CENTRO UNIVERSITÁRIO FEI

JONATHAN KENJI KINOSHITA

APRENDIZADO POR REFORÇO PROFUNDO COM REDES RECORRENTES APLICADO À NEGOCIAÇÃO DO MINICONTRATO FUTURO DE DÓLAR

São Bernardo do Campo

2023

JONATHAN KENJI KINOSHITA

APRENDIZADO POR REFORÇO PROFUNDO COM REDES RECORRENTES APLICADO À NEGOCIAÇÃO DO MINICONTRATO FUTURO DE DÓLAR

Dissertação de Mestrado, apresentada ao Centro Universitário FEI como pré-requisito para a obtenção do título de Mestre em Engenharia Elétrica. Orientado pelo o Prof. Dr. Reinaldo A. C. Bianchi.

São Bernardo do Campo

2023

Kinoshita, Jonathan Kenji.
APRENDIZADO POR REFORCO PROFUNDO COM REDES RECORRENTES APLICADO A NEGOCIACAO DO MINICONTRATO FUTURO DE DOLAR / Jonathan Kenji Kinoshita. São Bernardo do Campo, 2023.
88 f. : il.

Dissertação - Centro Universitário FEI.
Orientador: Prof. Dr. Reinaldo Augusto da Costa Bianchi.

1. Aprendizado por Reforço Profundo. 2. Deep Recurrent Q Network. 3. Redes Neurais Convolucionais. 4. Redes Neurais Recorrentes. 5. Long Short-Term Network. I. Bianchi, Reinaldo Augusto da Costa , orient. II. Título.

Elaborada pelo sistema de geração automática de ficha catalográfica da FEI com os dados fornecidos pelo(a) autor(a).

centro universitário FEI

APRESENTAÇÃO DE DISSERTAÇÃO
ATA DA BANCA EXAMINADORA

Mestrado

Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

PGE-10

Aluno(a): Jonathan Kenji Kinoshita **Matrícula:** 120102-9

Título do Trabalho: APRENDIZADO POR REFORÇO PROFUNDO COM REDES RECORRENTES APLICADO À NEGOCIAÇÃO DO MINICONTRATO FUTURO DE DÓLAR

Área de Concentração: Inteligência Artificial Aplicada À Automoção e Robótica

Orientador(a): Prof. Dr. Reinaldo Augusto da Costa Bianchi

Data da realização da defesa: 27/02/2023

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

A banca foi realizada no dia 27 de fevereiro de 2023 às 14:00 horas, iniciando pela apresentação do aluno, e seguiu para a arguição, onde o aluno respondeu a todas as questões de forma adequada demonstrando conhecimento pleno do tema. Foram sugeridas melhorias em relação ao texto para a versão final. A aprovação foi por unanimidade.

A Banca Julgadora acima-assinada atribuiu ao aluno o seguinte resultado:

APROVADO ☐ REPROVADO ☐

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Reinaldo Augusto da Costa Bianchi

Prof. Dr. Ricardo de Carvalho Destro

Prof. Dr. Thiago Pedro Donadon Homem

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Dedico este trabalho aos meus pais e amigos.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pelo dom da vida e de poder aproveitá-la da melhor maneira possível, mesmo no meio de tantos desafios e provações.

Ao meu orientador Prof. Dr Reinaldo Augusto da Costa Bianchi, por aceitar em me orientar e pela paciência que teve comigo durantes esses anos.

Ao meu amigo Dr. Douglas de Rizzo Meneghetti, que sem ele este trabalho nunca teria saído do lugar, e por sempre mostrar o melhor caminho a ser seguido.

Aos amigos do Colégio Agostiniano Mendel, local onde estudei e criei os laços de amizade mais fortes que tenho, obrigado pelo enorme companheirismo e apoio de sempre.

Aos meus amigos do *Rolê Gourmet* que sem o apoio inestimável deles, este trabalho não seria concluído.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“Se você não encontrar um jeito de ganhar dinheiro enquanto dorme, você vai trabalhar até morrer”

Warren Buffet

RESUMO

Recentemente há um aumento exponencial no uso de técnicas de aprendizado de máquina no mercado financeiro, principalmente para negociação de ações, na tentativa de prever o seu preço futuro. O objetivo desse projeto é desenvolver um sistema de negociação inteligente para o Minicontrato Futuro de Dólar, baseado no uso de aprendizado por reforço, usando o *Deep Recurrent Q learning*, um modelo de Redes Neurais Convolucionais combinadas com as Redes Neurais Recorrentes. O treinamento foi baseado em uma base da dados históricos do ativo e o agente realizou três ações: comprar, vender, manter o ativo, sempre visando o máximo retorno financeiro. Os experimentos realizados demonstraram que o sistema proposto teve um desempenho melhor do que as estratégias de *Buy and Hold*, um modelo baseado na *Deep Q Network*, um Fundo Cambial e uma estratégia baseada no indicador técnico MACD.

Palavras-chave: Aprendizado por Reforço Profundo. Redes Neurais Convolucionais. Redes Neurais Recorrentes. *Long Short-Term Network*. *Deep Recurrent Q Network*. Mercado Futuro.

ABSTRACT

Recently, there is an exponential increase in the usage of machine learning applied in the financial market, primaly for trade stocks, in an attempt to forecast the movement of their prices. The proposal of this research is developing an intelligent trade system for Mini US Dollar Future, based on Deep Recurrent Q Learning, a model that uses Convolutional Neural Networks combined with Recurrent Neural Networks. The training is based on the historical data from the asset and the agent performs its actions of buy, hold and sell, always aiming the maximum return. The experiments demonstrated that the propose system has had better outcomes than the traditional strategy Buy and Hold, a model based on the Deep q Network, an Exchange Fund and a strategy that utilized the technical indicator MACD as action generator.
Keywords: Deep Reinforcement Learning. Convolutional Neural Network. Recurrent Network. Long Short-Term Network. Deep Recurrent Q Network. Future Market.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ALGORITMOS

LISTA DE ABREVIATURAS

A2C	Advantage Actor-Critic
B3	Brasil,Bolsa,Balcão
CDB	Certificados de Depósito Bancário
CDI	Certificado de Depósito Interbancário
Cetip	Central de Custódia e de Liquidação Financeira de Títulos Privados
CNN	Redes Neurais Convolucionais (*Convolutional Neural Networks*)
Copom	Comitê de Política Econômica
DDQN	Rede Q Profunda Dupla (*Double Deep Q Network*)
DI	Depósito Interbancário
DL	Aprendizado Profundo (*Deep Learning*)
DQN	Rede Q Profunda (*Deep Q Network*)
DRQN	Rede Q Profunda Recorrente (*Deep Recurrent Q Network*)
ELU	Unidade Linear Exponencial (*Exponential Linear Unit*)
GDPG	Gated Deterministic Policy Gradient
GDQN	Gated Deep Reinforcement Q-Learning
GRU	Gated Recurrent Units
LCA	Letras de Crédito do Agronegócio
LCI	Letras de Crédito Imobiliário
LSTM	*Long Short-Term Memory*
LTN	Letra do Tesouro Nacional
MACD	Média Móvel Convergente e Divergente (*Moving Average Convergence Divergence*)
MDP	Processo de Decisão de Markov (*Markov Decision Process*)
MLP	Perceptron Multi Camadas (*Multilayer Percetron*)
MME	Média Móvel Exponencial
NLP	Linguagem Natural de Processamento (Natural Language Processing)
NTN-B	Nota do Tesouro Nacional Tipo B
PG	Policy Gradient
POMDP	Processo de Decisão Markoviano Parcialmente Observável (*Partially Observable Markov Decision Process*)
RBM	Máquinas de Boltzman Restritas (*Restricted Boltzman Machines*)

ReLu	Unidade Linear Retifcadora (*Rectifed Linear Unit*)
RNN	Redes Neurais Recorrentes (*Recurrent Neural Networks*)
RRL	Recurrent Reinforcement Learning
Selic	Sistema Especial de Liquidação e de Custódia
TDQN	Trading Deep Q Network

SUMÁRIO

1 INTRODUÇÃO

O mercado financeiro é considerado o coração da economia mundial. A natureza caótica, estocástica e não-linear do mercado (HSIEH, 1991), assim como sua evolução ao longo do tempo, fazem da predição de seu comportamento futuro uma tarefa desafiadora.

Um ativo financeiro é tudo que tem valor agregado e pode ser negociado no mercado financeiro, não são bens ou mercadorias e não possuem representação contratual além da documentação que os define. Seus valores derivam de uma reivindicação contratual do que representam. Entre os exemplos de ativos financeiros mais conhecidos estão as ações, moedas, títulos públicos, títulos privados, commodities e índices. Ativos são comumente chamados de papel.

Atualmente existem duas vertentes para o estudo de um ativo, uma visando o valor agregado à longo prazo, chamado de análise fundamentalista e a outra baseada no movimento dos preços, seu histórico e seus possíveis valores futuros, conhecido como análise técnica (LO; MAMAYSKY; WANG, 2000).

Técnicas de aprendizado de máquina têm mostrado sua capacidade em predizer o comportamento futuro dos preços dos ativos. Redes neurais artificiais e máquina de vetores de suporte são algumas das técnicas utilizadas com esse objetivo (GURESEN; KAYAKUTLU; DAIM, 2011; KARA; ACAR BOYACIOGLU; BAYKAN, 2011). Em um estudo com uma rede neural usando indicadores técnicos (médias aritmética e o estocástico), Chan e Foo Kean Teong (1995) demonstraram que a utilização de redes neurais resultou em um lucro maior do que o de sistemas baseados apenas nos tradicionais sinais técnicos.

Em muitos domínios, como o da visão computacional, métodos de aprendizado profundo têm demonstrado grande capacidade de abstrair atributos complexos de dados extremamente simples (HE et al., 2016; LECUN; BENGIO, Yoshua; HINTON, 2015). Pelo fato do mercado estar continuamente mudando e evoluindo devido a sua relação não linear, caos e a natureza estocástica (HSIEH, 1991), conseguir extrair informações desse ambiente se torna uma tarefa árdua. Modelos de Perceptron Multi Camadas (*Multilayer Percetron*) (MLP) (YONG; ABDUL RAHIM; ABDULLAH, 2017), Máquinas de Boltzman Restritas (*Restricted Boltzman Machines*) (RBM) (CAI; HU; LIN, 2012), *Long Short-Term Memory* (LSTM) (CHEN; ZHOU; DAI, 2015; FISCHER; KRAUSS, 2018), e Redes Neurais Convolucionais (*Convolutional Neural Networks*) (CNN) (GUNDUZ; YASLAN; CATALTEPE, 2017; PERSIO; HONCHAR, 2016) são algumas das técnicas de aprendizado profundo utilizadas para prever o mercado de ativos.

Desenvolver um sistema de negociação de ativos robusto requer certas características, como a resiliência, ou seja, deve ter a capacidade de se adaptar a qualquer momento do mercado, e gerar lucros consistentes mesmo em períodos de volatilidade.

Os avanços em aprendizado profundo da última década introduziram nossas topologias de redes neurais capazes de aprender automaticamente representações úteis para os dados, de acordo com a função de erro que tentam minimizar. Já as técnicas de aprendizado por reforço se concentram na solução de problemas de tomada de decisão sequencial, ao passo que não são focadas no aprendizado de representações ou na resolução de problemas cuja dimensionalidade dos dados seja alta.

A combinação dos dois campos dá-se o nome de Aprendizado por Reforço Profundo, cujas técnicas visam a criação de sistemas de tomadas de decisão sequenciais em ambientes de alta dimensionalidade de dados. Diversas pesquisas aplicam técnicas de Aprendizado por Reforço Profundo na negociação de ativos (LI; NI; CHANG, 2020; CARAPUÇO; NEVES; HORTA, 2018; LEE et al., 2019; XIONG et al., 2018).

Pela natureza do mercado financeiro, este dificilmente será totalmente explícito nos estados, para se tomar uma decisão é necessário um conhecimento mais amplo do ambiente, podendo assim ser considerado um Processo de Decisão Markoviano Parcialmente Observável, do ingle a sigla POMDP de Partially Observable Markov Decision Process (HAUSKNECHT; STONE, 2017). O objetivo deste trabalho foi investigar o funcionamento de um sistema de negociação de ativos, para realizar operações de compra e venda, sempre visando o maior retorno financeiro, baseado na Rede Q Profunda Recorrente (DRQN, do inglês *Deep Recurrent Q-Network*), a qual usa uma combinação entre Redes Neurais Recorrentes, do tipo *Long Short-Term Memory* (LSTM) com Redes Neurais Convolucionais (CNN, do inglês *Convolutional Neural Networks*).

Este trabalho está organizado da seguinte forma: o Capítulo 2 descreve a fundamentação teórica abordando conceitos da área de Derivativos e seu funcionamento, Inteligência Artificial, Aprendizado por Reforço; o Capítulo 3 apresenta uma breve revisão bibliográfica da área de pesquisa do projeto, apresentando trabalhos similares e os que influenciaram esta dissertação; o Capítulo 4 apresenta a metodologia utilizada na modelagem do problema proposto por este trabalho; o Capítulo 5 expõe como foi feita a implementação do trabalho e os experimentos realizados; o Capítulo 6 mostra os resultados experimentais obtidos e suas análises; o Capítulo 8 apresenta as conclusões do projeto e projetos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 ANÁLISE TÉCNICA E FUNDAMENTALISTA

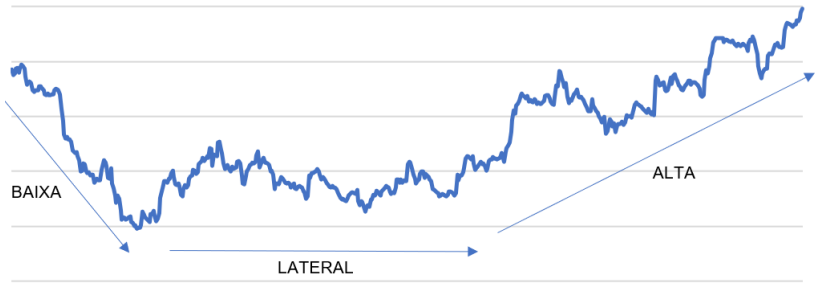
Os mercados financeiros têm um papel importante na organização econômica e social da sociedade moderna. Eles são constituídos de investidores que visam o lucro a partir da compra e venda de ativos. Em vista disso, os profissionais da área se baseiam em dois modelos de estudos antes de tomarem alguma decisão sobre certo ativo: a análise técnica e a análise fundamentalista.

Na análise fundamentalista, o objeto de estudo é a empresa, considerando a situação financeira dela, a perspectiva de crescimento, a credibilidade, o setor econômico de atuação, e com todas essas informações estima-se um valor de mercado dela. Normalmente esse estudo é voltado para um investimentos de médio a longo prazo.

Na análise técnica, o ativo é tratado como sendo uma série temporal onde os preços e seus indicadores técnicos representam todas as informações relacionadas ao papel, e que sinalizam o momento mais propício para a execução de estratégias de negociação. Profissionais da área acreditam que a previsão do comportamento futuro do mercado é possível apenas analisando o histórico e os indicadores dos preços do ativo.

O preço de um ativo apresenta apenas três movimentos: de alta, baixa e lateral. No movimento de alta há uma valorização do preço do ativo, no de baixa, acontece o contrário, há uma desvalorização. Já o movimento lateral é caracterizado pela estabilidade do preço. A Figura 1 exemplifica esses movimentos.

Figura 1 – Movimentos do preço de um ativo

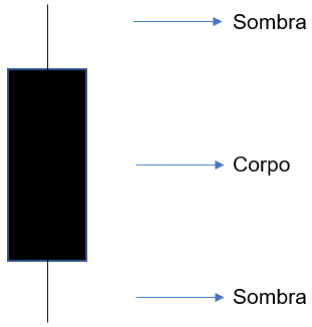


Fonte: Autor

A análise técnica é muito baseada no estudo dos gráficos de vela (*candlestick*), os quais ajudam os investidores a visualizar os possíveis pontos de entrada e saída de um ativo, a partir da identificação dos movimentos dos preços dele. Esses gráficos representam o preço de um ativo em um determinado período de tempo. A figura 2 ilustra uma vela desses gráficos, ela

é composta por um corpo, que é representado por um retângulo, e por suas sombras, que são representadas pelas linhas.

Figura 2 – Exemplo de uma vela de um gráfico de velas

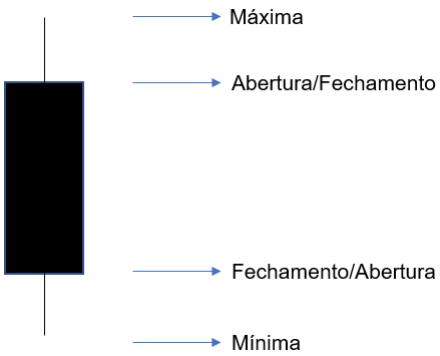


Fonte: Autor

Cada vela representa o histórico do preço de um ativo em um determinado período, que pode variar de minutos, horas, dias, semanas e até anos. Uma vela contém varias informações dentre elas (figura 3), os preços de:

a) Abertura: Preço que o ativo estava no início do período
b) Fechamento: Preço que o ativo estava no final do período
c) Máxima (Alta): Maior preço negociado no período
d) Mínima (Baixa): Menor preço negociado no período

Figura 3 – Informações presentes em uma vela



Fonte: Autor

A partir dessas informações, é possível inferir se no período da vela o ativo aumentou ou diminuiu de preço. Na figura 4, a vela vermelha representa o decaimento do preço, sendo o seu preço de fechamento inferior ao preço de abertura, e já na vela verde, o ativo teve uma valorização, pois seu preço de fechamento foi superior ao de abertura.

Figura 4 – Velas ilustrando o incremento e decaimento do preço de um ativo

Fonte: Autor

A partir do princípio da análise técnica, no qual todas as informações necessárias para a negociação estão contidas nos preços e nos indicadores técnicos do ativo, têm-se os algoritmos de negociação, que utilizam dessa premissa na geração de sinais de compra ou venda de um ativo, baseado em uma série de regras pre-estabelecidas pelo profissional da área. (RATNER; LEAL, 1999; KWON; KISH, 2002; PHAN HUY; CUONG, 2018). Os sinais podem ser gerados a partir de dados de longos períodos ou de flutuações recentes dos preços dos ativos. Com o desenvolvimento de programas baseados nesses algoritmos é possível automatizar todo o processo de compra e venda de um ativo (CHAN, 2009, 2013), conseguindo assim lidar com muita mais informação ao mesmo tempo e muito mais rápido que um negociador de ativos normal.

Algoritmos de negociação envolvem a introdução se uma série de regras e estratégias para a realização automática da negociação do ativo. Sendo assim, é natural que apareça a aplicação de técnicas de aprendizado de máquina nesses algoritmos, podendo deixar as estratégias mais flexíveis e até mesmo retirar a necessidade da configuração das regras dos algoritmos. Algumas das aplicações de aprendizado de máquina nos últimos foram em:

a) Previsão do preço futuro de um ativo: Utilizam da premissa da análise técnica, na qual a maioria das informações sobre os ativos estão refletidas nos seus preços, sendo assim, com a identificação de seu movimento, o preço futuro é facilmente previsto. Há muitos trabalhos relacionados aplicando várias técnicas de aprendizado de máquinas, ainda mais que as redes neurais conseguem captar e aprender varias relações não lineares (KROLLNER; VANSTONE; FINNIE, 2010; SHAH; ISAH; ZULKERNINE, 2019; DONGDONG et al., 2019).

b) Otimização de Portfólio: Inclui o processo de escolher o melhor ativo dentre vários, de acordo com os objetivos, como o maior retorno possível ao menor risco (MARKOWITZ, 1952).

c) Análise de sentimento: Um problema típico é a análise de informações externas e seus impactos nos preços dos ativos. Essas informações são geralmente extraídas de notícias de sites (LI et al., 2014) ou redes sociais (ROCHA et al., 2021) e analisadas usando técnicas de Linguagem Natural de Processamento (Natural Language Processing) (NLP) (ZHANG; WANG; LIU, 2018).

Neste sentido, teve a aplicação do Aprendizado por Reforço nestes algoritmos de negociação, o que pode permitir, por exemplo, a combinação da predição dos valores futuros do ativo com o gerenciamento do portfólio, se aproximando e alinhando cada vez mais com os objetivos de um investidor. (FISCHER, 2018; TERRY LINGZE MENG, 2019)

2.1.1 Média Móvel Convergente e Divergente

Dentre os indicadores utilizados nos algoritmos de negociação tem a Média Móvel Convergente e Divergente (*Moving Average Convergence Divergence*) (MACD) (APPEL, Gerald, 2005; APPEL, G., 1985). Ela é usada para mostrar a força, direção, mudança, direção e duração da tendência de um ativo nos estágios iniciais, e é baseada em duas Médias Móveis Exponenciais (MME), uma curta e uma longa, sendo a sua fórmula descrita na equação 1.

$$MME_t = \alpha(P_t) + (1 - \alpha)MME_{t-1} \tag{1}$$

Onde, t representa o período a ser utilizado na média móvel, e α os graus de decaimento; $\alpha = \frac{2}{(t+1)}$, e P_t é o preço de fechamento

Por exemplo, supondo que Média Móvel Exponencial curta seja de 12 períodos e a longa de 26 períodos, a formula descrita pela equação 1 se transforma nas equações 2 e 3, respectivamente.

$$MME_{12} = \frac{2}{(12+1)}P_{12} + (1 - \frac{2}{(12+1)})MME_{12-1} \tag{2}$$

$$MME_{26} = \frac{2}{(26+1)}P_{26} + (1 - \frac{2}{(26+1)})MME_{26-1} \tag{3}$$

O indicador técnico MACD é composto por três sinais calculados a partir do histórico de preços:

a) Linha MACD, que é a diferença entra a MME rápida e a MME lenta

$$\text{Linha MACD} = \text{MME}_{12} - \text{MME}_{26} \tag{4}$$

b) Sinal MACD, que é a MME da linha MACD

$$\text{Sinal MACD} = \text{MME}_9\text{, no caso, utilizado período de 9} \tag{5}$$

c) Histograma MACD, é a diferença entra a Linha MACD e o Sinal MACD

$$\text{Histograma MACD} = \text{Linha MACD} - \text{Sinal MACD} \tag{6}$$

Na figura 5 há um exemplo do indicador técnico MACD.

