÃO CENTRAL DA PESQUISA

Este trabalho estuda e discute diferentes métodos de aprendizado de máquina por reforço e redes neurais para se obter o melhor desempenho possível em operações de curta duração na bolsa de valores.

## 1.5 PANORAMA ECONÔMICO

Esta monografia se insere em dois contextos econômicos, o tecnológico e o financeiro. O tecnológico pelos diversos conceitos computacionais e matemáticos aplicados. E o setor financeiro, pelo fato de ter motivado toda essa pesquisa em torno da bolsa de valores e do mercado financeiro.

### 1.5.1 MERCADO

Este trabalho é contemporâneo de alguns cenários ainda mais fomentadores para a sua realização. Em relação ao panorama financeiro, a bolsa de valores atualmente se mostra mais atraente do que nunca para investidores brasileiros em razão da mínima histórica da taxa básica de juros até então (5,0%). Tal conjuntura ocasiona uma maior retirada de recursos da renda fixa para a renda variável, pelo fato de ser necessário passar por maiores riscos para se obter a mesma rentabilidade real que anteriormente era esperada. Outro ponto é a bolsa de valores brasileira ter registrado, pela primeira vez, mais de um milhão de investidores pessoa física, o que demonstra a expansão e notoriedade dessa nos últimos anos.

Ao contexto tecnológico e computacional, citam-se as técnicas e algoritmos aplicados e estudados serem objeto de uma intensa atividade de pesquisa durante os últimos anos, tendo o aprendizado por reforço e as redes neurais artificiais ganhado maior destaque devido às evoluções no processamento dos computadores e servidores, e da disseminação de trabalhos científicos por meio da internet. Este último evento pôde ser notado diversas vezes pelos vários artigos, sobre *Reinforcement Learning*, que eram divulgados na web no decorrer da pesquisa.

### 1.5.2 OPORTUNIDADES

O trabalho pode despertar o interesse de instituições financeiras, como fundos de investimento e bancos, que queiram aplicar o algoritmo em seus ambientes de operações ou adaptá-lo para cumprir outras tarefas relacionadas ao mercado financeiro.

No entanto, uma maior atenção poderá vir das pessoas físicas, pelo fato da pesquisa ter um enfoque em mini contratos futuros, ativos que não precisam de um muito dinheiro para serem operados. Além disso, o algoritmo não necessita de um grande conhecimento em finanças e na bolsa para ser executado, sendo trivial a quem deseja utilizá-lo para fins pessoais.

# 2 REVISÃO DA LITERATURA

Este capítulo é dedicado à revisão bibliográfica do trabalho, em que se discute pesquisas e artigos sobre operações de *day-trade* na bolsa de valores brasileira e os seus riscos. Em seguida, aborda-se os trabalhos que explicam o uso e os tipos de algoritmos de *trading* nas bolsas de valores. Além disso decica-se uma seção à literatura que trata de técnicas de aprendizado por reforço utilizadas no desenvolvimento nesse trabalho.

## 2.1 DAY-TRADE NO BRASIL

O Banco Central do Brasil (BCB, 2006) e a bolsa de valores, a B3 S.A. (B3, 2019), consideram as operações de *day-trade* como operações de compra e venda de um mesmo ativo realizadas na mesma data de negociação, por um mesmo participante (instituições ou pessoas físicas) e em uma mesma conta.

Em relação às estratégias utilizadas no mercado, Abe (2014, p. 42) descreve:

> O *day-trade* é uma modalidade onde o operador abre e fecha operações no mesmo dia. Muitos *day-traders* costumam utilizar gráficos de 5 minutos como principal meio de análise para efetuar suas operações. Outros preferem gráficos de 10 ou 15 minutos, e há ainda aqueles que combinem referências temporais como 5 e 15 minutos ou 5 e 30 minutos.

Gomes (2018) estudou três estratégias utilizadas por operadores profissionais e iniciantes em *day-trade*, avaliando sua eficácia e comparando-as com negociações com foco no longo prazo. O autor classifica os métodos como sendo de análise técnica, consistindo em achar padrões gráficos e técnicos, o primeiro sendo uma análise mais visual e a segunda baseada em indicadores fundamentados na movimentação dos preços. O autor concluiu que as estratégias, durante um ano de teste, tiveram um resultado negativo e inferior ao *buy-and-hold*, um tipo de investimento em ações para o longo prazo. Além disso o autor verificou que a especulação estudada minimizou os retornos positivos e não aproveitou a então valorização dos ativos.

Spritzer e Tauhata (2017) também pesquisaram sobre o uso da análise técnica no *day-trade* para dez ações da bolsa e seu resultado ao longo de dez anos. Os autores reforçam que o

método se baseia em verificar padrões passados no presente e agir de acordo, sendo estes padrões uma explicação de como o comportamento humano influenciou o preço do ativo, no entanto também pontuam alguns argumentos contra essa visão, tais como a falta de racionalidade da análise, o comportamento caótico do mercado e a falta de correlação e casualidade entre os padrões que o método deduz. Eles terminam por afirmar que após os testes a estratégia obteve um retorno quatro vezes menor que a taxa básica de juros (SELIC) e uma taxa de acerto média de cerca de 50%, similar à uma escolha aleatória de decisões. Spritzer e Tauhata (2017) citam que um possível futuro trabalho poderia estar no estudo dos impactos da inteligência artificial na bolsa.

Como elucidado nos últimos trabalhos, o *day-trader* atua com um alto risco operacional, podendo obter prejuízos significativos. Do Amaral (2015) estuda o comportamento dos especuladores frente ao mercado. Além de demonstrar o impacto da mente humana durante o *day-trade*, também discute as técnicas que os profissionais utilizam para trabalhar em cima de melhorias e que uma parcela dos maus resultados advém da própria cabeça do *day-trader*. A conclusão do autor é de que o desempenho de um operador está diretamente relacionado com sua postura psicológica, e também que há a existência de um amadurecimento e aprendizado que o especulador precisa passar para atingir uma consistência operacional.

No entanto, uma pesquisa recente da Fundação Getúlio Vargas (CHAGUE, DE-LOSSO e GIOVANNETTI, 2019) em que os autores obtiveram acesso a um banco de dados da Comissão de Valores Mobiliários (CVM), melhor exemplifica estatisticamente a baixa eficácia da especulação. Os dados continham informações sobre 19.646 pessoas que começaram a operar *day-trade* em mini contratos futuros de índice iBovespa, ativos muito utilizados por especuladores, entre 2013 e 2015. Por meio de análises nos retornos dos participantes apontou-se que proporção de dias positivos decresce conforme o número de dias operados aumenta. Como exemplo, só 15,5% daqueles que negociaram durante 2 a 50 dias obtiveram lucro, enquanto que isso só foi alcançado por 3,0% daqueles que operaram acima de 300 dias. Os pesquisadores afirmam que os resultados demonstram que não existe um aprendizado nos operadores e que há padrão similar a uma roleta de cassino, em que o número de jogadores com êxito diminui com o decorrer das rodadas. Outros levantamentos do trabalho mostram que apenas 1,1% dos 3,0% positivos ganharam mais que um salário mínimo por dia e que se

os custos de operação forem levados em conta, essa taxa diminui ainda mais. No geral de todos os dias, mais de 90% dos analisados terminaram com prejuízo. A conclusão traz a ideia de que é ilusório operar *day-trade* como meio de ganhar dinheiro e de não existe um aprendizado por trás da prática, dado que seu resultado se assimila à um jogo de azar. Termina por apontar que o número de especuladores tem crescido no país, devido aos diversos conteúdos que estão na internet e às recentes taxas de corretagem zeradas para mini contratos por algumas corretoras.

## 2.2 ALGORITMOS DE TRADING

Segundo Schwager (1989), a criação e desenvolvimento dos *algotradings* é creditada a *Edward Seykota*, que em 1970 criou um sistema automatizado que operava de acordo com indicadores de médias de preços em gráficos, como médias móveis e figuras gráficas, e, conforme Hartle (1992) e Cox (1993), *Seykota* obteve lucros expressivos durante essa década. O desenvolvimento desse algoritmo de negociação só foi possível após o início da utilização de computadores para registrar as ordens de compra e venda no mercado, sendo a bolsa de valores de Nova York (*NYSE*) a primeira a implementar esse sistema no começo dos anos 70. Estabeleceu-se, logo no início, que o objetivo da negociação algorítmica na bolsa de valores seria executar ordens utilizando instruções automatizadas e pré-programadas. Na década seguinte, os *algotradings* já estavam popularizados nas bolsas americanas, sendo responsáveis por uma parcela significativa das operações realizadas. Esta rápida disseminação relaciona-se ao aclamado trabalho de Fischer Black, Myron Scholes e Robert Merton na precificação de derivativos pelo modelo de Black-Scholes, o qual era um cálculo avançado que exigia rapidez que só um computador teria. Soma-se tal contexto à popularização da internet e a crescente disponibilidade comercial dos microprocessadores e computadores pessoais, que definiriam uma década de extremo avanço na eletrônica e nas linguagens de programação, cruciais para o desenvolvimento de algoritmos ainda mais eficientes e rápidos.

Conforme Zhang (2010) em meados da década de 90, os operadores começaram a imigrar do famoso pregão viva-voz em direção à negociação eletrônica. O aumento da liquidez nos mercados e os avanços tecnológicos dessa época criou as condições ideais para o desenvolvimento e propagação de algoritmos extremamente rápidos e com capacidade de

operar milhares de vezes no dia, os chamados *High Frequency Traders* (HFTs ou operadores de alta frequência). Os HFTs buscam lucros em pequenas oscilações de preço e em diversas operações durante o dia. Ainda segundo Zhang (2010) a negociação algorítmica em alta frequência foi responsável por volta de 78% de todo o volume negociado em 2009 na bolsa americana, em relação a quase zero em 1995.

