

CENTRO UNIVERSITÁRIO FEI
DAVI GUARACHO NASCIMENTO

**REDES NEURAIS CONVOLUCIONAIS APLICADAS À NEGOCIAÇÃO DE ATIVOS
NO MERCADO FINANCEIRO**

São Bernardo do Campo

2021

DAVI GUARACHO NASCIMENTO

**REDES NEURAIS CONVOLUCIONAIS APLICADAS À NEGOCIAÇÃO DE ATIVOS
NO MERCADO FINANCEIRO**

Dissertação de mestrado apresentada ao Centro
Universitário da FEI para obtenção do título de
Mestre em Engenharia Elétrica, orientada pelo
Prof. Dr. Reinaldo Augusto da Costa Bianchi.

São Bernardo do Campo

2021

GUARACHO NASCIMENTO, DAVI.
REDES NEURAIAS CONVOLUCIONAIS APLICADAS À
NEGOCIAÇÃO DE ATIVOS NO MERCADO FINANCEIRO / DAVI
GUARACHO NASCIMENTO. São Bernardo do Campo, 2020.
112 p. : il.

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

1. Redes Neurais Convolucionais. 2. Aprendizado Profundo. 3.
Análise Técnica. 4. Mercado de ações. I. Augusto da Costa
Bianchi, Reinaldo, orient. II. Título.

Aluno: Davi Guaracho do Nascimento

Matrícula: 119105-5

Título do Trabalho: Redes neurais convolucionais aplicadas à negociação de ativos no mercado financeiro.

Área de Concentração: Inteligência Artificial Aplicada à Automação e Robótica

Orientador: Prof. Dr. Reinaldo Augusto da Costa Bianchi

Data da realização da defesa: 11/12/2020

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

A banca foi realizada no dia 11 de dezembro às 09:00 horas, e se iniciou com a apresentação do aluno, que foi muito boa, 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.

São Bernardo do Campo, / / .

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Reinaldo Augusto da Costa Bianchi

Ass.: _____

Prof. Dr. Ricardo de Carvalho Destro

Ass.: _____

Prof. Dr. Leandro dos Santos Maciel

Ass.: _____

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

APROVADO ☒

REPROVADO ☐

VERSÃO FINAL DA DISSERTAÇÃO

**APROVO A VERSÃO FINAL DA DISSERTAÇÃO EM QUE
FORAM INCLUÍDAS AS RECOMENDAÇÕES DA BANCA
EXAMINADORA**

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Dedico este trabalho a Deus; sem ele eu não teria capacidade para desenvolver este trabalho.

AGRADECIMENTOS

A Deus, que me guia todos os dias e ilumina meu caminho, pelo dom da vida.

Aos meus pais, Adriana e Edmilson, por todo incentivo, amor incondicional, carinho e apoio durante o período do mestrado e por toda minha jornada acadêmica até então.

Ao meu orientador, Prof. Dr. Reinaldo Augusto da Costa Bianchi, pelo apoio, paciência e pelos valiosos ensinamentos e sua orientação que foram fundamentais para a elaboração deste trabalho.

Ao Centro Universitário FEI pelo apoio institucional.

A todos os meus colegas do Centro Universitário FEI, Prof. Dr. Ricardo de Carvalho Destro, Prof. Dr. Thiago Pedro Donadon, Victor Biazon e outros mais, pelas discussões, ideias, conselhos e sugestões neste texto.

À FAPESP, CAPES e CNPq, pelo apoio financeiro. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“Se você quer ter resultados melhores do que a maioria das pessoas, faça as coisas de forma diferente da maioria.”

John Templeton

“A volatilidade dos mercados é a maior aliada do verdadeiro investidor.”

Warren Buffet

“Eu seria mendigo nas ruas com uma caneca de lata se os mercados fossem eficientes.”

Warren Buffet

RESUMO

Previsão do mercado financeiro tem sido um desafio bastante popular nas pesquisas de Aprendizado de Máquina (AM). O desejo da maioria dos investidores é tomar decisões com base em critérios objetivos que venham a proporcionar maior retorno nas operações. Recentemente, estudos têm usado técnicas de Aprendizado Profundo (AP), como Redes Neurais Convolucionais (*Convolutional Neural Networks* - *CNN*), para realizar regressão nos preços ou classificação de sinal de negociação em ativos do mercado financeiro. Neste trabalho, é proposta uma arquitetura de sistema que utiliza uma CNN a fim de realizar a indicação da melhor operação para cada momento no mercado de ações, este sistema foi chamado *CNN Trading Classifier* (*CNN-TC*). Este sistema é composto por pré-processamento dos dados, classificação pelo modelo CNN e tomada de decisão no mercado. O mesmo foi avaliado com base em dados das bolsas de valores brasileira e americana em três períodos diferentes, para isto foram feitas avaliação estatística, utilizando as métricas de classificação acurácia, precisão, revocação e F1, e financeira com base nas classificações realizadas pelo modelo. Além disso foi realizado um teste em ambiente simulado utilizando o software MetaTrader a fim de atestar a eficácia desta abordagem. Os resultados mostram que o sistema teve resultados estatísticos e financeiros melhores na maioria das avaliações em comparação com o uso de outros modelos de AP e superou a estratégia *Buy and Hold* (*BH*) e retornos da renda fixa.

Palavras-chave: Redes Neurais Convolucionais. Aprendizado Profundo. Análise Técnica. Mercado de Ações.

ABSTRACT

Stock market forecasting has been a quite popular challenge in machine learning researches. Most investors want to make decisions based on criteria that will provide greater returns in their operations. Recently, studies have been using Deep Learning techniques, such as Convolutional Neural Networks (CNN), to perform price regression or trade signal classification in financial market. In this work, a system architecture that uses a CNN model is proposed to perform the indication of the best operation for each moment in the stock market, this system was called CNN Trading Classifier (CNN-TC). This system consists of data pre-processing, classification by the model and decision making in the market. It was evaluated based on data from the Brazilian and American stock market in three different periods. For this, statistical evaluation was performed, using the metrics of accuracy, precision, recall and F1 classification, and financial based on the classifications performed by the model. In addition, a test on a simulated environment using the MetaTrader software was evaluated in order to attest to the effectiveness of this approach. The results show that the system had better statistical and financial results in most evaluations compared to the use of other Deep Learning models and overcame the strategy Buy and Hold (BH) and fixed income returns.

Keywords: Convolutional Neural Networks. Deep Learning. Technical Analysis. Stock Market

LISTA DE ILUSTRAÇÕES

Figura 1	– Crescimento do volume financeiro anual negociado na bolsa Brasileira, em reais e em dólares	18
Figura 2	– CNN utilizada por Wang et al. (2018)	26
Figura 3	– Abordagem proposta por Tsai, Chen e Wang (2019) utilizando CNN	28
Figura 4	– CNN utilizada por Sezer e Ozbayoglu (2018)	29
Figura 5	– Sumário da revisão da literatura	31
Figura 6	– Conjunto de indicadores utilizados no trabalho	36
Figura 7	– Preço de fechamento no gráfico superior e preço de fechamento estacionário no inferior	40
Figura 8	– Representação do neurônio artificial	41
Figura 9	– Gráfico da função Sigmóide	42
Figura 10	– Gráfico da função TanH	42
Figura 11	– Representação de rede MLP	45
Figura 12	– Representação de uma LSTM	47
Figura 13	– Representação do processo de uma CNN	48
Figura 14	– Representação de conectividade global e esparsa das redes MLPs e CNNs .	49
Figura 15	– Exemplo do campo receptivo local das CNNs	50
Figura 16	– Compartilhamento de parâmetros. Conexões de mesma cor representam os elementos do filtro sendo compartilhados	51
Figura 17	– Representação de duas etapas do processo de convolução em uma matriz de entrada 6x6 com um filtro 3x3	52
Figura 18	– Exemplos de mapas de características	53
Figura 19	– Gráfico da função ReLU	54
Figura 20	– Representação do processo de <i>max-pooling</i>	55
Figura 21	– Representação da técnica de <i>Dropout</i>	56
Figura 22	– Representação do neurônio convolucional	57
Figura 23	– Representação da convolução nos pesos da entrada	58
Figura 24	– Representação da convolução nos pesos do filtro	59
Figura 25	– Comparação do método Adam com demais métodos de otimização	61
Figura 26	– Processo do sistema de negociação proposto	66
Figura 27	– Processo de remodelagem dos dados de entrada	67

Figura 28 – Rótulos com pontos de compra e venda. Dias de Compra marcados em verde, Venda em vermelho e dias Neutro não estão marcados.	69
Figura 29 – Arquitetura da CNN utilizada no sistema red	70
Figura 30 – Período de treinamento em azul, validação em roxo e de avaliação em amarelo	74
Figura 31 – Retorno do CDI entre 02/01/2015 e 01/01/2019	78
Figura 32 – Retorno do CDI entre 01/03/2006 e 01/01/2012	79
Figura 33 – Retorno do CDI entre 01/03/2007 e 01/01/2016	79
Figura 34 – Retorno sobre investimento para BOVA11	81
Figura 35 – Retorno sobre investimento para VALE3	82
Figura 36 – Retorno sobre investimento para PETR4	83
Figura 37 – Retorno sobre investimento para ITUB4	84
Figura 38 – Retorno sobre investimento para ABEV3	85
Figura 39 – Retorno sobre investimento para WEGE3	86
Figura 40 – Retorno sobre investimento para AAPL	87
Figura 41 – Retorno sobre investimento para WFC	89
Figura 42 – Retorno sobre investimento para JPM	90
Figura 43 – Retorno sobre investimento para TRV	91
Figura 44 – Retorno sobre investimento para XLE	92
Figura 45 – Retorno sobre investimento para XLF	94
Figura 46 – Retorno sobre investimento no teste em ambiente simulado	96
Figura 47 – Captura de tela do relatório do MetaTrader	96

LISTA DE TABELAS

Tabela 1	–	Ativos utilizados nos experimentos	71
Tabela 2	–	Métricas estatísticas para BOVA11	80
Tabela 3	–	Métricas estatísticas para VALE3	81
Tabela 4	–	Métricas estatísticas para PETR4	82
Tabela 5	–	Métricas estatísticas para ITUB4	84
Tabela 6	–	Métricas estatísticas para ABEV3	85
Tabela 7	–	Métricas estatísticas para WEGE3	86
Tabela 8	–	Métricas estatísticas para AAPL	87
Tabela 9	–	Métricas estatísticas para WFC	88
Tabela 10	–	Métricas estatísticas para JPM	89
Tabela 11	–	Métricas estatísticas para TRV	91
Tabela 12	–	Métricas estatísticas para XLE	92
Tabela 13	–	Métricas estatísticas para XLF	93
Tabela 14	–	Taxas de emolumentos B3 para negociação e liquidação	95
Tabela 15	–	Resultados financeiros das estratégias analisadas	97
Tabela 16	–	Médias das métricas estatísticas para mercado brasileiro	97
Tabela 17	–	Desvio padrão das métricas estatísticas para mercado brasileiro	97
Tabela 18	–	Médias das métricas estatísticas para mercado americano	98
Tabela 19	–	Desvio padrão das métricas estatísticas para mercado americano	98
Tabela 20	–	Matrizes de confusão para o ativo BOVA11	100

LISTA DE ALGORITMOS

Algoritmo 1 – Rotulação de amostras	68
Algoritmo 2 – Avaliação financeira	77

LISTA DE ABREVIATURAS

AAPL	Apple Inc.
ABEV3	Ambev S.A.
AG	Algoritmo Genético
AM	Aprendizado de Máquina
AP	Aprendizado Profundo
API	Interface de Programação de Aplicações (Application Programming Interface)
B3	Brasil, Bolsa, Balcão
BCB	Banco Central do Brasil
BH	Buy and Hold
BOVA11	iShares Ibovespa Fundo de Índice
BOVESPA	Bolsa de Valores do Estado de São Paulo
BR	Brasil
CDI	Certificado de Depósito Interbancário
CNN	Rede Neural Convolutacional (<i>Convolutional Neural Network - CNN</i>)
CNN-TA	Convolutional Neural Network - Technical Analysis
CNN-TC	CNN Trading Classifier
CSE	Colombo Stock Exchange
DBN	Rede de Crença Profunda (<i>Deep Belief Network - DBN</i>)
ETF	Exchange-Traded fund
EUA	Estados Unidos da América
GAN	Rede Adversarial Generativa (<i>Generative Adversarial Network - GAN</i>)
GRU	Unidade Recorrente Chaveada (<i>Gated Recurrent Unit - GRU</i>)
IBOVESPA	Índice Bovespa
ITUB4	Itaú Unibanco S.A.
JPM	JPMorgan Chase & Co.
LOG	Logistic Regression Classifier
LSTM	Rede de Memória de Curto Longo Prazo (<i>Long Short Term Memory - LSTM</i>)
MACD	Convergência e Divergência de Médias Móveis (<i>Moving Average Convergence Divergence - MACD</i>)
MFI	Índice de Fluxo Monetário(<i>Money Flow Index - MFI</i>)
MLP	Multilayer Perceptron

NSE	National Stock Exchange
NYSE	New York Stock Exchange
OBV	Volume de equilíbrio (<i>On-balance volume - OBV</i>)
PCA	Análise de Componentes Principais (Principal Component Analysis)
PETR4	Petrobras S.A.
PFE	Preço de Fechamento Estacionário
PLN	Processamento de Linguagem Natural
RAF	Random Forest
RBFNN	Rede Neural de Função Base Radial (<i>Radial Basis Function Neural Network - RBFNN</i>)
RBM	Máquina de Boltzmann Restrita (<i>Restricted Boltzmann Machine - DBM</i>)
ReLU	Rectified Linear Unit
RNA	Redes Neurais Artificiais
RNN	Rede Neural Recorrente (<i>Recurrent Neural Network - RNN</i>)
ROC	Taxa de Variação (<i>Rate of Change - RSI</i>)
RSI	Índice de Força Relativa (<i>Relative Strength Index - RSI</i>)
SRNN	Rede Neural Recorrente Simples (<i>Simple Recurrent Neural Network - SRNN</i>)
SVM	Máquina de Vetores de Suporte (<i>Support Vector Machines - SVM</i>)
TanH	Tangente Hiperbólica
TRV	The Travelers Companies
VALE3	Vale S.A.
W%R	Williams %R
WEGE3	Weg S.A.
WFC	Wells Fargo & Co.
XGBoost	Extreme Gradient Boosting
XLE	Energy Select Sector SPDR ETF
XLF	Financial Select Sector SPDR ETF

SUMÁRIO

1	INTRODUÇÃO	18
1.1	Motivação	20
1.2	Objetivos	21
1.3	Justificativa	21
1.4	Organização do trabalho	22
2	TRABALHOS CORRELATOS	23
3	FUNDAMENTAÇÃO TEÓRICA	32
3.1	Estratégias de Investimento	32
3.1.1	Análise Fundamentalista	32
3.1.2	Análise Técnica	34
3.1.3	Indicadores Técnicos	35
3.1.3.1	<i>Índice de Força Relativa (Relative Strength Index - RSI)</i>	37
3.1.3.2	<i>Williams %R (WR)</i>	37
3.1.3.3	<i>Índice de Fluxo Monetário (Money Flow Index - MFI)</i>	38
3.1.3.4	<i>Índice de Canal de Mercadoria (Commodity Channel Index - CCI)</i>	38
3.2	Série Estacionária	39
3.3	Redes Neurais Artificiais (RNA)	40
3.3.1	Rede Neural <i>Feedforward</i>	43
3.3.2	<i>Multilayer Perceptron (MLP)</i>	44
3.3.3	Rede de Memória de Curto Longo Prazo (<i>Long Short Term Memory - LSTM</i>)	46
3.3.4	Rede Neural Convolucional (<i>Convolutional Neural Network - CNN</i>)	47
3.3.4.1	<i>Camada convolucional</i>	48
3.3.4.2	<i>Camada de pooling</i>	54
3.3.4.3	<i>Dropout</i>	55
3.3.4.4	<i>Camadas totalmente conectadas</i>	56
3.3.4.5	<i>Retropropagação na CNN</i>	56
3.3.4.6	<i>Método de otimização</i>	59
3.3.4.7	<i>Inicialização de pesos</i>	61
3.3.5	Aprendizado Profundo (AP)	63
4	SISTEMA DE NEGOCIAÇÃO PROPOSTO	65
4.1	Arquitetura do sistema	65

4.2	Dados de entrada	67
4.3	Pré-processamento de dados	67
4.4	Rotulação dos Dados	68
4.5	Arquitetura da CNN utilizada no sistema	69
5	EXPERIMENTOS	71
5.1	Ativos utilizados nos experimentos	71
5.2	Desenvolvimento dos Modelos Comparados	72
5.3	Processo de Treinamento	73
5.4	Testes em Ambiente Simulado Utilizando o MetaTrader	74
5.5	Avaliação Estatística	75
5.5.1	Acurácia	75
5.5.2	Precisão	76
5.5.3	Revocação	76
5.5.4	F1	76
5.6	Algoritmo de Avaliação Financeira	76
6	RESULTADOS	80
6.1	BOVA11	80
6.2	VALE3	81
6.3	PETR4	82
6.4	ITUB4	83
6.5	ABEV3	84
6.6	WEGE3	85
6.7	AAPL	87
6.8	WFC	88
6.9	JPM	89
6.10	TRV	90
6.11	XLE	91
6.12	XLFX	93
6.13	Custos e Emolumentos	94
6.14	Teste em ambiente simulado	95
6.15	Discussões	96
7	CONCLUSÃO	101
7.1	Trabalhos Futuros	103

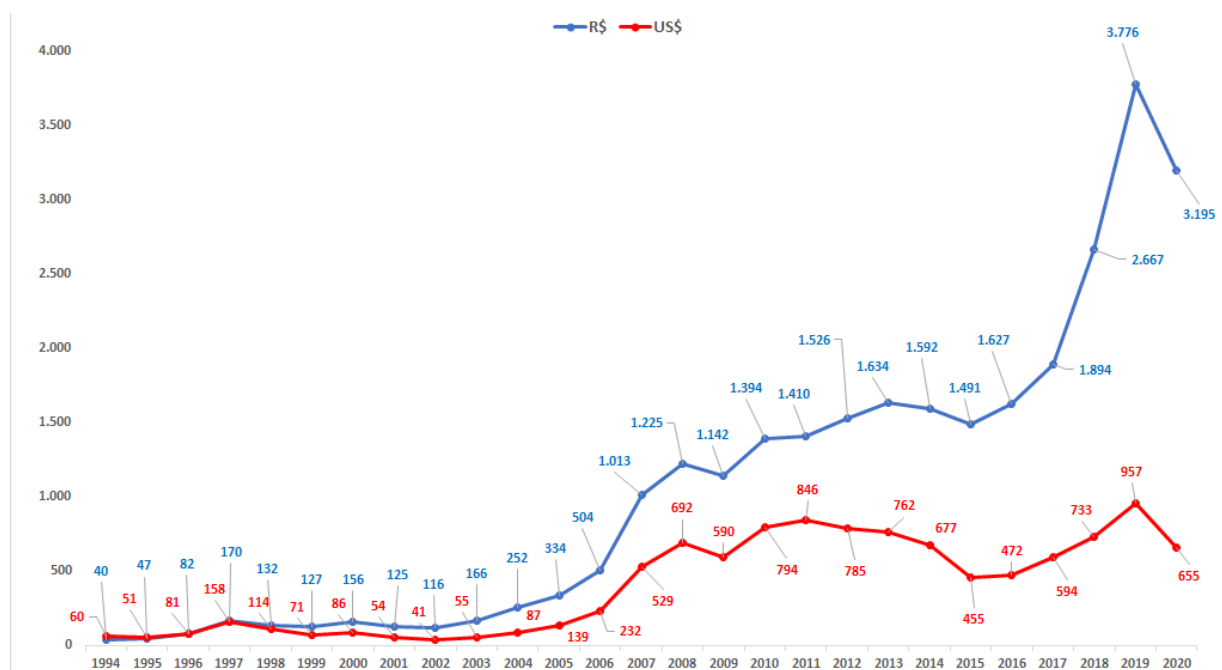
REFERÊNCIAS	105
------------------------------	------------

1 INTRODUÇÃO

O desejo da maioria dos investidores e pessoas que operam no mercado financeiro é basear suas decisões em critérios objetivos que venham a proporcionar maior retorno nas operações. Nos últimos anos, previsão do mercado financeiro tem sido um tema muito explorado na comunidade de Aprendizado de Máquina (AM) por pesquisadores na área de finanças. Na literatura, diversas técnicas de Aprendizado Profundo (AP) (HUANG; CHAI; CHO, 2020), alinhadas com a Análise Técnica, têm sido aplicadas buscando encontrar soluções para este desafio de forma que seja possível obter um retorno financeiro expressivo nas negociações de ativos financeiros.

O número de pessoas realizando operações no mercado financeiro, assim como o volume financeiro negociado, tem crescido nos últimos anos e como resultado, a busca pelo desenvolvimento de sistemas de decisão inteligentes está recebendo cada vez mais atenção nos mercados financeiros globalmente. No Brasil, o mercado de ações cresceu de forma significativa nos últimos anos. Dados fornecidos pela Economatica (2020) evidenciam o crescimento do volume financeiro negociado em ações entre 1994 e o primeiro semestre de 2020, conforme mostrado na Figura 1.

Figura 1 – Crescimento do volume financeiro anual negociado na bolsa Brasileira, em reais e em dólares



Fonte: adaptado de (ECONOMATICA, 2020)

Os preços das ações são naturalmente altamente voláteis, e escolher o momento ideal para comprar ou vender pode ser uma tarefa complexa, dependendo de muitas variáveis; esta é uma razão pela qual a estratégia *Buy and Hold (BH)* é a escolha preferida por muitos investidores. BH consiste em uma estratégia de comprar e manter ações ao longo do tempo, considerando que esta se valorizará no longo prazo. Desta forma o investidor não se preocupa com a volatilidade dos preços, poupando a si mesmo de tomar decisões baseado em suas emoções. Entretanto, esta metodologia de investimento exige acompanhamento de resultados e balanços das empresas, bem como o comportamento de seus preços, para decidir quando uma ação está barata/cara o suficiente que mereça ser comprada/vendida.

Segundo a Hipótese do Mercado Eficiente, a flutuação dos preços das ações desconta todas as informações conhecidas sobre elas (MALKIEL; FAMA, 1970), bem como as preocupações e expectativas do mercado com relação ao futuro da empresa. Já a teoria do do Passeio Aleatório (*Random Walk*) (FAMA, 1995), afirma que o mercado financeiro funciona de forma irracional, sendo os movimentos dos preços imprevisíveis, seguindo um "passeio aleatório". Sendo assim, um investidor individual muitas vezes não dispõe de tempo ou conhecimento suficiente para avaliar empresas do ponto de vista de seus fundamentos e recorrem à chamada Análise Técnica (BULKOWSKI, 2011). Esta estratégia consiste em realizar operações de compra e venda a fim obter ganhos de capital nas flutuações dos preços.

Na tentativa de auxiliar nesta tomada de decisão e prever tendências futuras no mercado de ações, os investidores aprofundaram os estudos em Análise Técnica nos últimos anos. Este tipo de análise tem por objetivo encontrar regiões de interesse no preço e momentos para compra e venda baseando-se no histórico de preços passados. Padrões de gráficos de *Candlestick* (EDWARDS; BASSETTI; MAGEE, 2012) e vários indicadores técnicos (COLBY; MEYERS, 1988) foram desenvolvidos usando as cotações históricas para encontrar o melhor ponto de entrada de operação. Entre os mais populares, pode-se citar os indicadores rastreadores de tendências, de volatilidade e outros baseados no volume financeiro negociado (COLBY; MEYERS, 1988).

Atualmente, existem estratégias de investimento sofisticadas que fazem o uso combinado da Análise Fundamentalista com a Análise Técnica para selecionar empresas e tomar decisões de momentos de operação no mercado. Um exemplo são os fundos de investimento quantitativos, os quais utilizam algoritmos computacionais para analisar dados das empresas e do mercado de ações.

Segundo Li (2019), atualmente aproximadamente 80% das negociações realizadas no mercado de ações norte americano são automatizadas, o que significa que cada vez mais o mercado tem reagido às manchetes e está mais propenso a oscilações bruscas de preços devido aos algoritmos que exercem grande influência no mercado.