Figura 5 – Os três sinais do MACD. A linha azul representa a Linha MACD, a vermelha representa o Sinal MACD, e o gráfico de barras representa o Histograma MACD

Fonte: Autor

A estratégia de negociação gerada a partir deste indicador técnico provém destes três sinais acima descritos é representado como:

a) Sinal de compra do ativo: Linha MACD > Sinal MACD

b) Sinal de venda do ativo: Sinal MACD > Linha MACD

2.2 BRASIL BOLSA BALCÃO

Brasil, Bolsa, Balcão, ou simplesmente B3, é a empresa que administra a Bolsa de Valores do Brasil, onde é organizado o mercado de compra e venda de ativos, e tem-se um ambiente no qual as transações ocorrem de forma transparente. A B3 é a responsável por garantir um lugar seguro e simplificado de compra e venda, fazendo a mediação dos negócios entre empresa e investidores. Ela surgiu a partir da fusão da BMFBovespa com a Central de Custódia e de Liquidação Financeira de Títulos Privados (Cetip) em 2017.

Atualmente, há certa de 382 empresas listadas na B3 (B3, 2022), as quais disponibilizam suas ações para serem negociadas na bolsa, possibilitando assim, investidores de se tornarem acionistas dessas empresas. Além do mercado de ações, a B3 também é responsável na negociação de ativos como:

a) Ativos de Renda Fixa Públicas e Privadas, como Letras de Crédito Imobiliário (LCI), Letras de Crédito do Agronegócio (LCA), Certificados de Depósito Bancário

(CDB), Letra do Tesouro Nacional (LTN), Nota do Tesouro Nacional Tipo B (NTN-B)

b) Derivativos, como Contrato Futuros, Contratos a Termo, Opções e Swaps

2.2.1 Derivativos

Derivativos são ativos cujos valores dependem dos valores de outros ativos, dentre eles, as moedas, índices, taxas de juros e commodities. Um dos objetivos da criação dos derivativos foi para que os agentes econômicos possam se proteger contra os riscos de oscilações de preços. O uso dos derivativos poderia ter amenizado as enormes perdas dos investidores com aplicações prefixadas e para empresas com dívidas flutuantes, decorrentes dos períodos de alta e baixa da taxa de juros que ocorreram nos últimos tempos no Brasil. Existem dois tipos de derivativos: os financeiros e os não financeiros. Os derivativos não financeiros estão associados, por exemplo, ao petróleo, café, soja. Os derivativos financeiros estão diretamente relacionados às taxas de juros, moedas e índices de Bolsa, sendo que, os mais negociados no mercado brasileiro são: termo, futuros, opções e swaps.

a) Termo: é uma operação de compra e venda de um ativo, na qual é realizado o fechamento da operação em um data futura. Na data do vencimento, o comprador paga ao vendedor o preço previamente acordado e recebe o ativo. Em alguns casos, o acerto financeiro no vencimento é feito pela diferença entre o preço a termo e à vista.

b) Futuros: são compromissos de compra ou de venda de determinado ativo, em uma data futura, e a um certo preço. Na data futura esses compromissos são executados e o comprador geralmente recebe a diferença do valor futuro de compra com o de venda.

c) Opções: São direitos. Existem dois tipos de opções:

   – Opções de compra (*call*): Representa o direito de comprar um ativo em determinada data por certo preço.

   – Opções de venda (*put*): Representa o direito de vender um ativo em determinada data por um certo preço.

   O mais comum é o investidor fechar a posição antes do vencimento da opção, sendo assim, o resultado da operação é a diferença entre os valores pagos e recebidos pelo direito.

d) Swap: é um contrato no qual as partes trocam indexadores de operações ativas (credora) e passivas (devedora), e não o valor real da operação. A parte que contrata o swap se compromete a pagar a perna ativa e receber a perna passiva, ou seja, pagar a rentabilidade da perna ativa e receber a rentabilidade da perna passiva. Para esses indexadores do swap é comumente utilizado as taxas cambiais, de juros, commodities e índices.

2.2.1.1 Futuros

Nas operações no mercado derivativo de futuro são negociados os contratos futuros que são acordos de compra e venda de ativos em uma data futura, ou seja, o investidor negocia hoje a expectativa do preço do ativo no futuro. Eles são muito usados em operações de:

a) *Hedge*, que é uma estratégia de investimento muito usada para proteger o valor do ativo em carteira contra a possível variação futura no preço dele
b) Arbitragem, que é quado o investidor tem como objetivo lucrar com a diferença de preços de um mesmo ativo em diferentes mercados
c) Especulação, que é quando o especulador tem como objetivo lucrar com a compra e venda do derivativo, baseado na sua aposta da tendencia futura do ativo. Sua participação no mercado é importante, porque contribui com a liquidez do mercado.

Os principais contratos futuros negociados na Bolsa são:

a) DI: Depósito Interbancário (DI) de um dia
b) Dol: Dólar comercial
c) DDI e FRA: cupom cambial sujo e limpo
d) IND: Ibovespa

Os contratos futuros de DI consistem na negociação da taxa de juros efetiva dos Depósito Interbancário (DI), do período compreendido entre a data de negociação, inclusive, e a data de vencimento, exclusive. Já os contratos futuros de dólar comercial se baseiam na negociação da taxa de cambio de reais por dólar dos Estados Unidos.

Os contratos de cupom cambial usam como métrica o diferencial entre a taxa DI e a variação cambial para determinado período. Nos contratos de cupom cambial pode-se encontrar dois tipos: Cupom cambial sujo e cupom cambial limpo

a) Contrato Futuro de DDI (cupom cambial sujo), em que a variação cambial é medida a partir do dólar PTAX do dia anterior. Sendo o dólar PTAX calculado através

das médias das cotações apuradas a partir de consultas às instituições financeiras autorizadas a realizar operações cambiais.

b) Contrato Futuro de FRA (cupom cambial limpo), no qual a variação cambial é medida a partir do dólar à vista do dia

Os contratos futuros do Ibovespa negociam o índice de ações da Bolsa de Valores do Brasil para o último dia de negociação do contrato. O Ibovespa representa o valor atual de uma carteira teórica das ações mais negociadas na bolsa de valores do Brasil.

Para adentrar no mercado, o investidor necessita abrir uma posição, no qual é executado uma ordem de compra ou venda do ativo. Na hora de sair do mercado (fechar posição), o investidor necessita fazer uma operação contrária à original, ou seja, se ele executou uma ordem de compra anteriormente, é necessário realizar uma ordem de venda, caso o investidor entrou no mercado com uma ordem de venda, é necessário uma ordem de compra para fechar a posição.

Uma das grandes vantagens de operar no mercado de futuro é a possibilidade de negociar valores bem maiores que os disponíveis em conta, o que é comumente chamado de operar alavancado. Desse modo, é possível, por exemplo comprar e vender um contrato futuro no valor R$ 250000,00 sem ter essa quantia em posse na conta, apenas uma margem de garantia, sendo que, apenas o lucro ou prejuízo dessa operação que é recebido ou pago pelo investidor. Margem de garantia é simular a uma margem caução, pois o valor utilizado como garantia pode ser usado para arcar com parte dos prejuízos ocorridos com a operação. Para negociar, por exemplo, um contrato futuro de dólar é necessário ter como margem de garantia um valor de R$ 750,00 na conta da corretora de valores (CLEAR, 2022a).

Com o objetivo de atrair mais investidores para o mercado futuro, criaram-se os minicontratos que correspondem a 20% de um contrato futuro, permitindo que a margem de garantia necessária para operar esse ativo seja menor que o contrato futuro correspondente. Dentre os minicontratos negociados, há o minicontrato de dólar sendo o tamanho dele equivalente a US$ 10000,00, com a sua cotação expressa em reais por R$ 1000,00, ou seja, se o dólar valer R$ 5,00, a cotação do minicontrato será de R$ 5000,00 e representará a compra ou venda de R$ 50000,00. As cotação dos minicontratos de dólar variam diariamente, e são representados por pontos (diferença na cotação do minicontrato), que equivalem a R$ 10,00 cada, se negociado apenas com 1 minicontrato, e esse valor é escalonável, isto é, se forem operados 2 minicontratos, cada ponto equivalerá a R$ 20,00. Portanto, se o dólar estiver cotado a R$ 5,240, o minicontrato será negociado no valor de R$ 5240,00 e havendo a venda com o dólar à R$ 5,250, o seu respectivo ativo valerá R$ 5250,00 e o ganho estimado será de 10 pontos, que equivale à R$ 100,00, com 1

minicontrato sendo negociado. Para conseguir negociar um minicontrato de dólar é necessário apenas R$ 150,00 como margem de garantia, sendo assim, com apenas essa quantia é possível compra e vender um ativo que equivale à US$ 10000,00 (CLEAR, 2022b) e receber ou pagar apenas os ganhos/prejuízos da operação.

Além do minicontrato de dólar existe também o minicontrato de índice. Em alguns momentos, os minicontratos de dólar e de índice apresentam um certo antagonismo nos movimentos dos preços, por exemplo, quando investidores estrangeiros de grande porte veem que o cenário nacional não está muito propício para a injeção de dinheiro, devido à alguma instabilidade econômica ou política, retiram seu dinheiro do mercado brasileiro, fazendo o dólar subir, do mesmo modo o contrário se aplica, se o Brasil volta a se tornar atrativo para investimentos, havendo a aplicação de dinheiro, o dólar tende a abaixar. A cotação do minicontrato de índice também é cotado por pontos, só que cada ponto equivale à apenas R$ 0,20.

2.2.2 Fundos de Investimentos

Fundos de investimentos reúnem os recursos de diversas pessoas, que são aplicados em conjuntos no mercado financeiro. Os ganhos obtidos com essas aplicações são divididos entre os participantes desse fundo, proporcionalmente com o valor depositada por cada integrante. A soma total do dinheiro dos investidores é chamado de patrimônio do fundo e é aplicado em ativos selecionados por uma equipe de profissionais especializados. Os investimentos podem ser bem sucedidos ou não, o que impacta diretamente no valor das cotas do fundo. A divisão do patrimônio do fundo é chamado de cotas, quando um investidor aplica seu dinheiro em um fundo, ele está adquirindo cotas, ou seja, se ele investir R$ 100000,00 em um fundo que tem cotas valendo R$ 100 reais cada, ele adquirirá mil cotas.

Existem vários tipos de fundo a depender em quais tipos de ativos eles aplicam e a estratégia utilizada. Dentre eles, os mais comuns são:

a) Fundo cambial: investem acima de 80% do patrimônio do fundo em ativos que sejam relacionados a outras moedas (XP, 2022). Os mais conhecidos fundos cambiais são os de dólar, que tentam acompanhar a cotação da moeda americana. Eles podem ser uma boa opção para investidores que desejam ficar expostos a uma variação cambial

b) Fundo de ações: investem no mínimo 67% do patrimônio em ações negociadas na bolsa de valores, sendo o principal fator de risco a variação do preço dos papéis incluídos na carteira do fundo

c) Fundo Multimercado: investem em aplicações de renda fixa, moedas, ações e derivativos, sem compromisso de concentração em nenhum em especial

d) Fundo de Renda Fixa: investem pelo menos 80% do patrimônio em ativos de renda fixa, sendo assim, o principal fator de risco é a variação da taxa de juros

e) Fundo imobiliário: investem em empreendimentos imobiliários. A grande vantagem é que o investidor consegue investir em imóveis sem ter que comprar eles diretamente. Eles são listados na bolsa de valores como as ações e distribuem rendimentos periodicamente que são isentos de imposto de renda

A principal vantagem de investir em fundo de investimentos é contar com a gestão de profissionais especializados que decidirão os melhores ativos à serem adquiridos. Além disso, devido ao fato do fundo ter vários ativos alocados, permite ao investidor ter uma carteira diversificada mesmo sem ter muito dinheiro para investir.

2.2.3 Índice Sharpe

Umas das métricas comumente usadas para avaliar um ativo ou fundo de investimento considerando suas rentabilidades é o Índice de Sharpe, que foi desenvolvido por William Sharpe (SHARPE, 1994) e possibilitou o cálculo da rentabilidade de um ativo considerando o seu risco. O cálculo do índice é dado por:

$$S = \frac{(R_i - R_f)}{\sigma_i} \tag{7}$$

Onde, S é o índice Sharpe, R_i é taxa de retorno do investimento analisado, R_f é a taxa livre de risco e σ_i é o desvio padrão do retorno esperado, também conhecido como volatilidade.

A taxa de retorno representa a rentabilidade obtida em determinado período e a taxa livre de risco é a rentabilidade de de um investimento com risco zero.

No Brasil, a taxa Sistema Especial de Liquidação e de Custódia (Selic) ou o Certificado de Depósito Interbancário (CDI) são usados como referência na taxa livre de risco.

a) Taxa Selic: Representa os juros básicos da economia brasileira. Seu movimento influencia todas as taxas de juros praticadas no país. A taxa média diária praticada nas operações de títulos públicos federais administrada pelo Banco Central equivale

à taxa Selic. Essas operações são empréstimos de curto prazo, que geralmente tem vencimento de um dia, realizados entre instituições financeiras que usam títulos públicos como garantia. Ela é definida pelo Comitê de Política Econômica (Copom), que é um orgão do Banco Central.

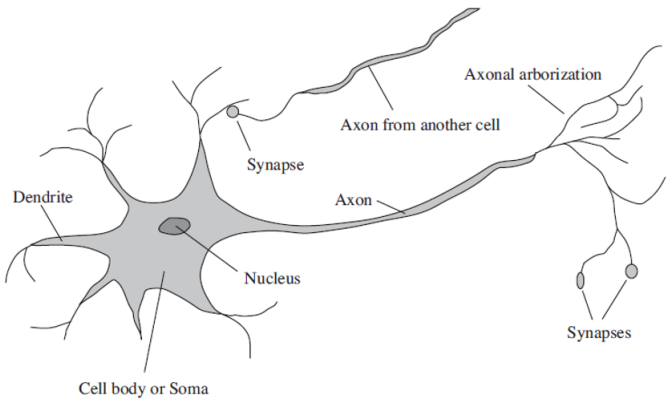
b) Taxa CDI: Representa empréstimos realizados entre instituições financeiras com seus próprios recursos para poder fechar seus balanços diários positivos. Na prática, a Taxa Selic e a CDI se equivalem.

Sendo assim, após o cálculo do Índice Sharpe, o melhor investimento será aquele que tiver o maior valor da razão Sharpe.

2.3 REDE NEURAL ARTIFICIAL

Algoritmos bio-inspirados têm se mostrado, ao longo dos tempo, um tremendo sucesso no campo da inteligência artificial. Estes algoritmos incluem os algoritmos evolucionários, a inteligência coletiva, as redes neurais e os sistemas imunológicos artificias (ENGELBRECHT, 2007). A rede neural artificial é uma forma de algoritmo bio-inspirado, que foi modelado baseado no sistema nervoso central do cérebro. A figura 6 apresenta um neurônio biológico. Se os sinais combinados recebidos pelos dendritos são forte o suficiente, o neurônio dispara um sinal de saída por um caminho chamado axônio.

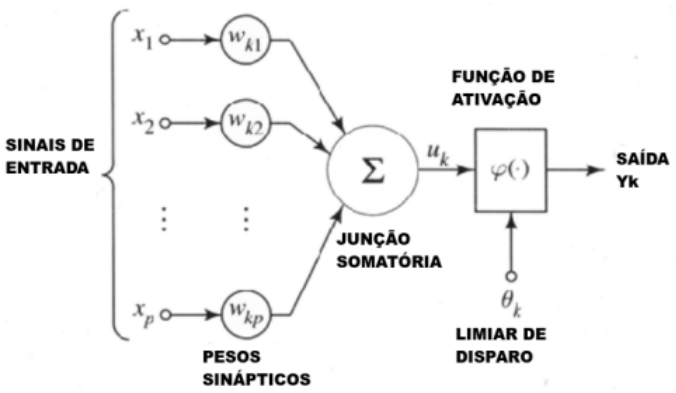
Figura 6 – Neurônio Biológico



Fonte: Russell et al. (2010)

O neurônio artificial (Figura 7) recebe sinais de outros neurônios e sobre certas condições o sinal é transmitido para outros neurônios artificiais. Sendo assim, os sinais de entrada chegam no neurônio artificial, há uma soma ponderada, e baseado na função de ativação há o envio de um sinal de saída para o próximo neurônio.

Figura 7 – Neurônio Artificial



Fonte: Adaptado de Rumelhart, Hinton e Williams (1986)

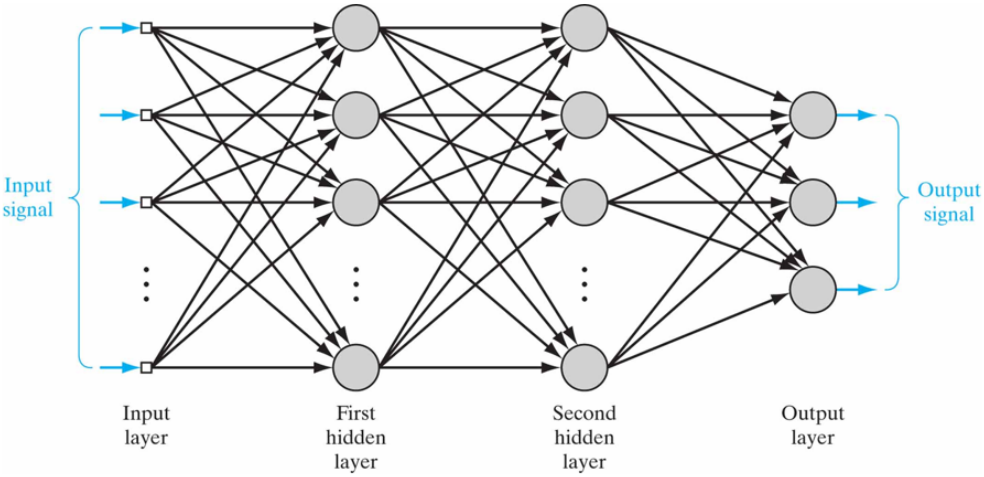
As redes neurais artificiais são formadas em camadas, possuindo um ou mais neurônios em cada. Normalmente possuem uma camada de entrada, camadas escondidas e uma camada de saída. Muitas arquiteturas de redes neurais artificias foram desenvolvidas, como as do tipo *feed-forward* e as recorrentes.

2.4 PERCEPTRON MULTI CAMADAS

As redes neurais do tipo *feed-forward* são capazes de moldar complexos modelos a partir de sua estrutura em camadas paralelas (OJHA; ABRAHAM; SNÁŠEL, 2017). Os elementos de processamento fundamentais dessas redes são os vários neurônios artificiais espalhados ao longo de várias camadas totalmente conectadas, sendo assim, qualquer elemento de uma dada camada alimenta todos elementos da camada seguinte. Um exemplo de rede desse tipo é a Perceptron Multi Camadas. As MLP são baseadas na comunicação entre três ou mais camadas: a de entrada, as ocultas e a de saída. A figura 8 representa uma rede MLP que possui duas camadas ocultas.

As redes Perceptron Multi Camadas normalmente são treinadas usando um algoritmo de retro-propagação (RUMELHART; HINTON; WILLIAMS, 1986), no qual os erros dos elementos processadores da camada de saída são retro-propagados para as camadas intermediárias. Sendo assim, as MLPs são treinadas aprendendo a corrigir os erros, portanto a desejada e correta saída do sistema deve ser conhecida.

Figura 8 – Estrutura de uma rede neural Perceptron Multi Camadas com duas camadas ocultas



Fonte: Haykin (2009)

2.5 APRENDIZADO PROFUNDO

Arquiteturas profundas são compostas de vários níveis de operações não-lineares, como nas redes neurais com várias camadas ocultas. Cada camada aplica uma transformação não-linear na entrada, repassando-a para a saída. O objetivo é aprender uma complicada e abstrata representação dos dados de uma maneira hierárquica, passando-os por múltiplas camadas transformadoras. A ideia de aprendizagem hierárquica em Aprendizado Profundo (*Deep Learning*) vem das áreas sensoriais primárias do neocórtex do cérebro humano (NAJAFABADI et al., 2015). Aprendizado profundo é um conjunto de técnicas de aprendizado de máquinas, que aprende em variados graus de abstração, onde características de alto-nível são definidas a partir de características de baixo-nível (AREL; ROSE; KARNOWSKI, 2010).

2.6 REDES NEURAIS CONVOLUCIONAIS

LeCun et al. (1989) aplicaram as rede neurais convolucionais no reconhecimento de dígitos, revolucionando o processamento de imagens. Desde então, melhores arquiteturas de CNN foram propostas para resolver as difíceis tarefas da visão computacional, como a rede AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, Geoffrey E., 2017), a VGGNet (SIMONYAN; ZISSERMAN, 2015), a GoogLeNet (SZEGEDY et al., 2014), e a ResNet (HE et al., 2016).

Redes neurais convolucionais podem extrair características locais, o qual é uma propriedade bem útil, já que variáveis espacialmente e temporalmente próximas geralmente são altamente relacionadas (LECUN; BENGIO, 1995). Sendo assim, elas são excelente extratores de

características dos dados de entrada. As redes neurais convolucionais possuem várias camadas, as quais podem ser caracterizadas como a camada convolucional, a de *pooling*, e a totalmente conectada.

2.6.1 Camada Convolucional

A camada convolucional é encarregada de realizar as operações de convolução nos dados. Supondo que a entrada da camada $l-1$ é uma matriz N x N, um filtro de convolução F x F é aplicado sobre ela. A entrada da camada l é dada pela Eq 8,

$$v_{i,j}^{l} = \delta(\sum_{k=0}^{F-1}\sum_{m=0}^{F-1} w_{k,m} V_{i+k,j+m}^{l-1}) \tag{8}$$

onde, $v_{i,j}^{l}$ é o valor da linha i e coluna j da camada l, $w_{k,m}$ é o peso da linha k e coluna m do filtro, e δ é a função de ativação.