Da Costa (2018) estudou a utilização e regulação jurídica dos HFTs na bolsa brasileira. O autor explica que estes algoritmos são um tipo específicos de *algotradings*, e são assim considerados quando têm a capacidade de processar os últimos dados e inserir ordens em um tempo inferior a 100 milissegundos. Da Costa (2018) pontua algumas vantagens da negociação algorítmica, tais como a rapidez, diminuição de erros humanos e do desgaste emocional, que como citado na última seção, interfere muito no desempenho do *day-trader.* Também cita que são algoritmos normalmente utilizados por mesas proprietárias e de operações de instituições financeiras, sendo alguns investidores institucionais notoriamente conhecidos no cenário internacional por atuarem como HFTs, executando operações como formadores de mercado e liquidez (*market makers*) ou negociando recursos próprios. Para enfatizar a atuação maçante de tais participantes internacionais no mercado brasileiro, aponta que estes são responsáveis por 49,2% do volume total negociado no mercado acionário e 38,4% no mercado de derivativos. O autor aborda algumas características associadas ao HFTs, como a alta sensibilidade na velocidade de comunicação com os sistemas de negociação (buscam uma latência mínima entre ambos), serem usualmente *day-traders* e o uso de tecnologias sofisticadas para implementar estratégias avançadas e maximizar os ganhos.

Sobre essas estratégias, da Costa (2018) ainda frisa que novas técnicas são desenvolvidas e identificadas ao longo do tempo, mas que geralmente são separadas entre estratégias passivas e direcionais, e a arbitragem. A primeira é relacionada aos formadores de mercado e consiste em prover liquidez ao mercado por meio de ordens limitadas no ativo alvo, isto é, o algoritmo não executa uma negociação imediata com base nas ordens que já estão no mercado, ele disponibiliza outras ordens para ter um maior leque de ofertas à disposição dos outros investidores. A direcional trata-se de uma negociação baseada em notícias, fundamentos das companhias e análises gráficas ou das ofertas que estão no mercado. Essa última abordagem talvez seja a mais veloz de todas dado que o algoritmo fica analisando as

mudanças do livro de ofertas do mercado a cada instante. Os algoritmos arbitradores atuam explorando discrepâncias ou ineficiências na precificação de ativos, por exemplo uma diferença de preços entre o contrato futuro de dólar e o seu respectivo mini contrato. Em todas estas estratégias, no entanto, o autor cita o uso característico de métodos quantitativos, técnica baseada em cálculos estatísticos para a tomada de decisões.

A pesquisa da FGV (CHAQUE, DE-LOSSO e GIOVANNETTI, 2019), já mencionada anteriormente, também cita o uso de algoritmos e HFTs no mercado atual, e comenta a difícil competição de especuladores pessoas físicas contra os negociadores algorítmicos no *day-trade*. O estudo argumenta que os grandes investimentos que as instituições tem feito para ampliar os lucros obtidos por seus algoritmos, podem influenciar ainda mais negativamente o desempenho dos *day-traders*. Um levantamento interessante da pesquisa mostra o crescimento que os HFTs tiveram na década atual: em 2012 dominavam 11,6% de todos os negócios dos mini contratos futuro (dólar e índice iBovespa), enquanto que em 2017 foram responsáveis por 41,9% de todas as operações realizadas nesse mesmo mercado. Os autores concluem analisando que o crescimento do uso dos algoritmos coincidiu com um aumento no prejuízo de investidores individuais, indicando uma correlação negativa entre o uso de *algotradings* e o lucro de *day-traders* pessoas físicas.

Zhang (2010) lembra também que, embora muito da negociação algorítmica se baseia em HFTs, alguns outros algoritmos merecem Fundos quantitativos, fundos mútuos e outros investidores institucionais utilizam *algotradings* para dividir ordens grandes em pequenas por meio de várias operações, afim de mitigar seu risco e diminuir o impacto que grandes lotes podem gerar nos preços do mercado.

## 2.3 TÉCNICAS DE APRENDIZADO POR REFORÇO

Segundo Sutton e Barto (2018), o aprendizado por reforço ("*Reinforcement Learning*", em inglês) consiste em uma técnica de aprendizado de máquina, na qual o algoritmo deve aprender quais ações tomar em um ambiente para maximizar um resultado acumulativo. O algoritmo não recebe quais ações deve tomar, devendo descobrir qual caminho de decisões leva à maior recompensa. Os autores explicam que o *Reinforcement Learning* é diferente dos

outros métodos de *Machine Learning*, sua diferença ao aprendizado supervisionado é que não há os dados de resposta desejados do sistema, mas também se opõe ao não-supervisionado pelo fato do RL não ser um paradigma em que se procure achar um modelo do comportamento do ambiente, sendo seu objetivo obter um conjunto de ações sequenciais que o levam a ter o melhor desempenho em uma certa tarefa. Sutton e Barto (2018) afirmam que o aprendizado por reforço é parte de um estudo de décadas na integração da inteligência artificial com estatística, otimização de processos e outros temas da matemática, além de haver aspectos da psicologia e neurociência em seus fundamentos. Os estudiosos comentam que o método tem como plano de fundo os processos de decisão de Markov, cuja função é definir a interação entre um agente aprendendo e os possíveis estados, ações e recompensas que este convive em um ambiente, características que fazem o *Reinforcement Learning* ser uma técnica de causa e efeito. Os estudiosos citam diversos exemplos, mas todos seguindo a mesma linha de raciocínio: o algoritmo (agente) observa o estado que ele está e toma uma decisão, em seguida a ação que esse realizou gera uma recompensa e o leva para outro estado, assim, ambos os parâmetros são retornados para o agente, que verifica se sua escolha foi positiva ou não. Por meio desse feedback que a recompensa gera para o algoritmo, ele é capaz de aprender, no decorrer do treino, as ações a se tomar em cada estado que se encontra para ter a maior recompensa acumulativa possível.

François-Lavet, Islam, et al. (2018) apontam que o objetivo do RL é encontrar a política de decisões que leva ao melhor resultado final. Os autores dividem os métodos existentes de *Reinforcement Learning* em três classes: baseados em valor, em um modelo ou em gradiente de política. Além disso também cita que cada técnica pode ser *offline*, em que primeiro todos os dados do ambiente são adquiridos e então ocorre o aprendizado por meio de um treino, ou *online*, na qual as informações são disponibilizadas sequencialmente e o algoritmo é treinado a cada passo. Em relação aos métodos baseados em valor, é explicado que a melhor política é definida a partir de um valor obtido por meio de uma função pré-definida, como o *Q-Learning* (QL) em que uma função chamada Q prevê a recompensa que cada ação pode ter em um certo estado (WATKINS, 1989). Já os algoritmos com base em gradiente de política (*policy gradient*, em inglês) otimizam a recompensa acumulada mediante variações dos gradientes estocásticos ascendentes e atualização dos parâmetros do sistema de acordo com um estimador de pontuação dos gradientes. Sobre os sistemas baseados em um modelo, os

autores discutem que essa abordagem recorre a um método baseado em um valor ou a uma política, mas ao contrário dos outros dois necessita de um modelo (já conhecido ou aprendido) que explique a dinâmica e as recompensas do ambiente.

François-Lavet, Islam, et al. (2018) se aprofundam principalmente na aplicação de redes neurais no aprendizado por reforço. A utilização de múltiplas camadas neurais de processamento, segundo os autores, mitiga instabilidades que alguns métodos de *Reinforcement Learning* podem apresentar e possibilita cobrir uma maior gama de dados do espaço estado-ação, facilitando o aprendizado em situações em que as informações são contínuas. O livro explica toda a recente evolução do *Deep Q-Learning* para o *Double Deep Q-Learning* e o *Dueling Q-Networks*, em que todas estas são melhorias do desempenho do QL a partir de redes neurais.

Um dos pioneiros nas pesquisas sobre aprendizado por reforço aplicado em operações na bolsa de valores é John Moody. Em Moody e Wu (1997) é apresentado resultados em que comprovam as vantagens de métodos de *Reinforcement Learning* aos de *Supervised Learning*, quando empregados em sistemas automatizados de *trading*. Em Moody e Saffell (2001), é comparada a performance entre o *Q-Learning* e um outro algoritmo de RL desenvolvido pelos autores (MOODY, WU, et al., 1998), baseado em *policy gradient* com uma rede recorrente, simulados em dados mensais do índice de ações S&P 500 e trimestrais de títulos públicos da dívida americana durante 25 anos, em que o algoritmo poderia ficar comprado ou vendido no ativo. Os autores obtiveram em ambos os métodos um retorno maior que o índice S&P, demonstrando que existe uma estrutura previsível nos preços das ações da bolsa americana, e também evitaram prejuízos durante as grandes quedas que ocorreram no período analisado. Outra conclusão que chegaram foi que o modelo desenvolvido foi superior ao QL, possivelmente por este ter problemas de processamento com um espaço estado-ação contínuo, uma adversidade que, segundo François-Lavet, Islam, et al. (2018), pode ser resolvida com o uso de uma rede neural em conjunto.