Na literatura, podemos encontrar algumas abordagens que utilizam diversas técnicas diferentes de AM para prever tendências ou preços futuros do mercado, tais como Máquina de Vetores de Suporte (*Support Vector Machines - SVM*) (CAVALCANTE et al., 2016), Rede Neural Recorrente (*Recurrent Neural Network - RNN*) (KRAUSS; DO; HUCK, 2017), Rede de Memória de Curto Longo Prazo (*Long Short Term Memory - LSTM*) (FISCHER; KRAUSS, 2018) e Rede Neural Convolucional (*Convolutional Neural Network - CNN*) (SEZER; OZBAYOGLU, 2018).

Dentre estes, a CNN é conhecida como um modelo de RNA eficaz para a extração de padrões em muitas áreas e tem sido o modelo de AP mais comumente adaptado para diversos problemas, tais como classificação e reconhecimento de objetos em visão computacional e Processamento de Linguagem Natural (PLN). Na competição chamada *ImageNet*, focada em visão computacional, a rede AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) alcançou uma taxa de sucesso de 60 a 65% e, mais recentemente, a GooLeNet (SZEGEDY et al., 2015), Inception (v3, v4) (SZEGEDY et al., 2016, 2017) e ResNet (v101, v152) (HE et al., 2016) alcançaram aproximadamente 75-80% de taxa de sucesso (CANZIANI; PASZKE; CULURCIO, 2016).

1.1 Motivação

O investimento no mercado financeiro requer gestão sobre muitas variáveis para que não se tenha prejuízo. Ainda que exista certa complexidade inerente a este tipo de investimento, devido à taxa de juros baixa no ano atual, a renda variável tem atraído muitos investidores devido à possibilidade de retornos melhores em comparação com a renda fixa.

O interesse em reduzir os riscos e manter um retorno expressivo é comum na maioria dos investidores, bem como o desejo destes em aumentar os retornos nas operações. Dessa forma, a melhoria de técnicas de previsão e de gestão e controle de risco auxilia fortemente na tomada de decisão.

Além disso, existem fundos de investimentos especializados em investimentos utilizando estratégias quantitativas. Esta estratégia consiste na utilização de modelos matemáticos e

algoritmos computacionais a fim de avaliar oportunidades de investimento de forma a alcançar resultados consistentes.

Sendo assim, o estudo de previsão com uso de RNAs aplicadas a finanças computacionais torna este trabalho relevante devido à carência de estudos nesta área no Brasil.

1.2 Objetivos

O objetivo deste trabalho é propor uma nova abordagem intitulada *CNN Trading Classifier (CNN-TC)* para analisar e classificar séries temporais utilizando um modelo de CNN visando identificar o melhor momento de negociação de ações no mercado financeiro a fim de construir uma estratégia de operação mais rentável.

A seguir são listados os objetivos específicos desta pesquisa:

- a) Desenvolver um sistema utilizando um modelo de AP dinâmico para previsão de mercados financeiros;
- b) Avaliar esta abordagem com base em métricas estatísticas;
- c) Comparar os resultados de experimentos simulados usando dados históricos passados do mercado financeiro com outros *benchmarks* (índices de referência comumente utilizados para avaliar o retorno de um investimento) a fim de avaliar a viabilidade deste sistema;
- d) Avaliar a aplicabilidade em cenário real simulado de operação utilizando um robô em uma plataforma de negociação simulada.

1.3 Justificativa

Esta pesquisa justifica-se pela sua contribuição para melhorar a relação de risco e retorno constante em aplicações no mercado de capitais e auxiliar na tomada de decisão no mercado de ações. Espera-se ainda que este trabalho auxilie no desenvolvimento de novas técnicas de previsão e predição com a utilização de outros modelos semelhantes.

O uso de RNA pode reduzir a incerteza em vários aspectos das estratégias de investimento, bem como melhorar a gestão do risco das mesmas. Neste trabalho será abordado, principalmente, a identificação da melhor oportunidade de operação com base na previsão de tendências de retornos, com avaliação financeira das operações realizadas pelas estratégias automatizadas.

1.4 Organização do trabalho

Além deste capítulo introdutório, o restante deste trabalho está estruturado em 6 capítulos da seguinte forma: o capítulo 2 apresenta a revisão bibliográfica e os trabalhos relacionados a esta proposta. O capítulo 3 discorre sobre a fundamentação teórica necessária para a compreensão e reprodução deste trabalho. Na sequência, no capítulo 4 é definido o sistema proposto. A metodologia de treinamento utilizada, bem como a descrição dos experimentos realizados são apresentadas no capítulo 5. O capítulo 6 apresenta os resultados acompanhada de uma discussão sobre os mesmos. Por fim, o capítulo 7 apresenta a conclusão deste trabalho e os trabalhos futuros.

2 TRABALHOS CORRELATOS

Neste capítulo serão apresentadas algumas abordagens para previsão de mercados financeiros que foram desenvolvidas utilizando modelos de AP. Algumas implementações tiveram como base a realização de regressão sobre de dados financeiros, enquanto outras foram baseadas no uso de indicadores técnicos para realizar classificação sobre séries temporais.

Bao, Yue e Rao (2017) propuseram uma abordagem de AP para prever o preço de fechamento do dia seguinte usando as transformadas *Wavelet* e auto-codificadores combinados com uma arquitetura LSTM a fim de tomar decisões no mercado financeiro. Os resultados fornecem evidências de que o modelo é capaz de prever séries temporais com boas métricas e desta auxiliar estratégias de investimento para superar o método de investimento BH e as abordagens usando sistema de negociação que implementam apenas LSTM e RNN puras.

Selvin et al. (2017) basearam sua pesquisa nas arquiteturas LSTM e CNN para prever os preços futuros das ações do mercado de ações da Índia. Os experimentos mostraram que a arquitetura CNN é capaz de identificar as mudanças nas tendências e foi o melhor modelo na comparação, o que evidencia a capacidade deste modelo de RNA em identificar padrões sobre séries temporais.

Sezer, Ozbayoglu e Dogdu (2017) propuseram um sistema utilizando MLP composto por quatro camadas para classificar o sinal de negociação no mercado de ações Dow Jones para cada dia. Os dados de entrada utilizados eram um conjunto de indicadores técnicos composto por: Índice de Força Relativa (*Relative Strength Index - RSI*), Williams %R e Convergência e Divergência de Médias Móveis (*Moving Average Convergence Divergence - MACD*). Segundo os autores, o sistema desenvolvido executou a tarefa de classificação de compra, venda ou neutro na indicação diária ao longo de um período de tempo e mostrou resultados promissores. Além disso, os resultados mostram que os indicadores selecionados contribuem para a tarefa de classificação de séries temporais e apresentam boa correlação com o movimento dos preços dos ativos.

Samarawickrama e Fernando (2017) fizeram uma comparação entre alguns modelos na previsão de preços futuros no mercado financeiro Colombo Stock Exchange (CSE). Os modelos escolhidos foram MLP, Unidade Recorrente Chaveada (*Gated Recurrent Unit - GRU*), LSTM e Rede Neural Recorrente Simples (*Simple Recurrent Neural Network - SRNN*). Como entrada dos modelos foram utilizados os preços de máxima, mínima e fechamento de dois dias consecutivos. Como conclusão, foi avaliado que LSTM e SRNN tiveram menos erros em geral,

exceto em casos específicos, onde o MLP teve melhores resultados. O modelo GRU teve o pior resultado em acurácia quando comparado com os outros modelos. Este trabalho evidencia que os modelos de RNNs são mais aplicáveis em tarefas de regressão.

Tsantekidis et al. (2017) propuseram um sistema utilizando CNN para análise do livro de ofertas dos ativos a fim de identificar a tendência do preço em instantes seguintes. O trabalho teve por objetivo propor uma abordagem capaz de realizar negociação de alta frequência, ou seja, obter lucro financeiro em pequenas variações do preço ao longo do dia. O sistema proposto foi comparado com o uso de SVM e MLP. Os resultados demonstram que a CNN apresenta resultados melhores para esta tarefa de classificação, alcançando precisão próxima de 70% frente a 60% dos demais. Este trabalho também demonstra que um modelo CNN pode ser usado para classificar tendência no mercado financeiro.

Nelson, Pereira e Oliveira (2017) utilizaram um modelo de LSTM na tarefa de predição de tendência futura para ações no mercado financeiro brasileiro utilizando indicadores técnicos. As variáveis de entrada somaram 180, considerando indicadores e histórico de preços (máxima, mínima, abertura e fechamento) em intervalos de 15 minutos e o modelo informava se o preço iria subir ou cair. Os resultados apresentam acurácia próxima de 55,9%. Além disso o retorno financeiro utilizando o sistema como tomada de decisão foi positivo no período de avaliação, enquanto o retorno dos ativos analisados foi negativo. Isto evidencia que o modelo LSTM pode ser utilizado para classificar tendência futuras sobre séries temporais.

Fischer e Krauss (2018) utilizaram um modelo de LSTM para realizar previsão no mercado financeiro. O estudo utilizou o índice S&P 500 entre os anos 1992 e 2005 e foram feitas comparações com modelos de Random Forest (RAF) e Logistic Regression Classifier (LOG). Como conclusão os autores afirmam que o modelo LSTM é o mais adequado para previsão de valores série temporal, superando modelos mais simples de forma evidente.

Jain, Gupta e Moghe (2018) desenvolveram três abordagens diferentes de regressão utilizando CNN, LSTM e CNN-LSTM para a tarefa de previsão de preços futuros. Todos os modelos foram baseados em convoluções 1D e os dados de entrada utilizados eram compostos de preços diários de abertura, máxima, mínima e fechamento dos últimos 5 dias úteis consecutivos. Os resultados mostraram que os modelos de RNA podem ser empregados na previsão de preços futuros de dados de ações e uma combinação de CNN-LSTM é capaz de fornecer resultados mais precisos quando comparados a cada arquitetura individualmente. Além disso, o trabalho reforça que um modelo CNN possibilita a identificação de padrões, enquanto o

LSTM complementa o resultado final realizando a previsão de valores de série temporal com base nos padrões destacados pela CNN.

Nobre e Neves (2019) em seu artigo utilizou transformadas de *Wavelets* com Extreme Gradient Boosting (XGBoost) e PCA para realizar previsão de tendências no mercado de ações. Para isto foram gerados 26 indicadores técnicos junto do histórico de preços de abertura, máxima, mínima e fechamento. Foi utilizado o PCA para extrair as variáveis mais relevantes e reduzir dimensionalidade. O conjunto de dados resultante foi então alimentado para um classificador binário XGBoost. Segundo os autores, os resultados obtidos mostram que este sistema é capaz de superar a estratégia BH em três dos cinco mercados financeiros analisados, com uma média de retorno de 49,26%, enquanto o BH teve em média 32,41%. Este trabalho demonstra que o movimento dos preços e melhor escolha dos indicadores, selecionando so que possuem maior correlação com os preços, podem ser aplicáveis em algoritmos de AM para previsão.

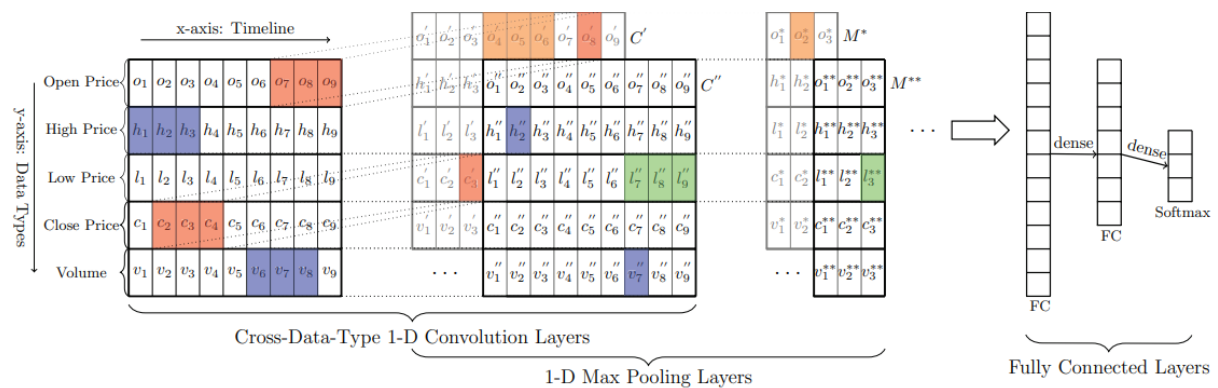
Cao e Wang (2019) desenvolveram dois sistema utilizando CNN e uma combinação CNN-SVM para realizar regressão nos valores do índice do mercado financeiro da Tailândia, *Thai Stock Exchange*. Os resultados mostraram grande eficiência na previsão dos valores do índice e o modelo combinado de CNN-SVM apresentou grande correlação com os preços reais e altas taxas de acurácia.

Sayavong, Wu e Chalita (2019) também desenvolveram um sistema de regressão baseado em CNN para realizar regressão em três ações do índice do mercado de ações da Tailândia. O sistema utilizou como dados de entrada os valores histórico de preços de abertura, máxima, fechamento e mínima dos ativos. Segundo os autores, os resultados mostraram que o modelo pode efetivamente identificar as tendências variáveis dos preços das ações e prever preços futuros com boas taxas de acurácia, podendo fornecer uma referência importante sobre o possível futuro comportamento do mercado.

Wang et al. (2018) desenvolveram uma abordagem utilizando modelo de CNN baseado em convoluções de tipos de dados separados. Os dados foram dispostos em formato de matriz 2D, cada linha representando uma série temporal de um elemento de entrada. Para as camadas convolucionais, foi utilizado um processo de escaneamento 1D; cada filtro passeia em cada tipo de dado, ao final da matriz, o filtro retorna ao início e segue para a próxima linha. A ideia era extrair padrões temporais que pudessem representar a direção do preço, sendo invariante à ordem de entrada dos elementos. A Figura 2 apresenta a CNN utilizada na proposta. Como se pode ver, o histórico de preços são dispostos nas linhas da matriz de entrada e as convoluções são realizadas em cada elemento de forma independente. As amostras são rotuladas com base no

preço do dia seguinte considerando a volatilidade recente nos preços. Como saída da rede tem-se as seguintes possibilidades: subir, cair ou neutro, caso o mercado não ultrapasse os valores de algumas das outras classes. A proposta foi avaliada no período entre 2010 e 2017, e os resultados mostraram que este sistema foi capaz de extrair informações mais generalizadas que demais modelos até então e bons resultados financeiros. Com base na informação da tendência, foi realizada uma avaliação financeira que demonstrou resultados superiores à estratégia BH. Este trabalho destaca a capacidade de redes do tipo CNN em aprender e identificar padrões em séries temporais que possam identificar situações de compra e venda e auxiliar em operações no mercado financeiro.

Figura 2 – CNN utilizada por Wang et al. (2018)



Fonte: Wang et al. (2018)

Hoseinzade e Haratizadeh (2019) sugeriram duas abordagens utilizando CNN para prever o preço de um dia futuro para o mercado de ações com base em indicadores técnicos e variáveis macroeconômicas. Estes modelos eram respectivamente CNNs 2D e 3D. A proposta era conseguir combinar dados de diversos mercados a fim de encontrar correlações entre eles para uma melhor previsão. Os modelos foram testados principalmente nos mercados S&P, Nasdaq e Dow Jones. Os resultados indicam que as CNNs são capazes de indicar a tendência futura do mercado como uma abordagem de classificação e que a combinação com dados de outros mercados pode auxiliar na previsão. Este trabalho demonstra que um modelo CNN pode ser usado para classificar séries temporais no mercado financeiro.

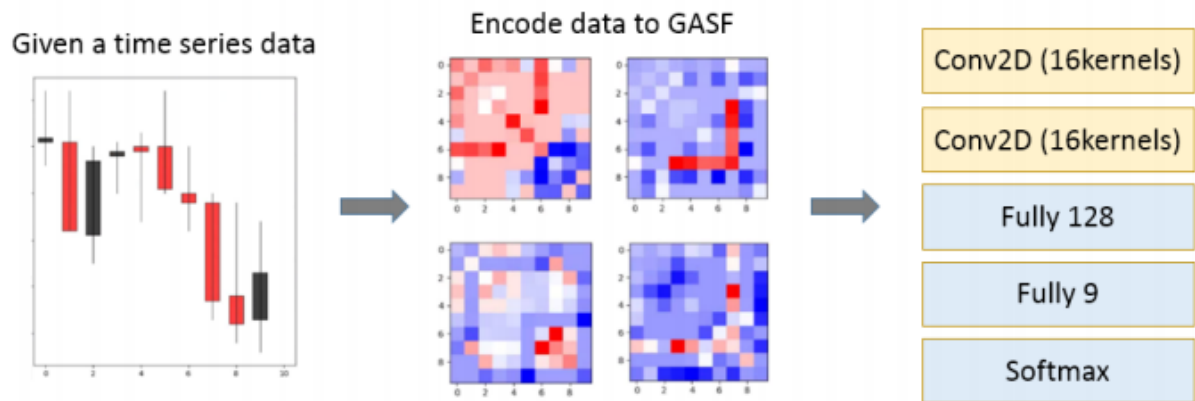
Pimenta et al. (2018) também realizaram pesquisa aplicada ao mercado de ações brasileiro. Fora empregados 11 indicadores técnicos subdivididos em osciladores, seguidores de tendência, sistemas de bandas e identificadores de divergências, treinando um modelo de Algoritmo Genético (AG) para aprender a direção das ações da Bolsa de Valores do Estado de São Paulo (BOVESPA). Além disso foi utilizada técnica de seleção de recursos para encontrar

correlações entre indicadores e o movimento dos preços. Dentre os indicadores selecionados estão RSI, W%R. Os resultados apresentaram boa acurácia da predição da direção da tendência, e superando expressivamente a estratégia BH em termos de retorno sobre investimento. Este trabalho evidencia que o uso de indicadores para realização previsão no mercado financeiro pode proporcionar resultados superiores às estratégias mais comuns de investimento.

Além disso, outros pesquisadores implementaram métodos de AP com base no reconhecimento de imagens. Velay e Daniel (2018) usaram CNN e LSTM para reconhecer padrões de gráficos comuns na análise técnica (BULKOWSKI, 2011). O objetivo era automatizar a detecção de tais padrões na tentativa de ajudar na tomada de decisão e fornecer uma previsão da movimentação dos preços no mercado de ações. O estudo utilizou os seguintes padrões: bandeiras de baixa e alta e topo e fundo duplos. Segundo os autores, o modelo LSTM alcançou a melhor taxa de revocação, chegando a 97% contra 73% da CNN.

Tsai, Chen e Wang (2019) propuseram o uso de CNN combinando uma técnica conhecida como *Gramian Angular Field (GAF)*, uma forma de representar série temporal em formato de imagens padronizadas, para classificação de padrões de gráficos. A arquitetura do sistema proposto é apresentada na Figura 3. O gráfico de *Candlestick* representa a série temporal dos preços. Cada série é submetida ao processamento do GAF gerando as imagens padronizadas para os preços de abertura, máxima, mínima e fechamento. As imagens então são utilizadas como entrada na rede para classificação. Esta é composta por 2 camadas convolucionais que extraem padrões das imagens, e a classificação é feita por camadas totalmente conectadas. Esses padrões de *Candlestick* são comumente usados para encontrar pontos de virada na tendência dos preços das ações e podem indicar uma operação de compra/venda (BULKOWSKI, 2011). Foram utilizados 8 padrões neste trabalho, sendo estes: estrela da manhã/noite, engolfo baixa/alta, martelo e martelo invertido, estrela cadente e enforcado. Os resultados mostraram uma precisão de aproximadamente 88% para esta classificação, o que pode ser aplicável em operações no mercado financeiro a fim de encontrar pontos de virada na tendência de preços e obter retornos financeiros superiores, antecipando momentos de queda e recomprando quando indicar reversão de tendência.

Figura 3 – Abordagem proposta por Tsai, Chen e Wang (2019) utilizando CNN



Fonte: Tsai, Chen e Wang (2019)

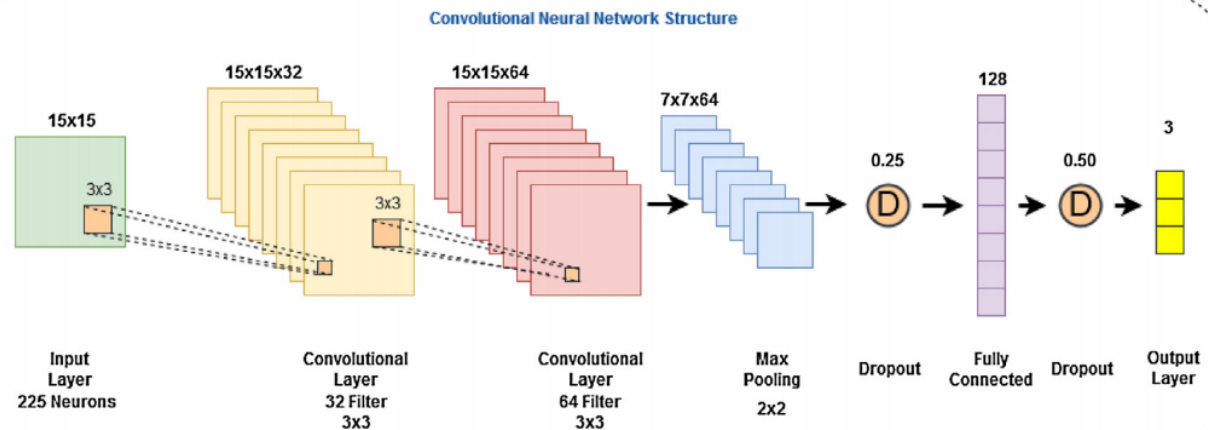
Em outro trabalho, Hiransha et al. (2018) fizeram a comparação entre 4 modelos de RNA: MLP, RNN, CNN e LSTM para prever o índice da National Stock Exchange (NSE), uma bolsa de valores da Índia, cujo índice principal é o New York Stock Exchange (NYSE). Em sua pesquisa o modelo CNN apresentou melhores resultados em relação aos outros algoritmos pois foi capaz de reagir mais rapidamente às mudanças abruptas do índice, evidenciando sua melhor capacidade na identificação de padrões sobre os preços.

Sezer e Ozbayoglu (2018) desenvolveram um sistema para negociação baseado em séries temporais usando uma abordagem de reconhecimento de imagem com CNN para classificar os melhores pontos de compra e venda. O sistema foi chamado *Convolutional Neural Network - Technical Analysis (CNN-TA)* e foi baseado no uso de 15 indicadores técnicos em 15 períodos subsequentes para gerar imagens em formato 2D de dimensões $[15 \times 15]$ que representam amostras diárias. A Figura 4 mostra a CNN utilizada na proposta. Esta é composta por 2 camadas convolucionais subsequentes e 1 de *max-pooling*, a fim de extrair padrões na matriz de entrada. São utilizados filtros bidimensionais de $[3 \times 3]$ nas camadas convolucionais e uma janela de *pooling* de $[2 \times 2]$. Ao final a classificação é realizada por meio de camadas totalmente conectadas. Todas as camadas utilizam a função de ativação ReLU, e a camada de saída utiliza a função *Softmax*. No treinamento foram utilizados 2 *Dropouts* a fim de melhorar a generalização da rede.

Cada imagem foi rotulada como compra, venda ou neutro, de acordo com um algoritmo de rotulagem baseado em janela deslizante. Os resultados foram comparados com a estratégia BH e outras estratégias de negociação por um longo período e indicam que o sistema CNN-TA

conseguiu superar todas as estratégias de negociação comumente comparáveis. Desta forma, este trabalho evidencia que um modelo CNN pode ser utilizado para classificar momentos de operação no mercado financeiro.

Figura 4 – CNN utilizada por Sezer e Ozbayoglu (2018)



Fonte: adaptado de (SEZER; OZBAYOGLU, 2018)

Zhang et al. (2019) utilizaram uma rede LSTM para previsões do preço de fechamento de ações. A abordagem foi construída de forma a minerar as distribuições de dados de uma determinada ação e gerar dados nas mesmas distribuições, onde uma rede Rede Adversarial Generativa (*Generative Adversarial Network - GAN*) junto de uma MLP discriminadora é empregada para diferenciar os dados reais dos dados gerados. O modelo LSTM foi responsável pela previsão do preço futuro. Os dados escolhidos para o estudo foram os valores diários do índice S&P 500 e diversas ações em uma larga faixa de dias de operação e tentou prever o preço diário de fechamento. O sistema proposto superou os modelos de RNA comparados, que foram uma LSTM pura, uma MLP e uma SVM, mostrando que LSTM são aplicáveis para regressão sobre séries temporais de preços de ações.