A figura 9 representa o filtro F x F sendo aplicado sobre os dados, para resultar no valor de $v_{1,1}$ da próxima camada. Normalmente a saída de cada filtro é passada por uma função de ativação não-linear antes de entrar na próxima camada, sendo a mais comumente utilizada a ReLu (KRIZHEVSKY; SUTSKEVER; HINTON, Geoffrey E, 2012).

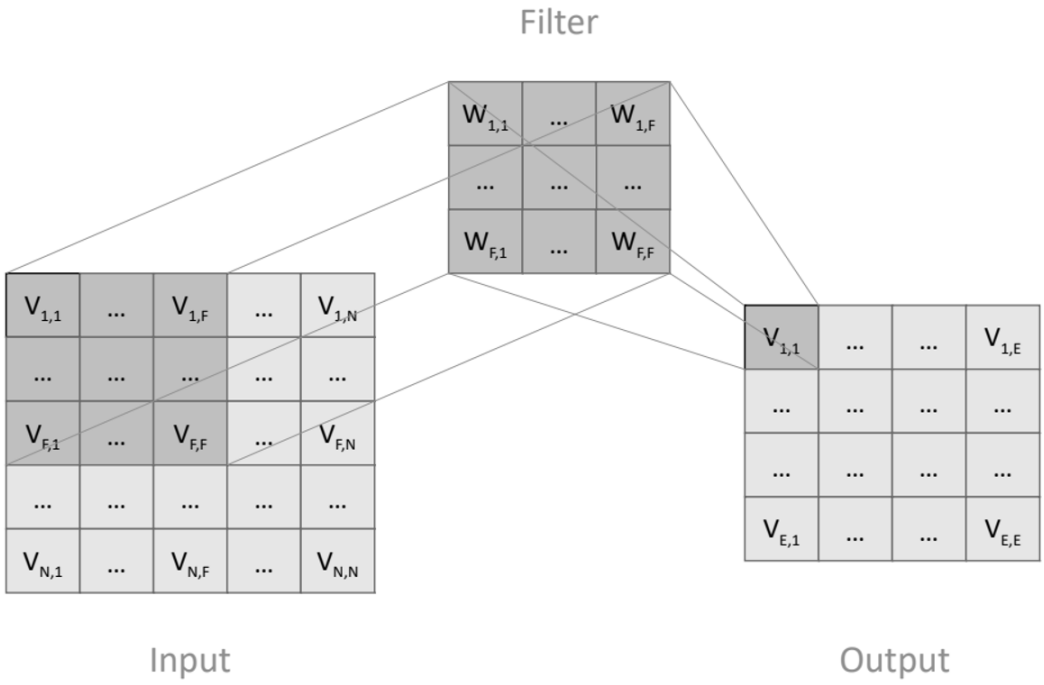
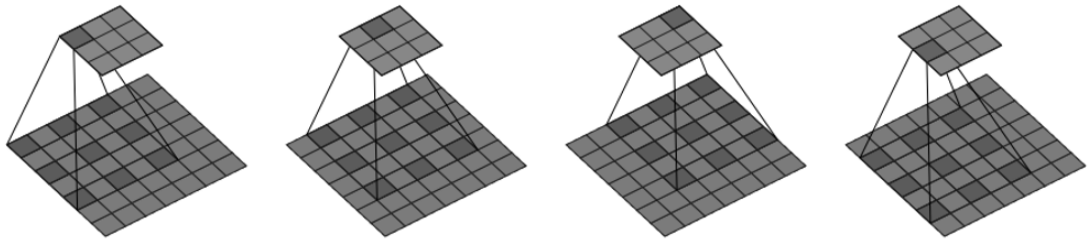


Figura 9 – Filtro F x F sendo aplicado no dado de entrada para a obtenção do valor $v_{1,1}$ da próxima camada

2.6.1.1 Camada de Convoluções Dilatadas

A camada convolucional apresentada pode ser substituída pela camada de convoluções dilatadas, na qual é inserido no filtro F x F espaços controlados por uma taxa de dilatação d, que normalmente indica que $d-1$ espaços serão inseridos no filtro, sendo $d=1$ correspondente à uma camada de convolução normal. Esse filtro dilatado se torna $F+(F-1)(d-1)$. A figura 10 representa a convolução com um filtro com $d=2$.

Figura 10 – Convolução com filtro 3 x 3 com uma taxa de dilatação de 2 sobre uma entrada 7 x 7



Fonte: Adaptado de: Dumoulin e Visin (2018)

As convoluções dilatadas são normalmente usadas para aumentar o campo receptivo da saída sem aumentar o tamanho do filtro ou das camadas convolucionais. Campo receptivo é a área captada na camada dada por cada entrada da camada seguinte. Na WaveNet (OORD et al., 2016) há a utilização de várias camadas de convoluções dilatadas para a produção de um som mais natural do que o método tradicional.

2.6.2 Camada de *Pooling*

A camada de *pooling* reduz o tamanho espacial dos dados, resultando em menos parâmetros e menos custo computacional no processo de treinamento. Sendo assim, essa camada auxilia no controle do problema de *overfitting*. *Overfitting* é quando o modelo se ajusta muito bem a um conjunto de dados, mas se mostra ineficaz para prever resultados a partir de dados inéditos. A função mais popular usada na camada de *pooling* é a maximização, que extrai o maior valor de uma janela escolhida.

2.6.3 Camada totalmente conectada

A camada totalmente conectada é uma rede MLP, que é responsável por converter os atributos extraídos nas camadas anteriores na saída final do modelo.

2.7 REDES NEURAIS RECORRENTES

Redes Neurais Recorrentes (*Recurrent Neural Networks*) (RNN) são um tipo de rede neural no qual os pesos computados no treinamentos são influenciados não apenas pela entrada atual da rede, mas também pelas saídas anteriores da mesma. Elas foram desenvolvidas para lidar com dados sequenciais e/ou temporais (BHATTACHARJEE; TOLLNER, 2016). Nas RNNs, a camada oculta além de receber dados da camada anterior, ela também lida com a saída passada.

As RNNs tem a capacidade de lembrar as observações decorridas, mantendo uma célula de estado que se atualiza a cada nova entrada. O principal problema da RNN é a dissipação dos gradientes, que durante o treinamento, eles tendem a zero, tornando o aprendizado lento ou até mesmo nulo. A partir dessa adversidade criou-se a *Long Short-Term Network* (HOCHREITER; SCHMIDHUBER, 1997)

2.7.1 *Long Short-Term Network*

A *Long Short-Term Network* foi desenvolvida para melhorar a habilidade das RNNs de relembrar as dependências de longo termo das observações passadas. Os $gates$ implementados nas unidades da LSTM resolveram o problema da dissipação dos gradientes, já que eles permitem que o gradiente mova de uma camada oculta para outra sem ser reduzido, permitindo que as primeiras camadas tenham a mesma qualidade de treinamento que as finais.

Como mostrado na figura 11 uma unidade da LSTM é composta de uma célula c_t, um *gate* de entrada i_t, um *gate* de esquecimento f_t e um *gate* de saida o_t. Esses três *gates* regulam o fluxo de informação dentro e fora da célula, fazendo com que a célula relembre de valores em variados intervalos tempos. Em um dado tempo t, a entrada da LSTM é definida por: valor de entrada x_t no tempo t, valor de saída h_{t-1} no tempo $t-1$ e o estado c_{t-1} no tempo $t-1$. A saída é definida por: valor de saída h_t no tempo t e o estado c_t no tempo t.

Numa LSTM o *gate* de esquecimento determina o impacto de c_{t-1} em c_t, o *gate* de entrada determina o impacto de x_t em c_t e o *gate* de saída controla o impacto de c_t sobre h_t. As formulas dos *gates* de esquecimento, de entrada, de saída são representadas por 9, 10, 11, respectivamente,

$$f_t = \sigma(W_f * h_{t-1} + W_f * x_t + b_f) \tag{9}$$

$$i_t = \sigma(W_i * h_{t-1} + W_i * x_t + b_i) \tag{10}$$

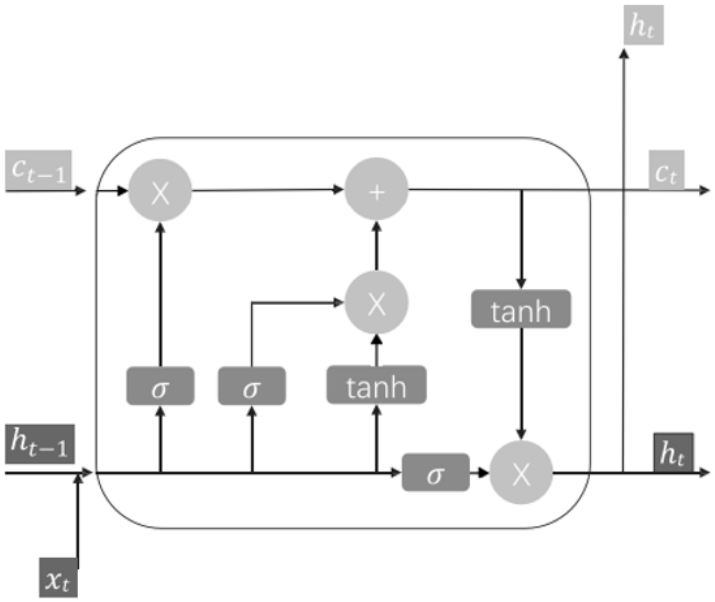


Figura 11 – Estrutura de uma *Long Short-Term Network*

$$o_t = \sigma(W_o * h_{t-1} + W_o * x_t + b_o) \tag{11}$$

onde nessas equações, σ representa a função de ativação sigmoide dada por: $\sigma(x) = 1/(1+e^{-x})$. W_f, W_i e W_o são os pesos e b_f, b_i e b_o são os termos bases dos *gates* de esquecimento, de entrada, de saída, respectivamente. A saída final da LSTM é definida a partir do *gate* de saída e do estado c_t, as equações 12, 13, 14 determinam esse valor de saída, onde $tanh$ representa a função de ativação tangente hiperbólica, definida por: $\tanh(x) = (e^z - e^{-z})/(e^z + e^{-z})$.

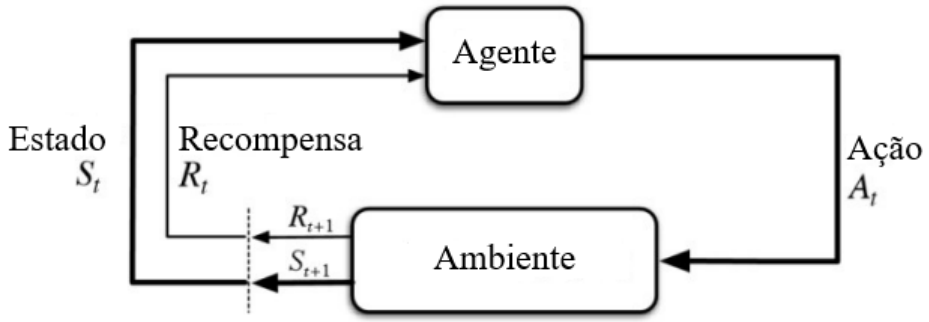
$$\overline{c_t} = tanh(W_c * h_{t-1} + W_c * x_t + b_c) \tag{12}$$

$$c_t = f_t * c_{t-1} + i_t * c_t) \tag{13}$$

$$h_t = o_t * tanh(c_t) \tag{14}$$

A LSTM aprende sozinha o que deve ser esquecido e adicionado à memória, e qual deve ser o valor de saída, sendo que cada uma das observações passadas influenciam nesse valor. Baseado nessas características, a LSTM ajuda em muito a resolver o problema da dissipação dos gradientes e permite que uma rede neural aprenda dependências de longa duração.

Figura 12 – Interação entre o Agente e o Ambiente no Aprendizado por Reforço



Fonte: Sutton e Barto (1998)

2.8 APRENDIZADO POR REFORÇO

A aprendizagem por reforço é uma abordagem computacional para entender e automatizar a aprendizagem direcionada para objetivos e a tomada de decisões (WATKINS, 1989). As inúmeras interações diretas com o ambiente e as recompensas recebidas por cada ação tomada, tem por objetivo conseguir a maior recompensa futura.

Em cada instante de tempo t, o agente recebe o estado atual s_t, escolhe uma determinada ação a_t, e a executa. Em seguida o ambiente envia para o agente uma recompensa $r_{t+1} = r(s_t, a_t)$ e para o estado seguinte $s_{t+1} = \delta(s_t, a_t)$, onde as funções δ e r são intrínsecas do ambiente e não necessariamente conhecidas pelo agente. O processo pode ser definido através do conjunto $< S,A,P,R >$, onde:

- Espaço de estados (S): Diversos atributos são usados para representar os estados. No mercado financeiro, na maioria dos casos, é utilizado o histórico dos preços de alta, baixa, abertura, fechamento e volume de negócios dos papeis, juntamente com indicadores técnicos relacionados (FISCHER; KRAUSS, 2018).
- Espaço de Ações (A): Em cada estado s $\in$ S, o agente deve escolher uma ação a de um conjunto de ações disponíveis no estado s
- Matriz de Probabilidade (P): É a probabilidade de o agente ir para o estado S_{t+1}, dado S_t e A_t
- Função Recompensa (R): É a recompensa recebida pela mudança no ambiente gerada pela ação tomada pelo agente. Quando $r(s_t, a_t)$ é positivo é considerado como lucro ou prêmio, e quando negativo é visto como custo ou punição

No Processo de Decisão de Markov (*Markov Decision Process*) (MDP), um estado é considerado de Markov se ele captura todas as informações relevantes do histórico. Portanto, as

funções r_{t+1} e s_{t+1} dependem somente do estado atual (s_t) e da ação (a_t), independendo dos estados e ações passadas. Além disso, em um MDP as ações tomadas pelo agente não influenciam apenas a recompensa imediata, mas também, as futuras ações, e portanto as recompensas futuras. O agente, quando toma uma ação, tem que aprender a equilibrar a recompensa imediata com a futura. A propriedade de Markov é importante para a aprendizagem por reforço, pois ela possibilita que as decisões sejam tomadas em função apenas do estado atual, ou seja, o estado futuro depende apenas do estado da decisão escolhida no presente.

Em alguns algoritmos de aprendizado por reforço envolvem a otimização de uma função valor. A função valor é uma função dos estados e quantifica para o agente o quanto bom é estar em determinado estado. A equação 15 expressa a função valor de um estado s sob uma política π.

$$V_\pi(s) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s \right] \tag{15}$$

Similar à função valor, tem-se a função ação-valor, o qual para uma política π expressa o valor de uma ação a em um estado s sob uma política π, e que é expressa pela equação 16.

$$Q_\pi(s,a) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s, A_t = a \right] \tag{16}$$

Onde (0 < γ < 1) é uma constante que determina o valor relativo das recompensas futuras versus imediatas.

Em problemas do mundo real, nem sempre o estado fica totalmente explícito para o agente, impossibilitando o estado (s) ser considerado de Markov, sendo assim, esse ambiente não pode ser modelado como uma MDP. Em vista disso, tem-se o POMDP que considera que o estado (s) representa apenas uma parte do ambiente, sendo então denominado como observação (HAUSKNECHT; STONE, 2017). O conjunto que define o POMDP passa a ser $< S,A,P,R,\Omega,O >$, onde S,A,P,R são similares aos do MDP, e o agente recebe uma observação $o \in \Omega$, que é gerada pelo estado s seguindo a distribuição de probabilidade $o \sim O(s)$.

2.8.1 *Q learning*

O algoritmo *Q learning* determina uma função Q (ação-valor), descrita pela equação 17

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma max_a Q(s',a') - Q(s,a)] \tag{17}$$

Onde s é o estado atual, a é a ação a ser realizada estando em s, r é o reforço recebido ao realizar a ação a no estado s, s' é o estado futuro, a' é a ação que será realizada estando no estado

s', γ é o fator de desconto (0 < γ < 1) e α é a taxa de aprendizado (0 $\leq$ α < 1). A implementação do algoritmo *Q learning* é descrito no algoritmo 1

Algoritmo 1 – Algoritmo *Q learning*

```
Inicia a tabela de valores Q, Q(s,a) ∀ s ∈ S, a ∈ A(s), arbitrariamente, e Q(terminal) = 0
para cada episodio faça
    Inicia estado S
    para cada passo de episódio ate S ser terminal faça
        Escolha A de S usando a política derivada de Q(ex, ϵ − greedy)
        Execute a ação A, observe R, S'
        Q(S,A) ← Q(S,A) + α[r + γ max_a Q(S',A') − Q(S,A)]
        S ← S'
    fim
fim
```

O *Q learning* consegue convergir para uma política ótima dado um espaço de estado simples e discreto, que pode ser acessado por uma tabela. Entretanto, para estados mais complexos e contínuos não há a possibilidade de serem representados em uma tabela. Sendo assim, faz-se a adoção das redes neurais artificias como aproximadores de função, que generalizam os estados desse ambiente complexo.

2.8.2 *Deep Q Network*

Em aplicações práticas de aprendizado por reforço, ambientes interessantes podem ter representações de estados complexas. Para simplificar o aprendizado em tais ambientes, pode-se empregar redes neurais profundas, capazes de aprender representações úteis para os dados de acordo com a função de erro a ser minimizada.

A Rede Q Profunda (*Deep Q Network*) (DQN) desenvolvida por Mnih et al. (2015) mostrou resultados bem expressivos quando aplicada em vários jogos de Atari 2600, tendo como entrada apenas os pixels da tela. Ela é composta por três camadas de redes neurais convolucionais seguida por duas camadas de redes totalmente conectadas, sendo que a de saída do modelo contém as ações possíveis.

Pelo fato de que com o uso de uma rede neural artificial para aproximar os valores da função Q (ação-valor) o processo de aprendizado se tornava instável, utilizou-se de duas redes neurais artificias além da adoção do *replay memory*.

O uso de uma segunda rede neural artificial chamada de *target network* para a geração dos valores $Q(s',a')$, além de sua atualização de parâmetros ser periódica, diferentemente da outra

rede (*main network*), faz com que o modelo se torne mais estável. O *replay memory* tem a função de descorrelacionar os dados de treinamento, o que contribui com a estabilidade e melhora a eficiência dos dados, já que reutiliza cada transição em muitas atualizações.

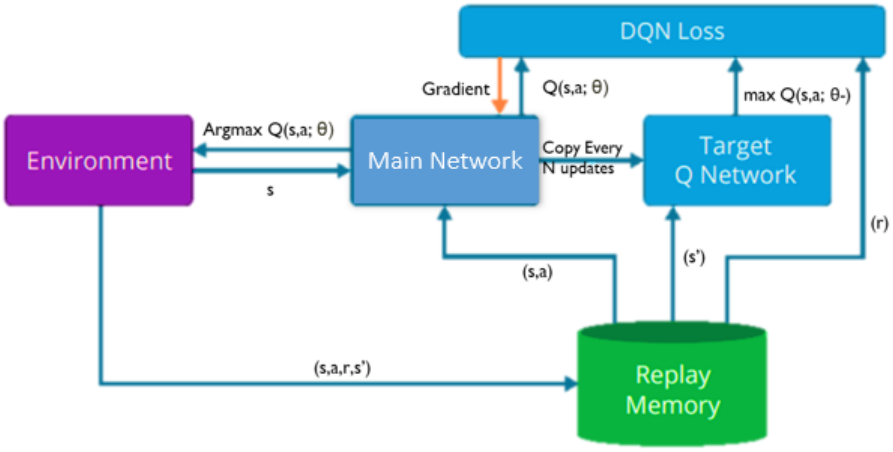
Para conseguir convergir, a DQN usa como estratégia de exploração do espaço de estados uma política gulosa ($\epsilon - greedy$), que faz com que o agente escolha uma ação de acordo com a probabilidade ϵ, podendo assim, ser uma ação aleatória ou baseada na função Q. Ao decorrer dos episódios, o valor de ϵ diminui gradativamente, resultando em maiores chances da ação ser fundamentada na função Q. Estratégia de exploração é necessária no aprendizado por reforço para o agente conseguir conhecer novas possibilidade de ações nos estados fornecidos.

A cada interação com o ambiente, o agente recebe o estado s_t, e de acordo com a probabilidade ϵ executa uma ação a_t, recebendo a recompensa r_t e o próximo estado s_{t+1}, formando assim, uma transição $e_t = (s_t, a_t, r_t, s_{t+1})$ que é armazenado no *replay memory* ($\mathcal{D}$). As transições do *replay memory* são amostradas aleatoriamente durante o treinamento. A equação 18 mostra a *loss function* durante o treinamento.

$$L_i(\theta_i) = E_{(s_t,a_t,r_t,s_{t+1})} \sim \mathcal{D}\left[(r + \gamma max_a Q(s',a';\theta^-) - Q(s,a;\theta_i))^2\right] \tag{18}$$

O processo de treinamento utilizado por Mnih et al. (2015) contento o *replay memory* e a atualização da *target network* a cada C passos está descrito no algoritmo 2 e ilustrado na figura 13.

Figura 13 – O algoritmo da DQN é composto de 3 principais componentes, a *Target Network*, a *Main Network* e o *Replay Memory*



Fonte: Adaptado de Nair et al. (2015)

Algoritmo 2 – Algoritmo do *Deep Q network*, adaptado de Mnih et al. (2015)

Inicia *Replay Memory* $\mathcal{D}$ com capacidade N
Inicia a função Q (ação-valor) (*Main Network*) com pesos aleatórios θ
Inicia a função $\overline{Q}$ alvo (ação-valor) (*Target Network*) com pesos $\theta^{-} = \theta$
para *episodio = 1 até M* **faça**
    Inicia a sequencia s_1
    para $t = 1$ *até* T **faça**
        Com probabilidade ϵ selecione uma ação aleatória a_t, caso contrario, selecione $a_t = argmax_a Q(s_t,a;\theta)$
        Execute a ação a_t no simulador e observe a recompensa r_t e o estado s_{t+1}
        Defina $s_{t+1} = s_t$
        Armazene transição (s_t,a_t,r_t,s_{t+1}) em $\mathcal{D}$
        Amostre um *minibatch* aleatório de transições (s_j,a_j,r_j,s_{j+1}) de $\mathcal{D}$
        Defina $y_i = \begin{cases} r_j & \text{se episodio terminar no passo } j+1 \\ r_j +_{a'} \overline{Q}(s_{j+1},a';\theta^{-}) & \text{caso contrário} \end{cases}$
        Realiza a descida de gradiente em $(y_j - Q(s_j,a_j;\theta))^2$ em relação a rede com os parâmetros θ
        A cada C passos $\overline{Q} = Q$
    fim
fim

A DQN leva em consideração o maior valor da função Q em todas as ações, mesmo não sendo a melhor ação a ser tomada dado o estado, ocasionando em uma superestimação nos valores Q. Para contornar tal problema desenvolveu-se a Rede Q Profunda Dupla (*Double Deep Q Network*) (DDQN) (HASSELT; GUEZ; SILVER, 2015).