Deng, Bao, et al. (2016), em linha com Moody e Saffell (2001), implementam o mesmo algoritmo de aprendizado por reforço, no entanto, utilizando agora *Deep Learning* e lógica *Fuzzy*, para operar índices de bolsas chinesas e commodities com dados dos preços a cada 1

minuto. Embora, os autores não afirmem, possivelmente seu algoritmo realiza algumas operações de *day-trade*, dado que há um baixo período de tempo entre os dados intradiários. Deng, Bao, et al. (2016) comenta que um dos motivos que os levaram à escolha de um sistema baseado em *policy gradient* foi a contraindicação de métodos baseado em valor para aprendizagem *online* (DENG, KONG, et al., 2015). O estudo comenta que por ser um algoritmo com técnicas robustas, exige um forte desempenho computacional. Por fim demonstram diversos resultados positivos que o *algotrading* conseguiu sendo superior a outros métodos de RL testados, como a rede recorrente de Moody e Saffell (2001), e conclui que o uso de *Reinforcement Learning* para *trading* se mostrou promissor, mas que um dos desafios da aplicação é escolher um período dos dados que seja ideal para treinar o algoritmo.

# 3 MODELO DESENVOLVIDO

Nesse capítulo são explicados o conceito da junção de inteligência artificial com redes neurais, o funcionamento de RNAs e a teoria do *Reinforcement Learning (RL)*. Estes temas ajudam a esclarecer os motivos que levaram à escolha do modelo *Deep Q-Learning (DQL ou DQN*, referente a *Deep Q-Networks*) ser utilizado nesse trabalho. Finalmente é apresentada a arquitetura por trás do modelo e seu funcionamento.

## 3.1 INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial, ou IA, é a transformação de uma tarefa humana em códigos que o computador consegue interpretar, dessa forma reproduzindo, e muitas vezes até melhorando, essa tarefa.

A ideia de computadores inteligentes vem antes mesmo da construção do primeiro computador, onde um dos pais da ciência da computação, Alan Turing, criou o Teste de Turing em 1950, para dizer se um computador realmente possui inteligência ou não. O teste é realizado colocando a máquina para conversar com um ser humano em linguagem de texto, separados um do outro. Se a pessoa não conseguir distinguir que está conversando com uma máquina ou um ser humano, conclui-se que a máquina tem mesmo uma inteligência. Esse teste, com algumas controvérsias, foi pela primeira vez vencido em 2014 (GRIFFIN, 2014).

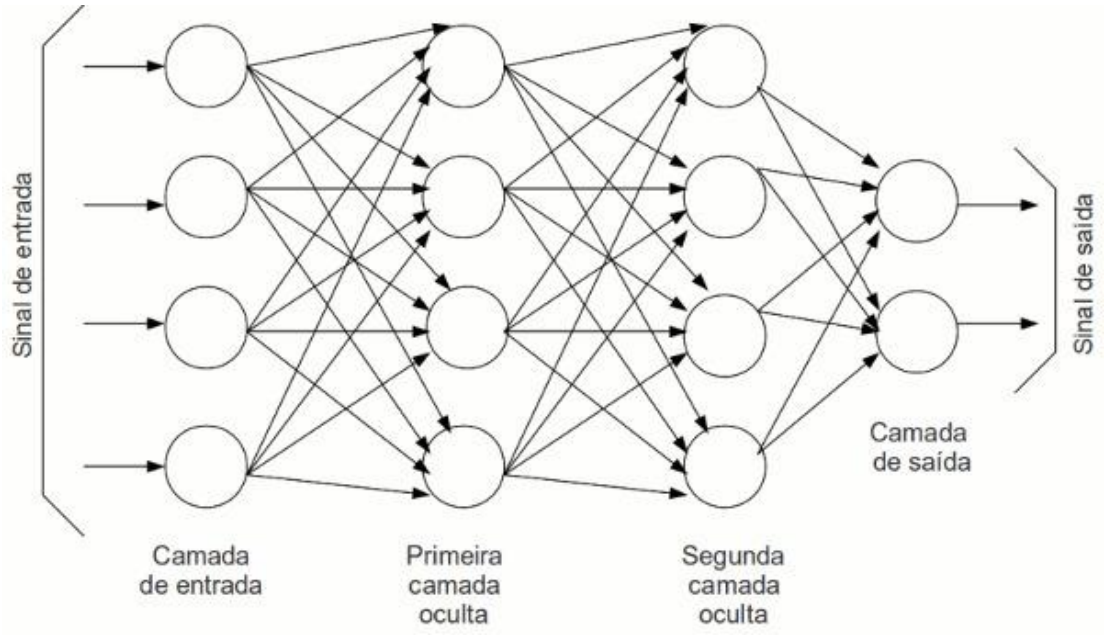
## 3.2 REDES NEURAIS ARTIFICIAIS

Segundo Haykin (2007), uma rede neural artificial pode ser entendida como uma “máquina” com diversas unidades de processamento, chamadas de neurônios ou unidades, projetada para modelar a maneira como o cérebro humano realiza tarefas e aprende. Os neurônios são conectados entre si com a função de armazenarem conhecimento a partir de dados ou de interações com um ambiente. A estrutura de uma rede pode conter uma ou várias camadas de unidades, podendo haver, no último caso, “camadas escondidas” (ou “ocultas”) que ligam a camada de entrada com a de saída, intermediando-as. Quanto à conexão e ao fluxo de informações, uma RNA multicamadas pode ser um *perceptron*, onde a propagação dos sinais é realizada para frente sem realimentação dos neurônios de uma camada anterior, ou

recorrente em que há retroalimentação dos dados para os próprios neurônios geradores dos sinais e aos demais (GAMBOGI, 2013).

Cada conexão entre os neurônios têm um valor numérico, conhecido como “peso”. Quando a rede está em operação, a primeira camada de nós recebe os dados de entrada e esses valores são multiplicados pelo peso de cada interligação. Cada neurônio da camada seguinte recebe a soma dessa multiplicação produzida por cada conexão ligada a ele, faz uma transformação algébrica nesse valor com o objetivo de fazer o processamento dessa informação, essa transformação se dá por meio de uma “função de ativação”, e envia o resultado para a próxima camada. Este processo é repetido por todas as camadas ocultas até que se alcance os neurônios da camada de saída, que, após mais uma transformação de ativação, geram os dados de saída. O número de unidades na saída é equivalente à quantidade de respostas desejadas (HAYKIN, 2007). A Figura 1 ilustra uma rede neural artificial de quatro camadas (três escondidas e 1 de saída).

Figura 1 — ilustração de uma rede neural artificial *feedforward*.




No início do treino, quando ainda não existe um aprendizado, os pesos são inicializados com números aleatórios pequenos. Com o decorrer do treino, esses valores são recursivamente alterados na intenção de fazer o modelo convergir à performance desejada. Portanto, o

aprendizado consiste na gradativa modificação dos pesos (HAYKIN, 2007). Há três principais paradigmas de aprendizado que são discutidos mais à frente.

## 3.3 APRENDIZADO DE MÁQUINA

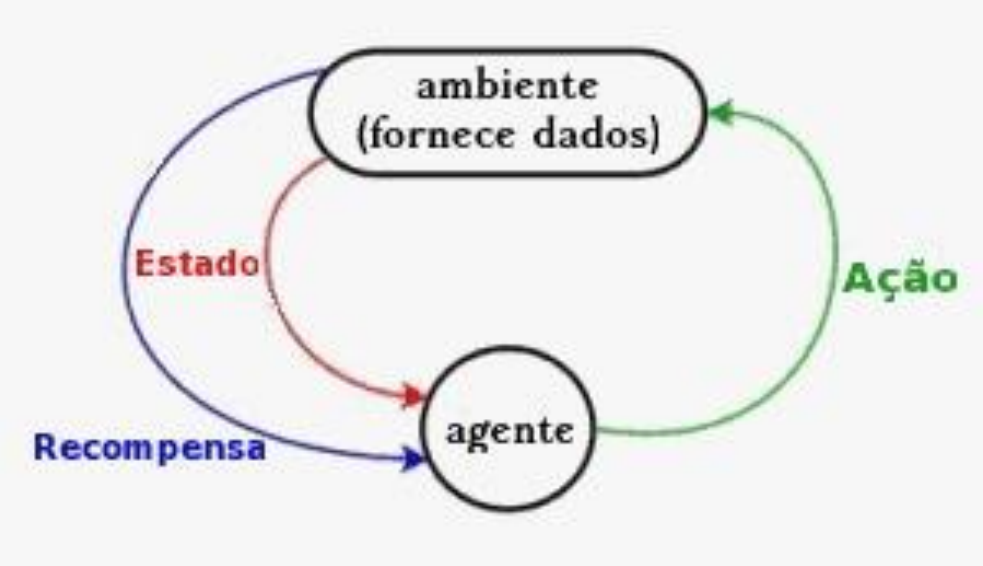
O aprendizado de máquina (em inglês: M*achine Learning*) é um subcampo da IA, que obtém sua inteligência apenas com o reconhecimento de padrões em dados. Foi definida em 1959 por Arthur Samuel como "o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados" (SAMUEL, 1959).