Alonso-Monsalve et al. (2020) desenvolveram um algoritmo híbrido chamado CLSTM que combina uma CNN para pré processar os dados de ações, sendo estes: preço e diversos indicadores técnicos. Em seguida, utilizou os resultados em uma rede LSTM para prever o preço do próximo instante no tempo. Os autores compararam os resultados com uma rede LSTM pura, uma Rede Neural de Função Base Radial (*Radial Basis Function Neural Network - RBFNN*) e uma CNN pura. Como conclusão, afirmam que o algoritmo proposto demonstrou melhores resultados que utilizando os modelos comparados. Este trabalho demonstra que a combinação CNN-LSTM para regressão pode ser aplicável para prever o preço futuro de

ações, onde a CNN extrai padrões de comportamento sobre os preços e a LSTM utiliza tais padrões para prever o preço em um instante seguinte.

Chung e Shin (2020) utilizaram CNN para previsão da flutuação dos índice de preços de ações da Coreia do Sul. Os parâmetros da rede foram otimizados utilizando AG. Os dados de entrada foram os dados históricos dos preços e volume financeiro negociado. O sistema proposto foi comparado com uma MLP. Segundo os autores, os resultados da proposta AG-CNN proposto superou os modelos de RNA comparados e demonstra a eficácia da abordagem híbrida de AG e CNN. Este trabalho reforça a capacidade da rede CNN em identificar padrões de mudanças em séries temporais.

Mehtab e Sen (2020) utilizaram rede CNN para realizar previsão do índice do mercado financeiro da Índia (NIFTY) por um período de 4 anos, de 2015 a 2019. Foram feitos vários testes com diferentes parâmetros com o objetivo de realizar ajustes finos dos parâmetros da rede. Segundo os autores, os resultados indicam claramente que o sistema de previsão baseado em CNN é o mais eficaz e preciso para prever o movimento dos valores do índice NIFTY com um horizonte de previsão semanal,

A Figura 5 apresenta um sumário dos trabalhos descritos neste capítulo. Os modelos mais comumente utilizados na literatura têm sido LSTM e CNN. As CNNs têm sido mais aplicadas para problemas de classificação e reconhecimento de padrões, enquanto as LSTMs têm sido a escolha preferida para regressão de séries temporais. Nota-se também que a maioria dos estudos têm sido realizados com ações, enquanto alguns expandiram as pesquisas para ETFs e índices de mercado.

Um algoritmo de negociação preciso depende de muitas variáveis e um modelo CNN tem se mostrado capaz de representar um conjunto de indicadores de análise técnica e outras variáveis de entrada em forma de séries temporais a fim de classificar padrões para possíveis sinais de negociação ou identificar padrões que possam prever preços futuros.

Além disso, os trabalhos mostram que a escolha dos indicadores técnicos financeiros pode ter grande influência no resultado do estudo e na capacidade dos modelos de RNA de realizarem previsão ou classificação. Analisando os trabalhos apresentados, pode-se concluir que aplicar modelos de AP utilizando indicadores técnicos financeiros pode servir como uma fonte de informação para tomada de decisão. As propostas mostram que RNAs são capazes de aprender tendências nos preços e extrair detalhes sobre o comportamento de cada ativo.

Figura 5 – Sumário da revisão da literatura

Ano	Autor	Título	Técnica	Ativo
2017	BAO; YUE; RAO;	A deep learning framework for financial time series using stacked autoencoders and long-short term memory	LSTM	Ações e Índices
	SELVIN et al.	Stock price prediction using LSTM, RNN and CNN-sliding window model	LSTM, CNN	Ações
	SEZER; OZBAYOGLU; DOGDU	An artificial neural network-based stock trading system using technical analysis and big data framework	MLP	Ações
	TSANTEKIDIS et al.	Forecasting stock prices from the limit order book using convolutional neural networks	CNN, SVM, MLP	Ações
	NELSON; PEREIRA; OLIVEIRA	Stock market's price movement prediction with LSTM neural networks	LSTM	Ações
	SAMARAWICKRAMA FERNANDO	A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market	MLP, GRU, LSTM, SRNN	Ações
2018	FISCHER; KRAUSS	Deep learning with long short-term memory networks for financial market predictions	LSTM	Ações
	JAIN; GUPTA; MOGHE	Stock Price Prediction on Daily Stock Data using Deep Neural Networks	CNN, LSTM, CNN-LSTM	Ações
	WANG et al.	Financial markets prediction with deep learning	CNN	Ações e Índices
	PIMENTA et al.	An automated investing method for stock market based on multiobjective genetic programming	AG	Ações
	VELAY; DANIEL	Stock chart pattern recognition with deep learning	CNN, LSTM	Ações
	HIRANSHA et al.	NSE stock market prediction using deep-learning models	MLP, RNN, CNN, LSTM	Índices
	SEZER; OZBAYOGLU	Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach	CNN	Ações e ETFs
2019	NOBRE; NEVES	Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets	XGBoost, PCA	Índices
	CAO; WANG	Stock price forecasting model based on modified convolution neural network and financial time series analysis	CNN, CNN-SVM	Ações
	SAYAVONG; WU; CHALITA	Research on Stock Price Prediction Method Based on Convolutional Neural Network	CNN	Ações
	HOSEINZADE; HARATIZADEH	CNNpred: CNN-based stock market prediction using a diverse set of variables	CNN	Ações
	TSAI; CHEN; WANG	Encoding candlesticks as images for patterns classification using convolutional neural networks	CNN	Ações
	ZHANG et al.	Stock market prediction based on generative adversarial network	GAN, MLP, LSTM	Ações e Índices
2020	ALONSOMONSALVE et al.	Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators	CNN	Criptomoedas
	CHUNG; SHIN	Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction	CNN, AG	Ações

Fonte: o autor

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresentará os conceitos básicos de estratégia de investimento, com maior foco na chamada Análise Técnica. Na sequência será apresentado um conceito de análise de série temporal baseado em estatística. Por fim as teorias sobre RNAs e modelos de AP serão abordados, os quais constituem a base deste trabalho.

3.1 Estratégias de Investimento

Tradicionalmente, dois métodos são utilizados para tomar decisão no mercado de ações: Análise Fundamentalista e Análise Técnica. O primeiro está interessado nos dados relacionados à empresa da respectiva ação, exigindo experiência na interpretação de dados financeiros, como balanços contábeis e fatores econômicos que interferem na empresa, bem como acompanhamento de notícias e fatos relevantes. A estratégia BH baseia-se em Análise Fundamentalista, selecionando empresas por meio da análise destes dados, deduzindo que as boas empresas continuarão a apresentar bons resultados e consequentemente o preço de suas ações aumentará ao longo do tempo. Por depender da interpretação de dados, uma abordagem de AP utilizando PLN é mais aplicável para este tipo de estratégia.

Por outro lado, a Análise Técnica está interessada em lucrar na variação dos preços das ações, geralmente em um curto espaço de tempo. Pode-se definir a Análise Técnica como o estudo da ação do mercado, primariamente por meio de uso de gráficos, com o objetivo de prever comportamento ou tendências futuras (LEMOS, 2017).

3.1.1 Análise Fundamentalista

Este tipo de estratégia busca investir em empresas do ponto de vista de seus resultados operacionais, tais como lucro financeiro e perspectiva de crescimento futuro. Segundo Damodaran e Cunha Serra (2000), há dezenas de modelos de avaliação de empresas, mas apenas duas abordagens: intrínseca e relativa. O valor intrínseco de um ativo é determinado pelos fluxos de caixa esperados durante a vida da empresa, ajustado pelo grau de incerteza por eles gerados. Um dos métodos mais utilizados para analisar empresas é o Fluxo de Caixa Descontado. Este é um método para avaliar a riqueza econômica de uma empresa dimensionada pelos retornos futuros esperados e descontados por uma taxa de atratividade que reflete o custo de oportunidade.

Ou seja, o valor de um ativo representa o seu montante justo hoje, enquanto preço significa por quanto o mercado o está negociando na data atual. Sendo assim, se o preço praticado é inferior ao valor, tem-se um investimento em potencial (DAMODARAN; CUNHA SERRA, 2000).

Já uma outra abordagem é conhecida como a Hipótese do Mercado Eficiente. Esta é uma teoria de investimento que fundamenta que os preços praticados no mercado sempre refletem todas as informações já conhecidas das empresas, bem como fatores que influenciam nas mesmas. Logo, segundo esta hipótese, não existem ações baratas ou caras, uma vez que o mercado eficiente sempre precifica a cotação de uma ação no seu valor justo (MALKIEL; FAMA, 1970).

Dentro da hipótese há três níveis distintos de eficiência:

- a) **Eficiência fraca:** defende que o mercado é eficiente em refletir todas as informações públicas disponíveis. Os retornos no mercado são independentes. Desta forma, retornos passados não ajudam a prever retornos futuros;
- b) **Eficiência semi-forte:** além de sustentar a hipótese fraca, sugere que as novas informações são absorvidas pelo mercado instantaneamente. Logo, os investidores não conseguem resultados acima do mercado com informações conhecidas;
- c) **Eficiência forte:** além de englobar as duas hipóteses descritas anteriormente, sustenta que os preços dos ativos refletem instantaneamente todo o tipo de informação. Seja ela pública ou privada.

Resumidamente, esta teoria afirma que é impossível obter retornos acima do mercado, uma vez que os preços já refletem todas as informações relevantes para os resultados futuros. Isto significa que a ação vai ser sempre levada ao seu valor justo, fazendo com que seja impossível para os investidores comprarem ações desvalorizadas ou vender ações sobrevalorizadas para alcançar lucros significativos. Isto também é defendido pela teoria do Passeio Aleatório (*Random Walk*) (FAMA, 1995), que afirma que as variações nos preços das ações independem uma da outra. Esta teoria defende que não se pode olhar para os movimentos passados de uma ação, padrão ou tendência para prever os movimentos futuros do mercado. O mercado funciona de forma irracional, sendo os movimentos dos preços imprevisíveis, seguindo um "passeio aleatório". Com isso, uma análise que se baseia nos movimentos passados no mercado financeiro não é eficiente.

3.1.2 Análise Técnica

A Análise Técnica acredita que é possível estimar o comportamento futuro do mercado analisando os históricos de preços anteriores. O objetivo é identificar oportunidades de negociação analisando tendências estatísticas e padrões gráficos com base no histórico de preços passados, uma vez que os movimentos dos preços são formados pela atuação de grandes investidores montando e desfazendo posições.

A Análise Técnica tornou-se uma ferramenta incrivelmente popular entre investidores para medir a força do mercado e prever a direção de curto/médio prazo para os mercados e as ações individualmente (COLBY; MEYERS, 1988). Isto se deve ao fato de ser uma abordagem mais simplista para as pessoas que não possuem um vasto conhecimento em análise de empresas e querem focar apenas no lucro financeiro, principalmente de curto prazo.

Os princípios da Análise Técnica podem ser descritos da seguinte forma (EDWARDS; BASSETTI; MAGEE, 2012; SHAN, 2012):

- a) As médias descontam tudo: pois estas refletem as atividades de investidores, incluindo aqueles possuidores de grandes previsões e melhores informações nas tendências dos eventos;
- b) Os preços se movem em tendências: o preço das ações se move em tendências, nas quais a mais importante é a tendência primária, de longo prazo. A segunda tendência é formada pelas correções que ocorrem em sentido contrário ao movimento primário. E a terceira é composta por tendências menores de curto prazo, dia a dia;
- c) A história se repete: os estudos apresentados por Dow consistem na crença de que os padrões identificados nas tendências e nas médias se repetem ao longo do tempo (SHAN, 2012).

Para Elder (2018), o preço é o líder do mercado: ele é o consenso máximo entre todos os participantes sobre o valor de determinado ativo. Na Análise Técnica, esses preços variam em tendências de alta ou de baixa e essa tendência só muda de fato com uma mudança de volume das negociações. Tais tendências podem ser observadas por meio de gráficos de *Candlestick*, onde cada Vela (*Candle*) representa a variação do preço em um determinado intervalo de tempo. Desta forma, padrões de gráficos de *Candlestick* (EDWARDS; BASSETTI; MAGEE, 2012)

são utilizados com o objetivo encontrar regiões de interesse no preço e momentos para realizar compras e vendas baseando-se no histórico de preços passados.

3.1.3 Indicadores Técnicos

Ao longo dos anos, com base em análises estatísticas, foram sendo desenvolvidas ferramentas para auxiliar na análise das tendências do mercado, os chamados indicadores técnicos. Vários indicadores técnicos foram desenvolvidos com o objetivo de indicar a operação mais rentável para o momento usando dados, como volume financeiro negociado e preços passados. Os indicadores são fórmulas matemáticas utilizadas para alertar que um comportamento tem probabilidade de ocorrer, confirmar as informações fornecidas por outros indicadores ou mesmo prever um comportamento futuro de uma série histórica do ativo. De acordo com Colby e Meyers (1988), pode-se classificar os indicadores nas seguintes categorias:

- a) **Rastreadores de Tendências:** a tendência é um valor central ou valor típico para uma distribuição de probabilidade (SILVA, 2009). A tendência dos preços no mercado de ações é verificada por meio do histórico de preços em um certo período. Por exemplo, pode-se verificar uma tendência de alta pela média aritmética do histórico de preços do ativo quando ela está ascendente.

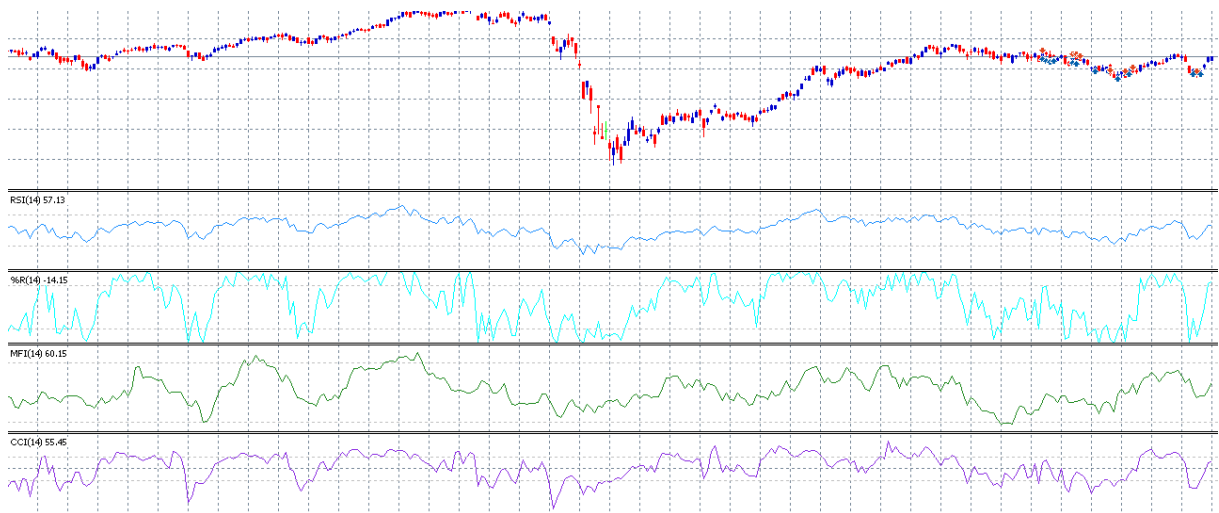
Os rastreadores de tendência são confiáveis durante períodos de tendência, porém não identificam nem o começo nem o fim das tendências perdendo, em média, um terço do comprimento total das tendências (SACHETIM, 2006);
- b) **Osciladores:** um oscilador é um indicador que flutua acima e abaixo de um centro, ou entre níveis preestabelecidos, sendo mais comum entre 0-100% (SACHETIM, 2006). Podem permanecer em níveis extremos, os pontos chamados de sobrecompra e sobrevenda. Possuem a característica de não sustentarem a mesma tendência por um longo período, por isso são denominados osciladores;
- c) **Indicadores de Volatilidade:** volatilidade é uma medida de dispersão dos preços de uma ação. Pode ser medida pela diferença entre o maior e o menor preço em um período, ou ainda pela variância. A variância expressa o quão longe, em geral, os seus valores se encontram do valor esperado. Serve para medir a intensidade e a frequência das oscilações nas cotações. A volatilidade pode ser utilizada como uma medida do risco;

- d) **Indicadores de Volume:** o volume de negociações ou de capital investido evidencia a força do mercado. É comumente utilizado para medir a força compradora ou a força vendedora. Quando existem variações de preço com volume crescente é sinal de consistência na variação do preço naquele sentido. Isso faz com que a reversão da tendência do preço seja mais difícil de ocorrer.

Majoritariamente, os indicadores consistem em uma série de dados que são derivados pela aplicação de uma fórmula. Estes números são interpretados visualmente como uma indicação de compra, venda ou neutro. Geralmente, é utilizada uma combinação de padrões gráficos e indicadores técnicos para tomar decisões de operação mais refinadas. Algumas abordagens que motivaram este trabalho utilizaram um conjunto de indicadores técnicos e um modelo de AP para reconhecer padrões em séries temporais a fim de classificar a melhor operação de negociação naquele momento.

Na figura 6 está representado o conjunto de indicadores selecionados para este trabalho juntamente com um gráfico de *Candelstick* (BULKOWSKI, 2011), de forma semelhante ao que é comumente utilizado pelas pessoas que operam no mercado financeiro. Este tipo de gráfico representa a variação de preços de um determinado ativo em uma unidade de tempo, no caso, diária. As fórmulas e interpretações dos respectivos indicadores técnicos são apresentadas na sequência.

Figura 6 – Conjunto de indicadores utilizados no trabalho



Fonte: própria autoria, figura obtida pelo software MetaTrader

3.1.3.1 Índice de Força Relativa (Relative Strength Index - RSI)

É um indicador de momento que mede a magnitude das recentes mudanças de preço para classificar como situação de sobrecompra ou sobrevenda. A interpretação tradicional deste indicador consiste em: se o valor do RSI estiver acima de 70 está sobrecomprado e abaixo de 30 está sobrevendido (WILDER, 1978). O cálculo é feito em 2 passos conforme descrito a seguir:

$$RSI_1 = 100 - \frac{100}{1 + \frac{GanhoMédio}{PerdaMédia}} \quad (1)$$

As variáveis *GanhoMédio* e *PerdaMédia* representam a média de dias que fecharam positivos e negativos, respectivamente, no período informado, sendo este geralmente de 14. Isto assegura que se na última janela de tempo analisada os preços apenas subiram o indicador chega próximo a 100, e se apenas caíram se aproxima de 0.

Uma vez que existam a quantidade dias de dados suficientes no histórico, o passo dois pode ser calculado conforme a Eq. (2):

$$RSI_2 = 100 - \frac{100}{1 + \frac{GanhoMédioAnterior - GanhoAtual}{PerdaMédiaAnterior - PerdaAtual}} \quad (2)$$

GanhoMédioAnterior e *PerdaMédiaAnterior* representam as mesmas médias mencionadas anteriormente considerando a partir do instante anterior ao atual.

3.1.3.2 Williams %R (WR)

Ou apenas W%R, este é um oscilador de inércia o qual reflete a alteração do preço de fechamento em relação ao valor mais alto para o período. O W%R oscila entre 0 e -100. A leitura de 0 a -20 é considerada como condição de sobrecompra, e de -80 a -100 considerada como sobrevendida. Geralmente é utilizado o período de 14 (WILLIAMS, 1998). A Eq (3) fornece a fórmula de cálculo do W%R.

$$W\%R = \frac{MáximoPeríodo - FechamentoAtual}{MáximoPeríodo - MinimaPeríodo} * -100 \quad (3)$$

As variáveis *MáximoPeríodo*, *MinimaPeríodo* e *FechamentoAtual* representam preço máximo, mínimo e preço de fechamento atual, respectivamente.

3.1.3.3 Índice de Fluxo Monetário (*Money Flow Index - MFI*)

Este indicador combina preço e volume para medir a pressão de compra e venda. Os valores são positivos quando o preço aumenta, ou seja, quando há pressão compradora e negativo quando os preços caem, o que indica pressão vendedora. Isto ocorre quando o Preço Típico (*Typical Price*) (TP) cresce ou cai. É geralmente aplicado para identificar reversões e extremos sendo utilizado para identificar sinais de sobrecompra e sobrevenda (QUONG; SOUDACK, 1989). O processo de cálculo é feito em 4 etapas, conforme descrito a seguir, onde i representa o período de cálculo, geralmente de 14:

Preço Típico - *Typical Price (TP)*:

$$TP = \frac{Máximo + Mínimo + Fechamento}{3} \quad (4)$$

Fluxo Monetário Bruto - *Raw Money Flow (RMF)*:

$$RMF = TP * Volume \quad (5)$$

Taxa de fluxo monetário - *Money Flow Ratio (MFR)*:

$$MFR = \frac{\sum_{i-t}^i RMF_{Positivo}}{\sum_{i-t}^i RMF_{Negativo}} \quad (6)$$

$$MFI = 100 - \frac{100}{1 + MFR} \quad (7)$$

3.1.3.4 Índice de Canal de Mercadoria (*Commodity Channel Index - CCI*)

Um indicador de momento usado para identificar o início de uma nova tendência ou avisar sobre condições extremas (sobrecompra e sobrevenda). Para isto, este indicador mede a correlação entre o preço atual e a média dos preços do período (LAMBERT, 1983). Sua fórmula é apresentada na Eq. (8):

$$CCI = \frac{TP * SMA_{20}(TP)}{0.015 * DesvioMédio} \quad (8)$$

A variável TP é calculada com a mesma fórmula utilizada no MFI. *SMA*, do inglês *Simple Moving Average*, consiste em uma média móvel de preços típicos do período.

3.2 Série Estacionária

Os preços das ações são altamente voláteis, no entanto, na maioria das vezes os preços seguem uma tendência de alta ao longo do tempo. Em outras palavras, o valor médio dos preços aumenta com o tempo, e é por isso que a estratégia BH é a estratégia de investimento preferida por muitos investidores.

Uma série temporal é dita estacionária quando ela se desenvolve ao longo do tempo ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Isso significa que propriedades estatísticas como média ou desvio padrão, não são significativas para os preços no longo prazo (SILVEIRA BUENO, 2008).

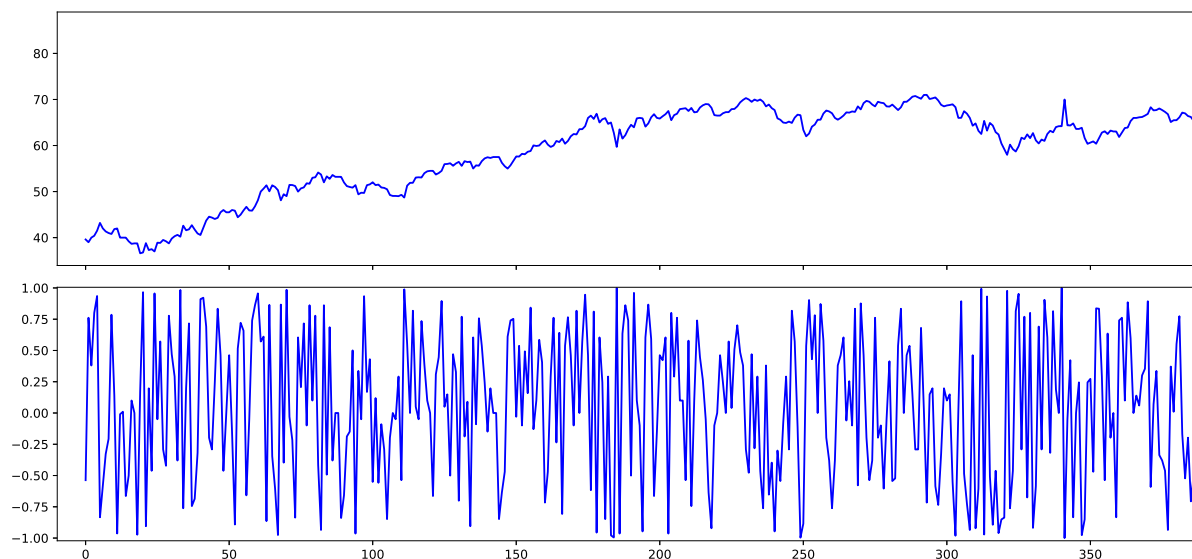
Para capturar padrões de séries temporais invariantes ao efeito temporal, como nas ações, pode-se utilizar o preço de forma estacionária. Dessa forma, torna-se mais fácil de prever o seu comportamento: basta deduzir que suas propriedades estatísticas serão as mesmas no futuro, como no passado.

Segundo Lee, Lee e Lee (2010), em um estudo aplicado a ações buscando encontrar evidências que os preços apresentam comportamento estacionário, este concluiu que a série de preços de ações apresenta estacionariedade em mercados financeiros de 32 países desenvolvidos e 26 em desenvolvimento. Assim, os índices reais de preços de ações são processos estacionários que são inconsistentes com a Hipótese de Mercado Eficiente. Isso mostra a presença de oportunidades lucrativas e de arbitragem entre os mercados de ações.