2.8.3 *Double Deep Q Network*

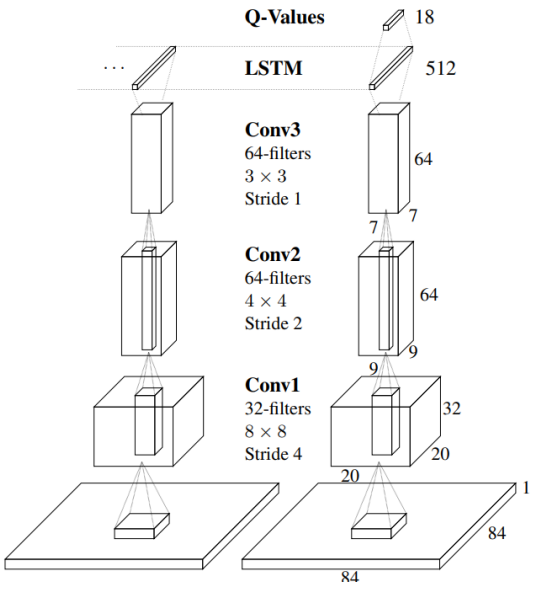
A *Double Deep Q Network* (HASSELT; GUEZ; SILVER, 2015) usa duas redes neurais com a mesma estrutura, assim como na DQN, e os parâmetros da *target network* são sincronizados com os da *main netwrok* periodicamente, promovendo assim mais estabilidade ao treinamento. Na DDQN as ações são selecionadas a partir da *main network* e a *target network* gera o valor Q para a ação.

2.8.4 *Deep Recurrent Q Network*

Para lidar com ambientes onde os estados (s) apresentam informações incompletas, Hausknecht e Stone (2017) desenvolveram a arquitetura do Rede Q Profunda Recorrente (*Deep Recurrent Q Network*) (DRQN). A estrutura da rede neural artificial e o algoritmo são praticamente os mesmos dos apresentados por Mnih et al. (2015), exceto que, há a substituição da primeira camada

totalmente conectada por uma camada do tipo *Long Short-Term Network* (HOCHREITER; SCHMIDHUBER, 1997), conforme a figura 14. Sendo assim, como as redes neurais do tipo recorrentes possuem capacidade de memória em relação às suas entradas anteriores, elas conseguem proporcionar uma melhor convergência dos valores Q, dado uma sequência de observações, pois são capazes de capturar as dependências temporais que podem existir, deixando a tomada de decisões melhor baseadas. Fazendo da DRQN uma boa abordagem para tarefas que requerem um modelamento da sequência de dados de entrada, como jogar um jogo ou mesmo desviar de objetos em um sistema de navegação autônoma de um robô (CHEN et al., 2021). Além dessa mudança na estrutura da rede neural, no treinamento é amostrado transições sequenciais, ao invés de transições aleatórias. A DRQN apresenta um custo computacional mais alto no treinamento, do que em relação à uma DQN, já que as redes recorrentes demandam de mais tempo de treinamento do que uma camada totalmente conectada. De modo geral, as DRQNs são uma poderosa ferramenta de aprendizado por reforço, que permitem que sistemas possam aprender e se adaptar à complexos ambientes.

Figura 14 – Arquitetura da rede neural usada na DRQN



Fonte: Hausknecht e Stone (2017)

2.9 CONCLUSÃO

Neste capítulo foi introduzido, a Brasil, Bolsa, Balcão (B3) e os ativos negociados nela, como os derivativos, dando ênfase nos contratos futuros. Além disso, foi apresentado como interpretar uma vela em um gráficos de velas de preço de um ativo, e discorrido como calcular um dos

indicadores técnicos utilizados na análise técnica, o MACD. E mesmo que de forma breve, foram introduzidos alguns conceitos de redes neurais artificiais, além de uma visão geral de aprendizado profundo, particularmente para redes neurais convolucionais e as do tipo recorrentes. Foi discorrido também o aprendizado por reforço, inclusive a sua versão profunda, a qual utiliza de redes neurais para a generalização dos estados. Dadas as técnicas básicas que serão usadas neste trabalho, o próximo capítulo irá expor os trabalhos apresentados nos últimos anos dessa área.

3 TRABALHOS RELACIONADOS

Apesar da existência da linha de pesquisa que visa a previsão dos preços e movimentos dos ativos utilizando o aprendizado profundo, esta seção dará ênfase em estudos que desenvolveram agentes negociantes de ativos, utilizando o aprendizado por reforço profundo, em especial a DRQN e a DQN. Apesar disso, também há a inclusão de um artigo relacionados ao uso de redes convolucionais dilatadas combinadas com as redes recorrentes. Todos os artigos apresentados contribuíram de alguma maneira com o projeto.

3.1 *APPLICATION OF DEEP REINFORCEMENT LEARNING ON AUTOMATED STOCK TRADING* (CHEN; GAO, 2019)

Inspirado no sucesso da *Deep Q Network* no Atari (MNIH et al., 2015), este artigo tenta aplicar a mesma ideia para desenvolver um agente que negocia ativos considerando o mercado de ativos como um jogo, no qual a busca de maximização da recompensa é o lucro. Uma variante da DQN é testada pelos autores: a *Deep Recurrent Q Network (DRQN)*. Para a realização do treinamento e teste é utilizado o histórico de preço de 19 anos do SP500 ETF, sendo apenas os cinco primeiro para treinamento e os demais anos para teste. Para efeito de comparação, o modelo proposto é comparado à estrategia de *buy and hold* e a um agente DQN.

Para o espaço de estados foi utilizado o preço de ajuste de fechamento sobre uma janela de 20 dias decorridos. No espaço de ações, o agente tem três opções: comprar, vender ou fazer nada. Já na função recompensa, foi utilizado a diferença do preço de ajuste do dia subsequente, baseado em qual ação o agente tomou.

O agente baseado na DRQN foi o melhor agente, superando ambos modelos de comparação, obtendo um retorno anual estimado em 22-23%.

3.2 *FINANCIAL TRADING AS A GAME: A DEEP REINFORCEMENT LEARNING APPROACH* (HUANG, 2018)

Há o emprego de um agente baseado na DRQN, com algumas modificações como: um *replay memmory* pequeno, a adoção de uma técnica de argumentação das ações para retirar a necessidade da exploração aleatória e além disso, para a realização da descida do gradiente, que passa a ser realizado em um intervalo de n passos, é amostrado do *replay memory* uma sequência bem longa.

No espaço de estados há a presença de 16 atributos (preço de abertura, alta, miníma, fechamento, volume e alguns indicadores técnicos) e o agente pode realizar apenas três ações: compra, venda e manter. Além disso, na função recompensa há o uso do retorno logaritmo do valor do portfólio.

O modelo foi aplicado no mercado de *forex* em 12 pares de moedas entre janeiro de 2012 e dezembro de 2017, e foi utilizado os preços com intervalo de 15 minutos. Em algumas paridades conseguiu-se obter um retorno anual de 60%, sendo a média total de 10%.

3.3 *ADAPTIVE STOCK TRADING STRATEGIES WITH DEEP REINFORCEMENT LEARNING METHODS* (WU et al., 2020)

Foi apresentado o modelo Gated Deep Reinforcement Q-Learning (GDQN), no qual há a utilização das redes recorrentes do tipo Gated Recurrent Units (GRU) para a extração automática das informações relevantes do mercado de ações.

No espaço de estados foi utilizado os preços de abertura, alta, miníma, fechamento, volume atrelados à alguns indicadores técnicos. O espaço de ações era composto por três ações: comprar, vender e manter. E na função recompensa foi utilizado o índice de Sortino (SORTINO; PRICE, 1994).

O modelo usou dados do início de 2008 ate o final de 2018 de 15 ativos dos Estados Unidos, Reino Unido e China, sendo os oito anos primeiros utilizados para treinamento. O melhor resultado do modelo GDQN foi em uma ação chinesa, obtendo um retorno de 171% e um índice de Sortino de 1.79. O modelo performou melhor que as estratégias de comparação, *Turtle Trading System* e a *Direct Reinforcement Learning Trading Strategy* (DENG et al., 2017). Além disso, o GDQN foi comparado com o modelo Gated Deterministic Policy Gradient (GDPG), mostrando que o GDPG é mais estável e proveu resultados um pouco melhores.

3.4 *DEEP REINFORCEMENT LEARNING FOR TRADING* (ZHANG; ZOHREN; ROBERTS, 2020)

Os autores implementaram os algoritmos da DQN, Policy Gradient (PG) (WILLIAMS, 1992) e a Advantage Actor-Critic (A2C) (MNIH et al., 2016) usando na construção da rede neural, as redes recorrentes do tipo LSTM.

No espaço de estados foi utilizado o preço de fechamento, retornos em variados períodos e alguns indicadores técnicos. Dependendo do algoritmo o espaço de ações variava entra discreto

e contínuo, no discreto era composto de três ações (-1,0,1) e no contínuo o valor da ação variava entre $[-1,1]$. Já a função recompensa era em função do lucro da operação atrelado ao seu risco.

Os dados utilizados variam entre 2005 e 2019, sendo que, os modelos eram retreinados a cada cinco anos. Além disso, os modelos foram comparados com outras estratégias como o *Buy and Hold*, e foi demonstrado que o algoritmo da DQN obteve o melhor desempenho dentre todos.

3.5 *AN APPLICATION OF DEEP REINFORCEMENT LEARNING TO ALGORITHMIC TRADING* (THÉATE; ERNST, 2021)

Os autores apresentaram o modelo Trading Deep Q Network (TDQN), que é baseado no algoritmo da DDQN.

Para o espaço de estados utilizou-se os preços de abertura, alta, miníma, fechamento, volume, além do posicionamento do agente em relação ao ativo, se esta comprado ou vendido. O agente pode comprar e vendar quantas ações achar necessárias ou simplesmente se abster. Em relação à função recompensa é utilizado o retorno diário, independente da ação executado pelo agente.

A base de dados consistia de 30 ações variadas, sendo que, os dados de treinamento iam de 01/01/2012 até 31/12/2017 e os de teste de 01/01/2018 até 31/12/2019. O modelo foi comparado à estrategias de *Buy and Hold*, *Sell and Hold*, seguidor de tendência com médias moveis (TF) e com a reversão à média com médias moveis (MR). Em relação às duas últimas estrategias, o TDQN superou com facilidade elas, mas com as outras duas o modelo apresentado obteve um desempenho semelhante.

3.6 *APPLICATION OF DEEP REINFORCEMENT LEARNING IN STOCK TRADING STRATEGIES AND STOCK FORECASTING* (LI; NI; CHANG, 2020)

Foi realizado uma comparação entre os três algoritmos: DQN, *Double* DQN e *Dueling* DQN.

O espaço de estados era composto de vários indicadores técnicos e o agente podia realizar apenas umas das três ações: de comprar, vender ou manter. Na função recompensa foi utilizado o lucro ou prejuízo obtido pelo agente decorrente de suas ações.

O modelo foi aplicado em 10 ações aleatórias seguindo a proporção de 4:6 entre treinamento e teste. Na comparação entre os três algoritmos, a DQN apresentou os melhores resultados, superando os algoritmos considerados a sua otimização (*Double* DQN e *Dueling* DQN).

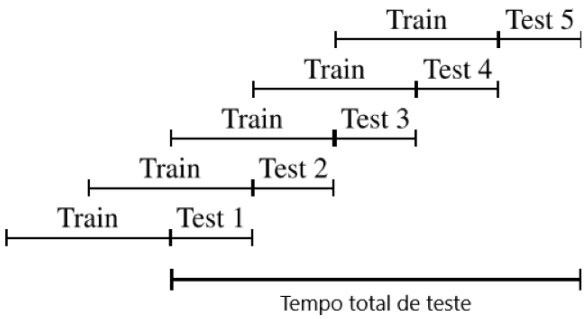
3.7 *REINFORCEMENT LEARNING APPLIED TO FOREX TRADING* (CARAPUÇO; NEVES; HORTA, 2018)

Há a apresentação de um modelo baseado no algoritmo da Double DQN aplicado no mercado de *Forex*

No espaço de estados há a implementação de um método de preprocessamento no valor dos preços dos pares de moedas, o posicionamento do agente em relação ao ativo, o valor do ganho/prejuízo da operação aberta pelo agente, o valor do portfólio em relação ao valor inicial, além disso, por ser uma aplicação no mercado forex que funciona 24/7, os autores inseriram também duas variáveis que representam as horas, já que hé certos períodos do dia que há um volume maior de negociações. No espaço de ações foi adotado apenas três ações possíveis (comprar, vender, manter), as quais dependem do estado do agente em relação ao ativo. Já a função recompensa variava em relação à ação tomada pelo agente, se ele fecha uma posição a recompensa recebida é o valor obtido com a operação, já se o agente abre ou mantém uma posição, a recompensa é a diferença entre o valor do ganho/prejuízo da operação em andamento no momento t com o momento $t + 1$.

Para o treinamento e teste do modelo foi utilizado dados de 2010 até 2017 do par EUR/USD, e adotado um sistema de janela deslizante, descrito na imagem 15, numa tentativa de amenizar a dificuldade de generalização do modelo treinado, fato este que ocorre devido à volatilidade existente no mercado. Como resultado, foi apresentado um retorno médio anual de 16,3±2,8%. Os autores enfatizaram que um dos maiores problemas da aplicação no mercado financeiro é a dificultado da generalização do modelo treinado, devido ao fato do mercado não ser estacionário

Figura 15 – Janela deslizante



Fonte: Adaptado de Carapuço, Neves e Horta (2018)

3.8 *DEEP Q-TRADING* (WANG et al., 2017)

O *Deep Q-trading* é baseado no algoritmo da DQN, onde foi adotado um aprendizado online, ou seja, a cada dia é adicionado transições s_t,a_t,r_t,s_{t+1} no *replay memory* que simulam as variadas ações possíveis, e após esse processo é realizado o treinamento padrão, há amostragem de um *batch* e a descida de gradiente resultante. Além disso, o custo por operação não foi considerado no estudo.

As diferenças de preços de fechamento diárias de uma janela de 200 dias foram usados no espaço de estados e o agente pode realizar apenas uma das três ações: comprar, vender ou manter. Na função recompensa foi usado o valor acumulado em 100 dias passados.

Foi utilizado no estudo duas ações, uma de Hong Kong e outra dos Estados Unidos. Cada base de dados apresenta 15 anos (01/01/2001 - 31/12/2015), sendo que para o treinamento inicial do modelo foi utilizado de 01/01/2001 até 21/12/2004, e para os testes e a realização do aprendizado online de 01/01/2005 até 31/12/2015. O modelo foi comparado com as estratégias de *Buy and Hold* e com um modelo que utiliza Recurrent Reinforcement Learning (RRL) (MOODY; WU, 1997; MOODY et al., 1998). Foi demonstrado que o *Deep Q-trading* apresenta um melhor desempenho que as outras estratégias de comparação, e ainda conseguiu se adaptar bem aos movimentos do mercado.

3.9 *REINFORCEMENT LEARNING IN STOCK TRADING* (DANG, 2019)

Foi comparado a DQN, *Double DQN* e a *Dueling Double DQN* com um modelo de aprendizado supervisionado (RNN-LSTM), que segue uma estratégia gulosa, ou seja, sempre que o modelo indicar que o preço do ativo irá subir, é realizada a compra dele, o mesmo acontece se o modelo indicar que preço irá descer.

O histórico dos preços dos ativos foi utilizado no espaço de estados, e o agente pode realizar apenas uma das três ações: comprar, vender ou manter. Na função recompensa foi empregado o lucro ou prejuízo obtido após realizar uma ação.

A base de dados era composta de mais de 7000 ações norte americanas. Para todos os ativos o treinamento era composto de dados diários de 01/01/2015 até 31/12/2016, e para os testes de 01/01/2017 até 10/11/2017. De modo geral, a DQN foi a que apresentou a maior média de lucro, mas também foi a que obteve a maior volatilidade. O modelo que utiliza o aprendizado supervisionado foi o que alcançou ganhos mais estáveis.

3.10 *APLICATION OF DEEP REINFORCEMENT LEARNING FOR INDIAN STOCK TRADING AUTOMATION* (BAJPAI, 2021)

Foi efetuado a comparação dos algoritmos da DQN , *Double DQN*, e *Dueling Double DQN* aplicados em 10 ações do mercado indiano.

O histórico de preços e volume das dez ações indianas foram usadas no espaço de estados, e divididos igualmente entre treinamento e teste. O espaço de ações consiste em três ações: comprar, vender e manter.

Na aplicação apresentada, a *Dueling Double DQN* foi a que apresentou os melhores resultados, seguida pela *Double DQN*.

3.11 *ROBUST FOREX TRADING WITH DEEP D NETWORK (DQN)* (SORNMAYURA, 2019)

Foi apresentado um modelo baseado na DQN e a sua aplicação no mercado de *forex*.

No espaço de estados foi utilizado o preço de fechamento, a sua diferença e médias móveis com variados períodos, além do uso de um indicador cíclico, como uma senoide. O agente podia realizar quatro ações: comprar, vender, manter e fechar. A função recompensa era baseado no valor adquirido com a realização de determinada ação.

O estudo se utilizou de 15 anos de dados das moedas EURUSD e USDJPY, sendo divididos entre treinamento (01/01/2001-31/12/2003) e teste (01/01/2004-31/12/2015). Nenhuma taxa de transação foi utilizada. O desempenho do modelo foi comparado com a estratégia de *Buy and Hold* e com um negociante de ativos experiente. Foi demonstrado que o modelo consegue obter melhores resultados que a estratégia de *Buy and Hold* quando aplicado em ambas moedas. Entretanto, em comparação ao negociante experiente, apenas quando o modelo foi aplicado no par EURUSD que se obteve melhores resultados.

3.12 *STOCK PRICE PREDICTION WITH CNN-LSTM NETWORK* (GUAN; LI; LU, s.d.)

Foi introduzido um modelo de aprendizado supervisionado combinado de CNN-LSTM para realizar a predição do preço da ação, aplicado como um sistema de negociação de ativos.

A principal base de dados provém do índice SP500 que varia desde janeiro de 2000 até abril de 2020. Para o treinamento foi utilizado 90% dos dados disponíveis e os restantes 10% foram usados na validação do modelo. Foi demonstrado que o uso de dilatações nas camadas

convolucionais resultam em uma melhora na previsão da tendência do ativo. Além disso, a adição de indicadores técnicos e econômicos beneficiou apenas a previsão à longo prazo, a curto prazo não apresentou nenhuma vantagem.

3.13 *IMPROVED METHOD OF STOCK TRADING UNDER REINFORCEMENT LEARNING BASED ON DRQN AND SENTIMENT INDICATORS ARBR* (ZHOU; TANG, 2021)

Os autores implementaram um sistema de negociação de ativo que era composto por 2 módulos, um baseado na DRQN e o outro no indicador de sentimento ARBR (DRQN-ARBR).

Utilizaram dados de 5 anos de uma ação chinesa, a Tapai Group, de janeiro de 2017 até julho de 2021. Os preços do ativo alimentavam os dois módulos simultaneamente, e cada um tinha como saída uma das 3 ações possíveis (compra, venda, fazer nada) e se ambas ações fossem iguais, o sistema executava no mercado a ação.

O modelo foi comparado com uma estratégia baseado no indicador MACD, numa LSTM, e com o modelo tendo apenas o módulo da DRQN. O modelo DRQN-ARBR foi o que obteve os melhores resultados dado o período de testes, seguido pela DRQN.

3.14 CONCLUSÃO

Com exceção da pesquisa de Guan, Li e Lu (s.d.), por não ser aprendizado por reforço, todas as outras apresentaram um estudo baseado na *Deep Q Network* ou em suas otimizações (DDQN, DRQN), demonstrando, com êxito, a sua possibilidade de ser empregada em um sistema de negociação de ativos.

De certa forma, todos os trabalhos impactaram neste trabalho, mas os mais relevantes foram os de Chen e Gao (2019) e Huang (2018), os quais aplicaram com sucesso a *Deep Recurrent Q Network*. Além disso, o estudo de Chen e Gao (2019) também expôs que a DRQN obteve melhores resultados que a DQN.

Apesar dos pesquisadores empregarem seus modelos em variados mercados como o *forex*, ações, índices (SP500), ainda não se teve uma aplicação nos derivativos, especificamente, os minicontratos futuros. O próximo capítulo discorrerá sobre a proposta de um sistema de negociação de minicontratos futuros.

4 PROPOSTA

Este trabalho tem como objetivo de propor um sistema baseado no aprendizado por reforço profundo, para realizar operações de compra e venda de um ativo, visando sempre o maior retorno financeiro.

Considerando o mercado de ativos como um Processo de Decisão Markoviano Parcialmente Observável, já que os os estados s representam apenas uma observação do ambiente, pois não conseguem explicitá-lo totalmente, a proposta consiste em utilizar um sistema fundamentado na *Deep Recurrent Q Network*, onde foi utilizado uma combinação das redes convolucionais dilatadas com as redes recorrentes do tipo LSTM. Sendo assim, as CNNs se encarregaram por extrair as características mais importantes do espaço de estado, em seguida, passando as informações obtidas para a LSTM, a qual ficou responsável por abstrair as informações contidas no tempo. Como demonstrado anteriormente nos trabalhos de Chen e Gao (2019) e Huang (2018), a DRQN proposta por eles apresentaram bons resultados quando usada como sistema de compra e venda de ativos. O ativo escolhido a ser negociado pelo modelo é o minicontrato futuro de dólar, e como a sua variação de preço é cotado em pontos, utilizou-se deles para a função recompensa no treinamento do modelo. A figura 16 ilustra o modelo da rede neural usado na DRQN.