Essa grande área de estudo é subdividida em três outros ramos: o aprendizado supervisionado ("*Supervised Learning*"), aprendizado não supervisionado *("Unsupervised Learning")* e aprendizado por reforço *("Reinforcement Learning").* A primeira requer dados de entrada e de saídas desejadas para o treinamento, tentando assim classificar ou prever resultados ainda sem um modelo conhecido. No segundo caso só há as informações da entrada, as saídas são desconhecidas, assim, tenta-se descobrir padrões nos dados para correlacioná-los e separá-los em grupos. O terceiro método é totalmente diferente dos outros dois: a ideia principal é de um algoritmo, conhecido como agente, inserido em um ambiente, que por meio de observações escolhe uma decisão a ser tomada, tendo esta uma recompensa retornada. O desafio do agente é descobrir padrões que consigam maximizar a recompensa a partir das ações escolhidas. Para tal, são necessárias informações sobre os possíveis estados que o agente ocupa no ambiente, as ações disponíveis em cada estado e a recompensa que cada ação gera, podendo o algoritmo melhorar estratégias passadas e explorar novas.

Segundo Sutton e Barto (2018), o *Reinforcement Learning* (RL) é inspirado por uma visão biológica no controle de processos estocásticos. Operar no mercado financeiro através de algoritmos inteligentes é considerado um problema de tomada de decisões em um espaço dinâmico e ao mesmo tempo determinístico e estocástico. Comparado a outros tipos de aprendizado, essa situação requer que o agente explore um ambiente desconhecido por si mesmo e, ao mesmo tempo, faça as escolhas corretas. Especular no mercado futuro pode ser otimizado com a seleção de uma política de decisões que gere o máximo de ganhos possíveis, o que é a finalidade do método. Casos similares ocasionaram a aplicação do aprendizado por

reforço em diversas tarefas tais como o sistema de navegação de Robôs, jogos de videogame e até controle de helicópteros, em que o algoritmo superou a performance de pessoas consideradas qualificadas na área, como cita Deng, Bao, et al. (2017). Para tanto, escolheu-se o RL como aprendizado dos algoritmos. Pela Figura 2 pode-se entender melhor o funcionamento do método, como mencionado, o algoritmo (agente) executa uma ação no ambiente e este retorna para o agente o estado em que se encontra e a recompensa que a ação gerou.

Figura 2 — lógica do funcionamento do *Reinforcement Learning*.




Os primórdios do aprendizado por reforço são relacionados à formulação da equação de Bellman e aos estudos dos Processos de Decisão de Markov, em meados da década de 1950. Os primeiros temas se contextualizam na simplificação de uma complicada escolha em uma sequência de decisões a partir de diversas funções recursivas. Tal método é conhecido como programação dinâmica. O aprendizado por reforço só seria teorizado na década de 1980, quando se associou a programação dinâmica à aprendizagem computacional (SUTTON e BARTO, 2018). Desde então, diversos trabalhos acadêmicos começaram a ser desenvolvidos, os quais resultaram na criação do método *Q-Learning*.

## 3.4 Q-LEARNING

O *Q-Learning* (QL) é um método que utiliza uma tabela, conhecida como *Q-Table*, com os valores das recompensas esperadas que cada ação gera em cada estado. Esses valores são

calculados pela função *Q* que é uma otimização da equação de Bellman e seu uso recursivo. A finalidade dessa tabela é guiar o agente para a melhor ação em cada estado. O QL, ao contrário da programação dinâmica, não necessita de um conhecimento prévio de um modelo que explique os dados, para poder maximizar a recompensa. Outro aspecto importante é que, além de simples de ser implementado, sua performance supera a de outros algoritmos de RL em aplicações que ocorrem mudanças no comportamento do ambiente e na política de decisões, que o algoritmo toma, possa se alterar (WATKINS, 1989).

A desvantagem da *Q-Table* é que, para problemas com um número contínuo de estados, o esforço computacional e a ineficácia do modelo são maiores, especialmente em uma aplicação na bolsa de valores, aonde há infinitas possibilidades. Para solucionar tal empecilho propõe-se também o seu uso aliado a uma rede neural profunda, uma técnica conhecida como *Deep Q-Learning*.

## 3.5 DEEP Q-LEARNING

O *Deep Learning* é usualmente utilizado quando se tem um grande número de dados para analisar, como em reconhecimento de imagens e áudios, e se baseia na utilização de redes neurais multicamadas para melhorar o processamento. Assim, todo algoritmo *Deep* consiste no uso de RNAs.

O *Q-Learning* além de ser simples, é uma técnica poderosa para decidir como agir em cada situação em que o agente (algoritmo interagindo com o ambiente) se encontra. No entanto, como citado, um espaço de estados infinito torna sua aplicação inviável. Para tanto o *Deep Q-Learning,* abreviado como DQL ou DQN (*Deep Q-Network*), substitui a *Q-Table* por uma aproximação da função *Q* realizada pela rede neural, dessa forma, conseguindo generalizar esses ambientes contínuos onde o agente pode se encontrar.

## 3.6 CRITÉRIO DE ESCOLHA DO ALGORITMO

A aplicação do *Deep Q-Learning* se apresenta como uma técnica de tomar decisões rapidamente e poder trabalhar em um espaço contínuo, em que podem existir infinitos estados e ações. Além disso, o resultado do DQN nos treinos é relacionado com a quantidade

de dados disponíveis, assim, quanto mais dados, melhor será seu desempenho, sendo os casos onde se tem muita informação histórica algo favorável ao algoritmo, ao invés de se tornar um impasse como no *Q-Learning*.

Ainda que Deng, Bao, et al. (2016) e Moody e Staffel (2001) citem que o QL não seja uma técnica ideal para *trading*, devido o limitado espaço estado-ação que consegue processar e a instabilidade em treinos *online*, o uso de redes neurais multicamadas contribui para um melhor desempenho quando há dados contínuos e uma melhor estabilidade no sistema (FRANÇOIS-LAVET, ISLAM, et al., 2018).

Ademais, o algoritmo também é muito aplicado em jogos pelo fato de aprender tomando decisões erradas e corretas, um método interessante a ser implementado nas especulações de preços na bolsa de valores, já que muito do aprendizado humano no *day-trade* é obtido após muita prática. Portanto, este trabalho desenvolve e implementa uma lógica de especulação por meio do *Deep Q-Learning*.

# 4 METODOLOGIA

Neste capítulo explica-se o funcionamento do mercado e do *day-trade* no Brasil. Em seguida discute-se os dados e as ferramentas que são utilizados no trabalho. As três últimas seções são dedicadas a explanar como o algoritmo é implementado nessa pesquisa.

## 4.1 FUNCIONAMENTO DO MERCADO

Nesta seção é descrita a estrutura do sistema financeiro brasileiro, os princípios do mercado de capitais e da bolsa de valores.

### 4.1.1 SISTEMA FINANCEIRO

Um sistema financeiro é fundamental em qualquer sociedade econômica contemporânea. Define-se como o conjunto de instituições, produtos e instrumentos que viabilizam a transferência de recursos e ativos financeiros entre poupadores e tomadores. Sua importância para o funcionamento e crescimento da economia de uma nação pode ser entendida pela captação de recursos por agentes tomadores e investimento para agentes poupadores.

O sistema financeiro brasileiro é denominado Sistema Financeiro Nacional e é segmentado em 4 grandes áreas (mercados):

a) mercado monetário: resume-se em transferências de recursos, geralmente com prazo de um dia, realizadas somente por instituições financeiras e o Banco Central do Brasil (BC ou Bacen), para garantir a liquidez da economia;

b) mercado de crédito: atua nesse mercado instituições que intermediam empréstimos para o consumo ou capital de giro para empresas;

c) mercado de câmbio: negocia-se a troca de moedas estrangeiras pela nacional;

d) mercado de capitais: tem como propósito a canalização de recursos para empresas, instituições financeiras e pessoas, a partir de operações de compra e venda de valores

mobiliários com investidores e intermediários. É constituído pelas bolsas de valores, corretoras e outras instituições financeiras (EDUCACIONAL BM&FBOVESPA, 2019).

### 4.1.2 MERCADO DE CAPITAIS E A BOLSA DE VALORES

Segundo a Comissão de Valores Mobiliários (CVM, 2014), autarquia responsável pelas normas e fiscalização no mercado de capitais, a relação que se estabelece neste mercado é que os investidores, ao emprestarem seus recursos diretamente para empresas ou instituições financeiras, adquirem títulos chamados de valores mobiliários.

Um valor mobiliário, também chamado de papel no mercado financeiro, é definido pela lei brasileira como qualquer título ou contrato de investimento que represente o direito de participação, parceria ou remuneração, cujos rendimentos e condições estabelecidas advém de quem captou o recurso. Pode-se citar ações, títulos de dívida pública e debêntures como exemplos de valores mobiliários, comumente referidos como ativos financeiros.

O local onde ocorre as negociações no mercado de capitais é chamado de bolsa de valores, que consiste em um mercado organizado onde registram-se as operações de compra e venda de diversos tipos desses papéis. Atualmente no Brasil a única bolsa existente é a B3, sediada em São Paulo.

As primeiras bolsas com características modernas surgiram em meados século XV, durante a expansão comercial na Europa. A palavra 'bolsa' ganhou seu sentido comercial e financeiro em 1487, quando mercadores e comerciantes passaram a se reunir na casa da família Van der Burse (cujo brasão continha o desenho de três bolsas) em Bruges, na Bélgica, a fim de realizar seus negócios de compra e venda de moedas, letras de câmbio e metais preciosos.

Dentro da bolsa de valores há a segmentação de mercados com diferentes características de valores mobiliários, como ações, títulos de renda fixa e contratos derivativos.