O Preço de Fechamento Estacionário (PFE) pode ser calculado utilizando o método da primeira diferença. Este método de diferenciação consiste em calcular a diferença do preço no período T com o valor do período anterior $T - 1$. Ao se fazer a primeira diferenciação, retira-se o efeito acumulado da série e mostra-se apenas a variação do período T contra o período $T - 1$ ao longo de toda a série.

Em seguida, para utilizar estes valores como entrada, pode-se utilizar uma função de normalização, como a Tangente Hiperbólica (TanH). Como resultado, é obtida a média do preço de fechamento em torno do ponto zero, como mostra a Figura 7.

Figura 7 – Preço de fechamento no gráfico superior e preço de fechamento estacionário no inferior



Fonte: o autor

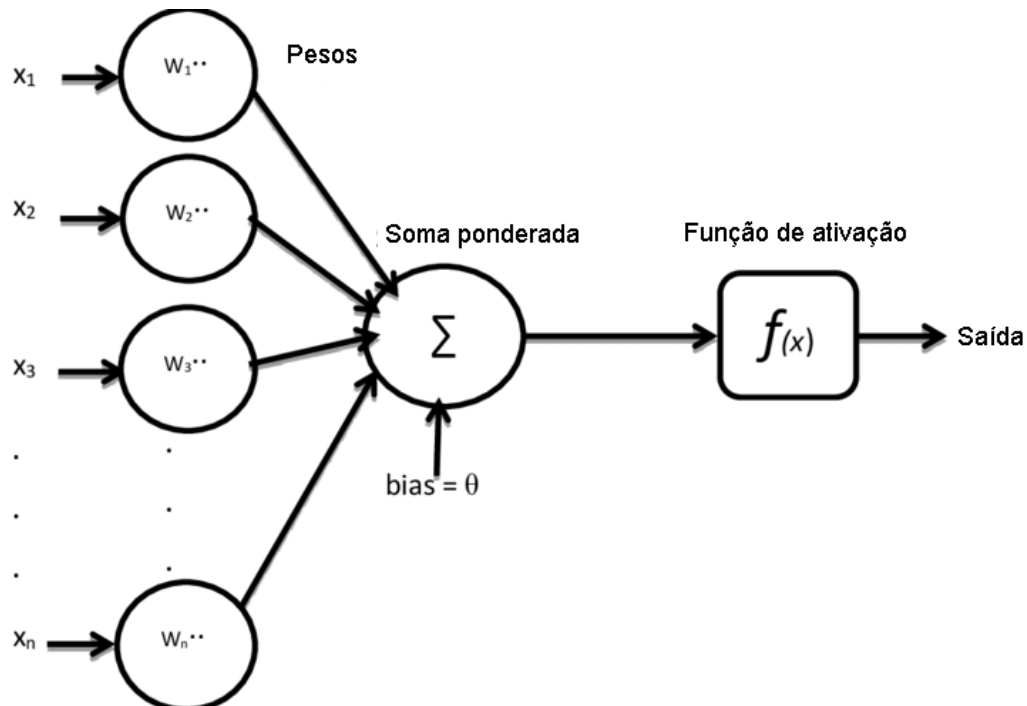
3.3 Redes Neurais Artificiais (RNA)

As pesquisas em Redes Neurais Artificiais (RNA)s iniciaram-se com a proposta de McCulloch e Pitts (1943) em um estudo sobre o comportamento do neurônio biológico a fim de criar um modelo matemático que pudesse representá-lo. Este modelo inicialmente proposto era capaz de separar entradas de forma binária.

O neurônio artificial é um modelo matemático que busca simular comportamento e funções do neurônio biológico. Dessa forma, os dendritos são substituídos pelas entradas, cujas ligações com o corpo celular são realizadas através dos pesos, simulando as sinapses neurais. Os estímulos são captados pelas entradas e processados no corpo celular pela função de soma ponderada das entradas e o *bias*.

O limiar de disparo do neurônio biológico é substituído pela função de ativação, e o resultado deste processo é repassado para a sequência da rede, substituindo o axônio biológico (RUSSELL; NORVIG, 2002). A Figura 8 exibe a representação do neurônio artificial de McCulloch e Pitts (1943).

Figura 8 – Representação do neurônio artificial



Fonte: adaptado de (YACIM; BOSHOF, 2018)

Pode-se matematicamente definir a saída do neurônio de McCulloch e Pitts (1943) pela Eq. (9).

$$y_i = f\left(\sum_{j=1}^n W_{ij}x_j + b_i\right) \quad (9)$$

onde:

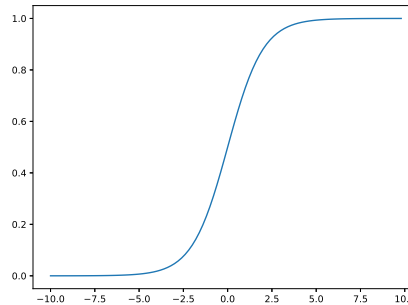
- a) x_j representa os sinais de entrada;
- b) W_{ij} representa os pesos associados às entradas. Estas entradas são multiplicadas por seus respectivos pesos, podendo gerar valores positivos (excitatórios) ou negativos (inibitórios);
- c) b representa o *bias* do neurônio;
- d) Σ representa o somatório dos sinais calculados pelos produtos das entradas por seus pesos;
- e) f representa a função de ativação;
- f) y_i representam a saída calculada do neurônio artificial i .

Dentre as principais funções de ativação utilizadas estão:

- a) **Função Sigmóide:** esta função, representada pela Eq. (10), com o comportamento gráfico da Figura 9, pode assumir valores entre 0 e 1;

$$f(x) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \quad (10)$$

Figura 9 – Gráfico da função Sigmóide

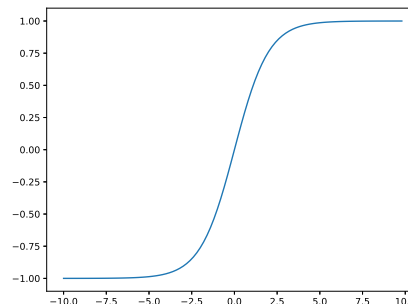


Fonte: o autor

- b) **Função Tangente Hiperbólica (TanH):** esta função, representada pela Eq. (11), com o comportamento gráfico da Figura 10, pode assumir valores entre -1 e 1;

$$f(x) = \frac{1 - \exp -x}{1 + \exp -x} \quad (11)$$

Figura 10 – Gráfico da função TanH



Fonte: o autor

Uma RNA é uma arquitetura composta por uma grande quantidade de neurônios artificiais (NIGRIN, 1993), formando um sistema paralelo e distribuído (HAYKIN, S.S., 1999). O neurônio artificial é a unidade de processamento básica de uma RNA. Em uma RNA, os neurônios são dispostos em uma ou mais camadas e interligados por um certo número de conexões (pesos), geralmente unidirecionais. Estas conexões armazenam o conhecimento representado no modelo e ponderam a entrada recebida por cada neurônio da rede através de uma função matemática (BRAGA, 2000).

A topologia de uma RNA é definida pelo número de camadas, de neurônios, grau de conectividade entre os neurônios e funções de ativação. Além disso, as redes podem possuir conexões de retroalimentação, nas quais o neurônio pode receber a saída de outros da mesma camada (ou posteriores) ou de si próprio. O paralelismo é uma das principais características de uma RNA, uma vez que este cria a possibilidade de desempenho superior em relação aos modelos algorítmicos da área de inteligência artificial clássica (HAYKIN, S.S., 1999).

As principais aplicações de RNAs podem ser divididas da seguinte forma:

- a) Classificação, devido à capacidade de abstração e generalização de conhecimento;
- b) Aproximação de funções ou análise de regressão, incluindo a previsão de séries temporais;
- c) Processamento de dados, incluindo filtragem e agrupamento.

A classificação é uma tarefa que consiste em dividir os dados em classes predefinidas. Pode ser um problema de classificação binário, onde existem apenas duas classes, ou multi-classes, onde existem três ou mais classes. Já a tarefa de regressão consiste em estimar o valor futuro de uma série baseando-se em dados e comportamentos históricos. Agrupamento de dados consiste na tarefa de computar funções a fim de separar dados que possuam características semelhantes.

3.3.1 Rede Neural *Feedforward*

As redes *feedforward* são a base para muitos modelos de RNAs que têm sido propostos aos longo dos anos. Para Kriesel (2007), a saída de uma rede *feedforward*, para qualquer entrada, é calculada passando o sinal, a partir da camada de entrada até a camada de saída, somente uma vez, ou seja, cada neurônio em uma camada possui conexão direta, somente, com os neurônios da próxima camada.

A publicação inicial de McCulloch e Pitts (1943) consiste na base deste tipo de RNA, e Hebb (1949) propôs um modelo básico de auto-organização para tal rede. Na sequência Rosenblatt (1958) desenvolveu o modelo *Perceptron*, o primeiro capaz de realizar o aprendizado supervisionado.

Para que uma RNA seja capaz de realizar inferências sobre os dados, é necessário que a mesma seja treinada. Segundo S. Haykin (2007) o aprendizado de uma RNA consiste em um processo pelo qual os parâmetros de uma rede neural são adaptados através de um processo

contínuo de estimulação pelo ambiente no qual a rede está inserida. Este processo pode ser classificado como:

- a) **Aprendizado Supervisionado:** um conjunto de dados é apresentado como entrada para a rede e, associado a este conjunto de entrada, é apresentado um conjunto de saída esperada. O algoritmo de aprendizado busca ajustar os pesos das conexões entre os neurônios de tal forma que, para dada uma entrada, a rede seja capaz de calcular uma saída o mais próximo possível da saída desejada;
- b) **Aprendizado Não Supervisionado:** neste tipo de treinamento é utilizado apenas as informações das entradas e realiza-se o aprendizado através de agrupamentos de dados e conceitos de vizinhança. Muito utilizado para encontrar padrões entre os dados cujas estruturas não sejam bem conhecidas.

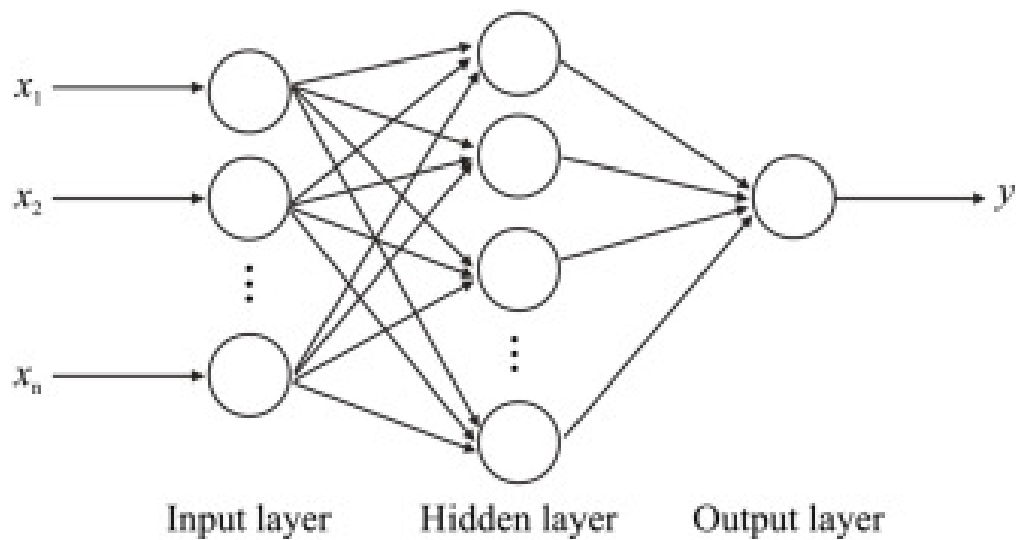
3.3.2 *Multilayer Perceptron (MLP)*

A rede *Multilayer Perceptron (MLP)* representa uma generalização do *Perceptron* proposto por Rosenblatt (1958). Sua arquitetura consiste em várias camadas compostas por neurônios artificiais onde cada neurônio de uma camada possui ligação com todos os neurônios da camada seguinte.

A camada de entrada recebe os dados, realiza as sinapses e repassa o sinal para a camada seguinte, e assim sucessivamente até a camada de saída. Neste modelo podem haver uma ou mais camadas intermediárias, que é onde a maior parte do processamento é realizada. Não existem conexões entre a saída de um neurônio e algum outro neurônio localizado em uma camada anterior, o que significa que esta é uma arquitetura de rede *feedforward*. A Figura 11 apresenta a arquitetura de uma rede MLP como exemplo.

Para que este modelo seja capaz de representar o conhecimento adquirido, é necessário realizar ajustes nos pesos de ligação entre os neurônios. Durante o treinamento, a informação inicial flui para a frente através da rede até a camada de saída. Isso é chamado de propagação direta. Para se realizar os ajustes dos pesos utiliza-se um algoritmo chamado Retropropagação, proposto por Rumelhart, Hinton e Williams (1986). Este algoritmo busca retropropagar o erro entre a saída desejada e a saída obtida visando calcular o gradiente a fim de minimizar o erro total da rede, no decorrer do treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 11 – Representação de rede MLP



Fonte: Fath, Madanifar e Abbasi (2020)

Para realizar ajustes a fim de melhorar o aprendizado da rede, há alguns parâmetros que merecem atenção:

- a) **Taxa de aprendizagem:** taxas de aprendizagem altas implicam em mudanças maiores nos pesos. Deve-se buscar um meio termo para evitar grandes oscilações e garantir um treinamento em tempo adequado;
- b) **Quantidade de épocas:** indica a quantidade de vezes que o conjunto de dados deve ser apresentado à rede durante o treinamento. Uma maior quantidade de épocas implica em um maior aprendizado por parte da rede para aquele conjunto de dados, porém reduz a capacidade de generalização da rede;
- c) **Tamanho de lote:** define o número de exemplos de treinamento usados em uma época. Lotes maiores proporcionam treinamento mais rápido, entretanto isto pode causar uma generalização deficiente para problemas mais complexos;
- d) **Quantidade de camadas:** deve-se tomar cuidado, pois é nestas camadas que os mapeamentos não lineares entre entradas e saídas são processados. Um grande número de camadas pode causar o problema de desvanecimento do gradiente, impossibilitando da rede aprender.

3.3.3 Rede de Memória de Curto Longo Prazo (*Long Short Term Memory - LSTM*)

Rede Neural Recorrente (*Recurrent Neural Network - RNN*) é um tipo de RNA especializada em reconhecimento de dados sequenciais, como séries temporais. Uma de suas principais vantagens nesta tarefa, é o reconhecimento de longas cadeias de sequências, as quais seriam de difícil reconhecimento em redes não especializadas no reconhecimento deste tipo de dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

A decisão de uma RNN na etapa de tempo $t - 1$ afeta a decisão que alcançará um momento mais tarde na etapa de tempo t . Assim, as redes recorrentes têm duas fontes de entrada, o presente e o passado recente, que se combinam para determinar como respondem a novos dados. Todos os modelos compartilham a necessidade de relembrar entradas e saídas passadas para prever o próximo estado das redes e adaptar a si mesmos (GOODFELLOW; BENGIO; COURVILLE, 2016).

O modelo de rede LSTM foi proposto inicialmente por Hochreiter e Schmidhuber (1997), este consiste em um tipo especial de RNN capaz de aprender conexões de longo prazo, ou seja, esta rede é capaz de memorizar valores, sendo bem adequada para prever séries temporais com intervalos de tempo de duração desconhecido. Além disso, a ideia deste modelo foi concebida na busca de resolver o problema de explosão ou desvanecimento do gradiente em RNNs.

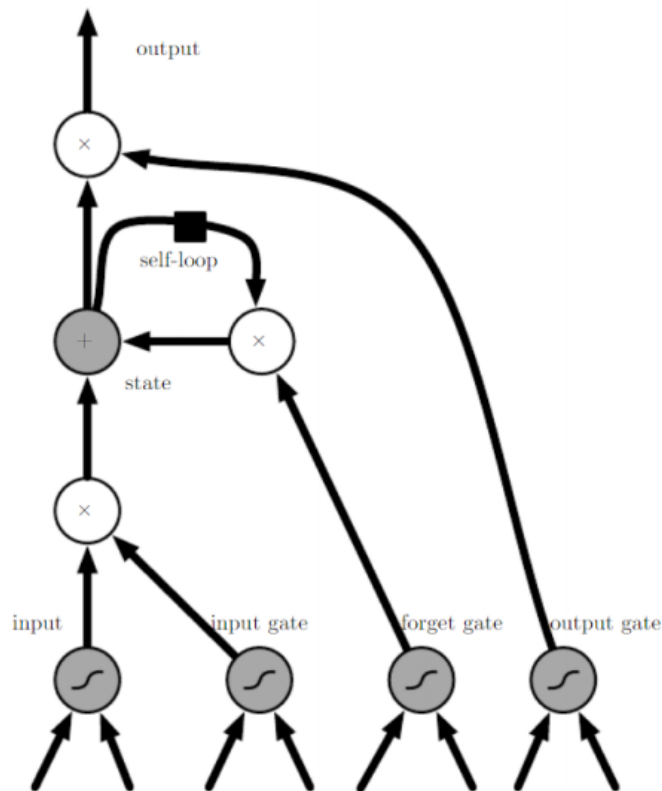
Uma de suas principais características em relação a outros modelos de RNN foi a inclusão de unidades especiais conhecidas como portões (*gates*). Essas unidades calculam os pesos que os conectam de forma a evitar o desvanecimento do gradiente (GOODFELLOW; BENGIO; COURVILLE, 2016).

Como toda RNN, as LSTMs têm a capacidade tanto de lembrar quanto de esquecer o estado anterior quando essa informação não for mais necessária. Ao longo do tempo de treinamento, a rede tem a capacidade de aprender o que esquecer exatamente, mecanismo que é executado através dos parâmetros do *forget gate*.

Dessa forma, os valores do estado anterior, a memória atual e a entrada são combinadas para formar a saída da unidade (ou célula), mecanismo que se mostrou bastante eficiente no aprendizado de longas dependências dos termos de uma sequência. A Figura 12 ilustra os principais elementos que estruturam uma célula LSTM.

O modelo LSTM foi desenvolvido há mais de duas décadas, mas até hoje é muito utilizado graças à expansão da capacidade computacional, a qual que era um empecilho até alguns anos atrás para as técnicas de AP.

Figura 12 – Representação de uma LSTM



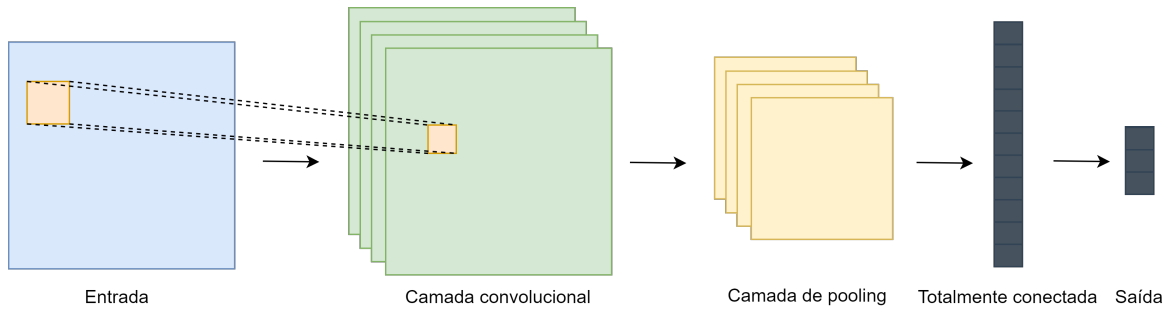
Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)

3.3.4 Rede Neural Convolucional (*Convolutional Neural Network - CNN*)

LeCun et al. (1990) introduziu as CNNs em um trabalho aplicado a um conjunto de dados para reconhecimento de letras manuscritas. Este modelo é inspirado no córtex visual dos animais e consiste em um tipo especializado de RNA para processamento de dados que possuem uma topologia conhecida como forma de grade (GOODFELLOW; BENGIO; COURVILLE, 2016), sendo os exemplos mais comuns dados como imagem ou séries temporais. Diversas abordagens baseadas em CNN têm sido desenvolvidas, tais como: AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), GoogLeNet (SZEGEDY et al., 2015), Inception (v3, v4) (SZEGEDY et al., 2016, 2017) e ResNet (HE et al., 2015).

Uma CNN básica é representada na Figura 13. Esta é composta por camadas convolucionais, funções de ativação, camadas de *pooling* e camadas totalmente conectadas, as quais serão detalhadas na sequência. Em uma primeira etapa ocorre a extração de recursos da imagem de entrada e na sequência as informações aprendidas são utilizadas para classificação.

Figura 13 – Representação do processo de uma CNN



Fonte: o autor

3.3.4.1 Camada convolucional

A camada convolucional é utilizada para encontrar padrões nos dados de entrada, como borda, forma ou cor em uma imagem. Esta camada substitui a multiplicação de matrizes típica das RNAs *Feedforward*, como a MLP, pela operação de convolução. Matematicamente a convolução de duas funções f e g é uma função h , que recebe um parâmetro x , para um deslocamento y , dada pela Eq.(12) (GOODFELLOW; BENGIO; COURVILLE, 2016; BRACEWELL; BRACEWELL, 1986).

$$(f * g)(x) = h(x) = \int f(x)g(y - x) dx \quad (12)$$

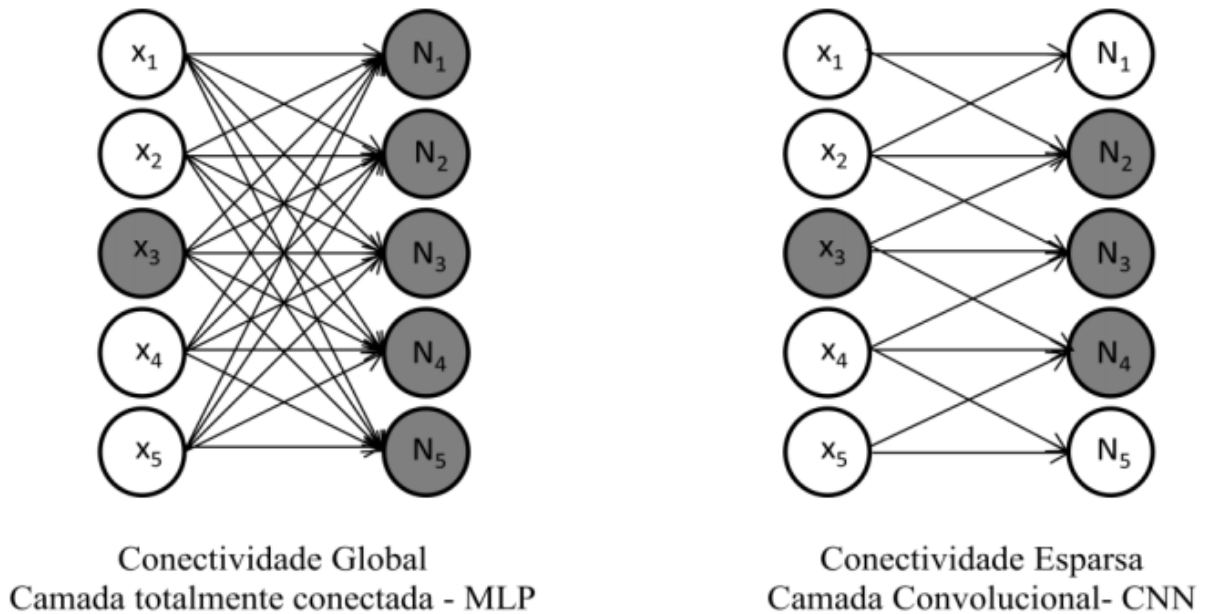
Na prática, este processo é realizado utilizando uma matriz, conhecida como filtro convolucional (também chamada de *kernel*), que é multiplicada, passo a passo (parâmetro chamado *stride* que define o tamanho do passo), pela matriz de entrada, resultando em um mapa de convolução como saída.

As redes CNN possuem algumas características que as diferenciam de modelos mais simples, como as redes MLPs por exemplo. Dentre estas características pode-se citar:

- a) **Conectividade esparsa:** este tipo de conectividade acontece devido ao fato de que o filtro que é utilizado no processo de convolução apresenta dimensões muito menores do que a matriz de entrada. Isto permite uma correlação espacial local através da aplicação de um padrão de conectividade local entre neurônios de camadas adjacentes (GOODFELLOW; BENGIO; COURVILLE, 2016). Na Figura 14 é possível verificar os dois tipos de conectividades (global e esparsa) para os dois tipos de arquiteturas (MLP e CNN). Observando a entrada x3 é possível verificar que quando o modelo faz uso da multiplicação de matrizes (conectividade global –

MLP) todas as unidades de saída (N1 a N5) são afetadas pela entrada x3. Isto não ocorre quando o modelo faz uso da camada convolucional, neste caso utilizando um filtro de tamanho 3 por exemplo, apenas três unidades da camada de saída (N2, N3 e N4) são afetadas pela entrada x3.