Figura 16 – Estrutura da rede neural

Fonte: Autor

4.1 ESPAÇO DE ESTADOS

Em um instante de tempo t, as seguintes informações são observáveis pelo agente:

a) preço de abertura do minicontrato do dólar;
b) preço na alta (máxima) do minicontrato do dólar;
c) preço na baixa (mínima) do minicontrato do dólar;
d) preço de fechamento do minicontrato do dólar;
e) volume negociado de minicontrato do dólar;
f) preço de fechamento do minicontrato do índice Ibovespa;
g) volume negociado de minicontrato de índice Ibovespa;

h) posicionamento do agente, em forma ortogonalizada:
   - comprado ([1,0,0]): no qual o agente compra o ativo com a expectativa de lucrar na sua venda;
   - vendido ([0,1,0]), no qual o agente espera lucrar com a descida do preço do ativo;
   - fora do mercado ([0,0,1]), onde o agente não possui nenhum ativo em posse.

i) valor acumulado com a operação (compra/venda) em aberto (lucro/prejuízo);

A presença do preço de fechamento e volume negociado do minicontrato de índice Ibovespa se justifica, uma vez que, em alguns momentos, os minicontratos de índice e dólar apresentam movimentos antagônicos em seus preços, também há a presença do lucro ou prejuízo do agente referente à sua operação em aberto (comprado ou vendido).

A representação de um estado s_t incorpora os 11 valores das variáveis supracitadas nos últimos 30 instantes de tempo, resultando em uma representação matricial de dimensões 30×11 e está representado na figura 17.

Figura 17 – Representação de um estado s no instante t

	Abertura	Alta	Baixa	Fechamento	Volume	Fechamento (Índice)	Volume (Índice)	Lucro/Prejuízo	Posicionamento do Agente		
t											
t-1											
t-2											
t-3											
t-4											
t-5											
t-6											
t-7											
t-8											
t-9											
t-10											
t-11											
t-12											
t-13											
t-14											
t-15											
t-16											
t-17											
t-18											
t-19											
t-20											
t-21											
t-22											
t-23											
t-24											
t-25											
t-26											
t-27											
t-28											
t-29											

Fonte: Autor

Considerando o que foi enfatizado no trabalho de Carapuço, Neves e Horta (2018), que uma das maiores dificuldade encontrada foi a generalização do modelo, devido ao mercado não ser estacionário, foi realizado uma normalização dos dados de entrada. Com exceção das três

colunas referentes ao posicionamento do agente, os dados foram normalizados entre 0 e 1 usando a normalização min-max de acordo com a equação 19. Depois desse processo, a matriz com todos os seus atributos é enviada para o agente, como um estado s.

$$z_i = \frac{xi - min(x)}{max(x) - min(x)} \tag{19}$$

4.2 ESPAÇO DE AÇÕES

Adotou-se um espaço de ações discreto de 3 ações, sendo cada ação aplicada diretamente sobre o papel. As ações são: compra, venda e manter posição atual. Se duas ações subsequentes forem iguais (e.g. comprar - comprar), a segunda ação é considerada como manter (comprar - manter). O agente pode operar comprado ou vendido.

4.3 AMBIENTE

O ambiente fornece para o agente as informações necessárias paras a sua aprendizagem. A partir dele são armazenadas o histórico das operações que juntamente com os preços do ativo, enviam o reforço ideal para o agente, além dos estados s. Também há a presença do valor do portfólio do agente, ou seja, o quanto de pontos ou que o agente conseguiu acumular com as operações realizadas.

Em cada episódio (de treinamento ou teste), o agente começa com o seu portfólio zerado, ou seja, possuindo zero pontos de dólar. Toda vez que o agente completa uma operação (lucrativa ou não), o valor adquirido é somado ao seu portfólio, já descontado as taxas transacionais existentes, que serão explicitadas na próxima seção.

4.4 REFORÇO

Como o objetivo é ter o maior retorno financeiro, a função recompensa é apenas o lucro ou prejuízo obtido na operação concluída, subtraindo-se as taxas transacionais pré determinadas pela B3.

No período que foi realizado este trabalho as taxas estipuladas pela B3, referentes a uma ação de compra ou venda de apenas 1 contrato de dólar, eram formadas pela taxa de registro e de emolumentos que estão discriminadas na Figura 18, e somadas correspondiam à US$ 1,05, mas

como 1 minicontrato de dólar equivale à 20% do contrato de dólar, as taxas também seguem o mesmo princípio, resultando em US$ 0,22.

Figura 18 – Taxa de registro e emolumentos que compõem o custo pela compra ou venda de 1 contrato de dólar. Sendo o ADV o número de contratos negociados no mês anterior.

ADV		Emolumentos (US$)	Taxa de registro	
De	Até		Componente variável (US$)	Componente fixo (US$)*
1	2.800	0,398	0,652	0,0319502

Fonte: https://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/listados-a-vista-e-derivativos/moedas/tarifas-de-dolar-dos-estados-unidos/futuros-de-dolar/, acessado em dezembro, 2021

O agente só recebe a recompensa quando a posição dele é fechada, por exemplo, se houve uma compra, ele só ganhará o reforço quando vender, e o valor da recompensa será o valor do preço de venda menos o de compra, descontadas as taxas. Vale ressaltar que para os treinamentos e testes do modelo não foi considerado a taxa do imposto de renda de 15% que incide sobre o lucro gerado no mês, optou-se por considerá-lo apenas na análise dos resultados.

5 MATERIAIS E MÉTODOS

O projeto foi implementado utilizando a Anaconda (ANACONDA, 2022) que é uma ferramenta de gerenciamento de pacotes que possibilita a criação de vários ambiente virtuais e tem sua distribuição gratuita, além de ser *open-source*. Focado para a computação científica, a Anaconda é usada em projetos de ciência de dados, aprendizado de máquina, análise preditiva e projetos científicos em geral. A versão do Python usada foi a 3.8, em conjunto com a biblioteca Keras (KERAS, 2022), para a implementação das redes neurais e o *framework* Tensorflow (TENSORFLOW, 2022) para a implementação do algoritmo da DRQN.

A biblioteca Keras utilizada é baseada no Tensorflow e foi desenvolvida para possibilitar rápida implementação das redes neurais profundas, com uma interface modular, intuitiva e extensível.

Os modelos foram treinados e testados em um computador com um AMD Ryzen 5 3600X, 32 GB de RAM, contendo uma placa de video GTX 1660 com 6 GB de memoria, em sistema operacional Windows 10. O maior problema do aprendizado por reforço profundo é o custo computacional, o que tornou o estudo extremamente extenso.

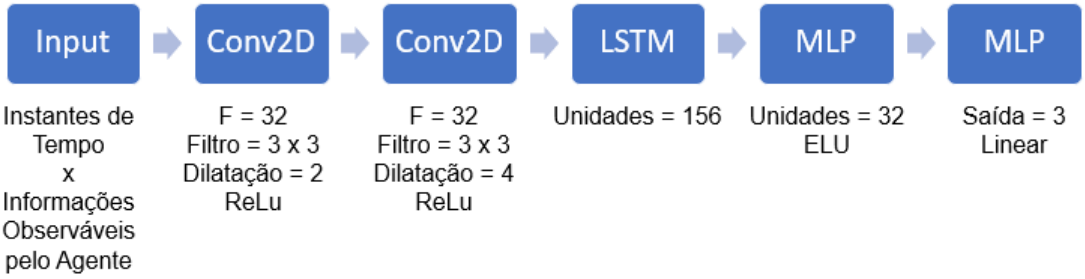
5.1 ARQUITETURA DA REDE

Como apresentado anteriormente, este trabalho empregou uma rede neural composta por CNNs, LSTM e MLP. Optou-se pela adoção das convoluções 2D para captar as inter-relações dos atributos, pois com a convolução 1D isso não seria possível. As camada das redes convolucionais possuíam 32 filtros, sendo o filtro uma matriz 3x3, com uma taxa de dilatação de 2 na primeira camada e 4 na segunda. A função de ativação usada nelas foi a função linear retificada, ReLu (LECUN; BENGIO, Y.; HINTON, 2015). A LSTM foi composta de 156 unidades, seguida de uma camada totalmente conectada com 32 unidades com função de ativação ELU (CLEVERT; UNTERTHINER; HOCHREITER, 2016). Como o modelo tem apenas três ações, a saída da rede neural contém 3 neurônios com função de ativação linear. A Figura 19 ilustra a arquitetura da rede utilizada.

5.2 DADOS UTILIZADOS E TREINAMENTO

Para os experimentos realizados, foram utilizados os preços históricos do minicontrato de dólar e índice, intervalados de 15 minutos, sendo assim, cada ação realizada pelo agente teve um hiato de

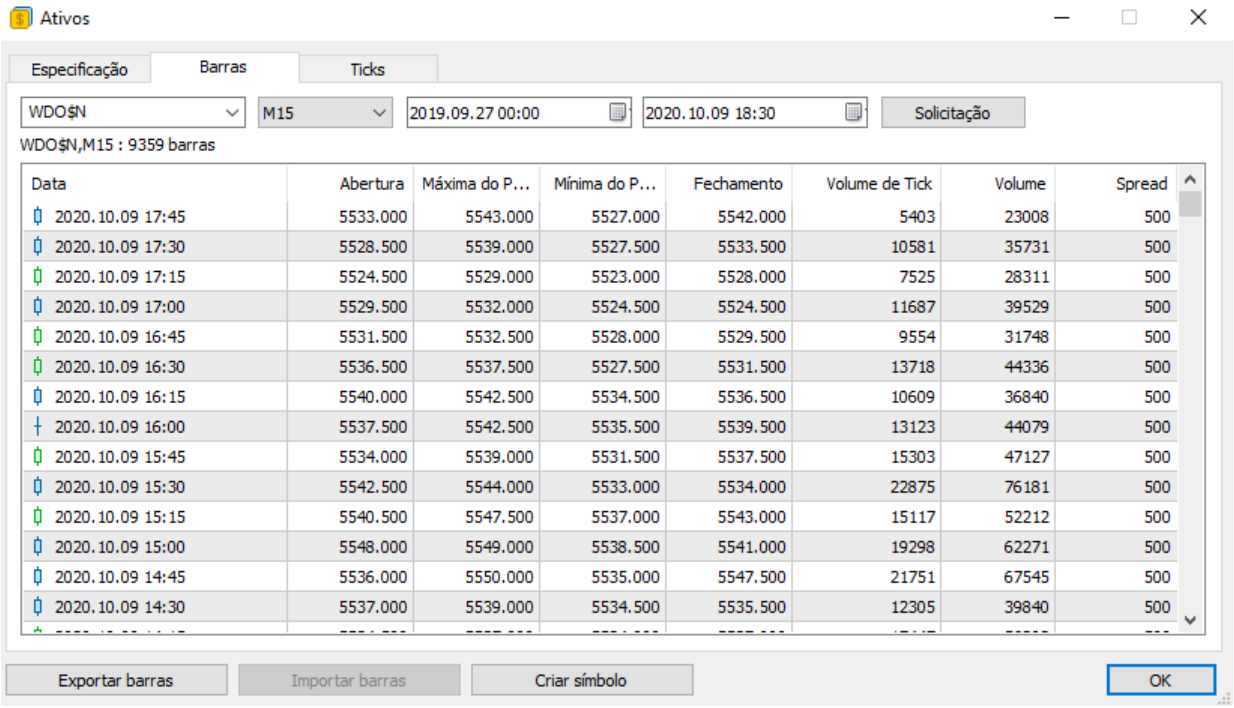
Figura 19 – Arquitetura da rede



Fonte: Autor

15 minutos. Os dados foram retirados da plataforma Metatrader 5 (METATRADER, 2022) com a função *Exportar barras* presente na aplicação (Figura 20) e englobam o período de 27/09/2019 até 09/10/2020.

Figura 20 – Exportar dados Metatrader 5



Data	Abertura	Máxima do P...	Mínima do P...	Fechamento	Volume de Tick	Volume	Spread
2020.10.09 17:45	5533.000	5543.000	5527.000	5542.000	5403	23008	500
2020.10.09 17:30	5528.500	5539.000	5527.500	5533.500	10581	35731	500
2020.10.09 17:15	5524.500	5529.000	5523.000	5528.000	7525	28311	500
2020.10.09 17:00	5529.500	5532.000	5524.500	5524.500	11687	39529	500
2020.10.09 16:45	5531.500	5532.500	5528.000	5529.500	9554	31748	500
2020.10.09 16:30	5536.500	5537.500	5527.500	5531.500	13718	44336	500
2020.10.09 16:15	5540.000	5542.500	5534.500	5536.500	10609	36840	500
2020.10.09 16:00	5537.500	5542.500	5535.500	5539.500	13123	44079	500
2020.10.09 15:45	5534.000	5539.000	5531.500	5537.500	15303	47127	500
2020.10.09 15:30	5542.500	5544.000	5533.000	5534.000	22875	76181	500
2020.10.09 15:15	5540.500	5547.500	5537.000	5543.000	15117	52212	500
2020.10.09 15:00	5548.000	5549.000	5538.500	5541.000	19298	62271	500
2020.10.09 14:45	5536.000	5550.000	5535.000	5547.500	21751	67545	500
2020.10.09 14:30	5537.000	5539.000	5534.500	5535.500	12305	39840	500

Fonte: Metatrader 5

A tabela 1 apresenta um trecho da base de dados obtida após a junção dos preços históricos de ambos minicontratos. Como descrito no capítulo anterior, há a adição apenas dos preços de fechamento e o volume negociado pertencentes ao minicontrato de índice.

Cada linha representa o valor do preço de abertura, o maior e o menor valor alcançado, o de fechamento e o volume negociado em um período de 15 minutos. A figura 21 apresenta todos os preços de fechamento da base de dados obtida.

Tabela 1 – Estrutura dos dados utilizados

Data	Tempo	Abertura	Alta	Baixa	Fechamento	Volume	Fechamento (Índice)	Volume (Índice)
03/08/2020	09:00	5235,0	5242,5	5216,5	5239,5	137665	103840	237021
03/08/2020	09:15	5239,5	5246,0	5234,0	5241,5	96845	103955	221627
03/08/2020	09:30	5242,0	5252,0	5238,5	5250,0	88794	103885	178212
03/08/2020	09:45	5250,0	5271,0	5249,0	5271,0	153845	103795	192113
03/08/2020	10:00	5271,0	5293,0	5267,0	5291,5	183569	103680	480949

Fonte: Autor

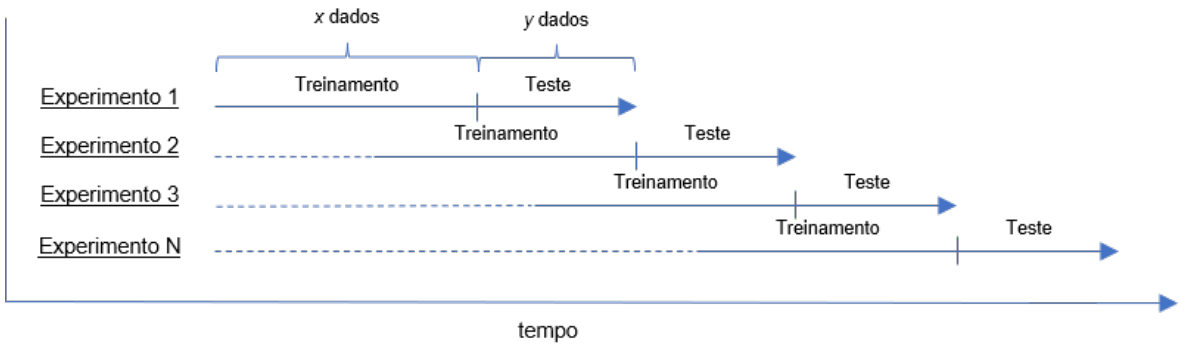
Figura 21 – Cotação do preço de fechamento do minicontrato de dólar

Fonte: Autor

Foi empregado o sistema de treinamento de janela deslizante, que após uma série de testes chegou-se à proporção de 14:4, ou seja, foram consideradas 14 semanas de treinamento para quatro de teste, portanto, após o primeiro experimento, os dados utilizados para treinamento e teste se deslocam quatro semanas, para assim ser realizado o experimento seguinte. A Figura 22 ilustra o sistema de treinamento mencionado. O treinamento possuiu, em média, 2400 dados enquanto o teste 700. O tamanho das janelas foi escolhido de forma a se balancear o número de experimentos realizados com o extenso tempo de treinamento dos modelos, uma vez que, para cada janela, um novo modelo foi treinado, com tempo médio de conclusão de 50 a 60 horas.

Com a base de dados formada e dividida seguindo a estrutura da Tabela 1, o ambiente gera a matriz $30x11$ do estado s, que contém os preços de abertura, alta, baixa, e os de fechamento e volume negociado de ambos minicontratos (dólar e índice), além de adicionar o posicionamento do agente em forma ortogonalizada, e o valor acumulado com a operação (compra/venda) em aberto. No início do processo de treinamento, os dois últimos atributos (posicionamento, e valor acumulado) possuem os valores de "fora do mercado"([0,0,1]) e zero, respectivamente, e são

Figura 22 – Janela deslizante, com $x = 2400$ e $y = 700$



Fonte: Autor

mantidos até completar as 30 observações necessárias. Uma vez completa e normalizada dentro do intervalo $[0 : 1]$, ela é enviada para o agente que devolve uma ação, e o ambiente reenvia o próximo estado e a recompensa. A atualização da matriz é dado pela eliminação do dado da última linha e a adição de um mais recente na primeira linha. E após 2400 interações, que é o tamanho da base de dados de treinamento, é finalizado um episódio. Cada experimento foi treinado por 4500 episódios.

A cada interação (passo) com o ambiente é armazenado a transição (s_t, a_t, r_t, s_{t+1}) no *replay memory*, e levando em consideração o estudo de Huang (2018), a cada n passos é amostrado transições sequenciais, calculado o erro e realizado a atualização dos pesos da rede neural, o que fez com que o tempo de treinamento fosse reduzido a uma taxa n.

Assim como na DQN, na DRQN empregada, há o uso de duas redes neurais, a *target network*, sendo seus parâmetros atualizados ao final de cada episódio, e a *main network*, cujo parâmetros são atualizados constantemente. Para garantir uma exploração adequada do espaço de estados foi usado o $\epsilon - greedy$, com o ϵ variando de 1 até 0,1.

No processo de treinamento do modelo foi aplicado o otimizador Adam (KINGMA; BA, 2014) com taxa de aprendizagem de 0,00025, $\gamma = 0{,}99$ e diferentemente da DRQN de Hausknecht e Stone (2017) que a cada instante de tempo utilizava vários episódios com transições sequenciais na amostragem do *replay memory*, neste projeto esse processo ocorreu a cada 4 passos e foi utilizado apenas um episódio contendo 128 transições sequenciais. Fatos estes, que resultou em um menor custo computacional e consequentemente em um menor tempo de treinamento. A arquitetura final da rede neural e os hiperparâmetros foram definidos empiricamente, depois de vários testes, resultou-se no modelo apresentado, por ter o melhor desempenho.

5.3 EXPERIMENTOS REALIZADOS

Foram realizados inicialmente testes para avaliar o impacto da proporção da janela deslizante e foram realizados testes em relação ao tamanho do espaço de estado s, referente à quantidade de instantes de tempo que cada estado irá apresentar para o agente.

Posteriormente, foram realizados testes para investigar e analisar o desempenho do modelo proposto.

5.3.1 Experimentos investigativos da janela deslizantes

Os primeiros experimentos realizados foram para investigar o impacto do tamanho da proporção da janela deslizante referente ao treinamento e foi utilizado como base de teste o período entre 03/02/2020 e 28/02/2020. O outros parâmetros utilizados foram os mesmos descritos na seção anterior. A única diferença efetiva foi a proporção da janela deslizante. A tabela 2 mostra os experimentos realizados. Para cada experimento a DRQN foi treinada 3 vezes, ou seja, obtiveram-se 3 redes neurais para efeito de comparação.

Tabela 2 – Plano de testes para o parâmetro da janela deslizante

Experimento	Treinamento	Tamanho da proporção (Treinamento)
1A	06/12/2019 - 31/01/2020	8 semanas
2A	22/11/2019 - 31/01/2020	10 semanas
3A	08/11/2019 - 31/01/2020	12 semanas
4A	25/10/2019 - 31/01/2020	14 semanas
5A	11/10/2019 - 31/01/2020	16 semanas
6A	27/09/2019 - 31/01/2020	18 semanas
7A	13/09/2019 - 31/01/2020	20 semanas

Fonte: Autor

5.3.2 Experimentos investigativos do tamanho do estado s

Definida a estrutura da janela deslizante, foi realizados experimentos para investigar a influência do tamanho dos instantes de tempo do estado s (Figura 17). A tabela 3 mostra os experimentos realizados. Foi utilizado a proporção de 14:4 da janela deslizantes e o período de testes de 03/02/2020 até 28/02/2020. Para cada experimento a DRQN foi treinada 3 vezes, ou seja, obtiveram-se 3 redes neurais para efeito de comparação.

Tabela 3 – Plano de testes para o determinar o tamanho do estado s

Experimento	Tamanho do estado s
1B	10 x 11 dados
2B	20 x 11 dados
3B	25 x 11 dados
4B	30 x 11 dados
5B	35 x 11 dados

Fonte: Autor

5.3.3 Experimentos investigativo da convolução dilatada

Tendo como base o trabalho de Guan, Li e Lu (s.d.), no qual a utilização de convoluções dilatadas contribuíram para a melhora da previsão da tendência do ativo. Foram realizados experimentos para investigar se essa melhora de desempenho se aplicava no modelo proposto. A tabela 4 ilustra os testes investigativos realizados. Foi utilizado a proporção de 14:4 da janela deslizantes e o período de testes de 03/02/2020 até 28/02/2020, além disso, para o espaço de estados foi utilizado o tamanho de $30x11$, ou seja, o numero de instantes de tempo do estado s utilizado foi de 30.