Os contratos derivativos, como o nome sugere, são títulos que derivam do preço de outros ativos, podendo ser financeiros, como ações e moedas estrangeiras ou materialistas, como boi, milho e metais, e estabelecem contratualmente o pagamento futuro de uma certa quantia

baseada no valor presente do ativo de referência ou a entrega deste no preço e data determinados no acordo. A criação de tais papéis foi motivada pela necessidade de produtores ou comerciantes se protegerem contra riscos de variações bruscas nos preços do mercado, como em períodos de escassez ou desvalorização da moeda nacional. Existem três ramos dos derivativos:

a) opções, cujo comprador do título detém o direito, e não a obrigação, de receber, para opções de compra, ou entregar, para opções de venda, o ativo de referência após certo tempo em um valor determinado;

b) swaps, contratos que se negocia a troca (swap, em inglês) do índice de rentabilidade entre dois ativos;

c) títulos a termo, no qual há a obrigação de se pagar ou receber o papel futuramente na quantia acertada entre as partes.

Os títulos a termo ainda podem ser separados em contratos de termo e contratos futuros. Ambos os títulos seguem o fundamento apresentado acima, a divergência, porém, está na forma em que a data de vencimento do contrato é apresentada. Nos contratos futuros os compromissos são ajustados diariamente às expectativas do mercado frente ao preço futuro do bem, denominado ajuste diário, um meio de garantir que as partes honrarão o acordo. Os contratos futuros são os mais negociados e populares, baseando o chamado mercado futuro (EDUCACIONAL BM&FBOVESPA, 2019).

### 4.1.3 MERCADO FUTURO

Segundo a B3, o mercado futuro deve ser entendido como uma evolução do mercado a termo devido aos ajustes diários e pela melhor definição contratual da data de vencimento, tornando-os mais negociados em bolsa.

Atualmente são negociados diversos desses títulos:

a) contratos agropecuários: têm como ativo-objeto commodities agrícolas, tais como, boi gordo, milho, etanol ou açúcar;

b) contratos de energia e climáticos: têm como objeto de negociação energia elétrica, gás natural, créditos de carbono e outros;

c) contratos de ações: tem como referência ações de empresas;

d) contratos financeiros: têm seu valor de mercado referenciado em alguma taxa ou índice financeiro, como taxas de câmbio, juros e inflação, ou índices de ações e outros.

Os contratos futuros podem ser negociados com a finalidade de se proteger contra riscos de flutuações nos preços do mercado (*hedge*), aproveitar discrepâncias na formação de preços dos diversos ativos, mercadorias e vencimentos (arbitragem), ou operar na tendência de preços no mercado (especulação).

A especulação tem como propósito básico obter lucro. Diferentemente dos *hedgers*, os especuladores não têm nenhuma negociação no mercado físico que necessite de proteção, e diferente dos arbitradores, assumem riscos quando colocam seu capital no mercado em busca de lucro. Tal atividade ganhou impulso com a popularização de plataformas de negociações eletrônicas que facilitaram o contato e o processo de se negociar ativos.

Os futuros agrícolas, de energia e climáticos são, na maioria das vezes, operados por *hedgers*, geralmente instituições e produtores rurais buscando se proteger contra possíveis flutuações nos preços do mercado em questão. Os futuros de ações e financeiros, são operados pelas três categorias de operadores, no entanto, a maioria das negociações são realizadas por especuladores (EDUCACIONAL BM&FBOVESPA, 2019).

Os contratos referenciados em ações, começaram a ser negociados em meados de 2019, possuem diversos papéis, dado o grande leque de empresas no mercado de ações e são primeiramente focados no desempenho da companhia. Tais motivos explicam o fato dos títulos financeiros terem o maior volume de operações, pois se relacionam com a economia

nacional e externa, um cenário muito maior do que empresas e matérias primas, e portanto atraem um maior número de participantes do mercado.

### 4.1.4 *DAY-TRADE*

Conforme mencionado, um especulador sempre procura obter ganhos no mercado financeiro. No entanto, tal prática não deve ser confundida com a de um investidor. Embora ambos sejam rotulados da mesma maneira diversas vezes, o investidor aplica seu dinheiro em um certo ativo visando o longo prazo, acima de 1 ano, levantando suas conclusões em análises e aspectos da empresa, governo e cenário macroeconômico que influenciam seu investimento, tais como saúde financeira, número de vendas e projetos futuros de uma companhia, conceitos que normalmente levam um tempo para amadurecer e trazer os resultados esperados. Já o especulador foca nos preços recentes do título no mercado, através da análise de gráficos e dos negócios realizados no dia, e notícias com efeito imediatista, para 'especular' um valor futuro do ativo, para poder comprar o papel em um preço considerado baixo e vender em um valor alto, obtendo assim lucro. Tais operações são, no geral, para o curto e médio prazo, como alguns meses, dias e até mesmo minutos e segundos, devido tais análises apresentarem sentido durante um tempo reduzido.

A especulação que dura mais de um dia deixa o operador exposto às notícias e eventos econômicos que podem ocorrer durante o período em que não está se negociando, como acontecimentos negativos para a economia durante a noite, fora do horário de negociação da bolsa, fazendo no dia seguinte o papel abrir em queda, causando grandes prejuízos à negociação. Logo, muitos especuladores optam por operações que começam e terminam dentro do horário de negociação dos ativos na bolsa, essa prática é conhecida como *day-trade*.

### 4.1.5 CRITÉRIO DE ESCOLHA DO ATIVO

O *day-trade* se baseia na ideia de quanto menos durar uma operação, menos exposição e risco o especulador terá, logo negócios que duram minutos ou segundos são mais comuns. Para conseguir comprar e vender nos patamares que deseja, o mais indicado é negociar um papel que tenha liquidez, isto é, tenha sempre vários outros participantes do mercado pretendendo operar nos mesmos níveis. Um mercado com maior volume torna-se mais confiável e ocasiona

uma movimentação mais rápida dos preços, que é interessante para o especulador. O mercado de ações tem alguns ativos com alto número de negociações, como Petrobrás e Vale S.A., no entanto a variação é de centavos e a especulação deve ser feita com uma grande quantidade para obter ganhos expressivos, podendo não haver interessados suficientes em comprar ou vender uma grande quantia de papeis para o especulador no preço desejado. O mercado futuro, porém, se apresenta como uma solução a tal empecilho, pois apresenta contratos com variações de 1 a 5 reais por quantidade e com alta liquidez. Portanto, a maioria dos especuladores que praticam o *day-trade* optam por contratos futuros.

A partir 2006, a BM&F (atual B3) com intuito de aumentar as operações no mercado futuro e de inserir mais pessoas físicas nas negociações criou os "minicontratos" de dólar futuro (WDO) e índice iBovespa Futuro (WIN), que têm as mesmas características que os contratos padrão ("cheios"), mas são 20% do preço deste, o que dá mais flexibilidade a um especulador que não tem uma grande quantia financeira para operar contratos cheios. Os ativos que basearam esses dois tipos de minicontratos são o dólar comercial futuro (DOL), cujo valor é referenciado na moeda de mesmo nome, e o índice iBovespa futuro (IND), mais conhecido como índice futuro, que tem seu preço derivado do índice iBovespa, um indicador do mercado de ações. Atualmente os minicontratos apresentam uma boa liquidez, movimentam um volume financeiro diário maior que a maioria dos outros ativos, cerca de 20 bilhões de reais na média, segundo dados da B3, e um baixo custo para operar *day-trade*, o que faz a maioria dos especuladores optarem por estes ativos. Outro aspecto importante desses papeis é a possibilidade de vender o título sem tê-lo e depois comprá-lo, isso gera uma grande flexibilização no operacional do *day-trader*, viabilizando operações com os preços em queda também. A lógica de tal prática seria vender antes de abaixar e zerar a posição após a caída para lucrar com essa diferença (EDUCACIONAL BM&FBOVESPA, 2019).

O WIN e o WDO são operados em pontos, o mini índice varia de 5 em 5 pontos no mínimo e cada mudança dessa representa uma variação R$ 1,00 por contrato. No mini dólar, no entanto, cada alteração é de pelo menos 0,5 em 0,5 ponto, e gera uma diferença de R$ 5,00 por lote. Por sua variação ter um maior impacto financeiro, o dólar futuro pode se apresentar como uma forma de lucrar mais rápido, mas leva à noção de ser mais arriscado também. Mesmo assim, em questões operacionais, o WDO pode ser considerado mais atrativo do que

o WIN, por ser influenciado por menos fatores econômicos o dólar leva em consideração principalmente a situação da economia do Brasil e dos EUA. No caso do mini índice, além desses motivos, o seu valor pode ser induzido por aspectos microeconômicos, como resultados de uma grande empresa, distribuição de dividendos ou falência de uma companhia. Ademais o mini dólar tem variações não tão bruscas quanto o WIN para se trabalhar. Portanto, este trabalho é desenvolvido e aplicado aos contratos de mini dólar futuro.