Figura 14 – Representação de conectividade global e esparsa das redes MLPs e CNNs

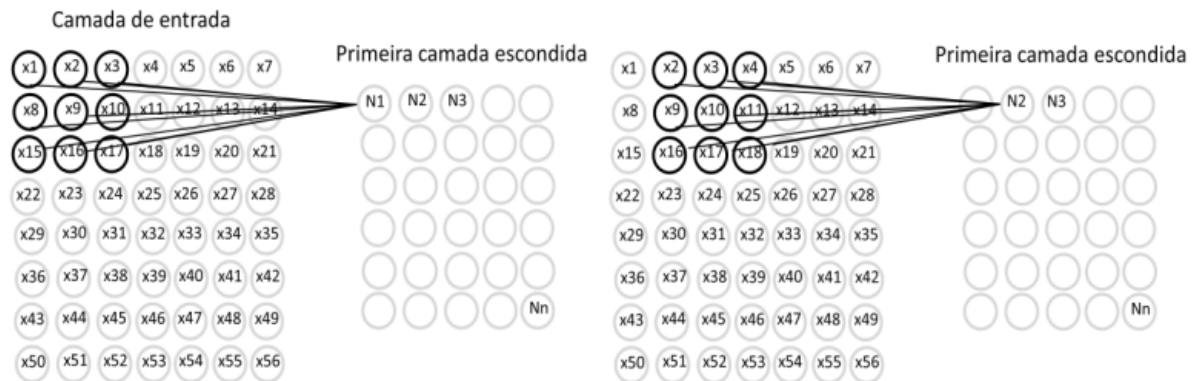


Fonte: adaptado de Goodfellow, Bengio e Courville (2016)

Ao processar uma imagem, a entrada pode ter milhares *pixels*, mas é possível detectar pequenos recursos significativos, como arestas, com filtros que ocupam apenas dezenas *pixels*. Isso significa que é necessário armazenar menos parâmetros, o que reduz os requisitos de memória do modelo e melhora sua eficiência estatística (GOODFELLOW; BENGIO; COURVILLE, 2016).

- b) **Campos receptivos locais:** cada região na matriz de entrada que representa o tamanho do filtro utilizado corresponde a um único neurônio na camada escondida. Cada deslocamento do filtro ao longo da matriz de entrada é representado por um novo neurônio na camada escondida. Este processo pode ser visualizado na Figura 15. Em um primeiro momento, o resultado do filtro de formato $[3 \times 3]$ é representado pelo neurônio N1 da camada escondida. A região destacada em preto na matriz de entrada representa o campo receptivo local do neurônio de saída N1. Avança-se um passo no processo de convolução, e a nova posição do filtro é representada pelo neurônio N2.

Figura 15 – Exemplo do campo receptivo local das CNNs

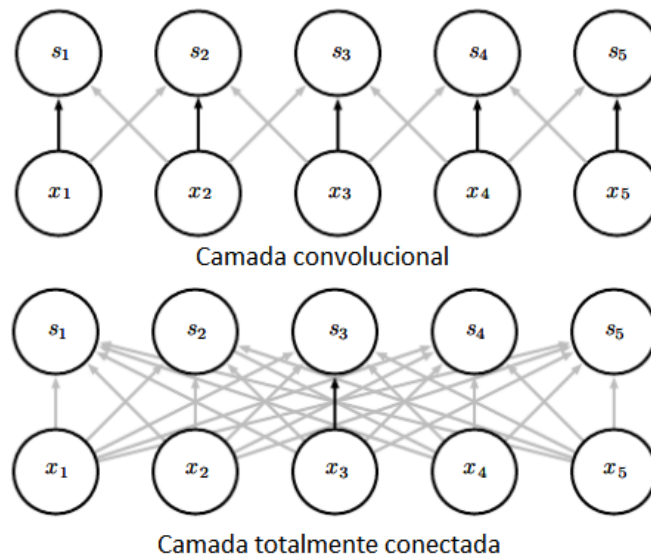


Fonte: adaptado de Nielsen (2015)

- c) **Compartilhamento de parâmetros:** em uma rede MLP, cada elemento da matriz de peso é usado exatamente uma vez ao calcular a saída de uma camada. Na CNN, cada membro do filtro é usado em todas as posições da entrada. Isto faz com que os padrões que ocorrem com frequência na matriz de entrada e que estejam localizados em qualquer parte da entrada possam ser aprendidos e em vez de aprender um conjunto separado de parâmetros para cada local, aprende-se apenas um conjunto buscado nos dados de entrada. Isto reduz o número total de parâmetros treináveis e a quantidade de multiplicações realizadas. Na Figura 16 pode ser visualizado como acontece o compartilhamento de parâmetros nas camadas convolucionais. As setas pretas indicam as conexões que usam um determinado parâmetro em dois neurônios diferentes. Na camada convolucional as setas pretas indicam os usos do elemento central de um filtro de 3 elementos. Devido ao compartilhamento de parâmetros, este parâmetro único é usado em todos os locais de entrada. Já na camada totalmente conectada (semelhante a uma rede MLP), a única seta preta indica o uso do elemento central da matriz de peso.
- d) **Representação equivariante:** o compartilhamento de parâmetros faz com que a camada convolucional possa ter uma propriedade chamada equivariância à translação. Isto significa que uma função é equivariante significa que se a entrada muda, a saída muda da mesma maneira, ou seja, uma função $f(x)$ é equivariante a uma função g se $f(g(x)) = g(f(x))$. Ao processar dados de série temporal, isso significa essa convolução produz uma espécie de linha do tempo que mostra quando diferentes características aparecem na entrada. Ao mover-se um evento mais tarde na entrada, a mesma representação do mesmo aparecerá na saída, logo mais tarde. Si-

milamente com imagens, a convolução cria um mapa 2-D de onde certos recursos aparecem na entrada. Caso mova-se o objeto na entrada, sua representação moverá a mesma quantidade na saída (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 16 – Compartilhamento de parâmetros. Conexões de mesma cor representam os elementos do filtro sendo compartilhados

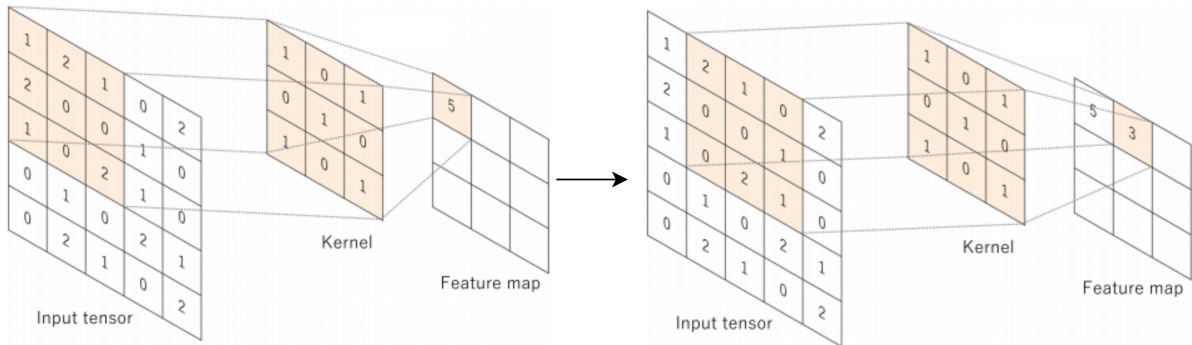


Fonte: adaptado de Goodfellow, Bengio e Courville (2016)

O filtro é formado por pesos que geralmente são inicializados utilizando algum método de distribuição aleatória e são atualizados a cada nova entrada durante o processo de Retropropagação. Os métodos de inicialização mais utilizados são os propostos por Glorot e Bengio (2010) e He et al. (2015). Ambos os métodos têm por objetivo uma inicialização que busca distribuir o gradiente de forma semelhante ao longo das camadas da rede inicializando os pesos tendo em mente o tamanho da camada anterior, o que ajuda a atingir um mínimo global da função de custo de forma mais rápida e eficiente.

Estes filtros são responsáveis por destacarem detalhes marcantes nos dados de entrada. Em uma imagem por exemplo, bordas e formas podem ser ressaltadas pelos filtros no processo de convolução. Isto permite que a rede demande um nível mínimo de pré-processamento quando comparada a outros algoritmos de classificação de imagens (LECUN et al., 1990). A Figura 17 representa o processo de convolução em uma matriz.

Figura 17 – Representação de duas etapas do processo de convolução em uma matriz de entrada 6x6 com um filtro 3x3



Fonte: adaptado de Yamashita et al. (2018)

Para dados de entrada tridimensionais, como imagens RGB, o processo de convolução é realizado em cada canal e os resultados são somados, resultando no mapa de convolução do filtro específico. À medida que o filtro se desloca, através da entrada, é calculado o produto entre os pesos do filtro e os pixels correspondentes da entrada. Sendo assim, para dados de entrada com apenas 1 canal, pode-se matematicamente descrever o processo de convolução da CNN pela Eq (13) (GOODFELLOW; BENGIO; COURVILLE, 2016).

$$c(I, K)_{x,y} = \sum_{m=1}^h \sum_{n=1}^w I(x+m, y+n)K(m, n) + b \quad (13)$$

onde:

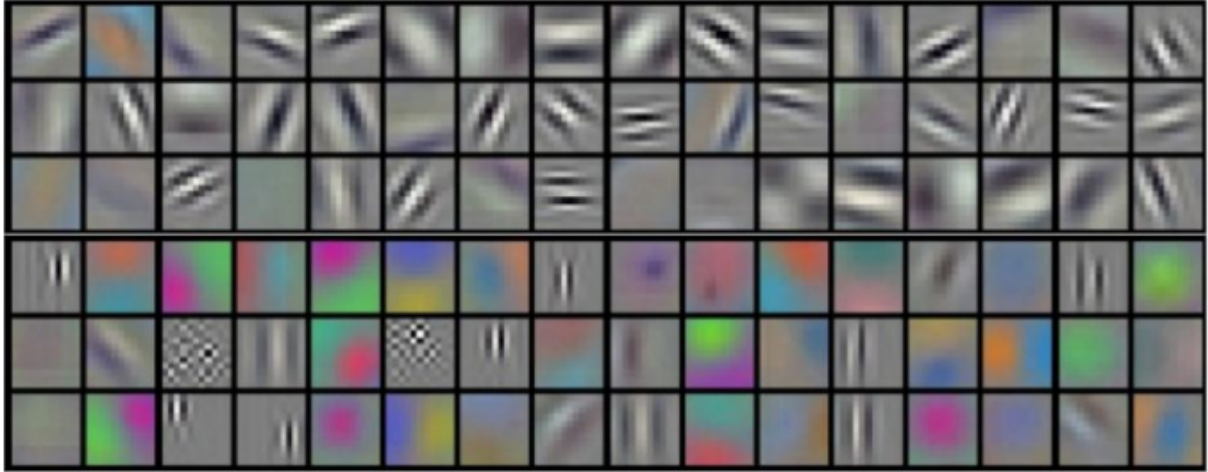
- a) I representa uma matriz como entrada;
- b) K um filtro (*kernel*) convolucional bidimensional que passeia sobre a matriz I ;
- c) h representa a altura da matriz I ;
- d) w representa a largura da matriz I ;
- e) b representa o *bias* da camada convolucional;

Dessa forma, o valor de cada pixel com posição (x, y) da nova imagem c é obtido mediante à somatória do produto dos pesos do filtro K pelos valores dos pixels correspondentes à posição do filtro na imagem de entrada I . Através das iterações durante o processo de treinamento, a rede vai aprender os pesos dos filtros, que vão se ativar quando virem algum tipo específico de característica em alguma posição espacial nos dados de entrada.

O resultado deste processo gera os mapas de características. Este mapa consiste em um conjunto de dados gerados pela extração de um determinado filtro sobre a entrada durante a convolução e quantifica a presença da característica buscada. Na Figura 18 são mostrados os

mapas de características gerados por 96 filtros aprendidos na primeira camada de convolução da proposta no trabalho de Krizhevsky, Sutskever e Hinton (2012).

Figura 18 – Exemplos de mapas de características



Fonte: Krizhevsky, Sutskever e Hinton (2012)

Além disso, uma função de ativação é usada no final do processo de convolução para fornecer normalização dos dados. As funções mais comumente utilizadas são: *Rectified Linear Unit (ReLU)*, *Sigmoid*, *Tangente Hiperbólica (TanH)* e *Softmax*. Dentre estas quatro funções de ativação, a ReLU é a mais conhecida pois resolve o problema de desvanecimento do gradiente e problemas de esparsidade (GLOROT; BORDES; BENGIO, 2011).

- a) **Função ReLU:** esta função retorna 0 se receber qualquer entrada de valor negativo, e para qualquer valor positivo é retornado o próprio valor de volta. A Eq. (14) representa a função, com o comportamento no gráfico da Figura 19;

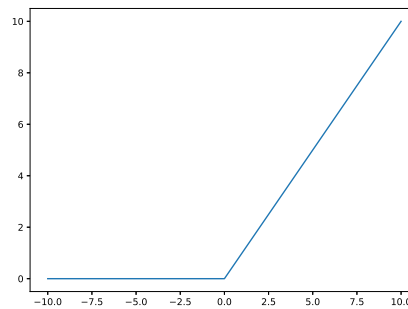
$$f(x) = \max(0, x) \quad (14)$$

- b) **Função Softmax:** esta função tem por objetivo forçar a saída de uma RNA a representar a probabilidade dos dados serem de uma das classes definidas. Utilizando demais funções, a saída seria um simples valor numérico onde o maior indica a classe vencedora. A Eq.(15) apresenta a fórmula desta função:

$$\phi_i = \frac{\epsilon_i^z}{\sum \epsilon_j^z} \quad (15)$$

onde i representa o índice do neurônio de saída sendo calculado, j representa os índices de todos os neurônios de uma camada e z os valores do vetor de saída.

Figura 19 – Gráfico da função ReLU



Fonte: o autor

3.3.4.2 Camada de pooling

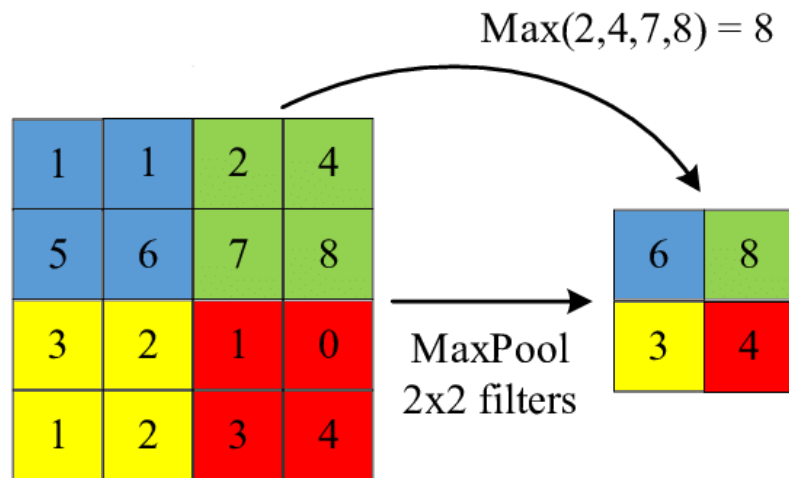
A camada de *pooling* é usada para obter invariância às transformações da imagem e representações mais compactas dos dados, reduzindo a dimensão da matriz e preservando os recursos extraídos da imagem (BOUREAU; PONCE; LECUN, 2010; GOODFELLOW; BENGIO; COURVILLE, 2016), além disso esta sumarização de dados serve para diminuir a quantidade de pesos a serem aprendidos pela rede. Uma camada de *pooling* recebe cada saída do mapa de características da camada convolucional e prepara um mapa de características condensadas. Esta operação envolve deslizar um filtro bidimensional sobre cada canal do mapa de recursos e resumir os recursos situados na região coberta pelo filtro. Os tipos de operação de *pooling* existentes são:

- a) *maximum pooling*: também chamado de *pooling* máximo ou *max-pooling*, este tipo seleciona o maior valor na área do *kernel*;
- b) *average pooling*: mantém o valor médio dos valores que aparecem no filtro;
- c) *minimum pooling*: semelhante ao *maximum pooling*, neste tipo o valor mínimo é armazenado em vez do máximo;

A Figura 20 apresenta o processo de *max-pooling* com *pool-size* (dimensões do filtro de *pooling*) de $[2 \times 2]$ em uma matriz $[4 \times 4]$. Essencialmente, neste tipo de operação, a matriz tem seu tamanho reduzido pela metade.

Com isso, diversos mapas de características podem ser criados, contendo diversas características específicas. O resultado do uso de uma camada de *pooling* são mapas de características amostrados ou agrupados em uma versão resumida dos recursos detectados na entrada. Sendo assim, pequenas mudanças na localização do padrão na entrada detectada pela camada convolucional resultarão em um mapa de recurso agrupado com o recurso no mesmo local.

Figura 20 – Representação do processo de *max-pooling*



Fonte: adaptado de Kamencay et al. (2017)

3.3.4.3 Dropout

As RNAs profundas possuem múltiplas camadas escondidas não lineares, as quais as tornam modelos capazes de aprender relações complexas entre entradas e saídas. No entanto, é comum que os dados apresentem desvios causados por erros de medição ou fatores aleatórios, de modo que existirão no conjunto de treino, mas não em dados de teste reais (SRIVASTAVA et al., 2014).

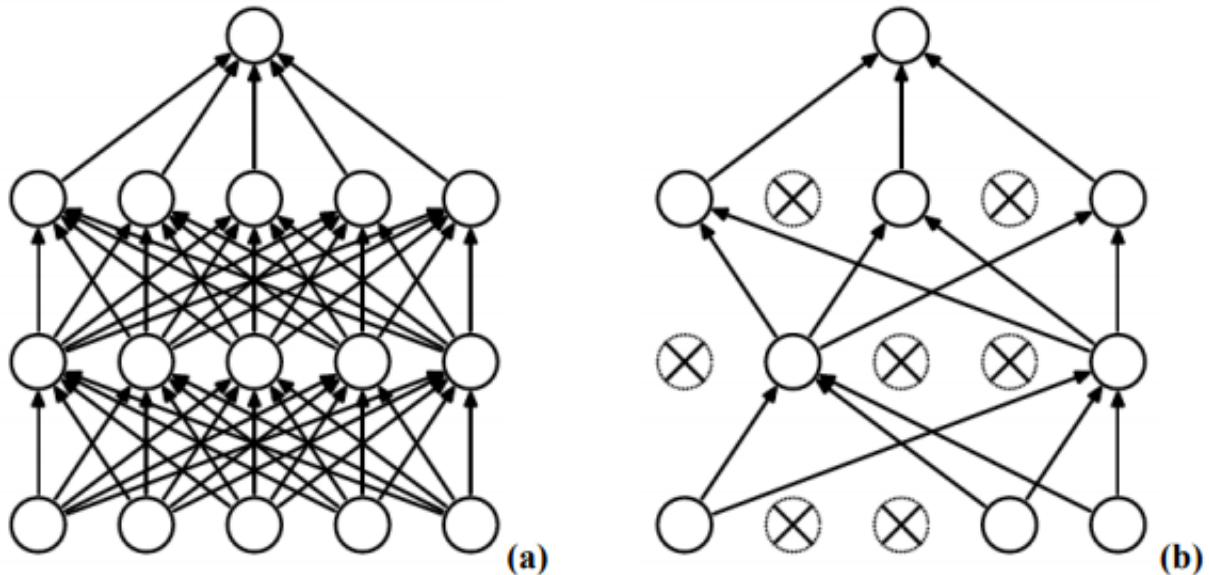
Ademais, quando conjunto de dados de treinamento é pequeno ou limitado a rede pode se ajustar demasiadamente, de modo que será especializada em reconhecer somente os padrões com os quais foi treinada e não vai conseguir uma boa generalização. Este caso é chamado de *overfitting*, ou sobreajuste. Um modelo sobreajustado apresenta alta precisão quando testado com seu conjunto de dados de treinamento, porém não capaz de fazer uma boa representação da realidade.

O *Dropout* é uma técnica para solucionar o problema de sobreajuste dos pesos da RNA (SRIVASTAVA et al., 2014). A ideia principal é desativar a ligação de neurônios da RNA aleatoriamente durante o treinamento para evitar mínimos locais e ajudar o algoritmo de aprendizado a convergir para um modelo mais generalizado e capaz de classificar dados futuros não rotulados (HINTON et al., 2012).

A Figura 21 representa este conceito. No modelo (a) é apresentada uma rede com todas as camadas interconectadas. No modelo (b) a representação do *Dropout* aplicada em determi-

nado instante. A cada entrada durante o treinamento, uma unidade pode ser removida temporariamente da rede de forma aleatória juntamente com todas as suas conexões de entrada e saída.

Figura 21 – Representação da técnica de *Dropout*



Fonte: adaptado de Srivastava et al. (2014)

3.3.4.4 Camadas totalmente conectadas

Por último, as camadas finais de uma CNN são geralmente compostas por camadas de rede totalmente conectadas, normalmente uma rede MLP, utilizada para executar a classificação dos dados.

Na literatura também podem ser encontradas abordagens que utilizam outras redes conectadas ao final, tais como a proposto de Jain, Gupta e Moghe (2018), o qual utilizou uma camada LSTM antes da saída da rede para realizar regressão.

3.3.4.5 Retropropagação na CNN

Uma das diferenças do processo de retropropagação das redes CNN para redes MLP é que a passagem do erro para trás durante a retropropagação também usa convolução. O processo de cálculo dos gradientes será detalhado na sequência (LECUN et al., 1990; GOODFELLOW; BENGIO; COURVILLE, 2016; NG et al., 2015; SOLAI, 2018).

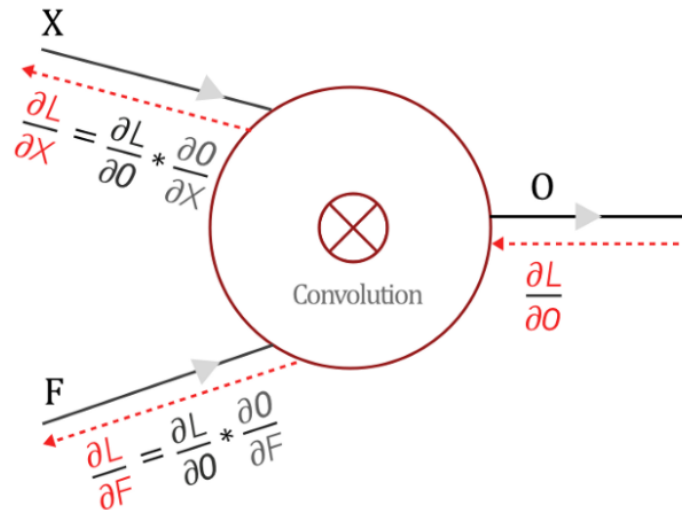
Formalmente, segundo Goodfellow, Bengio e Courville (2016) o processo de convolução se dá pela Eq. (16), onde usa-se o filtro rotacionado em 180 graus, no sentido de que conforme o valor de m aumenta, o índice na entrada aumenta, mas o índice no filtro diminui:

$$c(I, K)_{x,y} = \sum_m \sum_n I(m, n) K(x - m, y - n) \quad (16)$$

Entretanto, o cálculo mais comumente utilizado no processo de propagação das camadas convolucionais é o descrito na Eq. (13), que é formalmente chamado de correlação cruzada, mas é comum usar-se o nome de convolução.

A Figura 22 representa um neurônio convolucional, onde X uma imagem de entrada, F representa o filtro convolucional e O a saída do processo de convolução.

Figura 22 – Representação do neurônio convolucional



Fonte: adaptado de (SOLAI, 2018)

Sendo $\frac{\partial O}{\partial X}$ e $\frac{\partial O}{\partial F}$ os gradientes locais e $\frac{\partial L}{\partial O}$ o erro da camada anterior no processo de retropropagação que deve ser repassado para a camada seguinte, deve-se calcular $\frac{\partial L}{\partial X}$ e $\frac{\partial L}{\partial F}$.