Tabela 4 – Plano de testes investigar o impacto das convoluções dilatadas

Experimento	Camada de Convoluções
1C	Sem Dilatação
2C	Com taxa de dilatação de 2 na primeira e na segunda camada
3C	Com taxa de dilatação de 2 na primeira e de 4 na segunda camada

Fonte: Autor

5.3.4 Experimentos investigando o impacto do indicador técnico MACD no espaço de estados

Foi realizado um estudo também em relação ao impacto da agregação do indicador técnico MACD no espaço de estados. Sendo assim, além do preço de fechamento e o volume do minicontrato de índice, foi adicionado ao espaço de estados a Linha MACD, o Sinal MACD e o Histograma MACD. A tabela 5 ilustra os testes investigativos realizados. Foi utilizado a proporção de 14:4 da janela deslizantes e o período de testes de 03/02/2020 até 28/02/2020. O número de instantes de tempo do estado s foi de 30, portanto, o estado s foi representado por uma matriz $30x14$.Nas camadas de convoluções foi utilizado convoluções dilatadas com a taxa de dilatação de 2 na primeira camada e de 4 na segunda.

Tabela 5 – Plano de testes investigar o impacto das convoluções dilatadas

Experimento	Espaço de estado s
1D	Sem MACD
2D	Com MACD

Fonte: Autor

5.3.5 Experimentos com o modelo proposto

Com os parâmetros definidos, sendo a janela deslizante de 14:4 e o tamanho do estado s de $30x11$, utilizando convoluções dilatadas com a taxa de dilatação de 2 na primeira camada e de 4 na segunda, e utilizando apenas os preços do minicontrato de dólar e seu volume com o preço de fechamento do minicontrato de índice e seu volume. A Tabela 6 exibe a divisão da base de dados para os experimentos executados.

Tabela 6 – Plano de testes

Experimento	Treinamento	Teste
1	27/09/2019 - 03/01/2020	06/01/2020 - 31/01/2020
2	25/10/2019 - 31/01/2020	03/02/2020 - 28/02/2020
3	22/11/2019 - 28/02/2020	02/03/2020 - 27/03/2020
4	20/12/2019 - 27/03/2020	30/03/2020 - 24/04/2020
5	17/01/2020 - 24/04/2020	27/04/2020 - 22/05/2020
6	14/02/2020 - 22/05/2020	25/05/2020 - 19/06/2020
7	13/03/2020 - 19/06/2020	22/06/2020 - 17/07/2020
8	09/04/2020 - 17/07/2020	20/07/2020 - 14/08/2020
9	08/05/2020 - 14/08/2020	17/08/2020 - 11/09/2020
10	05/06/2020 - 11/09/2020	14/09/2020 - 09/10/2020

Fonte: Autor

Para efeito de comparação foi utilizado uma DQN que foi treinada identicamente à DRQN, sendo as únicas diferenças existentes: a rede neural, onde há a substituição da camada LSTM por uma camada totalmente conectada, mantendo as mesmas 156 unidades; e o *minibatch*, que ao invés de ser amostrado sequencialmente, ele é amostrado aleatoriamente.

Ambos modelos, DQN e DRQN, foram treinados 6 vezes em cada experimento, ou seja, dado o Experimento 1, por exemplo, obtiveram-se 6 redes neurais para a DQN e mais 6 para a DRQN. Em todos os experimentos executados foram considerados a negociação de apenas 1 minicontrato de dólar, tanto para os treinamentos quanto para os testes.

6 RESULTADOS

As primeiras seções deste capítulo se dedicam a mostrar o estudo investigativo da influência da janela deslizante e o tamanho do espaço de estado s. Além disso, há a exposição dos resultado referentes à influência das convoluções dilatadas e a adição do indicador técnico MACD ao espaço de estados. Por seguinte, é discorrido e analisado os resultados do plano de testes proposto na subseção 5.3.5

6.1 INFLUÊNCIA DA PROPORÇÃO DA JANELA DESLIZANTE

A tabela 7 mostra resultados obtidos a partir do plano de testes da tabela 2. O período de teste engloba as datas entre 03/02/2020 e 28/02/2020. Os valores mostrados por cada rede se referem à quantidade de pontos que o modelo obteve ao final do período de teste, sendo apresentado a média referentes às três redes neurais treinadas em cada experimento, juntamente com o seu desvio padrão.

Tabela 7 – Resultado dos experimentos investigando a influência da janela deslizante

Experimento	Treinamento	Rede 1	Rede 2	Rede 3	Desvio Padrão	Média
1A	8 semanas	-98	-292	21	157.99	-123.0
2A	10 semanas	-77	-135	187	171.63	-8.33
3A	12 semanas	86	166	-27	96.97	75.00
4A	14 semanas	135	102	156	27.15	130.67
5A	16 semanas	95	148	151	31.50	131.33
6A	18 semanas	187	198	91	58.86	158.67
7A	20 semanas	255	-23	182	144.13	138.00

Fonte: Autor

Nos experimentos 1A e 2A, obtiveram-se médias negativas, e um desvio padrão apresentado relativamente alto, fato que sugere que 8 e 10 semanas não foram suficientes para o modelo encontrar uma política financeiramente lucrativa.

Apenas a partir de 12 semanas que os modelos apresentaram médias de pontos positivas. No experimento 3A apesar de ter uma média positiva, os outros modelos conseguiram supera-lá e alguns ainda apresentaram um desvio padrão mais baixo. Nos experimentos 4A, 5A, 6A e 7A as médias ficaram bem próximas, sendo a maior diferença nos valores do desvio padrão. Apesar de no experimento 7A apresentar uma rede com o maior resultado de pontos em relação à todos os resultados apresentados na tabela, e uma média coerente com o resto dos experimentos, este experimento apresentou um desvio padrão alto.

Os experimentos 4A, 5A e 6A tiveram um desempenho similar e apresentaram os desvios padrões mais baixos dentre todos os experimentos realizados. O experimento 6A apesar de ter apresentado a maior média o seu desvio foi o mais alto. Já entre os experimentos 4A e 5A os resultados foram muito similares, ambos conseguiram encontrar uma política rentável e apresentaram um desvio padrão com uma diferença entre eles bem baixa.

6.2 INFLUÊNCIA DO NÚMERO DE INSTANTES DE TEMPO DO ESTADO s

A tabela 8 apresenta os resultados obtidos a partir do plano de testes da tabela 3. Como mencionado na subseção 5.3.2 foi utilizado a proporção de 14:4 na janela deslizante e o período de teste foi de 03/02/2020 até 28/02/2020. Os valores apresentados por cada Rede se referem ao total de pontos obtido pela rede treinada ao final do período de teste, sendo a média e o desvio padrão referente às três redes neurais treinadas em cada experimento.

Tabela 8 – Resultado dos experimentos investigando a influência do tamanho do estado s

Experimento	Estado s	Rede 1	Rede 2	Rede 3	Desvio Padrão	Média
1B	$10x11$	84	-580	-337	335.95	-277.67
2B	$20x11$	43	122	147	54.29	104.00
3B	$25x11$	182	-52	227	149.79	119.00
4B	$30x11$	134	102	156	27.15	130.67
5B	$35x11$	125	177	144	26.31	148.67

Fonte: Autor

O experimento 1B, que contém apenas 10 instantes de tempo no estado s, apresentou uma média negativa, chegando a ter uma rede que obteve ao final do período de teste um resultado final de 580 pontos negativos.

A partir do experimento 2B as médias ficaram positivas e próximas. Com exceção do experimento 3B que teve um desvio padrão alto em relação aos demais, os outros três experimentos tiveram um desvio padrão próximos. Os experimentos 4B e 5B apresentaram médias similares, assim como, seus desvios padrões.

6.3 INFLUÊNCIA DAS CONVOLUÇÕES DILATADAS

A tabela 9 apresenta os resultados obtidos a partir do plano de testes da tabela 4. Como dito na subseção 5.3.3 foi utilizado uma proporção de 14:4 na janela deslizantes, o período de testes engloba o período entre 03/02/2020 até 28/02/2020 e o número de instantes de tempo do estado

s adotado foi de 30. Foram treinadas três redes neurais e os valores descritos por cada rede referem-se à quantidade de pontos elas adquiriram ao final do período de teste.

Tabela 9 – Resultado dos experimentos investigando a influência das convoluções dilatadas

Exp.	Camada de Convoluções	Rede 1	Rede 2	Rede 3	Desvio Padrão	Média
1C	Sem Dilatação	93	149	125	28.09	122.33
2C	2 em ambas	116	113	171	32.65	133.33
3C	2 e 4	134	102	156	27.15	130.67

Fonte: Autor

Os três experimentos obtiveram resultados bem parecidos, nenhuma das nove redes treinadas apresentou um resultado muito discrepante em relação às demais. O experimento 2C foi o que apresentou o maior desvio padrão e média, e o 3C foi o experimento com o desvio padrão mais baixo.

6.4 INFLUÊNCIA DA ADIÇÃO DO MACD NO ESPAÇO DE ESTADOS

A tabela 10 apresenta os resultados obtidos a partir do plano de testes da tabela 5. Como detalhado na subseção 5.3.4 a proporção da janela deslizante foi de 14:4, foram utilizados 30 instantes de tempo no espaço de estados s, além da adoção de convoluções dilatadas, sendo a taxa da primeira camada de 2 e da segunda de 4. O período de testes foi de 03/02/2020 até 28/02/2020. Foram treinadas 3 redes neurais para cada experimento e os valores descritos por cada rede são referentes à quantidade de pontos obtidos ao final do período de teste.

Tabela 10 – Resultado dos experimentos investigando a influência da adição do MACD no espaço de estados

Experimento	Espaço de Estados	Rede 1	Rede 2	Rede 3	Desvio Padrão	Média
1D	Sem MACD	134	102	156	27.15	130.67
2D	Com MACD	-12	-81	47	64.07	-15.34

Fonte: Autor

Com base no experimento 2D, a adição do indicador técnico no espaço de estados fez com a algumas redes tivessem resultados negativos, levando à uma média negativa.

6.5 DISCUSSÃO DOS PARÂMETROS E MODELOS COMPARATIVOS

Se no experimento 7A da tabela 7, referente à janela deslizante, fosse considerado que o resultado da Rede 2 foi um ponto fora da curva, por ter um valor muito discrepante com o restante das

redes, os resultados presentes na tabela, de modo geral, sugeririam que o aumento nas semanas no treinamento influência positivamente o resultado do modelo. Já que a partir de 14 semanas o valor da média das redes aumentaria, sem alterar muito o valor do desvio padrão.

Em relação aos resultados obtidos na tabela 3, referente ao número de instantes do estado s, os experimentos 4B e 5B tiveram desempenhos similares, sugerindo que o aumento dos instantes de tempo, auxiliam o agente a ter uma melhor percepção do ambiente. Já com relação às convoluções dilatadas, apesar dos resultados serem parecidos (tabela 9), o uso delas resultou em um acréscimo de 10 pontos na média das redes, o que permite inferir que as convoluções dilatadas podem ter uma influência positiva no resultado do modelo, já que elas conseguem abstrair informações da combinação de dados antigos com os mais recentes.

A adição do indicador técnico MACD no espaço de estados fez que com as redes treinadas tivessem pontos negativos ao final do período de teste, resultando em um decréscimo no desempenho do modelo. Fato que sugeriu que a inclusão MACD apenas contribuiu com ruídos para o espaço de estados, ou seja, os dados adicionais atrapalharam o agente a tomar decisões, não sendo assim uma boa opção para se adicionar no espaço de estados.

Mas com base nestes resultados obtidos a partir dos experimentos realizados para investigar a proporção da janela deslizante e o tamanho do estado s, pelo simples fato de que nos resultados apresentados na tabela 7, a janela de treinamento com 14 semanas apresentou um desvio padrão menor do que o de 16, sugeriu que a melhor proporção é a de 14 semanas. Para o tamanho do estado s, considerando que quanto maior o numero de instantes de tempo no espaço s maior o tempo de treinamento, e como os resultados dos experimentos 4B e 5B foram muito próximos, o experimento 4B foi o que apresentaria o menor tempo de treinamento. Em relação à inclusão do MACD é evidente, com base nos resultados da tabela 10, que a adição dele contribuiu para um decréscimo no desempenho.

O resto deste capítulo se dedicou à apresentação dos resultados obtidos utilizando a abordagem proposta (seção 5.1 e 5.2), e levando em consideração o plano de testes presente na tabela 6. Além disso, foi realizado uma série de comparações:

a) Com uma estratégia conservadora: *Buy and Hold*
b) com uma estratégia utilizando um algoritmo de negociação baseado em um indicador técnico, o MACD
c) Com o fundo cambial com o maior rendimento no ano de 2020 (OLIVEIRA, 2020): BB TOP DÓLAR FI CAMBIAL LP
d) Com a DQN treinada

Inicialmente foi considerado apenas o valor do retorno final na comparação, e posteriormente foi utilizado como métrica de desempenho o Índice Sharpe. Além disso, houve uma comparação quando aplicado o Imposto de Renda.

Vale ressaltar que, em todos os resultados, já foram descontadas as taxas transacionais estabelecidas pela B3, e em relação ao *Buy and Hold* como foi a compra a venda de apenas 1 minicontrato de dólar, o desconto foi irrisório.

Também foi analisado a evolução do montante de pontos adquiridos após cada experimento. Além disso, a partir da análise do retorno de pontos por experimento, foram efetuados testes de comparação estatística entre a DQN e a DRQN para que se pudesse validar os resultados brutos.

6.6 COMPARAÇÃO COM O *BUY AND HOLD*

A estratégia *Buy and Hold* é o ato de comprar um ativo e mantê-lo em carteira por um longo prazo de forma a se beneficiar dos rendimentos e valorização do papel que por ventura venha a apresentar no futuro. Foi considerada a compra de apenas 1 minicontrato de dólar no primeiro momento de teste e a sua venda no final do período em questão, tendo em média 4 semanas entre a ação de compra e venda.

A tabela 11 apresenta as médias das quantidades finais de pontos presentes em seus portfólios das 6 redes neurais testadas em cada experimento pelo modelo DRQN.

Tabela 11 – Resultado dos experimentos *Buy and Hold* x DRQN

Experimento	*Buy and Hold*	DRQN
1	224	131,83
2	221	143.75
3	603,5	515.57
4	468,5	206.50
5	-26	190.33
6	-207	290.50
7	108,5	106.17
8	52,5	267.83
9	-95,5	382.0
10	224,5	-96.70
SOMA	1574	2137,78
Média	157.40	213.78

Fonte: Autor

Considerando-se apenas o total dos pontos obtidos em cada período de teste como métrica, os pontos obtidos pela DRQN foram 27.58% superiores que a estratégia do *Buy and Hold*. O

período de testes englobou a pandemia mundial decorrente do coronavírus, o que fez com que o preço do dólar subisse muito e abruptamente, e como consequência beneficiando a estratégia do *Buy and Hold* neste período. E mesmo assim a DRQN conseguiu superar essa estratégia.

Em um exemplo de *Buy and Hold* que realizasse a compra de um minicontrato no primeiro passo de teste do experimento 1 (06/01/2020) e sua venda ao final do experimento 10 (09/10/2020), o lucro resultante seria de 1477 pontos ao longo de 40 semanas. Neste caso, a DRQN ainda acumularia 35,95% mais pontos que o *Buy and Hold*.

6.7 COMPARAÇÃO COM A ESTRATÉGIA QUE UTILIZA O INDICADOR TÉCNICO MACD

A estratégia utilizando um algoritmo de negociação tendo como base o MACD, foi implementado seguindo a descrição do exemplo descrito na subseção 2.1.1. Sendo assim, no MACD utilizado, a Média Móvel Exponencial curta foi de 12 períodos, a longa foi de 26 períodos e a da linha MACD foi de 9 períodos (NELOGICA, 2022), e os sinais de compra e de venda foram gerados a partir de:

a) Sinal de compra do ativo: Linha MACD > Sinal MACD
b) Sinal de venda do ativo: Sinal MACD > Linha MACD

A tabela 12 mostra as médias das 6 redes neurais treinadas da DRQN frente ao algoritmo de negociação usando MACD. É perceptível que o MACD não conseguiu ser uma estratégia rentável, ou seja, a soma de todos os pontos dos experimentos terminou negativa. Em apenas três experimentos dos dez que o MACD conseguiu terminar positivo. Em todos os experimentos a DRQN obteve mais pontos que o MACD.

6.8 COMPARAÇÃO COM A DQN

A tabela 13 apresenta as médias das quantidades finais de pontos presentes em seus portfólios das 6 redes neurais testadas em cada experimento por cada modelo (DQN e DRQN).

Em comparação com a DQN, a DRQN foi melhor em 339,71%, o que sugere que o emprego de camadas recorrentes e treinamento com estados que incorporam informações histórias de interação com o ambiente melhoram o desempenho do modelo. Além disso, a DRQN foi a que teve os ganhos mais estáveis, apenas em um experimento que ela terminou com o portfólio negativo, enquanto que na DQN, esse fato aconteceu três vezes.

Tabela 12 – Resultado dos experimentos MACD x DRQN

Experimento	MACD	DRQN
1	-104	131,83
2	-56	143.75
3	26	515.57
4	-123	206.50
5	-170	190.33
6	68	290.50
7	-10	106.17
8	268	267.83
9	-95	382.0
10	-268	-96.70
SOMA	-464,00	2137,78
Média	-46,40	213.78

Fonte: Autor

Tabela 13 – Resultado dos experimentos DQN x DRQN

Experimento	DQN	DRQN
1	66,83	131,83
2	-105.17	143.75
3	-127.0	515.57
4	169.17	206.50
5	229.50	190.33
6	22.50	290.50
7	-72.50	106.17
8	152.83	267.83
9	54.67	382.0
10	95.33	-96.70
SOMA	486,17	2137,78
Média	48,62	213.78

Fonte: Autor

6.9 COMPARAÇÃO COM O FUNDO CAMBIAL BB TOP DÓLAR FI CAMBIAL LP

Este fundo busca acompanhar a variação do dólar americano e agregar rentabilidade, servindo-se de oportunidades oferecidas no mercado financeiro americano. No mínimo, 80% de seu patrimônio líquido está relacionado diretamente ou sintetizados, via derivativos, à variação do Dólar (BB, 2022).

A figura 23 ilustra a rentabilidade do Fundo Cambial no período entre 06/01/2020 até 09/10/2020, que no caso rendeu 37,67%, ou seja, se o investidor comprasse cotas desse fundo, no valor total de R$ 100000,00 e resgatasse em 09/10/2020, ele teria um retorno de R$ 37670,00.

Ao final do período de testes a DRQN obteve 2137,78 pontos e levando em conta que para converter estes pontos para Reais, é apenas multiplicar por 10, já que é negociado apenas 1 minicontrato de dólar, tem-se um valor final de R$ 21377,80. Considerando o que foi explicado na subseção 2.2.1.1 e o valor intrínseco do minicontrato de dólar de US$ 10000,00, sendo o valor do dólar de R$ 5,525 no dia 09/10/2020 (UOL, 2020b), para efeitos de cálculos, têm que o minicontrato de dólar equivale a R$ 55250,00. Similarmente, considerando esse mesmo valor de R$ 55250,00 como investimento inicial para o Fundo Cambial, obtêm-se um retorno de R$ 20812,67.

Sendo assim, a DRQN teve um retorno 2,72% maior que o Fundo Cambial BB TOP DÓLAR FI CAMBIAL LP

Figura 23 – Rentabilidade do Fundo Cambial BB TOP DÓLAR FI CAMBIAL LP no período entre 06/01/2020 até 09/10/2020

Fonte: https://maisretorno.com/fundo/bb-top-dolar-fi-cambial-lp, acessado em 20/01/2023

6.10 ÍNDICE SHARPE COMO MÉTRICA DE DESEMPENHO

Do mesmo modo que na seção 6.9, onde foi convertida a rentabilidade do Fundo Cambial para Reais, no cálculo do Índice Sharpe foi adotado o mesmo princípio, já que este índice é rentabilidade do investimento sobre seu risco. Portando, o investimento inicial utilizado para o cálculo das rentabilidades foi de R$ 55250,00. Sendo assim, a tabela 14 mostra o Índice Sharpe, a rentabilidade total, a Volatilidade, e a rentabilidade por experimento, calculadas para o *Buy and Hold*, MACD, DQN, DRQN. A taxa livre de risco utilizada foi a Selic que no ano de 2020 estava à 2% ao ano (UOL, 2020a).

Tabela 14 – Calculo do Índice Sharpe

	Buy and Hold	MACD	DQN	DRQN
	Rentabilidade	Rentabilidade	Rentabilidade	Rentabilidade
Exp.1	4.05%	-1.88%	1.21%	2.39%
Exp.2	4.00%	-1.01%	-1.90%	2.60%
Exp.3	10.92%	0.47%	-2.30%	9.33%
Exp.4	8.48%	-2.23%	3.06%	3.74%
Exp.5	-0.47%	-3.08%	4.15%	3.44%
Exp.6	-3.75%	1.23%	0.41%	5.26%
Exp.7	1.96%	-0.18%	-1.31%	1.92%
Exp.8	0.95%	4.85%	2.77%	4.85%
Exp.9	-1.73%	-1.72%	0.99%	6.91%
Exp.10	4.06%	-4.85%	1.73%	-1.75%
Rentabilidade Total	28.49%	-8.40%	8.80%	38.69%
Volatilidade	4.25%	2.52%	2.07%	2.85%
Sharpe	6.23	-4.12	3.29	12.86

Fonte: Autor

Para o cálculo da rentabilidade total foi utilizado como base o número total de pontos obtidos ao final dos 10 experimentos, levando em conta o investimento inicial de R$ 55250,00. Na rentabilidade por experimento foi considerado os pontos conquistados ao final de cada período de teste em questão, nos casos da DQN e a DRQN foi calculada a partir da médias das 6 redes neurais treinadas para cada modelo.

As rentabilidades do Fundo Cambial Fundo Cambial BB TOP DÓLAR FI CAMBIAL LP não estão descritas na tabela 14, pois as rentabilidades históricas divulgadas são mensais, sendo assim, não equivalem aos períodos dos experimentos. Portanto, foi realizado o cálculo do Índice Sharpe do Fundo Cambial contemplando o período de Janeiro/2020 até Outubro/2020. A tabela 15 mostra a rentabilidade mensal do Fundo Cambial (RETORNO, 2023), com a rentabilidade total no período, juntamente com a volatilidade e o Índice Sharpe calculado.