## 4.2 DESCRIÇÃO E APRESENTAÇÃO DOS DADOS

Para o treinamento e validação dos algoritmos deste trabalho, utiliza-se dados sobre todos os negócios intradiários ocorridos no mercado futuro de 02/07/2018 até 21/08/2019 obtidos no site da B3 (B3 MARKET DATA, 2019). As seguintes variáveis foram extraídas dos arquivos:

a) preço dos negócios de WDO realizados;

b) horário da negociação;

c) quantidade negociada de WDO;

d) preço dos negócios do índice iBovespa futuro (IND);

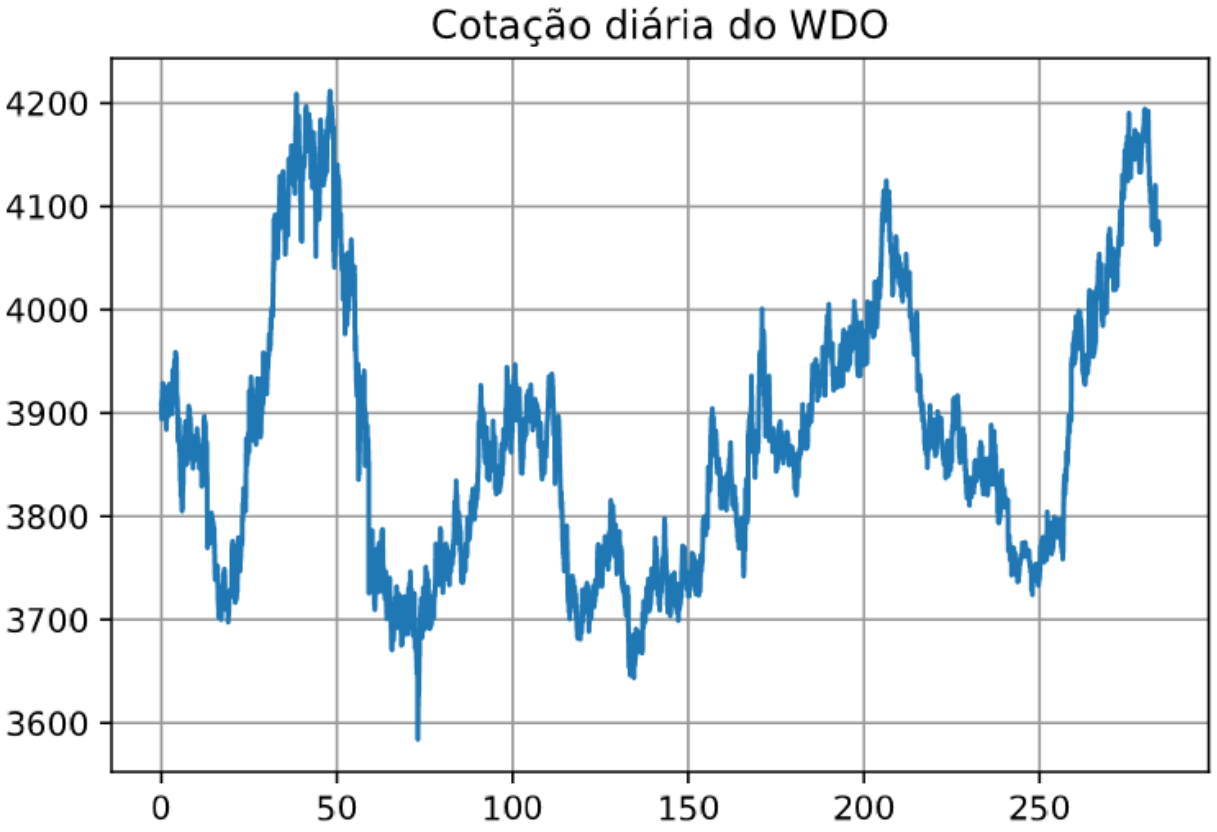
e) preço dos negócios do índice S&P futuro (ISP).

O IND e o ISP espelham a situação e variação da bolsa brasileira e americana, respectivamente, e apresentam uma forte correlação com o dólar, podendo ser utilizados como parâmetros para antecipar um movimento do WDO.

A Figura 3 ilustra a cotação do ativo pelo número de dias. Percebe-se que houve uma grande movimentação no preço do ativo, explicada principalmente por fatores políticos econômicos do Brasil e do exterior, tais como as eleições presidenciais de 2018 no Brasil, resquícios da greve dos caminhoneiros de 2018 e a guerra comercial entre EUA e China. Treinar e testar um algoritmo em momentos de instabilidade e incerteza no mercado como este, torna-se

interessante para analisar quais padrões o algoritmo aprendeu, já que há uma enorme variação de valores sem uma tendência macro definida.

Figura 3 — Preços do mini dólar no período coletado.




Entre essas informações não há uma periodicidade definida e o horário que cada negócio é realizado é medido em milissegundos pela bolsa, podendo um negócio sair a qualquer momento a partir de 1 ms. Entretanto, um período para analisar os dados precisa ser estabelecido para o algoritmo ter um tempo de atuação fixo e poder determinar um padrão. Levando-se em consideração a alta liquidez do mini dólar, inicialmente escolheu-se coletar os dados de 5 em 5 segundos para haver uma rápida atuação pelo algoritmo. No entanto, quanto menor essa “janela de tempo” mais parâmetros e variáveis terão que ser treinadas e calculadas, resultando em um grande esforço computacional, o qual exige uma alta capacidade de processamento não existente nos computadores utilizados nesse. Para tanto, a periodicidade utilizada entre as janelas foi de 5 em 5 minutos, o que ainda exige uma boa computação, mas que pode ser treinada em um tempo razoável.

Com os parâmetros citados acima e com um período de tempo estabelecido, pode-se calcular os seguintes parâmetros, a cada 5 minutos, para utilizar como entrada para o modelo:

a) preço ponderado pela soma da quantidade;

b) soma da quantidade negociada;

c) preço máximo;

d) preço mínimo.

Tais variáveis levam à rede um melhor entendimento do que ocorreu durante os 5 minutos em que não esteve recebendo informações sobre o mercado, assim, ao invés de receber apenas o preço negociado mais recente, também é alimentada por dados que remetem a outros negócios no período.

No site da B3 ainda havia uma base de dados referente à todas as ordens de venda e compra inseridas no mercado. O uso de tais informações seria ainda mais relevante para o aprendizado do algoritmo, pois passaria ao algoritmo dados que podem levar ao aprendizado de uma análise de fluxo das ordens. No entanto, trata-se de um arquivo diário pesado, de aproximadamente 500 MB cada, não sendo prático para o desenvolvimento deste trabalho.

## 4.3 DESCRIÇÃO DAS FERRAMENTAS UTILIZADAS

Para pôr em prática o algoritmo utilizou-se a linguagem de programação *Pytho*n. O *Python* foi criado em 1989 e apenas recentemente começou a ser amplamente utilizado, principalmente, em pesquisas científicas, matemática computacional e inteligência artificial, por possuir uma sintaxe muito simples de se programar e pelo grande número de bibliotecas disponíveis. Nesse trabalho utilizou-se a versão 3.7.4 do *Python* e as suas bibliotecas *TensorFlow* e *Keras*, direcionadas à inteligência artificial e às redes neurais, que permitem criar facilmente um modelo neural e treiná-lo de acordo com os parâmetros definidos. Além dessas, outras bibliotecas foram aplicadas para manipulação de arquivos e operações matemáticas, tais como o *Pandas*, *Numpy* e *Scikit Learn*.

## 4.4 TREINAMENTO, VALIDAÇÃO E TESTE

Utiliza-se 250 dos 284 dias coletados para treinar o algoritmo e o resto para validar e testar o modelo treinado, uma proporção de aproximadamente 88 e 12% dos dados, aproximadamente. Embora seja comum a maioria dos pesquisadores optarem por uma proporção 70-30, não existe um consenso quanto a isso, sendo tal relação uma “regra do polegar” e não uma limitação. Além disso essa quantidade de dias treinador se aproxima dos dias úteis do ano, logo, seria conveniente afirmar que o treinamento se deu em dados de 1 ano.

O treino é realizado durante um certo número de épocas ou episódio, sendo cada época um aprendizado nos 250 dias separados. Um dos princípios do aprendizado por reforço se baseia na exploração de novas decisões e no aproveitamento do aprendizado. Essa exploração de novas políticas de decisões pode ser implementada mediante um parâmetro denominado e*psilon-greedy*, que controla a aleatoriedade do treino. Quanto mais randômico, mais políticas serão exploradas e menor o aproveitamento e vice-versa. Inicialmente deseja-se um algoritmo que teste inúmeras ações e possiblidades de estado, para ao longo do treino o sistema ir filtrando cada vez mais estes testes e ir executando o que já aprendeu, logo, utiliza-se um e*psilon-greedy* decrescendo sua aleatoriedade no decorrer dos episódios treinados.

Além disso, existe ainda a taxa de desconto que pondera a importância dada às recompensas futuras em relação as recompensas mais recentes e a taxa de aprendizado que atualiza os pesos e parâmetros do modelo buscando a conversão deste (OLIVEIRA e PEREIRA, 2009). Tais mudanças somente são realizadas no fim de uma época.

## 4.5 ATRIBUTOS DO ALGORITMO

O foco do trabalho é principalmente em pessoas físicas, portanto o algoritmo foi restringido a posicionar-se em no máximo em 5 contratos, podendo comprar ou vender quantas quantidades quiser dentro desse limite. Como comentado anteriormente, existe a possibilidade nos mini contratos de se operar alavancado, havendo em algumas corretoras a disponibilidade de ser necessário ter apenas R$ 25,00 por contrato, uma alavancagem de 25 vezes o valor real do papel (RICO, 2019). Embora não haja um estudo acadêmico ou alerta do

Bacen ou da CVM sobre o assunto, a maioria das corretoras recomendam ter em conta R$ 1.000,00 por contrato como uma folga para eventuais prejuízos e chamadas de margem (CLEAR, 2019). Seguindo por este raciocínio então, considera-se que o algoritmo arrisca no máximo R$ 5.000,00, um valor apropriado para especuladores individuais. Entre outros paradigmas operacionais seguidos cita-se a posição zerada sempre no começo e no final do dia para configurar somente um *day-trade*, e a possibilidade de permanecer zerado quando julgar necessário.