Primeiro será abordado o cálculo de $\frac{\partial L}{\partial F}$. O e F são matrizes, e $\frac{\partial O}{\partial F}$ é o gradiente local da saída O e relação ao filtro F . Isto significa que deve-se primeiro diferenciar a matriz O em relação à uma matriz F . Na Eq. (17) tem-se como exemplo a posição $[1, 1]$ da matriz O :

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \quad (17)$$

E as derivadas em respeito a F_{11} , F_{12} , F_{21} e F_{22} são:

$$\frac{\partial O_{11}}{\partial F_{11}} = X_{11}; \frac{\partial O_{11}}{\partial F_{12}} = X_{12}; \frac{\partial O_{11}}{\partial F_{21}} = X_{21}; \frac{\partial O_{11}}{\partial F_{22}} = X_{22} \quad (18)$$

Isto significa que $\frac{\partial O}{\partial F}$ será uma derivada parcial da matriz O em relação à matriz F . Usando a regra da cadeia, para cada elemento de F , calcula-se:

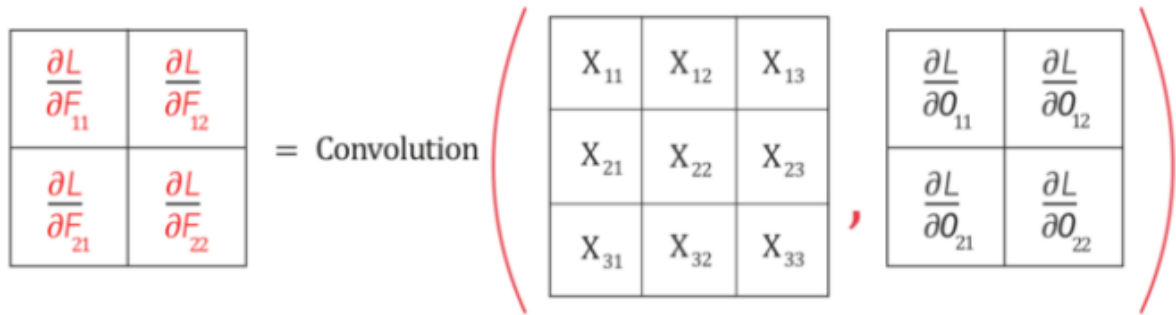
$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^m \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i} \quad (19)$$

Como exemplo, para a posição $[1, 1]$ tem-se:

$$\frac{\partial L}{\partial F_{11}} = \frac{\partial L}{\partial O_{11}} * X_{11} + \frac{\partial L}{\partial O_{12}} * X_{12} + \frac{\partial L}{\partial O_{21}} * X_{21} + \frac{\partial L}{\partial O_{22}} * X_{22} \quad (20)$$

Desta forma, chega-se à conclusão que este processo representa a correlação cruzada entre a entrada X e o gradiente do erro da próxima camada $\frac{\partial L}{\partial O}$, conforme representado na Figura 23.

Figura 23 – Representação da convolução nos pesos da entrada



Fonte: adaptado de (SOLAI, 2018)

Semelhantemente, após o processo de diferenciação para a posição $[1, 1]$, com base na Eq. (17), as derivadas em respeito a X_{11} , X_{12} , X_{21} e X_{22} são:

$$\frac{\partial O_{11}}{\partial X_{11}} = F_{11}; \frac{\partial O_{11}}{\partial X_{12}} = F_{12}; \frac{\partial O_{11}}{\partial X_{21}} = F_{21}; \frac{\partial O_{11}}{\partial X_{22}} = F_{22} \quad (21)$$

O cálculo de $\frac{\partial L}{\partial X}$ pode ser representado pela Eq. (22):

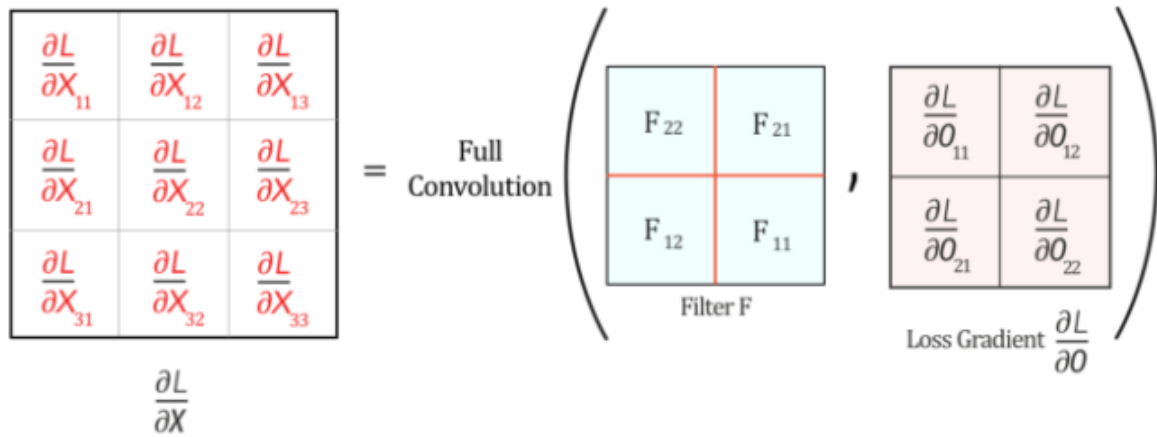
$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^m \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i} \quad (22)$$

E como exemplo, para as posições $[1, 1]$, $[1, 2]$ e $[1, 3]$:

$$\begin{aligned}
\frac{\partial L}{\partial X_{11}} &= \frac{\partial L}{\partial O_{11}} * F_{11} \\
\frac{\partial L}{\partial X_{12}} &= \frac{\partial L}{\partial O_{11}} * F_{12} + \frac{\partial L}{\partial O_{12}} * F_{11} \\
\frac{\partial L}{\partial X_{13}} &= \frac{\partial L}{\partial O_{12}} * F_{12}
\end{aligned} \tag{23}$$

Para obter os gradientes da matriz de entrada $\frac{\partial L}{\partial X}$, é necessário rotacionar o filtro F em 180 graus e calcular a convolução completa do filtro pelos gradientes da saída em relação ao erro $\frac{\partial L}{\partial O}$, conforme representado na Figura 24.

Figura 24 – Representação da convolução nos pesos do filtro



Fonte: adaptado de (SOLAI, 2018)

Em suma, $\frac{\partial L}{\partial F}$ consiste em uma correlação cruzada entre uma entrada X e o gradiente de erro $\frac{\partial L}{\partial O}$. E para calcular $\frac{\partial L}{\partial X}$, é realizada uma convolução entre um filtro F rotacionado em 180 graus e o gradiente de erro $\frac{\partial L}{\partial O}$. Para calcular os gradientes das camadas de pooling os gradientes podem ser calculados seguindo o mesmo procedimento de usar a regra da cadeia de derivadas.

3.3.4.6 Método de otimização

Os algoritmos de aprendizagem profunda exigem otimização em muitos contextos, por exemplo, a realização da inferência e retropropagação nas CNN implica a resolução de um problema de otimização. O treinamento da rede é uma questão importante e que a depender da

metodologia adotada, pode envolver alto custo computacional. Diversas técnicas de otimização foram desenvolvidas para resolver tal problema.

O treinamento é feito utilizando o algoritmo de retropropagação proposto por Rumelhart, Hinton e Williams (1986). O termo retropropagação é muitas vezes mal interpretado como o único algoritmo responsável pela aprendizagem dos pesos das redes neurais multicamadas. Na verdade, este somente se refere ao método para calcular o gradiente, enquanto outro algoritmo é responsável pela aprendizagem utilizando este gradiente, também chamado de método de otimização.

Neste trabalho foi utilizado o método de otimização Adam (KINGMA; BA, 2014). Este consiste em um método de otimização de funções objetivas estocásticas, calculando taxa de aprendizagem adaptativa individualmente para cada parâmetro a partir de estimativas do primeiro e segundo momentos dos gradientes. O algoritmo calcula uma média móvel exponencial do gradiente e do gradiente quadrado e os parâmetros β_1 e β_2 controlam as taxas de decaimento dessas médias móveis. O seu nome deriva de *adaptive moment estimation*, que significa estimação de momento adaptativo. Este método combina vantagens de dois métodos, o AdaGrad (DUCHI; HAZAN; SINGER, 2011) e o RMSprop (HINTON et al., 2012).

A seguir, são apresentadas as equações de atualização (KINGMA; BA, 2014):

$$(m_t)_i = \beta_1(m_{t-1})_i + (1 - \beta_1)(\nabla L(W_t))_i \quad (24)$$

$$(v_t)_i = \beta_2(v_{t-1})_i + (1 - \beta_2)(\nabla L(W_t))_i^2 \quad (25)$$

$$(W_t)_i = (W_{t-1})_i - \alpha \frac{\sqrt{1 - (\beta_2)_i^t}}{1 - (\beta_1)_i^t} \frac{(m_t)_i}{\sqrt{(v_t)_i} + \epsilon} \quad (26)$$

onde:

- a) α : taxa de aprendizado. A proporção em que as ponderações são atualizadas (valor padrão de 0,001). Valores maiores (por exemplo, 0,3) resultam em um aprendizado inicial mais rápido;
- b) β_1 : taxa de decaimento exponencial para as estimativas do primeiro momento (valor padrão de, 0,9);
- c) β_2 : taxa de decaimento exponencial para as estimativas do segundo momento (valor padrão de 0,999).
- d) $\nabla L(W_t)$: é o gradiente da função de erro avaliada no vetor de pesos W_t ;

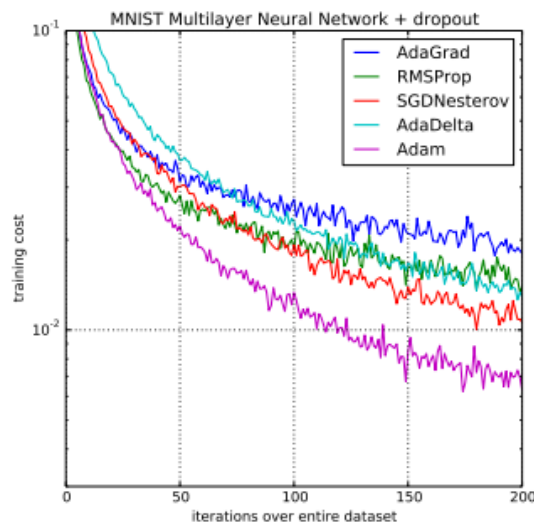
- e) m_t : média móvel exponencial do gradiente;
- f) v_t : o gradiente quadrado;
- g) ϵ : número com valor muito pequeno para evitar qualquer divisão por zero (valor padrão 10^{-8}).

Os autores deste método demonstraram sua eficácia aplicando o algoritmo no treinamento na base de dados do trabalho de LeCun et al. (1990). Dentre os benefícios deste algoritmo, os autores citam as seguintes características:

- a) Computacionalmente eficiente;
- b) Bem adequado para problemas grandes em termos de dados e/ou parâmetros;
- c) Adequado para problemas com gradientes muito barulhentos/ou esparsos.

Conforme mostrado na Figura 25, os autores demonstram que a convergência atende às expectativas da análise teórica frente a outros algoritmos já existentes.

Figura 25 – Comparação do método Adam com demais métodos de otimização



Fonte: Kingma e Ba (2014)

3.3.4.7 Inicialização de pesos

A inicialização dos pesos da rede é crucial para uma boa convergência em uma quantidade de tempo razoável. Se os valores dos pesos são muito pequenos, a variância do sinal de entrada começa a diminuir assim que passa através das camadas da rede. Se os pesos são muito grandes, então a variância dos dados de entrada tende a aumentar rapidamente a cada passagem pelas camadas e, eventualmente, tornam-se inutilizáveis.

Para garantir que a rede possa funcionar corretamente, faz-se necessário certificar-se de que os pesos estejam em um intervalo razoável. Este processo é feito de maneira aleatória ou utilizando distribuição gaussiana.

Em cada passagem pelas camadas, é desejável que se permaneça a mesma variância nos pesos. Isso ajuda a evitar problemas de sobreajustes bem como que os pesos se tornem muito grandes ou muito pequenos. Com base nesta ideia surgem os trabalhos de Glorot e Bengio (2010) e He et al. (2015), mencionados anteriormente.

Dado um neurônio y com entrada x e pesos W , o cálculo da variância é dado na Eq. (27):

$$var(y) = var(W_1x_1) + var(W_2x_2) + \dots + var(W_nx_n) \quad (27)$$

Para a variância dos termos dentro dos parênteses, tem-se:

$$var(W_ix_i) = E[x_i]^2 var(W_i) + E[W_i]^2 var(x_i) + var(W_i)var(x_i) \quad (28)$$

onde $E[\]$ representa o valor esperado de uma variável. Tendo os pesos uma distribuição de média zero, elimina-se o $E[\]$, e pode-se escrever:

$$var(y) = \sum_{i=1}^n var(x_i)var(W_i) \quad (29)$$

O objetivo é manter a variação semelhante ao longo de todas as camadas. Em outras palavras, deve-se ter como objetivo tornar a variância de x seja a mesma que y . Tendo assim:

$$var(y) = var(x) \Leftrightarrow var(W_i) = \frac{1}{fan_{in}} \quad (30)$$

onde fan_{in} é a quantidade de neurônios de entrada da camada. Generalizando para todos os pesos W , com o objetivo de manter a mesma variância para os pesos de entrada em ambos os processos de propagação e retropropagação, tem-se as fórmulas das Eqs. (31) e (32), que são as formulas das das inicializações com distribuição normal e uniforme de Glorot e Bengio (2010), respectivamente:

$$var(W) = \frac{1}{\frac{fan_{in} + fan_{out}}{2}} = \frac{2}{fan_{in} + fan_{out}}; \sigma(W) = \sqrt{\frac{2}{fan_{in} + fan_{out}}} \quad (31)$$

$$var(Uniform(-limit, limit)) = \frac{limit^2}{3} \quad (32)$$

$$var(W) = \frac{2}{fan_{in} + fan_{out}} \Leftrightarrow limit = \sqrt{\frac{6}{fan_{in} + fan_{out}}}$$

Outro método de inicialização de pesos foi proposto por He et al. (2015). Em seu trabalho, o autor estuda a eficácia do método de inicialização descrito anteriormente utilizando a função ReLU. As Eqs. (33) e (34), apresentam as formulas da inicializações com distribuição normal e uniforme de He et al. (2015), respectivamente:

$$var(W) = \frac{2}{fan_{in}}; \sigma(W) = \sqrt{\frac{2}{fan_{in}}} \quad (33)$$

$$\begin{aligned} var(Uniform(-limit, limit)) &= \frac{limit^2}{3} \\ var(W) = \frac{2}{fan_{in}} &\Leftrightarrow limit = \sqrt{\frac{6}{fan_{in}}} \end{aligned} \quad (34)$$

A proposta surgiu dado que a função ReLU transforma os números negativos em 0, removendo cerca de metade da variância. Sendo assim, a proposta dos autores é de dobrar a variância dos pesos para compensar isto.

3.3.5 Aprendizado Profundo (AP)

As técnicas de Aprendizado Profundo (AP) aperfeiçoaram a forma que modelos de AM convergem e generalização os dados. O primeiro relaciona-se ao aprendizado em reconhecer padrões. Este fator é importante, pois se a rede não aprendeu as relações existentes, significa que não convergiu e não atingirá bons resultados. Já o segundo tema, trata sobre reconhecimento de padrões não existentes nos dados de treinamento e assim identificar valores fora do padrão, também chamado de abstração.

Os modelos de Aprendizado Profundo (AP) constituem uma área de AM, composto por métodos estatísticos que tentam modelar abstrações de alto nível de dados. O grande avanço do AP é a melhoria do aprendizado de representação e da capacidade de abstração (LECUN; BENGIO; HINTON, 2015).

O aprendizado de representação é um conjunto de técnicas para descobrir matematicamente as representações de dados para detecção ou classificação por meio da extração de características. Os métodos de AM tradicionais não são suficientes para problemas mais complexos, como problemas de visão computacional.

Os modelos de AP resolvem o problema do aprendizado de representação introduzindo representações que são expressas em vários níveis de abstração por várias camadas de processamento (GOODFELLOW; BENGIO; COURVILLE, 2016). Assim, o AP está evoluindo o

estado da arte na solução de problemas que resistiram às melhores tentativas da comunidade de inteligência artificial por muitos anos, como reconhecimento de fala, de objetos e classificação de imagens e padrões (LECUN; BENGIO; HINTON, 2015).

Existem vários modelos de AP, dentre eles pode-se citar: Rede Neural Convolucional (*Convolutional Neural Network - CNN*), Rede de Crença Profunda (*Deep Belief Network - DBN*), Máquina de Boltzmann Restrita (*Restricted Boltzmann Machine - DBM*) e Rede de Memória de Curto Longo Prazo (*Long Short Term Memory - LSTM*). Na literatura, CNN, DBN e RBM são usadas principalmente nas abordagens de classificação e reconhecimento de imagens (WU; CHEN, 2015; KARPATHY et al., 2014; CIRESAN et al., 2011), enquanto RNNs, como LSTM, são focadas em análise de dados de séries temporais, processamento de linguagem natural e reconhecimento de fala (KARIM et al., 2017; GRAVES; JAITLEY; MOHAMED, 2013).

4 SISTEMA DE NEGOCIAÇÃO PROPOSTO

Neste capítulo, será apresentada a metodologia para a criação do sistema de tomada de decisão proposto neste trabalho, baseado na aquisição e pré processamento dos dados e sua utilização no treinamento e classificação pelo sistema proposto.

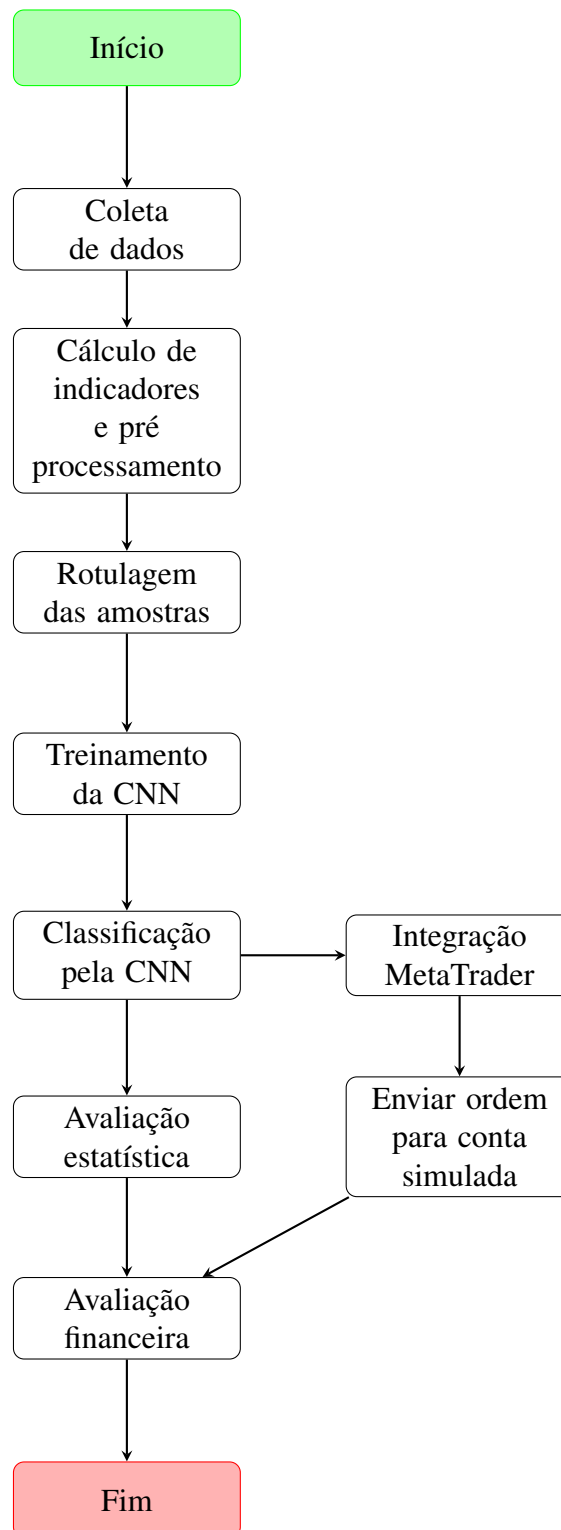
Inicialmente é descrito na seção 4.1 a arquitetura do sistema proposto, seguido pelos detalhes da coleta dos dados e do pré processamento nas seções 4.2 e 4.3, respectivamente. O método de rotulagem das amostras é apresentado na seção 4.4. Por fim a arquitetura do modelo de CNN utilizado é descrita na seção 4.5.

4.1 Arquitetura do sistema

Neste trabalho, é proposta uma abordagem aplicando uma arquitetura de CNN a fim de identificar os melhores momentos de compra e venda no mercado financeiro. O sistema foi chamado CNN Trading Classifier (CNN-TC), pois o mesmo é baseado na classificação de séries temporais a fim de identificar padrões que possam indicar a operação mais rentável no momento.

O processo do sistema de negociação automatizado proposto é apresentado na Figura 26. Os dados históricos são coletados e passam por um pré-processamento. Neste processo, é feito o cálculo de indicadores técnicos e a normalização dos dados. Cada amostra é então rotulada utilizando o método da seção 4.4 e os dados são utilizados para o treinamento da RNA. O modelo CNN executa a classificação de operação como compra/venda/neutro para cada dia do período de avaliação, e o resultado é utilizado para as avaliações estatística e financeira.

Figura 26 – Processo do sistema de negociação proposto



Fonte: o autor

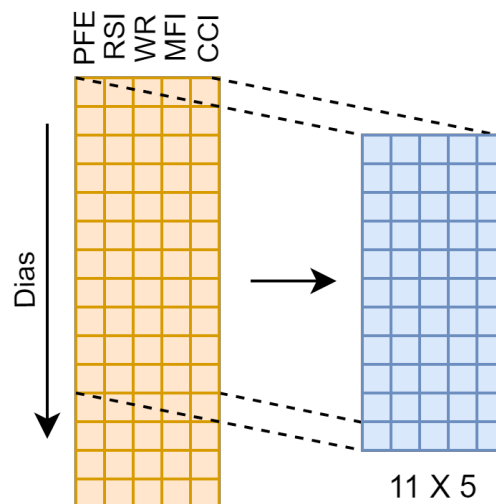
4.2 Dados de entrada

A base de dados inicialmente é composta por: data, volume financeiro, preços de abertura, máxima, mínima e fechamento. Como elementos de entrada da CNN, são utilizados o Preço de Fechamento Estacionário (PFE) e os indicadores técnicos previamente abordados na seção 3.1.2, dando um total de 5 elementos. Foi utilizada a Biblioteca de Análise Técnica em Python (PADIAL, acessado em 01/03/2020) para calcular os valores dos indicadores. O conjunto de elementos de entrada definido para esta abordagem foi avaliado empiricamente e escolhido o que forneceu a melhor precisão no processo de aprendizado.

4.3 Pré-processamento de dados

O conjunto de dados é dividido em amostras diárias utilizando uma abordagem de janela deslizante de tamanho 11, onde cada amostra representa um dia e é composta pelos dados dos 10 dias anteriores e o dia atual, o formato dos dados que entrarão na rede é de matrizes bidimensionais de $[11 \times 5]$, como mostrado na Figura 27.

Figura 27 – Processo de remodelagem dos dados de entrada



Fonte: o autor

Na sequência, os valores de cada coluna são normalizados entre 0-1 para cada amostra, usando a normalização min-max, de acordo com a Eq. 35. Cada coluna da matriz (eixo x) representa um recurso de entrada e cada linha (eixo y) representa um dia da série temporal.

Dessa forma o processo de convolução extrairá padrões sobre as representações de cada dia na matriz de entrada, cada dia será visualizado de forma independente.

$$y_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (35)$$

4.4 Rotulação dos Dados

É utilizado o processo de treinamento supervisionado e todas as amostras diárias são rotuladas como compra, venda ou neutro, de acordo com o Algoritmo 1 usando como base os preços de fechamento. Este algoritmo consiste em uma implementação de janela deslizante baseada na proposta de Sezer e Ozbayoglu (2018), e tem por objetivo encontrar topos e fundos na oscilação dos preços para as identificar os melhores pontos para operações de compra e venda.