O retorno total do Fundo Cambial foi diferente do mostrado na seção 6.9, pois foi referenciado o mês de outubro/2020 inteiro, sendo que no anterior se refere apenas à rentabilidade até a data de 09/10/2020.

Tendo em vista todos os Índices Sharpe calculados na tabela 14 a DRQN é a que se torna a melhor opção de investimento dentre as disponíveis, o MACD foi a pior, já que apresentou uma razão Sharpe negativa. Mesmo considerando o Índice Sharpe do Fundo Cambial apresentado na tabela 15, a DRQN continua sendo a melhor escolha, apesar da diferença entre ambos índices ser baixa.

Tabela 15 – Calculo do Índice Sharpe para o Fundo Cambial BB TOP DÓLAR FI CAMBIAL LP e suas rentabilidades mensais

Fundo Cambial BB	Rendimento
Janeiro	6.92%
Fevereiro	5.06%
Março	14.38%
Abril	5.20%
Maio	-0.96%
Junho	1.32%
Julho	3.69%
Agosto	5.09%
Setembro	2.92%
Outubro	2.06%
Retorno Total	42.75%
Volatilidade	3.92%
Sharpe	10.40

Fonte: Autor

6.11 INCIDÊNCIA DO IMPOSTO DE RENDA

Todos os resultados expostos até aqui consideraram apenas a incidência dos custos transacionais estabelecidas pela B3 e explicitas na seção 4.4, com exceção do *Buy and Hold* por usar apenas 1 minicontrato de dólar suas taxas transacionais se tornam irrisórias.

A Tabela 16 mostra a soma dos pontos convertidos em reais (seção 6.9) e os valores finais caso o IR fosse aplicado. No caso do Fundo Cambial foi levado em conta o investimento inicial de R$ 55250,00, seguindo a mesma lógica descrita na seção 6.9. Como a estratégia usando o MACD como base apresentou um resultado final negativo, ele não foi considerado na comparação quando aplicado o Imposto de Renda.

Tabela 16 – Incidência do Imposto de Renda (IR)

	Buy and Hold	Fundo Cambial	DQN	DRQN
Resultado	R$ 15740,00	R$ 20812,67	R$ 4861,70	R$ 21377,80
IR	R$ 2361,00 (15%)	R$ 4162,53 (20%)	R$ 729,26 (15%)	R$ 3206,67 (15%)
Final	R$ 13379,00	R$ 16650,14	R$ 4132,45	R$ 18171,13

Fonte: Autor

Levando em conta a incidência do imposto de renda, mesmo que no Fundo Cambial a taxa seja de 20%, a DRQN conseguiu acumular um valor final maior em relação à todos os modelos comparativos.

6.12 DISCUSSÃO E TESTES ESTATÍSTICOS

Considerando apenas a soma final total de pontos adquiridos em todos os experimentos, a DRQN conseguiu superar todos os modelos de comparação, o *Buy and Hold*, MACD, DQN e quando transformado o total de pontos em rentabilidade, conseguiu também superar, mesmo que por pouco, um Fundo Cambial, que é gerido por profissionais altamente especializados.

Utilizando como métrica de desempenho o Índice Sharpe, a DRQN foi a que obteve a razão Sharpe mais alta, sendo assim, quando comparado a rentabilidade considerando todo o risco envolvido, a DRQN conseguiu superar todos as outras estratégias. Até mesmo considerando a incidência do imposto de renda ao final do período total de testes, a DRQN obteve um montante final maior que os outros modelos comparativos.

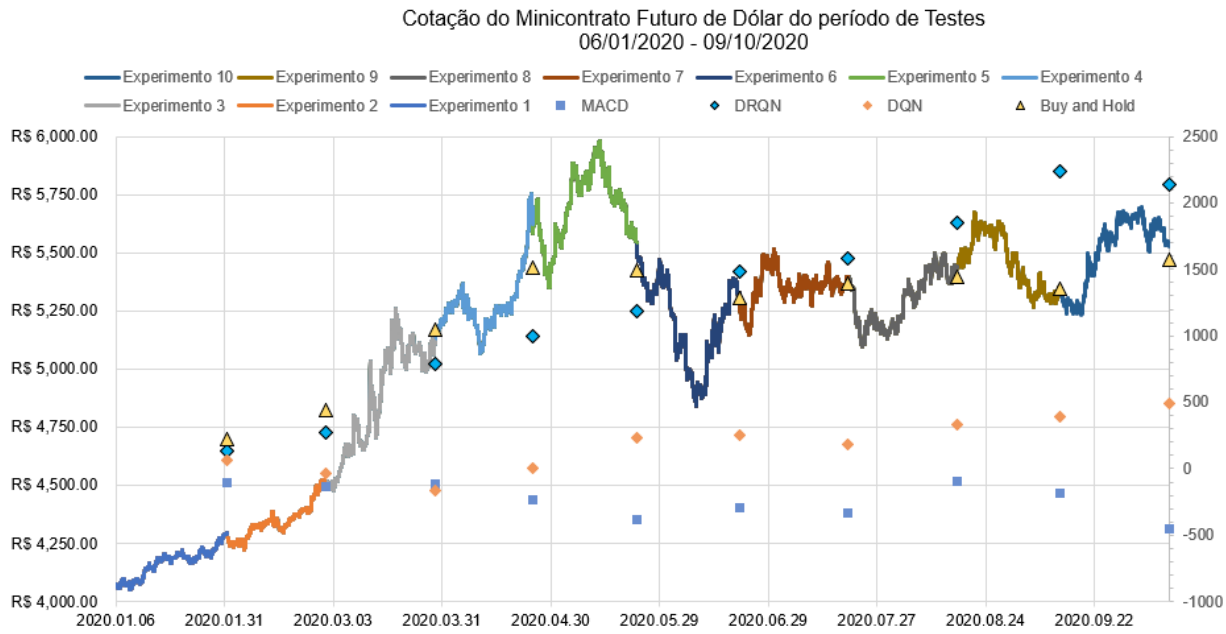
A seguir, foi apresentado uma análise geral dos resultados baseado apenas na evolução do montante de pontos ao longo dos experimentos, e após isso, com o auxílio de testes estatísticos foi realizado uma análise mais profunda, sendo o escopo principal o retorno de pontos em cada experimento. Dado que no Fundo Cambial BB TOP DÓLAR FI CAMBIAL LP o período dos rendimentos divulgados não equivalem com o dos experimentos realizados, a comparação da evolução do montante de pontos acumulados, e a análise dos pontos por experimento, não se fez necessária.

6.12.1 Análise da evolução do montante de pontos

Na Figura 24 foi ilustrado a cotação do minicontrato de dólar, juntamente com a evolução do montante de pontos das estratégias *Buy and Hold*, MACD, DQN, em comparação com o DRQN. O gráfico está seccionado pelos experimentos descritos na tabela 6. Cada marcador representa a quantidade de pontos acumulados após cada experimento.

Até o final do experimento 5 o *Buy and Hold* apresentou um melhor desempenho que a DRQN, já que acumulou mais pontos nesse período de tempo. Devido principalmente ao fato da cotação do dólar ter disparado nesse período decorrente da crise do COVID, fato que favoreceu a estratégia do *Buy and Hold*, a qual é puramente dependente do aumento da cotação do ativo. A partir do experimento 6 a DRQN conseguiu obter melhores resultados, culminando em um montante maior de pontos ao final do período de testes. Considerando apenas o começo e o fim de cada período dos experimentos a cotação do dólar não teve um aumento abrupto que favorecesse o *Buy and Hold* como aconteceu nos experimentos anteriores.

Figura 24 – Cotação do minicontrato de dólar, seccionado por experimento, com a evolução do portfólio de todos os modelos comparativos, com exceção do Fundo Cambial



Fonte: Autor

Enquanto isso, a DQN e a estratégia do MACD apresentaram um desempenho diferenciado em relação ao *Buy and Hold* e a DRQN. A estratégia do MACD não conseguiu se adaptar aos movimentos do minicontrato de dólar passando a maior parte do tempo com o montante de pontos no negativo.

Já a DQN apresentou um desempenho mediano, não foi tão alto quanto a DRQN e o *Buy and Hold* e não foi tão baixo quanto o MACD, a DQN até o experimento 3 estava com o montante de pontos negativo, mas a partir do experimento 4 ela conseguiu se manter positiva até o último experimento, fato este decorrente à DQN ter provavelmente aprendido a como lidar com essas mudanças rápidas no preço do minicontrato de dólar, já que pelo uso da janela deslizante, elas fizeram parte do seu treinamento. O mesmo se aplica no caso da DRQN, ela conseguiu abstrair no seu treinamento como se comportar com essas mudanças, e isso é percebido no seu montante de pontos acumulados, que aumenta de maneira a terminar com mais pontos que o restante das estratégias de comparação.

6.12.2 Análise por testes estatísticos

Mas apenas os resultados e análises do montante de pontos acumulados ao longo do período não são suficientes para avaliar se a DRQN consegue ter um desempenho melhor, principalmente

em relação à DQN. Para aferir estatisticamente se as duas médias de distribuições são distintas foi aplicado em cada experimento o teste T de Student (STUDENT, 1908), um valor p menor que um determinado limiar ($p < 0.05$) indica que os dois modelos demonstraram diferenças de desempenho estatisticamente significativas.

Para a aplicação do teste de T de Student é necessário que os valores finais dos portfólios das redes neurais treinadas sigam uma distribuição normal, e para aferir tal fato foi aplicado o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965), que quando $p > 0.05$ significa que a distribuição é semelhante a uma distribuição normal.

A Tabela 17 apresenta os pontos obtidos por cada rede neural treinada para a DRQN e a Tabela 18 referente aos valores da DQN.

Tabela 17 – Resultados das redes neurais treinadas da DRQN

Exp.	Rede 1	Rede 2	Rede 3	Rede 4	Rede 5	Rede 6	Desvio Padrão	Média
1	133	168	92	187	184	27	57.23	131.83
2	134.5	184	158	102	156	128	25.98	143.75
3	678.4	540	512	638	578	147	174.09	515.57
4	421	378	245	187	82	-74	169.05	206.50
5	167	176	276	165	246	112	54.72	190.33
6	166	120	67	768	377	245	235.26	290.50
7	-61	121	103	252	314	-92	148.27	106.17
8	439	40	532	164	42	390	194.81	267.83
9	403	326	305	418	385	455	51.87	382.00
10	-114.2	-71	-101	-53	-154	-87	32.33	-96.70

Fonte: Autor

Tabela 18 – Resultados das redes neurais treinadas da DQN

Exp.	Rede 1	Rede 2	Rede 3	Rede 4	Rede 5	Rede 6	Desvio Padrão	Média
1	97	81	35	48	93	47	24.35	66.83
2	-140	-71	-133	-115	-90	-82	25.88	-105.17
3	-363	-247	102	-199	89	-144	170.62	-127.00
4	287	252	352	9	168	-53	146.73	169.17
5	306	246	211	145	247	222	48.25	229.50
6	-137	-78	153	54	177	-34	115.92	22.50
7	-132	-23	-166	21	-74	-61	62.73	-72.50
8	536	236	128	-108	53	72	199.74	152.83
9	-24	-51	255	-125	252	21	147.11	54.67
10	111	121	99	71	88	82	17.04	95.33

Fonte: Autor

Baseado nas Tabelas 17 e 18, a Tabela 19 apresenta os valores p calculados de ambos os testes estatísticos, Shapiro-Wilk e T de Student. Em todos os experimentos o valor p de Shapiro-Wilk apresentou um valor > 0.05, significando que todos seguem uma distribuição normal. E em relação ao T de Student caso um experimento apresentasse $p \leq 0{,}05$, o valor é exibido em negrito na tabela, indicando diferença significativa no desempenho de um modelo comparado ao outro.

Tabela 19 – Testes estatísticos realizados

	DRQN/DQN	
Experimento	Shapiro-Wilk (p)	T de Student (p)
1	0.359	**0.043**
2	0.649	**<0.001**
3	0.680	**<0.001**
4	0.546	0.717
5	0.544	0.258
6	0.116	**0.045**
7	0.962	**0.032**
8	0.547	0.380
9	0.506	**<0.001**
10	0.705	**<0.001**

Fonte: Autor

Para facilitar a análise, a Tabela 20 apresenta todos os resultados dos modelos comparativos, *Buy and Hold*, MACD, DQN e DRQN. Sendo que, na DQN e na DRQN foi apresentado a média das 6 redes neurais treinadas.

Tabela 20 – Resultado dos experimentos

Experimento	*Buy and Hold*	MACD	DQN	DRQN
1	224	-104	66,83	131,83
2	221	-56	-105.17	143.75
3	603,5	26	-127.0	515.57
4	468,5	-123	169.17	206.50
5	-26	-170	229.50	190.33
6	-207	68	22.50	290.50
7	108,5	-10	-72.50	106.17
8	52,5	268	152.83	267.83
9	-95,5	-95	54.67	382.0
10	224,5	-268	95.33	-96.70

Fonte: Autor

A estratégia usando o indicador técnico MACD apresentou os piores retornos nos experimentos 1, 4, 5 e 10, mas conseguiu se adaptar bem às flutuações da cotação do dólar apenas no experimento 8, apresentando o maior resultado dentre todos.

O *Buy and Hold* apresentou os maiores retorno até o experimento 4, devido principalmente à alta abrupta dos preços do minicontrato de dólar. Nos experimentos 6 e 9 apresentou os piores desempenhos, decorrente de serem períodos nos quais o minicontrato de dólar apresentou um movimento de baixa, ou seja seus preços abaixaram.

A influência da utilização da janela deslizante é perceptível nos experimentos 6 e 7. No experimento 6 a base de treinamento da DRQN apresentava apenas movimentos de alta e baixa, com poucos movimentos laterais, sendo assim, como no período de teste deste experimento a cotação do minicontrato de dólar apresentou apenas movimentos de alta e baixa, a DRQN conseguiu apresentar o maior número de pontos ao final do experimento. Da mesma maneira, no experimento 7, a DRQN que foi treinada majoritariamente com movimentos de alta e baixa dos preços teve o seu retorno reduzido já que neste experimento, o período de teste apresentou, mais para o final, um movimento lateral. Dito isso, o uso da janela deslizante proporcionou que a DRQN tivesse sempre na sua base de treinamento, os movimentos mais recentes dos preços do minicontrato de dólar, auxiliando o agente a se adaptar melhor aos possíveis movimentos dos preços, aumentando assim, o retorno de pontos.

Com base nos resultados da Tabela 20 e analisando juntamente com o preço do minicontrato de dólar da Figura 24, percebeu-se uma relação entre os movimentos dos preços e o desempenho da DRQN, que apresentou um melhor retorno quando teve movimentos de alta e baixa mais acentuados, como foi o caso, principalmente, dos experimentos 3, 6 e 9. Em movimentos laterais como apresentado no experimento 7, a DRQN não se comportou igual aos outros movimentos.

Nos experimentos 4, 5 e 8, apesar de não ser possível validar estatisticamente a superioridade de um modelo em relação ao outro, por apresentarem um $p > 0.05$, é plausível inferir que, dado o período de teste de cada experimento, ambos modelos (DQN e DRQN) conseguiram encontrar uma política rentável, já que nenhum apresentou um montante de pontos negativo no final do período. Com exceção do experimento 4, os outros dois conseguiram até superar a estratégia do *Buy and Hold*.

Nos outros sete experimentos, por apresentarem $p \leq 0.05$ foi possível aferir qual modelo (DQN ou DRQN) apresentou um melhor desempenho. Nos experimentos 1, 2, 3, 6, 7 e 9, a DRQN teve um desempenho superior à DQN, o que corrobora com o fato de que o emprego de camadas recorrentes auxiliam no aumento do desempenho do modelo. No experimento 10 a

DRQN não conseguiu superar a DQN e nem o *Buy and Hold*, apresentando um portfólio negativo juntamente com a estratégia do MACD. Apesar de ter um movimento de alta no início do período, a DRQN, neste caso, não conseguiu encontrar uma estratégia que se adaptasse ao movimento mais equilibrado do preço, apresentado mais no final do período.

Dos dez experimentos realizados com a DRQN, nove deles conseguiram encontrar uma política lucrativa, e dentre estes, é possível aferir com embasamento estatístico, que em seis experimentos, a DRQN superou a DQN. E destes seis experimentos, em dois a DRQN superou ainda o *Buy and Hold*.

7 CONCLUSÃO

Este trabalho investigou a aplicação da *Deep Recurrent Q-Network* na negociação do minicontrato futuro de dólar, objetivando o maior retorno financeiro possível. Mesmo considerando todas as taxas envolvidas na negociação do ativo e até mesmo a incidência do Imposto de Renda ao final do processo, o modelo apresentado conseguiu encontrar uma política rentável. Conseguindo ter um Índice Sharpe superior às outras estratégia de comparação, ou seja, a rentabilidade da DRQN considerando seu risco foi melhor que que as outras estratégias, superando até a razão Sharpe do Fundo Cambial mais rentável de 2020. E mesmo tendo a crise mundial decorrente da COVID em 2020, a DRQN conseguiu se adaptar às mudanças abruptas que se teve nos preços dos minicontrato de dólar.

Foi utilizado o sistema de janela deslizante para a realização dos treinamentos e testes, e comparado com a estratégia conservadora de *Buy and Hold*, com uma estratégia baseado no indicador técnico MACD e com um modelo de DQN, que foi treinada exatamente igual a DRQN, com exceção das amostragens do *replay memory* e a arquitetura da rede neural utilizada. A DRQN apresentou resultados bem promissores. Considerando apenas os valores finais do montante de pontos ao final do período de testes como métrica de análise, a DRQN foi a que apresentou os melhores resultados, superando o *Buy and Hold*, MACD, a DQN e um Fundo Cambial. Levando em conta apenas os experimentos nos quais a análise estatística conseguia validar se o modelo DRQN era melhor que a DQN, a DRQN foi superior em 6 dos 7 experimentos.

Também pode-se observar a influência do uso da janela deslizante, pois analisando a evolução do montante de pontos adquiridos em cada experimento, observou-se que tanto a DQN quanto a DRQN, que após as mudanças abruptas dos preços do minicontrato de dólar começaram a fazer parte de seus treinamentos, elas conseguiram encontrar uma melhor estratégia para comprar e vender.

Os testes realizados para encontrar os parâmetros da janela deslizante e o número de instantes de tempo do estado s sugeriram que o aumento de ambos implica em uma melhora no desempenho do modelo. Como tema de pesquisa para trabalhos futuros, propõe-se a analisar e investigar os parâmetros máximos dessa melhora, que seria onde eles deixam de ajudar o modelo e começam a prejudicar o desempenho, já que com dados demasiados, eles podem deixar de serem significativos. Além disso, a adoção de convoluções dilatadas, a partir do estudo investigativo apresentado, contribuiu de certa maneira com o acréscimo no desempenho do

modelo, por outro lado, a inclusão do indicador MACD no espaço de estados fez com que o modelo apresentasse um decréscimo no resultado.

Também seria de grande valor para próximos trabalhos, investigar outros períodos da vela de preços, foi utilizado neste trabalho o período de 15 minutos, talvez aumentando o período, os ganhos fossem maiores, já que com um período maior tem-se menores ruídos nos dados, ou seja o estado s apresentaria indícios mais concretos para o agente realizar uma ação mais certeira.

Além disso, propõe-se a investigar o uso desse modelo proposto em outros ativos, como na negociação do minicontrato de índice, ações e até mesmo em algumas criptomoedas. Assim como, a utilização do algoritmo da DDQN, ao invés, da DQN na implementação da DRQN.

REFERÊNCIAS

ANACONDA. **Anaconda: The World's Most Popular Data Science Platform**. Disponível em: https://www.anaconda.com/. Acesso em: 11 dez. 2022.

APPEL, G. **The Moving Average Convergence-divergence Trading Method: Advanced Version**. [S.l.]: Scientific Investment Systems, 1985. Disponível em: https://books.google.lk/books?id=pzAbnQAACAAJ.

APPEL, Gerald. **Technical Analysis: Power Tools for Active Investors**. First. [S.l.]: FT Press, 2005. ISBN 0131479024.

AREL, I; ROSE, D C; KARNOWSKI, T P. Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]. **IEEE Computational Intelligence Magazine**, v. 5, n. 4, p. 13–18, nov. 2010. ISSN 1556-603X. DOI: 10.1109/MCI.2010.938364. Disponível em: http://ieeexplore.ieee.org/document/5605630/. Acesso em: 21 out. 2021.

B3. **B3: Brasil, Bolsa, Balcão**. Disponível em: https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm. Acesso em: 11 dez. 2022.

BAJPAI, Supriya. **Application of deep reinforcement learning for Indian stock trading automation**. [S.l.: s.n.], 2021. arXiv: 2106.16088 [q-fin.TR].

BB. **Informe Mensal: BB Cambial Dolar LP FIC FI**. Dez 2022. Disponível em: https://www.bb.com.br/docs/pub/siteEsp/dtvm/dwn/inf04128893.pdf.

BHATTACHARJEE, Natalia V.; TOLLNER, Ernest W. Improving management of windrow composting systems by modeling runoff water quality dynamics using recurrent neural network. en. **Ecological Modelling**, v. 339, p. 68–76, nov. 2016. ISSN 03043800. DOI: 10.1016/j.ecolmodel.2016.08.011. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0304380016303106. Acesso em: 28 out. 2021.

CAI, Xianggao; HU, Su; LIN, Xiaola. Feature extraction using Restricted Boltzmann Machine for stock price prediction. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE). Zhangjiajie, China: IEEE, mai. 2012. P. 80–83. ISBN 9781467300896 9781467300889 9781467300872. DOI: 10.1109/CSAE.2012.6272913. Disponível em: http://ieeexplore.ieee.org/document/6272913/. Acesso em: 13 out. 2021.

CARAPUÇO, João; NEVES, Rui; HORTA, Nuno. Reinforcement learning applied to Forex trading. en. **Applied Soft Computing**, v. 73, p. 783–794, dez. 2018. ISSN 15684946. DOI: 10.1016/j.asoc.2018.09.017. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S1568494618305349. Acesso em: 3 ago. 2021.