No mercado há sempre um spread de compra e venda entre as ofertas do mercado e a possibilidade de inserir uma ordem a ser executada em certo preço. Nesse trabalho se desenvolveu um *algotrading* agressor de ordens, que compra ou venda na melhor oferta do que estiver disponível no momento da decisão, não necessitando a inserção de ordens no mercado, a chamado *apregoação*.

Os custos de corretagem e custódia foram considerados como zero, dado que algumas corretoras de investimento não cobram taxas em cima de *day-trade* em mini contratos. Ademais, as taxas operacionais cobrada pela bolsa de valores para derivativos de dólar não foram consideradas por falta de praticidade na sua implementação: seu cálculo é realizado com base no valor de fechamento do dólar no mês anterior e pela quantidade média diária de contratos operados nos últimos 21 dias, sendo este um parâmetro variável e dependente de quanto o algoritmo opera por dia (BM&FBOVESPA, 2016). Além disso não se descontou o Imposto de Renda dos resultados. Segundo a legislação fiscal brasileira, o IR para operações de *day-trade* é de 20% em qualquer retorno positivo auferido, assim, o lucro líquido será 80% do que calculado no treino e teste (CONGRESSO NACIONAL, 2000).

## 4.6 ESTRUTURA DA REDE NEURAL

Os resultados positivos são obtidos com uma rede neural *feedforward* multicamadas com *backpropagation*. Conta-se com multicamadas intermediárias com unidade linear retificada (ReLU) como função de ativação e uma de saída com função de ativação linear. O número de neurônios em cada camada diminui em relação à anterior, tendo a última o número de

unidades correspondente à quantidade de ações possíveis ao algoritmo. Utilizou-se o otimizador Adam e o erro quadrático médio (MSE) como função custo.

As entradas da rede são citadas na seção descritiva dos dados e tem uma periodicidade de 5 minutos no mercado entre cada leitura.

# 5 RESULTADOS E DISCUSSÃO

Para os experimentos, utilizou-se os parâmetros da tabela 1:

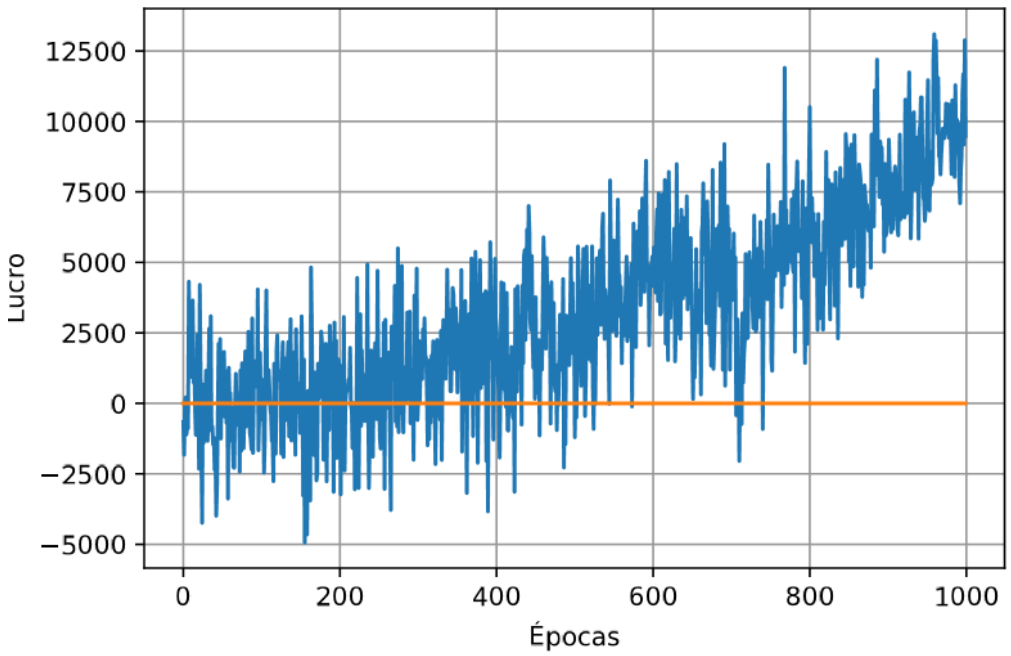
Tabela 1 — Parâmetros do algoritmo *Deep Q-Learning*.

| Parâmetro | Valor |
|---|---|
| Taxa inicial de exploração | 1,0 |
| Taxa de decrescimento por época | $10^{-3}$ |
| Fator de desconto | 0,98 |
| Taxa de aprendizado | $10^{-4}$ |
| Número de *hidden layers* | 7 |
| Número de épocas treinadas | 1.000 |

Os parâmetros acima foram ajustados diversas vezes até atingir uma boa convergência no treinamento. Todos resultados aqui discutidos foram implementados somente em um ambiente de simulação no *Python*, logo, o algoritmo não teve contato com uma plataforma da bolsa nem com o mercado real.

A Figura 4 ilustra os lucros obtidos por episódio no treino. Fica nítido que houve um aprendizado do algoritmo em relação aos dados, sendo o resultado aprendido aproximadamente R$ 11.760,00 nos 250 dias, um lucro médio diário de cerca de R$ 47,04 por dia, um retorno não alto e bruto de custos operacionais da bolsa de valores. No entanto, se for considerado que o máximo arriscado ou especulado diariamente eram R$ 5.000,00, como já mencionado, o algoritmo obteve um retorno médio por volta de 135% no período e de 0,94% diariamente. Todas essas taxas são de juros simples, não compostos, já que o percentual incide sempre no mesmo principal (R$ 5.000,00).

Figura 4 — Retorno por episódio no treino.




Analisando a Figura 4 novamente, percebe-se uma curva de aprendizado crescente, uma característica indicativa de que realmente houve um aprendizado e acúmulo de experiência durante o treino.

O teste, para validação do treino, foi realizado em 34 dias e obteve um resultado de cerca de R$ 114,50. A tabela 2 informa melhor os retornos:

Tabela 2 — Resultados do teste.

| Informação | Valor |
|---|---|
| Lucro total | R$ 114,50 |
| Lucro médio diário | R$ 3,37 |
| Retorno total | 0,23% |
| Retorno médio diário | 0,07% |

O teste não obteve o mesmo sucesso que o treino em termos de retorno, ainda que positivo. Uma análise dia a dia dos resultados, pelas Figuras 5 e 6, ajuda a esclarecer melhor o desempenho do treinamento e de sua validação:

Figura 5 — Retornos dia a dia no treino.

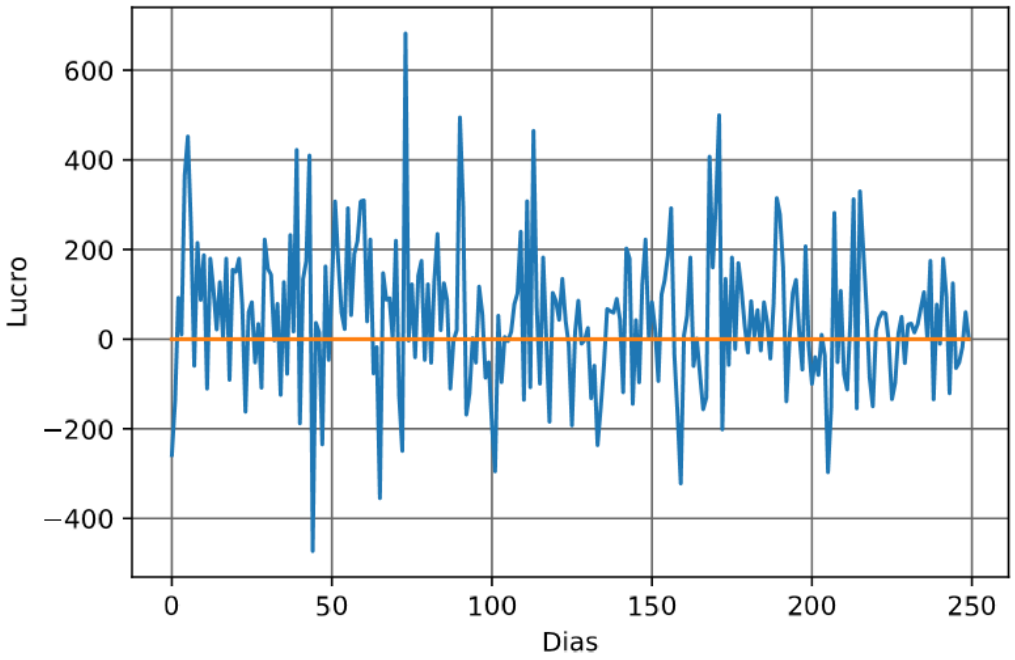
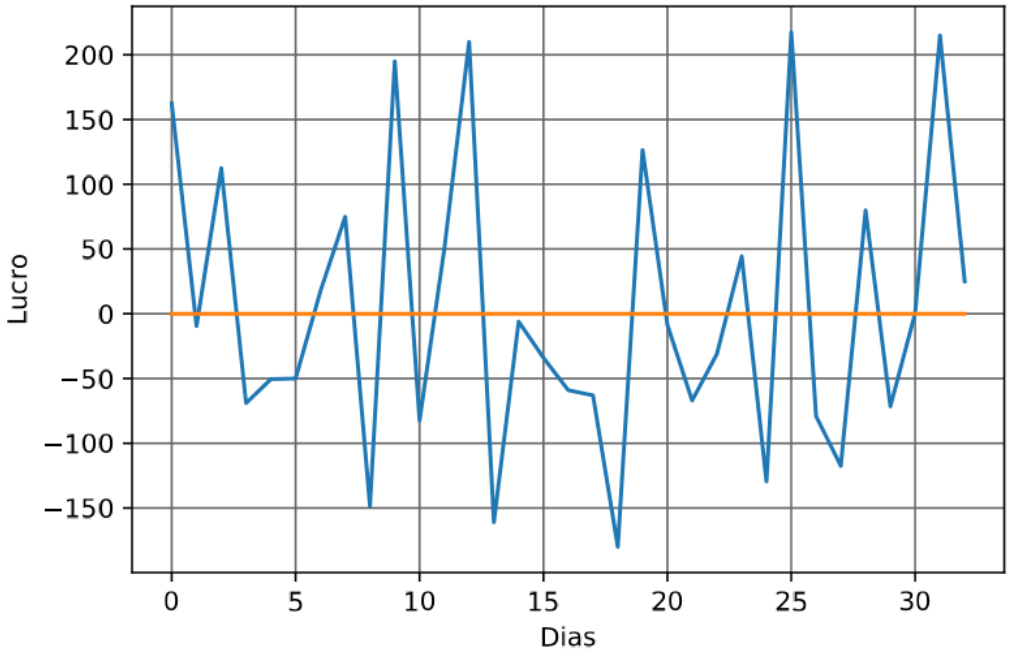