Algoritmo 1 – Rotulação de amostras

Entrada: preços

Saída: rótulos

inicioJanela \leftarrow 0

tamanhoJanela \leftarrow 11

fimJanela \leftarrow *inicioJanela* + *tamanhoJanela*

diasTotais \leftarrow Comprimento(preços)

enquanto *contador* \leq *diasTotais* **faça**

valorMin \leftarrow Min(preços[*inicioJanela*:*fimJanela*])

valorMax \leftarrow Max(preços[*inicioJanela*:*fimJanela*])

para *i* in preços[*inicioJanela*:*fimJanela*] **faça**

se preços[*i*] = *valorMax* AND preços[*i*] NOT NULO **então**

rótulos[*i*] = "Neutro";

rótulos[*i* + 1] = "Venda";

fim

senão se preços[*i*] = *valorMin* AND preços[*i*] NOT NULO **então**

rótulos[*i*] = "Neutro";

rótulos[*i* + 1] = "Compra";

fim

senão

rótulos[*i*] = "Neutro";

fim

inicioJanela \leftarrow *fimJanela* + 1

fimJanela \leftarrow *inicioJanela* + *tamanhoJanela*

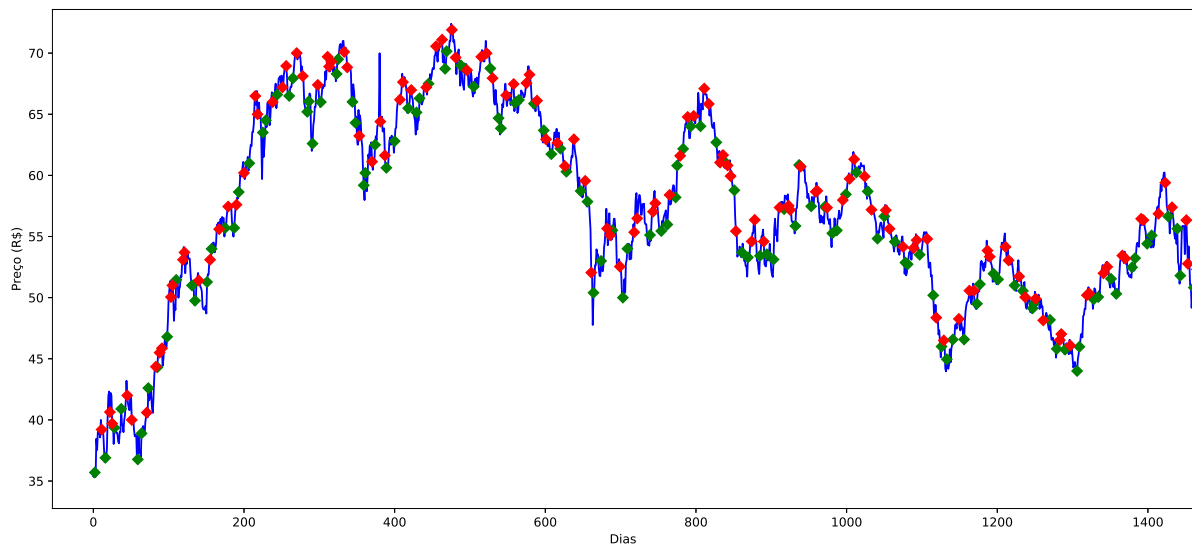
contador \leftarrow *fimJanela*

fim

fim

O método de rotulagem divide a série temporal em segmentos menores. Começando no primeiro dia, avança-se usando uma janela deslizante com passo de 11. Para cada segmento, os dias subsequentes após os valores mínimo e máximo no segmento atual são marcados como compra e venda, respectivamente, e os outros dias são marcados como neutro. Esse processo se repete até o final da série. Como resultado, os pontos de virada na tendência do preço são marcados como um ponto de entrada de negociação. A Figura 28 mostra os pontos de compra e venda gerados a partir do Algoritmo 1 na base de dados do BOVA11.

Figura 28 – Rótulos com pontos de compra e venda. Dias de Compra marcados em verde, Venda em vermelho e dias Neutro não estão marcados.



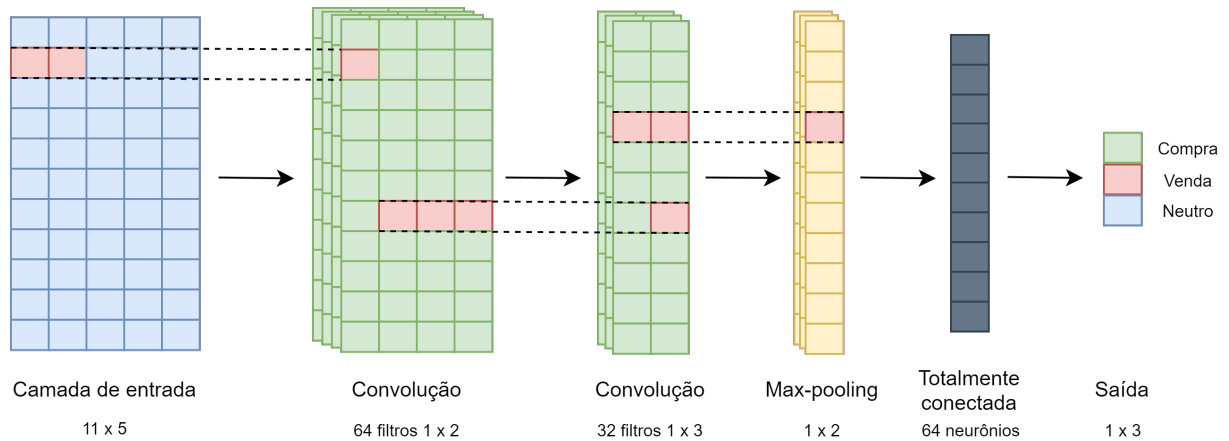
Fonte: o autor

4.5 Arquitetura da CNN utilizada no sistema

A arquitetura da rede CNN utilizada é apresentada na Figura 29. O modelo foi implementado utilizando a estrutura do Keras (CHOLLET et al., 2015) e a base das funções do TensorFlow (MARTÍN ABADI et al., 2015) utilizando a linguagem Python. A CNN está estruturada da seguinte forma: a entrada é uma matriz bidimensional de formato $[11 \times 5]$, conforme descrito anteriormente, duas camadas convolucionais (64 *kernels* de $[1 \times 2]$ e 32 *kernels* de $[1 \times 3]$), uma camada de *pooling* máximo (*pool-size* de $[1 \times 2]$), uma camada totalmente conectada (64 neurônios) e uma camada de saída de tamanho 3. Também foi utilizado um *Dropout* de 0,5 na rede antes das camadas totalmente conectadas.

Todas as camadas utilizam a função de ativação ReLU, e a camada de saída utiliza a função *Softmax*. Além disso, todas as camadas são definidas para utilizar a inicialização uniforme de pesos proposta por He et al. (2015), descrita na seção 3.3.4.7.

Figura 29 – Arquitetura da CNN utilizada no sistema red



Fonte: o autor

As camadas convolucionais capturam os padrões ao longo do tempo para cada recurso de entrada, as camadas de *pooling* máximo extraem recursos maiores e reduzem a dimensionalidade dos dados. As camadas totalmente conectadas realizam a classificação do sinal de negociação.

Todos os filtros convolucionais varrem apenas ao longo do eixo x da matriz. Uma vez que um *kernel* finaliza a varredura em uma linha, este passa para a próxima e reinicia a varredura desde o início. Isto é crucial para a abordagem pois todo este processo, no final, combina as características extraídas de cada dia de entrada de forma resumida, sendo possível a identificação de padrões mais destacados. Como resultado, a dimensão de recursos diminui de 5 para 1 no final deste processo.

A arquitetura final da CNN utilizada foi definida empiricamente, após extensos testes com vários hiper parâmetros e arranjos de camadas diferentes com o objetivo de melhorar a classificação das séries temporais com base nos trabalhos estudados, resultando no rede apresentada na Figura 29. Esta configuração foi escolhida pois foi a que apresentou os melhores resultados em termos estatísticos e de retorno financeiro durante os experimentos.

5 EXPERIMENTOS

Este capítulo apresentará os detalhes da metodologia de avaliação utilizada. A seção 5.1 apresenta os ativos selecionados para os experimentos. A seção 5.2 apresenta os modelos de RNA utilizados para comparação nos experimentos. Na sequência a seção 5.3 apresenta o processo de treinamento utilizado, seguida pela seção 5.5 que detalha as métricas estatísticas utilizadas na avaliação dos resultados. Por fim, as seções 5.6, 5.4 detalham a metodologia utilizada nas operações e o teste em ambiente real simulado.

5.1 Ativos utilizados nos experimentos

Os experimentos foram realizados inicialmente em ativos da Brasil, Bolsa, Balcão (B3), principalmente em ativos que compõe o Índice Bovespa (IBOVESPA), o qual consiste em uma carteira teórica composta pelas ações mais negociadas da B3, o que reflete o desempenho do mercado de ações brasileiro no geral.

Sendo assim, foram utilizados os preços diários de 6 ativos do mercado financeiro brasileiro (BR) (1 ETF e 5 ações) e 6 ativos do mercado norte americano (EUA) (2 ETFs e 4 ações). Os históricos foram obtidos pelo site do Yahoo Finance ¹. A Tabela 1 apresenta os ativos selecionados para os experimentos bem como os períodos de início e final de coleta dos dados utilizados.

Tabela 1 – Ativos utilizados nos experimentos

País	Código	Descrição	Setor	Início	Fim
BR	BOVA11	iShares Ibovespa Fundo de Índice	ETF do índice Ibovespa	01/12/2008	01/01/2019
BR	VALE3	Vale S.A.	Mineração	01/12/2008	01/01/2019
BR	PETR4	Petrobras S.A.	Petróleo	01/12/2008	01/01/2019
BR	ITUB4	Itaú Unibanco S.A.	Banco	01/12/2008	01/01/2019
BR	ABEV3	Ambev S.A.	Bebidas e alimentos	01/12/2008	01/01/2019
BR	WEGE3	Weg S.A.	Bens industriais	01/12/2008	01/01/2019
EUA	AAPL	Apple Inc.	Tecnologia	01/12/2008	01/01/2019
EUA	WFC	Wells Fargo & Co.	Banco	01/12/2008	01/01/2019
EUA	JPM	JPMorgan Chase & Co.	Banco	01/01/2000	01/01/2012
EUA	TRV	The Travelers Companies	Seguros	01/01/2001	01/01/2016
EUA	XLE	Energy Select Sector SPDR ETF	ETF setor de Energia	01/12/2000	01/01/2012
EUA	XLFX	Financial Select Sector SPDR ETF	ETF setor Financeiro	01/12/2001	01/01/2016

Fonte: o autor

¹<https://finance.yahoo.com>, acessado em Setembro de 2020.

Os ativos do mercado brasileiro foram escolhidos pelo seguinte critério: o BOVA11 pois este replica o portfólio do IBOVESPA; as ações por serem algumas das com maiores ponderações na carteira teórica do IBOVESPA do terceiro quadrimestre de 2020, as chamadas *blue chips*, e sendo uma de cada setor e com histórico suficiente para a avaliação.

Cada ação representava a seguinte ponderação no portfólio do IBOVESPA ²: VALE3 com 10,64% do setor de mineração; ITUB4 com 6,17%, sendo o maior banco privado do Brasil; PETR4 sendo 5,24%, a maior petrolífera brasileira; ABEV3 com 3,11%, do setor de bebidas e alimentos; e WEGE3 com 2,45%, sendo uma das maiores produtoras de máquinas elétricas, automação e tintas industriais.

No mercado norte americano, foram selecionadas uma das maiores empresas da bolsa americana de tecnologia (NASDAQ), a Apple Inc. (AAPL), e o banco Wells Fargo & Co. (WFC) por ser um dos maiores bancos americanos. As demais ações e ETFs foram utilizadas para comparação com o sistema CNN-TA, o qual realizou experimentos nas mesmas. Tais ativos foram JPM, uma empresa de serviços financeiros em gerais sediada nos EUA. TRV, uma empresa do setor de seguros, também sediada nos EUA. E por último, as ETFs XLE e XLF que de forma semelhante ao BOVA11, consistem em uma carteira teórica de ativos, no caso, dos setores de energia e financeiro, respectivamente.

5.2 Desenvolvimento dos Modelos Comparados

O CNN-TC foi comparado com um modelo de CNN semelhante e base deste trabalho, o CNN-TA (SEZER; OZBAYOGLU, 2018). Também foram desenvolvidos outros dois modelos de RNAs, uma rede MLP, a fim de comparar a classificação de operação utilizando um modelo de RNA mais simples, e uma LSTM, muito utilizada em análises de séries temporais, como mencionado no capítulo 2. Esta comparação tem por objetivo testar a eficácia desta abordagem de classificação de operação frente à utilização de modelos mais simples ou semelhantes em termos de métricas estatísticas e retorno financeiro das estratégias.

A implementação do CNN-TA foi a melhor possível de se obter dadas as informações fornecidas no artigo. Também foram feitos ajustes finos para os principais hiper parâmetros a fim de obter o melhor resultado possível deste sistema.

²http://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm, acessado em Setembro de 2020

Outras duas RNAs receberam os mesmos dados de entrada do CNN-TC e foram treinados com o mesmo método de rotulagem deste trabalho. O modelo MLP foi implementado com 3 camadas, tendo 110, 64 e 3 neurônios, respectivamente. O modelo LSTM foi implementado com 1 camada recorrente de 32 unidades e 2 camadas totalmente conectadas de 64 e 3 neurônios, respectivamente. Semelhantemente ao CNN-TC, todas as camadas utilizaram a função de ativação ReLu e a última utilizou *Softmax*.

5.3 Processo de Treinamento

Cada sessão de treinamento foi realizada com tamanho de lote de entrada de 128 amostras. Foi escolhido este tamanho de lote para que o treinamento da rede fosse mais lento a fim melhorar o aprendizado e que permitisse uma generalização mais eficiente; por se tratar de análise de série temporal, nem sempre há características especificamente diferentes o suficiente nas amostras para que a rede seja capaz de abstrair rapidamente.

Foi utilizado o algoritmo de otimização Adam (KINGMA; BA, 2014), com taxa de aprendizado de 1^{-4} . Este é um método de otimização que calcula taxas de aprendizado adaptáveis para cada parâmetro. Foi escolhido o algoritmo Adam pois conforme demonstrado na seção 3.3.4.6, este demonstra ser um algoritmo computacionalmente mais eficiente que seus antecessores, tais como Adadelta e Adagrad. Como métricas de treinamento foram utilizados acurácia e F1 ponderado.

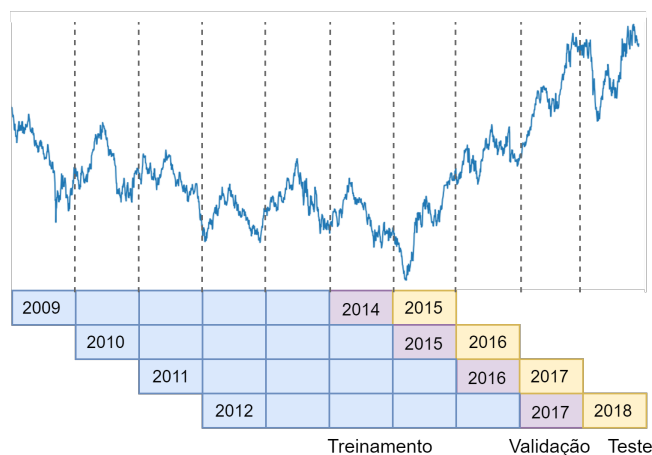
Para otimizar o treinamento e evitar o problema de sobreajuste, o recurso do Keras de parada antecipada foi utilizado, monitorando o decrescimento de erros no conjunto de validação. A quantidade de pontos de compra/venda são menos frequentes que os da classe neutro, ou seja, as classes são desbalanceadas. Sendo assim para melhorar o treinamento, foi utilizado o recurso de amostras ponderadas. O peso de cada amostra é calculado e utilizado para configurar o Keras para prestar menos atenção à classe neutro, a fim de identificar melhor as características mais das classes compra e venda.

As sessões de treinamento foram realizadas em janelas deslizantes com abordagem de reciclagem de conhecimento, conforme representado na Figura 30. Foi escolhida essa configuração para evitar o problema de diferentes comportamentos de mercado em diferentes períodos. A configuração de treinamento mais comum é baseada em janela estática, mas, historicamente, o comportamento do mercado de ações é composto por períodos de ascensão e queda ao longo

do tempo. Se treinarmos a rede apenas em um período de alta e testarmos em um período de queda, a classificação poderá ser afetada negativamente.

Esta metodologia de treinamento é baseada na que foi utilizada por Sezer e Ozbayoglu (2018). Durante o treinamento, foram retirados seis anos do conjunto de dados e divididos 20% para validação cruzada. O ano subsequente da amostras de dados foi utilizado para as avaliações estatística e financeira. Em seguida, foi avançado 1 ano, e o modelo foi retreinado e reavaliado novamente com os novos dados.

Figura 30 – Período de treinamento em azul, validação em roxo e de avaliação em amarelo



Fonte: o autor

5.4 Testes em Ambiente Simulado Utilizando o MetaTrader

Após os treinamentos dos modelos e os testes com dados históricos, foi realizado um teste em ambiente real de forma simulada a fim de se avaliar os resultados desta pesquisa aplicando o conhecimento adquirido pelo modelo em classificar as operações de negociação para cada dia na B3.

Para isto, foi utilizado o software MetaTrader 5 ³, uma plataforma de negociação comumente utilizada para negociação por meio de robôs automatizados. A integração do modelo com o MetaTrader foi feita utilizando sua Interface de Programação de Aplicações (Application Programming Interface) (API) com o Python ⁴. As operações foram realizadas em lotes de 10 quantidades, como descrito na seção 5.6.

³<https://www.metatrader5.com/pt>, acessado em 01/08/2020

⁴https://www.mql5.com/pt/docs/integration/python_metatrader5, acessado em 01/08/2020

O objetivo é validar a abordagem desenvolvida neste trabalho em um ambiente real, verificando sua eficácia e rentabilidade com dados vindos da plataforma. Foi feita coleta diária dos dados de mercado utilizando a API do MetaTrader 5 no horário de leilão de fechamento do mercado à vista da B3, executada a classificação e tomada a decisão, de forma semelhante ao método adotado na avaliação financeira com os dados históricos.

5.5 Avaliação Estatística

Devido às classes serem desbalanceadas, as métricas estatísticas para cada classe foram calculadas e o valor final de cada métrica consiste em uma média ponderada das 3 classes. Cada conjunto de dados foi avaliado separadamente, calculando as métricas acurácia, precisão, revocação e F1 ponderados. Para isto foi utilizada a biblioteca Scikit Learn do Python ⁵. A seguir é apresentado o método de cálculo base para cada métrica para melhor compreensão (POWERS, 2020), onde:

- a) VP: verdadeiro positivo, por exemplo é classificado compra em um dia que foi rotulado como compra na base de dados;
- b) VN: verdadeiro negativo, por exemplo não é classificado venda/neutro em um dia que foi rotulado como compra;
- c) FP: falso positivo, por exemplo é classificado compra em um dia que foi rotulado como venda na base de dados;
- d) FN: falso negativo, por exemplo a classificação deveria ser compra mas foi classificado venda;

5.5.1 Acurácia

Esta métrica indica um desempenho geral do modelo, levando em conta o número de predições corretas, dividida pelo total de predições realizadas. Resumidamente é mensurado: dentre todas as classificações, quantas o modelo classificou corretamente. O cálculo da mesma se dá pela Eq. (36).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (36)$$

⁵https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html, acessado em 01/11/2020

5.5.2 Precisão

A precisão mensura o quão preciso é o modelo em relação aos positivos previstos, quantos deles são realmente verdadeiros positivos. Em outras palavras, quantas classificações estão corretas. A fórmula é apresentada na Eq. (37)

$$Precisão = \frac{VP}{VP + FP} \quad (37)$$

5.5.3 Revocação

A métrica Revocação calcula quantos dos positivos reais o modelo foi capaz de identificar, rotulando-o como positivo (verdadeiro positivo). Este calculo é feito utilizando a fórmula na Eq. (38).

$$Revocação = \frac{VP}{VP + FN} \quad (38)$$

5.5.4 F1

O F1 consiste em uma média harmônica entre Precisão e Revocação a fim de mensurar a eficácia utilizando um único número. O cálculo é feito pela Eq. (39).

$$F1 = \frac{2 * Precisão * Revocação}{Precisão + Revocação} \quad (39)$$

5.6 Algoritmo de Avaliação Financeira

Foram simuladas operações de negociação nos conjuntos de dados de teste a cada sessão de treinamento de acordo com o Algoritmo 2. Esta avaliação se dá realizando as simulações de compra e venda no momento da classificação, utilizando o preço de fechamento do ativo.

O desempenho do CNN-TC foi comparado à estratégia BH e com os resultados obtidos utilizando os demais modelos. Vale ressaltar que a avaliação foi feita com a estratégia *long-only*, o que significa que busca ganho financeiro comprando o ativo e esperando que o seu preço aumente para vender.

Algoritmo 2 – Avaliação financeira

Entrada: preços, rótulos, tamanhoLote, valorTotal
Saída: históricoLucro

```

para  $i$  in tamanho(rótulos) faça
  se rótulos[ $i$ ] = "Compra" então
    se qtdAtual = 0 então
      podeComprar  $\leftarrow$  valorTotal/preços[ $i$ ]
      qtdComprado  $\leftarrow$  podeComprar – (podeComprar%tamanhoLote)
      valorComprado = preços[ $i$ ] * qtdComprado;
      qtdAtual  $\leftarrow$  qtdComprado
      caixa  $\leftarrow$  valorTotal – valorComprado
      históricoLucro[ $i$ ]  $\leftarrow$  valorComprado + caixa
    fim
  senão
    totalAtual  $\leftarrow$  qtdAtual * preços[ $i$ ]
    históricoLucro[ $i$ ]  $\leftarrow$  totalAtual + caixa
  fim
fim
senão se rótulos[ $i$ ] = "Venda" então
  se qtdAtual > 0 então
    totalAtual  $\leftarrow$  qtdAtual * preços[ $i$ ]
    valorTotal  $\leftarrow$  totalAtual + caixa
    caixa  $\leftarrow$  0
    qtdAtual  $\leftarrow$  0
    históricoLucro[ $i$ ]  $\leftarrow$  valorTotal
  fim
  senão
    históricoLucro[ $i$ ]  $\leftarrow$  valorTotal
  fim
fim
senão
  se qtdAtual > 0 então
    totalAtual  $\leftarrow$  quantidade * preços[ $i$ ]
    históricoLucro[ $i$ ]  $\leftarrow$  totalAtual
  fim
  senão
    históricoLucro[ $i$ ]  $\leftarrow$  valorTotal
  fim
fim
fim

```

No mercado de financeiro brasileiro, o lote padrão para compra de ações é de múltiplos de 100 e para ETFs é múltiplos de 10, enquanto no mercado norte americano não existe lote mínimo, podendo até comprar frações de ação em algumas corretoras; nos experimentos foi adotada a mesma quantidade utilizada no mercado brasileiro. Além disso, para as ações do mercado financeiro americano não foi considerada a variação cambial.

As operações foram executadas de acordo com a quantidade de lote descrita acima: no primeiro sinal de negociação, a classificação de ação obtida para o dia atual é tomada. Se o for neutro, nenhuma ação será tomada. Se for classificado como compra, será comprado no fechamento daquele dia com todo o valor disponível em conta (isso quando ainda não tiver sido comprado), o restante irá para o caixa. Se for venda, todos os ativos serão vendidos (quando ainda não estiverem vendidos). O valor inicial da avaliação foi de R\$ 10.000.

Além disso, a título de comparação dos resultados financeiros frente a investimentos de renda fixa, que são considerados mais simples pela maioria dos investidores, foi calculado o retorno do Certificado de Depósito Interbancário (CDI) acumulado nos períodos de simulação.

O CDI ⁶ consiste em um título emitido pelos bancos para a realização de operações de empréstimo entre si, em uma modalidade de curtíssimo prazo. Da mesma forma, os investidores podem emprestar o seu dinheiro para as instituições financeiras, e a mesma devolve o valor corrigido pela taxa acordada no ato do empréstimo na data de vencimento do título. Geralmente estes títulos são indexados com base em uma porcentagem do CDI. Para esta comparação foi considerada uma taxa de 100% do CDI.

Os cálculos foram realizados utilizando uma calculadora disponibilizada pelo Banco Central do Brasil (BCB) ⁷. A Figura 31 apresenta o cálculo do CDI no primeiro período de avaliação financeiro, de 02/01/2015 a 01/01/2019. Na Figura 32 é apresentado o resultado o retorno do CDI entre 01/03/2006 e 01/01/2012. Por último, a Figura 33 apresenta o valor de 01/03/2007 à 01/01/2016.