CHAN, E. **Algorithmic Trading: Winning Strategies and Their Rationale**. [S.l.]: Wiley, 2013. (Wiley Trading). ISBN 9781118460146. Disponível em: https://books.google.com.br/books?id=WAlFDwAAQBAJ.

CHAN, E. **Quantitative Trading: How to Build Your Own Algorithmic Trading Business**. [S.l.]: Wiley, 2009. (Wiley Trading). ISBN 9780470466261. Disponível em: https://books.google.com.br/books?id=NZlV0M5Ije4C.

CHAN, K.C.C.; FOO KEAN TEONG. Enhancing technical analysis in the forex market using neural networks. In: PROCEEDINGS of ICNN'95 - International Conference on Neural Networks. Perth, WA, Australia: IEEE, 1995. v. 2, p. 1023–1027. ISBN 9780780327689. DOI: 10.1109/ICNN.1995.487561. Disponível em: http://ieeexplore.ieee.org/document/487561/. Acesso em: 3 ago. 2021.

CHEN, Kai; ZHOU, Yi; DAI, Fangyan. A LSTM-based method for stock returns prediction: A case study of China stock market. In: 2015 IEEE International Conference on Big Data (Big Data). Santa Clara, CA, USA: IEEE, out. 2015. P. 2823–2824. ISBN 9781479999262. DOI: 10.1109/BigData.2015.7364089. Disponível em: http://ieeexplore.ieee.org/document/7364089/. Acesso em: 13 out. 2021.

CHEN, Lin; GAO, Q. Application of Deep Reinforcement Learning on Automated Stock Trading. **2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)**, p. 29–33, 2019.

CHEN, Yu'an et al. DRQN-based 3D Obstacle Avoidance with a Limited Field of View. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.: s.n.], 2021. P. 8137–8143. DOI: 10.1109/IROS51168.2021.9635949.

CLEAR. **Clear Corretora**. Disponível em: https://corretora.clear.com.br/investimentos/contratos-cheios/. Acesso em: 11 dez. 2022.

CLEAR. **Minicontratos de índice e dólar: Clear corretora**. [S.l.: s.n.]. Disponível em: https://corretora.clear.com.br/investimentos/minicontratos/. Acesso em: 23 dez. 2022.

CLEVERT, Djork-Arné; UNTERTHINER, Thomas; HOCHREITER, Sepp. **Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)**. [S.l.: s.n.], 2016. arXiv: 1511.07289 [cs.LG].

DANG, Quang-Vinh. Reinforcement Learning in Stock Trading. In: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE, APPLIED MATHEMATICS AND APPLICATIONS. Hanoi, Vietnam: [s.n.], 2019. Disponível em: https://hal.archives-ouvertes.fr/hal-02306522.

DENG, Yue et al. Deep Direct Reinforcement Learning for Financial Signal Representation and Trading. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 3, p. 653–664, 2017. DOI: 10.1109/TNNLS.2016.2522401.

DONGDONG, Lv et al. An Empirical Study of Machine Learning Algorithms for Stock Daily Trading Strategy. **Mathematical Problems in Engineering**, v. 2019, p. 1–30, abr. 2019. DOI: 10.1155/2019/7816154.

DUMOULIN, Vincent; VISIN, Francesco. **A guide to convolution arithmetic for deep learning**. [S.l.: s.n.], 2018. arXiv: 1603.07285 [stat.ML].

ENGELBRECHT, Andries P. **Computational intelligence: an introduction**. 2nd ed. Chichester, England ; Hoboken, NJ: John Wiley & Sons, 2007. OCLC: ocn133465571. ISBN 9780470035610.

FISCHER, Thomas; KRAUSS, Christopher. Deep learning with long short-term memory networks for financial market predictions. en. **European Journal of Operational Research**, v. 270, n. 2, p. 654–669, out. 2018. ISSN 03772217. DOI: 10.1016/j.ejor.2017.11.054. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0377221717310652. Acesso em: 13 out. 2021.

FISCHER, Thomas G. Reinforcement learning in financial markets - a survey. en. **undefined**, 2018. Disponível em: https://www.semanticscholar.org/paper/Reinforcement-learning-in-financial-markets-a-Fischer/922864ede84bc49be4ac676951278a9b568b6383. Acesso em: 3 ago. 2021.

GUAN, Yushi; LI, Peiyao; LU, Cheng. Stock Price Prediction with CNN-LSTM Network. [S.l.]. Disponível em: https://gavinguan95.github.io/files/Stock_Price_Prediction_with_CNN-LSTM_Network.pdf.

GUNDUZ, Hakan; YASLAN, Yusuf; CATALTEPE, Zehra. Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. en. **Knowledge-Based Systems**, v. 137, p. 138–148, dez. 2017. ISSN 09507051. DOI: 10.1016/j.knosys.2017.09.023. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0950705117304252. Acesso em: 13 out. 2021.

GURESEN, Erkam; KAYAKUTLU, Gulgun; DAIM, Tugrul U. Using artificial neural network models in stock market index prediction. en. **Expert Systems with Applications**, v. 38, n. 8, p. 10389–10397, ago. 2011. ISSN 09574174. DOI: 10.1016/j.eswa.2011.02.068. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0957417411002740. Acesso em: 11 out. 2021.

HASSELT, Hado van; GUEZ, Arthur; SILVER, David. Deep Reinforcement Learning with Double Q-learning. en, set. 2015. Disponível em: https://arxiv.org/abs/1509.06461v3. Acesso em: 3 ago. 2021.

HAUSKNECHT, Matthew; STONE, Peter. Deep Recurrent Q-Learning for Partially Observable MDPs. **arXiv:1507.06527 [cs]**, jan. 2017. arXiv: 1507.06527. Disponível em: http://arxiv.org/abs/1507.06527. Acesso em: 3 ago. 2021.

HAYKIN, Simon S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009.

HE, Kaiming et al. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, jun. 2016. P. 770–778. ISBN 9781467388511. DOI: 10.1109/CVPR.2016.90. Disponível em: http://ieeexplore.ieee.org/document/7780459/. Acesso em: 11 out. 2021.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. **Neural Comput.**, MIT Press, Cambridge, MA, USA, v. 9, n. 8, p. 1735–1780, nov. 1997. ISSN 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. Disponível em: https://doi.org/10.1162/neco.1997.9.8.1735.

HSIEH, David A. Chaos and Nonlinear Dynamics: Application to Financial Markets. en. **The Journal of Finance**, v. 46, n. 5, p. 1839–1877, dez. 1991. ISSN 00221082. DOI: 10.1111/j.1540-6261.1991.tb04646.x. Disponível em: https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1991.tb04646.x. Acesso em: 3 ago. 2021.

HUANG, Chien Yi. **Financial Trading as a Game: A Deep Reinforcement Learning Approach**. [S.l.: s.n.], 2018. arXiv: 1807.02787 [q-fin.TR].

KARA, Yakup; ACAR BOYACIOGLU, Melek; BAYKAN, Ömer Kaan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. en. **Expert Systems with Applications**, v. 38, n. 5, p. 5311–5319, mai. 2011. ISSN 09574174. DOI: 10.1016/j.eswa.2010.10.027. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0957417410011711. Acesso em: 11 out. 2021.

KERAS. **Keras: the Python deep learning API**. Disponível em: https://keras.io/. Acesso em: 11 dez. 2022.

KINGMA, Diederik P.; BA, Jimmy. Adam: A Method for Stochastic Optimization. en, dez. 2014. Disponível em: https://arxiv.org/abs/1412.6980v9. Acesso em: 3 ago. 2021.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. In: PEREIRA, F. et al. (Ed.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2012. v. 25. Disponível em: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. en. **Communications of the ACM**, v. 60, n. 6, p. 84–90, mai. 2017. ISSN 0001-0782, 1557-7317. DOI: 10.1145/3065386. Disponível em: https://dl.acm.org/doi/10.1145/3065386. Acesso em: 3 ago. 2021.

KROLLNER, Bjoern; VANSTONE, Bruce; FINNIE, Gavin. Financial time series forecasting with machine learning techniques: A survey. English. In: PROCEEDINGS of the 18th European Symposium on Artificial Neural Networks (ESANN 2010). [S.l.: s.n.], 2010. P. 25–30. European Symposium on Artificial Neural Networks : Computational Intelligence and Machine Learning, ESANN 2010 ; Conference date: 28-04-2010 Through 30-04-2010. ISBN 2930307102.

KWON, Ki-Yeol; KISH, Richard J. Technical trading strategies and return predictability: NYSE. **Applied Financial Economics**, Routledge, v. 12, n. 9, p. 639–653, 2002. DOI: 10.1080/09603100010016139. eprint: https://doi.org/10.1080/09603100010016139. Disponível em: https://doi.org/10.1080/09603100010016139.

LECUN, Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition. en. **Neural Computation**, v. 1, n. 4, p. 541–551, dez. 1989. ISSN 0899-7667, 1530-888X. DOI: 10.1162/neco.1989.1.4.541. Disponível em: https://direct.mit.edu/neco/article/1/4/541-551/5515. Acesso em: 22 out. 2021.

LECUN, Yann; BENGIO, Y.; HINTON, Geoffrey. Deep Learning. **Nature**, v. 521, p. 436–44, mai. 2015. DOI: 10.1038/nature14539.

LECUN, Yann; BENGIO, Yoshua. Convolutional networks for images, speech, and time-series. In: **The handbook of brain theory and neural networks**. Edição: M.A. Arbib. [S.l.]: MIT Press, 1995.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. en. **Nature**, v. 521, n. 7553, p. 436–444, mai. 2015. ISSN 0028-0836, 1476-4687. DOI: 10.1038/nature14539. Disponível em: http://www.nature.com/articles/nature14539. Acesso em: 11 out. 2021.

LEE, Jinho et al. Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network. **IEEE Access**, v. 7, p. 167260–167277, 2019. ISSN 2169-3536. DOI: 10.1109/ACCESS.2019.2953542. Disponível em: https://ieeexplore.ieee.org/document/8901118/. Acesso em: 13 out. 2021.

LI, Xiaodong et al. News impact on stock price return via sentiment analysis. **Knowledge-Based Systems**, v. 69, p. 14–23, 2014. ISSN 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2014.04.022. Disponível em: https://www.sciencedirect.com/science/article/pii/S0950705114001440.

LI, Yuming; NI, Pin; CHANG, Victor. Application of deep reinforcement learning in stock trading strategies and stock forecasting. en. **Computing**, v. 102, n. 6, p. 1305–1322, jun. 2020. ISSN 0010-485X, 1436-5057. DOI: 10.1007/s00607-019-00773-w. Disponível em: http://link.springer.com/10.1007/s00607-019-00773-w. Acesso em: 3 ago. 2021.

LO, Andrew W.; MAMAYSKY, Harry; WANG, Jiang. Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. en. **The Journal of Finance**, v. 55, n. 4, p. 1705–1765, ago. 2000. ISSN 00221082. DOI: 10.1111/0022-1082.00265. Disponível em: http://doi.wiley.com/10.1111/0022-1082.00265. Acesso em: 3 ago. 2021.

MARKOWITZ, Harry. PORTFOLIO SELECTION*. **The Journal of Finance**, v. 7, n. 1, p. 77–91, 1952. DOI: https://doi.org/10.1111/j.1540-6261.1952.tb01525.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1952.tb01525.x. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x.

METATRADER. **MetaTrader 5, plataforma de negociação para Ações e Forex**. Disponível em: https://www.metatrader5.com/pt. Acesso em: 11 dez. 2022.

MNIH, Volodymyr et al. **Asynchronous Methods for Deep Reinforcement Learning**. [S.l.: s.n.], 2016. arXiv: 1602.01783 [cs.LG].

MNIH, Volodymyr et al. Human-level control through deep reinforcement learning. en. **Nature**, v. 518, n. 7540, p. 529–533, fev. 2015. ISSN 0028-0836, 1476-4687. DOI: 10.1038/nature14236. Disponível em: http://www.nature.com/articles/nature14236. Acesso em: 3 ago. 2021.

MOODY, John et al. Performance functions and reinforcement learning for trading systems and portfolios. **Journal of Forecasting**, v. 17, n. 56, p. 441–470, 1998.

MOODY, John E.; WU, Lizhong. Optimization of trading systems and portfolios. **Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr)**, p. 300–307, 1997.

NAIR, Arun et al. Massively Parallel Methods for Deep Reinforcement Learning, jul. 2015.

NAJAFABADI, Maryam M et al. Deep learning applications and challenges in big data analytics. en. **Journal of Big Data**, v. 2, n. 1, p. 1, dez. 2015. ISSN 2196-1115. DOI: 10.1186/s40537-014-0007-7. Disponível em: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7. Acesso em: 21 out. 2021.

NELOGICA. **Nelogica: O Indicador MACD**. Disponível em: https://www.nelogica.com.br/conhecimento/artigos/indicadores-estudo/indicador-macd. Acesso em: 26 dez. 2022.

OJHA, Varun Kumar; ABRAHAM, Ajith; SNÁŠEL, Václav. Metaheuristic design of feedforward neural networks: A review of two decades of research. en. **Engineering Applications of Artificial Intelligence**, v. 60, p. 97–116, abr. 2017. ISSN 09521976. DOI: 10.1016/j.engappai.2017.01.013. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0952197617300234. Acesso em: 19 out. 2021.

OLIVEIRA, Isaac de. **OS 25 fundos cambiais com O Melhor Retorno no ano - investimentos - estadão E-investidor - as PRINCIPAIS Notícias do Mercado Financeiro**. [S.l.]: Estadão E-Investidor - As principais notícias do mercado financeiro, out. 2020. Disponível em: https://einvestidor.estadao.com.br/investimentos/fundos-cambiais-melhor-retorno-janeiro-setembro/.

OORD, Aaron van den et al. **WaveNet: A Generative Model for Raw Audio**. [S.l.: s.n.], 2016. arXiv: 1609.03499 [cs.SD].

PERSIO, L.; HONCHAR, O. Artificial Neural Networks architectures for stock price prediction: comparisons and applications. en. **undefined**, 2016. Disponível em: https://www.semanticscholar.org/paper/Artificial-Neural-Networks-architectures-for-stock-Persio-Honchar/41487776364a6dddee91f1e2d2b78f46d0b93499. Acesso em: 13 out. 2021.

PHAN HUY, Tam; CUONG, Nguyen Thanh. Effectiveness of Investment Strategies Based on Technical Indicators: Evidence from Vietnamese Stock Markets ARTICLE INFO JEL Classification Keywords. **Journal of Insurance and Financial Management**, v. 3, abr. 2018.

RATNER, Mitchell; LEAL, Ricardo. Test of technical trading strategies in emerging equity markets of Latin America and Asia. **Journal of Banking Finance**, v. 23, p. 1887–1905, dez. 1999. DOI: 10.1016/S0378-4266(99)00042-4.

RETORNO, Mais. **Mais Retorno: BB TOP DOLAR FI CAMBIAL LP**. Jan. 2023. Disponível em: https://maisretorno.com/fundo/bb-top-dolar-fi-cambial-lp.

ROCHA, Rayana Souza et al. Sentiment Analysis of Twitter Data about Blockchain Technology. In: PROCEEDINGS of the 10th Euro-American Conference on Telematics and Information

Systems. Aveiro, Portugal: Association for Computing Machinery, 2021. (EATIS '20). ISBN 9781450377119. DOI: 10.1145/3401895.3401913. Disponível em: https://doi.org/10.1145/3401895.3401913.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. en. **Nature**, v. 323, n. 6088, p. 533–536, out. 1986. ISSN 0028-0836, 1476-4687. DOI: 10.1038/323533a0. Disponível em: http://www.nature.com/articles/323533a0. Acesso em: 19 out. 2021.

RUSSELL, Stuart et al. **Artificial intelligence: A modern approach**. [S.l.]: Prentice Hall, 2010. [Online; accessed 2022-12-10].

SHAH, Dev; ISAH, Haruna; ZULKERNINE, Farhana. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. **International Journal of Financial Studies**, v. 7, n. 2, 2019. ISSN 2227-7072. DOI: 10.3390/ijfs7020026. Disponível em: https://www.mdpi.com/2227-7072/7/2/26.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford University Press (OUP), v. 52, n. 3-4, p. 591–611, dez. 1965. DOI: 10.1093/biomet/52.3-4.591. Disponível em: https://doi.org/10.1093/biomet/52.3-4.591.

SHARPE, William F. The Sharpe Ratio. **The Journal of Portfolio Management**, Institutional Investor Journals Umbrella, v. 21, n. 1, p. 49–58, 1994. ISSN 0095-4918. DOI: 10.3905/jpm.1994.409501. eprint: https://jpm.pm-research.com/content/21/1/49.full.pdf. Disponível em: https://jpm.pm-research.com/content/21/1/49.

SIMONYAN, Karen; ZISSERMAN, Andrew. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. [S.l.: s.n.], 2015. arXiv: 1409.1556 [cs.CV].

SORNMAYURA, Sutta. Robust FOREX Trading with Deep Q Network (DQN). **ABAC Journal**, v. 39, 2019.

SORTINO, Frank A.; PRICE, Lee N. Performance Measurement in a Downside Risk Framework. **The Journal of Investing**, Institutional Investor Journals Umbrella, v. 3, n. 3, p. 59–64, 1994. ISSN 1068-0896. DOI: 10.3905/joi.3.3.59. eprint: https://joi.pm-research.com/content/3/3/59.full.pdf. Disponível em: https://joi.pm-research.com/content/3/3/59.

STUDENT. The probable error of a mean. **Biometrika**, JSTOR, p. 1–25, 1908.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement learning: an introduction**. Cambridge, Mass: MIT Press, 1998. (Adaptive computation and machine learning). ISBN 9780262193986.

SZEGEDY, Christian et al. **Going Deeper with Convolutions**. [S.l.: s.n.], 2014. arXiv: 1409.4842 [cs.CV].

TENSORFLOW. **TensorFlow**. Disponível em: https://www.tensorflow.org/. Acesso em: 11 dez. 2022.

TERRY LINGZE MENG, Matloob Khushi. Reinforcement Learning in Financial Markets. **Data**, v. 4, n. 3, p. 110, 2019. DOI: 10.3390/data4030110. Disponível em: https://www.mdpi.com/2306-5729/4/3/110/pdf.

THÉATE, Thibaut; ERNST, Damien. An application of deep reinforcement learning to algorithmic trading. en. **Expert Systems with Applications**, v. 173, p. 114632, jul. 2021. ISSN 09574174. DOI: 10.1016/j.eswa.2021.114632. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0957417421000737. Acesso em: 3 ago. 2021.

UOL. **Uol Economia:BC decide manter juros em 2% ao ano, menor patamar da história**. Out 2020. Disponível em: https://economia.uol.com.br/noticias/redacao/2020/10/28/bc-juros-selic-28-outubro.htm.

UOL. **Uol Economia:Bolsa tem queda de 0,45%, mas fecha semana no azul; dólar também cai**. Out 2020. Disponível em: https://economia.uol.com.br/cotacoes/noticias/redacao/2020/10/09/fechamento-dolar-bolsa-9-outubro.htm.

WANG, Y et al. Deep Q-trading. **Cslt.Riit.Tsinghua.Edu.Cn**, 2017. Disponível em: http://cslt.riit.tsinghua.edu.cn/mediawiki/images/5/5f/Dtq.pdf.

WATKINS, Christopher John Cornish Hellaby. **Learning from Delayed Rewards**. Mai. 1989. Tese (Doutorado) – King's College, Cambridge, UK. Disponível em: http://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf.

WILLIAMS, Ronald J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. en. **Machine Learning**, v. 8, n. 3-4, p. 229–256, mai. 1992. ISSN 0885-6125, 1573-0565. DOI: 10.1007/BF00992696. Disponível em: http://link.springer.com/10.1007/BF00992696. Acesso em: 3 ago. 2021.

WU, Xing et al. Adaptive stock trading strategies with deep reinforcement learning methods. **Inf. Sci.**, v. 538, p. 142–158, 2020.

XIONG, Zhuoran et al. Practical Deep Reinforcement Learning Approach for Stock Trading. **arXiv:1811.07522 [cs, q-fin, stat]**, dez. 2018. arXiv: 1811.07522. Disponível em: http://arxiv.org/abs/1811.07522. Acesso em: 13 out. 2021.

XP, Research. **Fundos Cambiais: Entenda o que São e Veja Como aplicar Nesse Investimento**. [S.l.: s.n.], ago. 2022. Disponível em: https://conteudos.xpi.com.br/aprenda-a-investir/relatorios/fundos-cambiais/.

YONG, Bang Xiang; ABDUL RAHIM, Mohd Rozaini; ABDULLAH, Ahmad Shahidan. A Stock Market Trading System Using Deep Neural Network. In: MOHAMED ALI, Mohamed Sultan et al. (Ed.). **Modeling, Design and Simulation of Systems**. Singapore: Springer Singapore, 2017. v. 751. P. 356–364. ISBN 9789811064623 9789811064630. DOI: 10.1007/978-981-10-6463-0_31. Disponível em: http://link.springer.com/10.1007/978-981-10-6463-0_31. Acesso em: 13 out. 2021.

ZHANG, Lei; WANG, Shuai; LIU, Bing. **Deep Learning for Sentiment Analysis : A Survey**. [S.l.]: arXiv, 2018. DOI: 10.48550/ARXIV.1801.07883. Disponível em: https://arxiv.org/abs/1801.07883.

ZHANG, Zihao; ZOHREN, Stefan; ROBERTS, Stephen. Deep Reinforcement Learning for Trading. en. **The Journal of Financial Data Science**, v. 2, n. 2, p. 25–40, abr. 2020. ISSN 2640-3943. DOI: 10.3905/jfds.2020.1.030. Disponível em: http://jfds.pm-research.com/lookup/doi/10.3905/jfds.2020.1.030. Acesso em: 3 ago. 2021.

ZHOU, Peng; TANG, Jingling. **Improved Method of Stock Trading under Reinforcement Learning Based on DRQN and Sentiment Indicators ARBR**. [S.l.]: arXiv, 2021. DOI: 10.48550/ARXIV.2111.15356. Disponível em: https://arxiv.org/abs/2111.15356.