Figura 6 — Retornos dia a dia no teste.




Pelas figuras percebe-se que há uma alta variabilidade entre os dados, havendo dias positivos e também negativos, principalmente no teste, aonde houve um retorno menor e uma maior variação entre os resultados diários. A falta de consistência aparente nos gráficos indica que o treino não foi capaz de generalizar todos os comportamentos caóticos do mercado futuro e se assimila até com um resultado aleatório.

Entretanto, o retorno final, em ambos os casos, foi positivo. Essa capacidade de ter lucro em meio uma alta volatilidade na rentabilidade, mostra uma vantagem importante de um algoritmo a um *day-trader* humano. O emocional do especulador, em momentos como esse, impacta muito seu comportamento e suas análises, enquanto que uma máquina continua tomando suas decisões com um viés totalmente matemático e analítico. Quando analisado em um período maior, 134% de lucro bruto em menos de um ano é extremamente atraente, ainda mais quando comparado à taxa básica de juros em 5,0% a.a., à rentabilidade de títulos públicos indexados ao CDI (4,9% a.a.) e o retorno médio do índice iBovespa nos últimos 12 meses (20,90%). No entanto, as oscilações de curto prazo é o que afeta a maioria dos investidores.

Há diversos ajustes e técnicas que se pode utilizar para melhorar ainda mais a sua generalização, mas acredita-se que o maior impacto possa vir de três alterações na implementação:

a) uma menor periodicidade entre os dados pode fazer o agente perceber mudanças e tendências no mercado mais rápido, e também a ter uma ação mais imediata;
b) utilizar uma classificação de tendência (alta, baixa ou neutro) em uma janela de preços, leva em conta uma análise técnica do que ocorreu no mercado e pode contribuir para o algoritmo ter um melhor *feedback* do estado em que o mercado se encontra;
c) implementação de métodos mais avançados e estáveis que o *Deep Q-Learning*, como o *Double Deep Q-Learning* e o *Dueling Q-Networks*.

Tais mudanças podem vir a ser os próximos passos do trabalho em busca de uma melhor performance e estabilidade nos resultados.

# 6 CONCLUSÕES

Utilizando o método *Deep Q-Learning* e uma base de dados de mais de um ano com todos os negócios intradiários do mini contrato de dólar futuro, implementou-se neste trabalho um *algotrading* para *day-trade*. Os resultados demonstrados comprovam que houve um aprendizado por parte do algoritmo e, mesmo apresentando certa instabilidade na sua generalização, foi possível obter retornos positivos em um ambiente simulado.

Este trabalho contribui para literatura de inteligência artificial e de mercado financeiro, já que não se encontrou nenhuma referência que aplique técnicas de aprendizado por reforço em operações de *day-trade*. Ademais, a utilização de redes neurais artificiais profundas expande ainda mais os temas e métodos abrangidos nessa pesquisa.

Este estudo pode ser melhorado com um melhor processamento computacional e com técnicas já citadas na seção anterior. Também pode ser estendido por meio de abordagens em outros ativos e implementações no mercado real da bolsa de valores.

# REFERÊNCIAS


B3. **Manual de procedimentos operacionais da câmara de compensação e liquidação da BM&FBovespa (câmra BM&FBovespa)**. B3 - Brasil Bolsa Balcão S.A. São Paulo, p. 211. 2019.

B3 MARKET DATA. B3 Market Data. **FTP BMF**. Disponivel em: <ftp://ftp.bmf.com.br/MarketData/BMF/>. Acesso em: 31 out. 2019.

BCB. **Resolução n° 3357, de 31 de março de 2006**. Banco Central do Brasil. Brasília, p. 5. 2006.

BELLMAN, R. E. A Markovian Decision Process. **Journal of Mathematics and Mechanics**, California, 18 abril 1957.

BERGAMO, Y. P. et al. **Accelerating reinforcement learning by reusing abstract policies**. Escola Politécnica, Universidade de São Paulo São Paulo. São Paulo, p. 12. 2011.

CHAGUE, F.; DE-LOSSO, R.; GIOVANNETTI, B. **Day trading for a living?** Escola de Economia, FGV / Departamento de Economia, Universidade de São Paulo. São Paulo, p. 16. 2019.

CVM. **Mercado de valores mobiliários brasileiro**. Comissão de Valores Mobiliários. Rio de Janeiro. 2014.

DA COSTA, I. S. **High frequency trading (HFT) em câmera lenta: compreender para regular**. Fundação Getúlioo Vargas. São Paulo, p. 333. 2018.

DENG, Y. et al. Sparse coding-inspired optimal trading system for HFT industry. **IEEE Trans. Ind. Informat**, Abril 2015. 467-475.

DENG, Y. et al. Deep Direct Reinforcement Learning for Financial. **IEEE Transactions on Neural Networks and Learning Systems**, 22 Janeiro 2016. 1-12.

DO AMARAL, R. P. **Comportamento dos investidores em operações daytrade**. Universidade Regional do Noroeste do Estado do Rio Grande do Sul. Ijuí, p. 29. 2015.

EDUCACIONAL BMF&BOVESPA. Apostila PQO. **Educacional BMF&Bovespa**, 2019. Disponivel em: <https://educacional.bmfbovespa.com.br/documentos/ApostilaPQO.pdf>. Acesso em: 18 abril 2019.

FRANÇOIS-LAVET, V. et al. Brief introduction to deep reinforcement learning. **Foundations and Trends in Machine Learning**, Boston, 11, 3 dez. 2018. 140.

GAMBOGI, J. A. **Aplicação de redes neurais na tomada de decisões no mercado de ações**. Universidade de São Paulo. 2013, p. 78. 2013.

GOMES, I. D. O. **Estratégias para operações de day trade na B3**. Fundação Getúlio Vargas. São Paulo, p. 51. 2018.

GRIFFIN, A. Turing test breakthrough as super-computer becomes first to convince us he is a human. **Independent**, 2014. Disponivel em: <https://www.independent.co.uk/life-

style/gadgets-and-tech/computer-becomes-first-to-pass-turing-test-in-artificial-intelligence-milestone-but-academics-warn-9508370.html>. Acesso em: 11 out. 2019.

HAYKIN, S. **Redes Neurais:** Príncipios e Prática. 2ª. ed. Porto Alegre: Bookman, v. 1, 2007. 898 p.

LOPES, R. S. **Aplicação de estratégias de high frequency trading no mercado brasileiro de dólar futuro**. Universidade de São Paulo. São Paulo, p. 59. 2018.

MOODY, J. et al. Performance functions and reinforcement learning for trading systems and portfolios. **J. Forecasting**, 17, 1998. 441-470.

MOODY, J.; SAFFELL, M. Learning to Trade via Direct Reinforcement. **IEEE Trans. Neural. Netw**, 2, Julho 2001. 875-889.

MOODY, J.; WU, L. Optimization of trading systems and portfolios. **'Decision Technologies for Financial Engineering'**, Londres, 1997. 23-35.

RICO. Day Trade com alavancagem. **Rico.com.vc**, 2019. Disponivel em: <https://www.rico.com.vc/rico-trader>. Acesso em: 19 nov. 2019.

SAMUEL, A. L. Some Studies in Machine Learning Using Game Checkers. **IBM Journal os Research and Development**, 1959. 210-229.

SCHWAGER, J. D. **Market Wizards:** Interviews with Top Traders. 1. ed. Nova York: John Wiley & Sons, v. 1, 1989.

SPRITZER, F. A.; TAHUATA, J. P. M. **Análise técnica para day trade: rentabilidade de indicadores no longo prazo**. Universidade Federal do Rio de Janeiro. Rio de janeiro, p. 51. 2017.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning:** An Introduction. 2ª. ed. Cambridge, Massachusetts: The MIT Press, v. 1, 2018.

WATKINS, C. J. C. H. **Learning from Delayed Rewards**. University of Cambridge. Cambridge. 1989.

ZHANG, X. F. **High-Frequency Trading, Stock Volatility, and Price Discovery**. Yale University. New Haven, Connecticut, p. 54. 2010.