Figura 31 – Retorno do CDI entre 02/01/2015 e 01/01/2019

Resultado da Correção pelo CDI	
<small>*A data informada não é dia útil, a data utilizada para este cálculo refere-se ao primeiro dia útil subsequente</small>	
Dados básicos da correção pelo CDI	
Dados informados	
Data inicial	02/02/2015*
Data final	02/01/2019*
Valor nominal	R\$ 10.000,00 (REAL)
% do CDI	100,00 %
Dados calculados	
Índice de correção no período	1,49626230
Valor percentual correspondente	49,626230 %
Valor corrigido na data final	R\$ 14.962,62 (REAL)

Fonte: o autor

⁶<https://www.sunoresearch.com.br/artigos/cdi/>, acessado em 05/11/2020

⁷<https://www3.bcb.gov.br/CALCIDADA0/publico/exibirFormCorrecaoValores.do?method=exibirFormCorrecaoValores&aba=5>, acessado em 05/11/2020

Figura 32 – Retorno do CDI entre 01/03/2006 e 01/01/2012

Resultado da Correção pelo CDI	
*A data informada não é dia útil, a data utilizada para este cálculo refere-se ao primeiro dia útil subsequente	
Dados básicos da correção pelo CDI	
Dados informados	
Data inicial	01/03/2006
Data final	02/01/2012*
Valor nominal	R\$ 10.000,00 (REAL)
% do CDI	100,00 %
Dados calculados	
Índice de correção no período	1,89618573
Valor percentual correspondente	89,618573 %
Valor corrigido na data final	R\$ 18.961,86 (REAL)

Fonte: o autor

Figura 33 – Retorno do CDI entre 01/03/2007 e 01/01/2016

Resultado da Correção pelo CDI	
*A data informada não é dia útil, a data utilizada para este cálculo refere-se ao primeiro dia útil subsequente	
Dados básicos da correção pelo CDI	
Dados informados	
Data inicial	01/03/2007
Data final	04/01/2016*
Valor nominal	R\$ 10.000,00 (REAL)
% do CDI	100,00 %
Dados calculados	
Índice de correção no período	2,43795887
Valor percentual correspondente	143,795887 %
Valor corrigido na data final	R\$ 24.379,59 (REAL)

Fonte: o autor

6 RESULTADOS

Neste capítulo serão apresentadas e comentadas as métricas obtidas em cada ativo para cada estratégia, e na sequência serão plotados os desempenhos financeiros de cada estratégia.

Cada ativo foi avaliado separadamente, utilizando acurácia, precisão, revocação e F1 ponderados, métricas comentadas na seção 5.5, a fim de mensurar a consistência das métricas estatísticas em diversas condições nos dados históricos. Os resultados apresentados são referentes a todo o período utilizado para a avaliação financeira.

6.1 BOVA11

A Tabela 2 apresenta as métricas estatísticas para o ativo BOVA11. Como pode ser visto, a acurácia do sistema CNN-TC apresentou-se consideravelmente melhor que os demais modelos, com uma vantagem de 13,35% para o segundo melhor nesta métrica. Em termos de precisão, pode-se notar que os sistemas CNN-TC e CNN-TA alcançaram taxas próximas de 80%, o que indica que estas abordagens foram capazes de identificar características específicas das classes na maioria dos casos. Baseando-se no F1 pode-se notar uma clara vantagem do CNN-TC em relação aos demais modelos para os experimentos realizados.

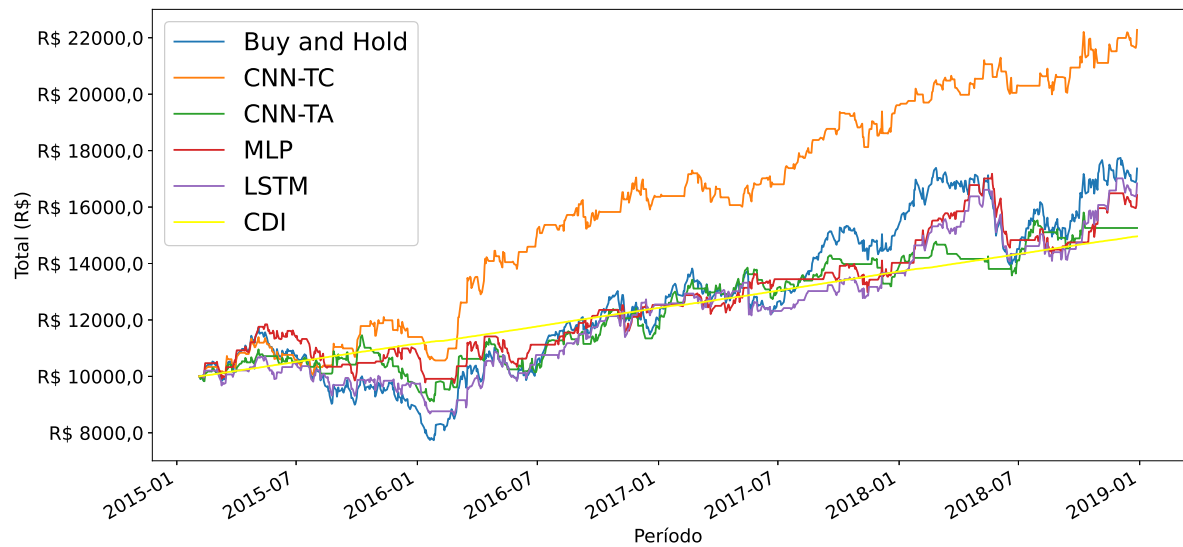
Tabela 2 – Métricas estatísticas para BOVA11

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	67,57%	80,45%	67,57%	71,30%
CNN-TA	45,28%	81,29%	45,28%	54,17%
MLP	54,22%	73,55%	54,22%	60,16%
LSTM	52,66%	72,63%	52,66%	58,20%

Fonte: o autor

Na Figura 34 é apresentado o retorno financeiro para cada estratégia. Avaliando os desempenhos obtidos, pode-se observar que o CNN-TC alcançou resultados bem superiores às demais estratégias, inclusive superando o BH. Considerando o retorno do CDI, que foi de 49,62%, pode-se observar também que todos os modelos obtiveram retornos acima deste *benchmark*.

Figura 34 – Retorno sobre investimento para BOVA11



Fonte: o autor

6.2 VALE3

Na Tabela 3 tem-se as métricas para o ativo VALE3. A precisão das abordagens baseadas em CNN foram próximas de 80%, porém a acurácia evidencia que o estratégia CNN-TC foi o que obteve maior número de classificações corretas. Já o CNN-TA, apesar de precisão elevada, a menor acurácia e o seu método de rotulagem não corroboraram, fazendo com que fossem realizadas vendas de forma antecipada e prejudicando o seu resultado financeiro.

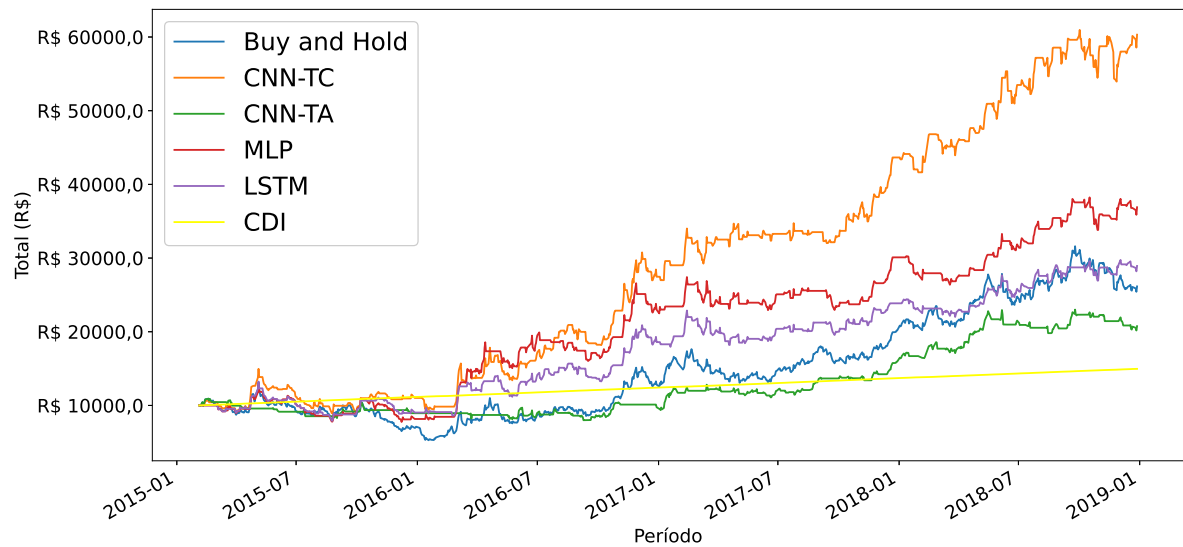
Tabela 3 – Métricas estatísticas para VALE3

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	65,28%	79,60%	65,28%	69,19%
CNN-TA	56,92%	81,01%	56,92%	59,91%
MLP	61,63%	77,65%	61,63%	66,26%
LSTM	62,25%	76,42%	62,25%	66,75%

Fonte: o autor

Na Figura 35 é possível visualizar que embora os modelos MLP e LSTM tenham sido inferiores nas métricas supradescritas, estes também obtiveram resultados financeiros satisfatórios, superando a estratégia utilizando o CNN-TA, bem como o BH. Além disso todas as RNAs superaram o CDI acumulado do período de avaliação. Portanto pode-se concluir que o método de rotulagem proposto neste trabalho corroborou para melhores resultados financeiros dos modelos comparados ao CNN-TA.

Figura 35 – Retorno sobre investimento para VALE3



Fonte: o autor

6.3 PETR4

Na Tabela 4 são apresentadas as métricas para o ativo PETR4. Para este ativo é evidente a vantagem do sistema CNN-TA, superando os demais em todas as métricas. O F1 do CNN-TA, que consiste em um número com o objetivo de ser uma média harmônica entre precisão e revocação, foi consideravelmente superior aos demais modelos, com uma vantagem de 6,59% sobre o CNN-TC. Vale destacar que o ativo apresentou movimentos abruptos no período de avaliação, podendo ter comprometido o desempenho dos modelos.

Tabela 4 – Métricas estatísticas para PETR4

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	57,87%	74,07%	57,87%	63,02%
CNN-TA	61,85%	81,21%	63,85%	69,61%
MLP	48,80%	70,28%	48,80%	55,23%
LSTM	50,89%	71,70%	50,89%	56,71%

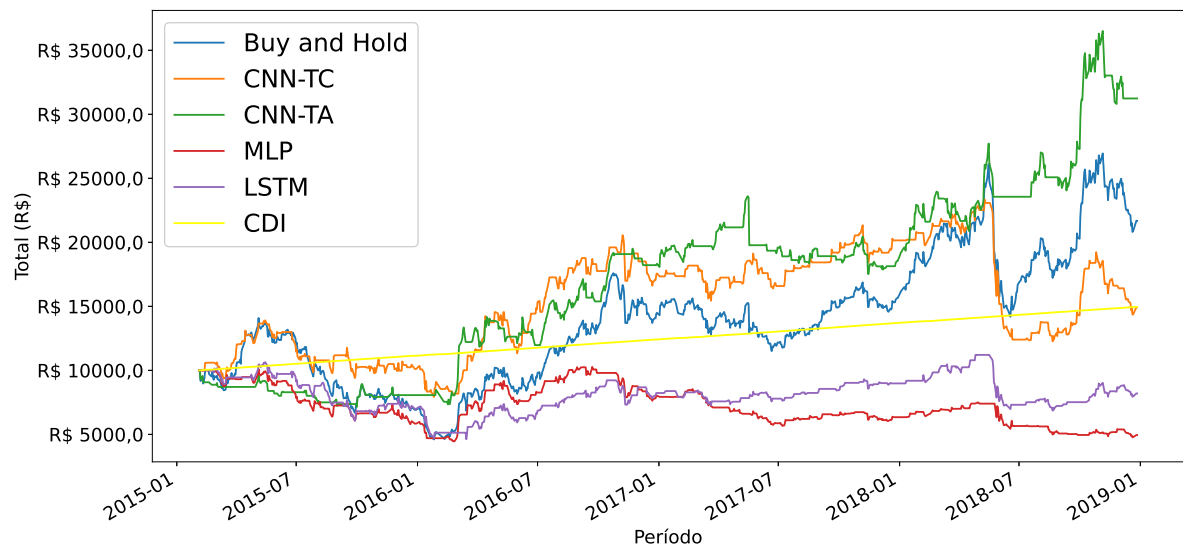
Fonte: o autor

Embora as métricas apresentadas na tabela acima tenham sido consideravelmente satisfatórias, na Figura 36 é possível visualizar que o retorno financeiro das estratégias MLP e LSTM não alcançaram resultados tão bons, apresentando prejuízos. O comportamento de ambos dos modelos aparenta ter sido influenciado negativamente por notícias relacionadas à empresa no

período analisado. Nota-se que nos momentos de alta, ambos os modelos não acompanham, realizando venda de forma antecipada.

Os sistemas CNN-TC e CNN-TA apresentaram comportamento próximos na maior parte do tempo, entretanto o CNN-TA foi capaz de antecipar a maior queda que ocorreu durante o período de avaliação e obteve assim um retorno financeiro maior. Pode-se então concluir que o método de rotulagem usado pelo CNN-TA foi o mais adequado para este ativo.

Figura 36 – Retorno sobre investimento para PETR4



Fonte: o autor

6.4 ITUB4

As métricas de ITUB4 são apresentadas na Tabela 5. Como se pode observar, neste ativo o sistema CNN-TC, proposto neste trabalho, alcançou resultados melhores em quase todas as métricas. Apesar de ter perdido em precisão para o CNN-TA, a acurácia do CNN-TC, que consiste em uma métrica a fim de mensurar o desempenho global do modelo, foi superior, evidenciando que este sistema acertou mais as classificações, no geral, o que ajudou o F1 a obter resultado estatístico superior.

Na Figura 34 é apresentado o gráfico do retorno financeiro das estratégias avaliadas. A tendência de alta no preço ao longo do período é evidente e muito semelhante ao comportamento do BOVA11. Pode-se observar que o CNN-TC alcançou resultados superiores, sendo capaz de antecipar os momentos de queda, garantindo maior rentabilidade. Todas as abor-

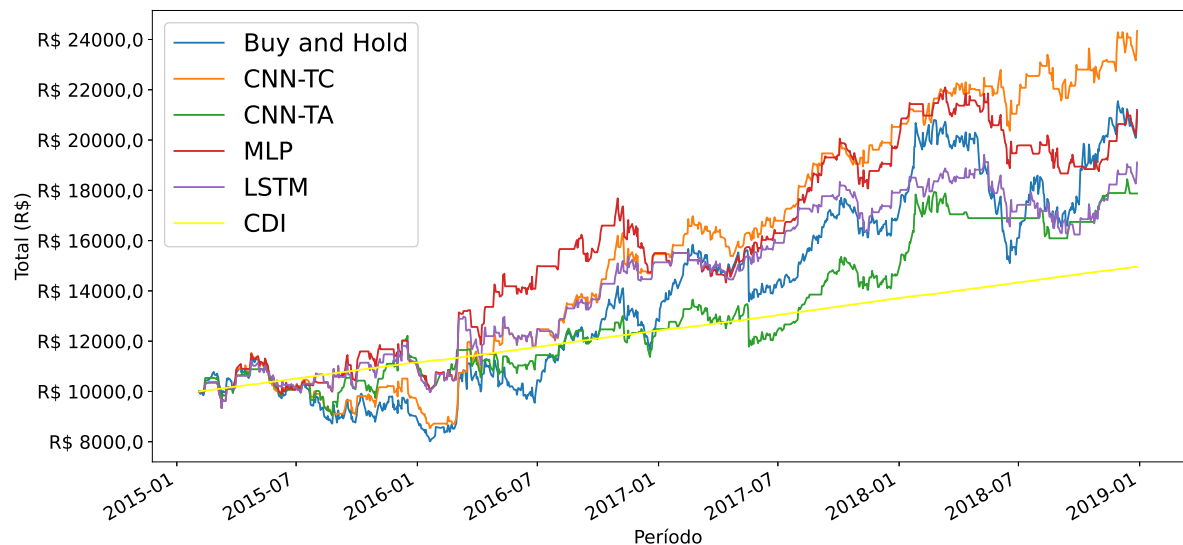
dagens superaram a rentabilidade do CDI, que foi de aproximadamente 49%. Entretanto, as estratégias CNN-TA e LSTM não superaram a estratégia BH. Conforme demonstrado na tabela de métricas, pode-se concluir que o método de rotulagem proposto neste trabalho, em conjunto com melhores métricas corroborou com os resultados do sistema CNN-TC e demais modelos que utilizaram este método, evidenciando sua capacidade de aproveitar tendência de alta nos preços.

Tabela 5 – Métricas estatísticas para ITUB4

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	63,09%	78,52%	63,09%	67,35%
CNN-TA	59,53%	80,41%	59,53%	66,88%
MLP	61,94%	77,42%	61,94%	66,66%
LSTM	60,33%	77,02%	60,33%	65,72%

Fonte: o autor

Figura 37 – Retorno sobre investimento para ITUB4



Fonte: o autor

6.5 ABEV3

A Tabela 6 apresenta as métricas para o ativo ABEV3. Para este ativo os sistemas utilizando CNN novamente obtiveram melhores resultados de classificação frente aos demais. Apesar de precisão levemente superior do CNN-TA, no geral o CNN-TC foi melhor, o que fica evidenciado pela vantagem em termos de F1. Acredita-se que por se tratar de um ativo que

o seu preço oscilou em uma faixa de preço sem grandes tendências de alta ou baixa possa ter influenciado nos resultados estatísticos dos demais modelos, e ressalta a capacidade das CNNs em identificar padrões de negociação.

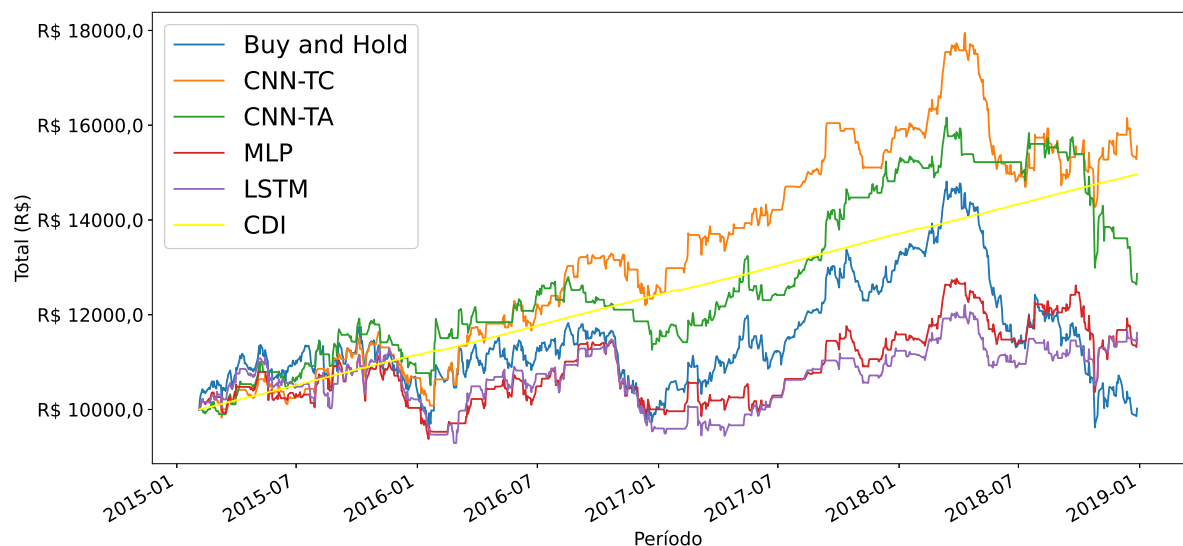
Tabela 6 – Métricas estatísticas para ABEV3

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	68,30%	78,93%	68,30%	71,56%
CNN-TA	60,45%	79,45%	60,45%	69,96%
MLP	57,14%	72,27%	57,14%	62,32%
LSTM	57,46%	73,45%	57,46%	62,54%

Fonte: o autor

Na Figura 38 pode-se visualizar o desempenho das estratégias de negociação para ABEV3. Apesar do ativo não ter apresentado tendência clara ao longo do período de avaliação, os sistemas baseados em CNN foram capazes de alcançar resultados financeiros acima dos demais. Todos as estratégias, no final, superaram a estratégia BH, que apresentou um leve prejuízo no período avaliado. Somente o sistema CNN-TC foi capaz de superar o CDI.

Figura 38 – Retorno sobre investimento para ABEV3



Fonte: o autor

6.6 WEGE3

Na Tabela 7 pode-se observar as métricas para o ativo WEGE3. As acurácias dos sistemas de CNN-TC e LSTM foram superiores. Pode-se notar também que em termos de precisão

que todos os modelos que utilizaram o método de rotulagem proposto neste trabalho apresentaram taxas superiores a 75%, o que indica que na maioria dos casos, a classificação de verdadeiros positivos foi bem sucedida. Pode-se concluir, em termos gerais, que a abordagem que foi menos eficaz nas classificações foi o CNN-TA.

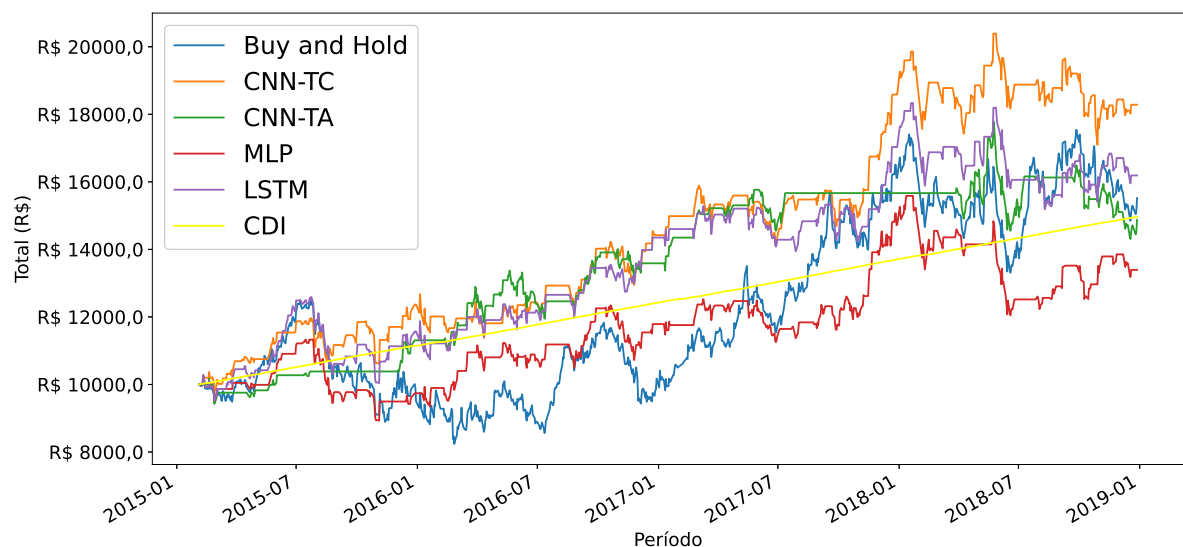
Tabela 7 – Métricas estatísticas para WEGE3

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	58,71%	78,21%	58,71%	63,56%
CNN-TA	51,40%	69,93%	51,40%	58,34%
MLP	54,89%	76,67%	54,89%	60,62%
LSTM	55,58%	76,21%	55,58%	61,30%

Fonte: o autor

Como se pode observar na Figura 39, os resultados do sistemas CNN-TC foram superiores. Os sistemas CNN-TC e LSTM apresentaram métricas próximas, o que também é refletido no resultado financeiro onde ambos tiveram desempenho próximos na maior parte do tempo. O CNN-TA também acompanhou a rentabilidade do CNN-TC, porém chegando ao final do período de avaliação, o mesmo não foi capaz de identificar padrões de negociações rentáveis. É possível notar que os períodos de alta no preço não são acompanhados pela maioria das estratégias, principalmente o MLP, onde é previsto venda de forma antecipada, interferido no retorno financeiro das estratégias.

Figura 39 – Retorno sobre investimento para WEGE3



Fonte: o autor

6.7 AAPL

A Tabela 8 apresenta as métricas para o ativo AAPL. Como se pode observar, as taxas de acurácia dos modelos não foram tão boas quanto os demais ativos até aqui comentados, ficando na maioria próximas de 50%. Porém vale destacar que em termos de precisão, o CNN-TC novamente se destacou nesta métrica, o que evidencia sua melhor capacidade de identificação de pontos operação corretos, o que consequentemente corroborou para melhor taxa de F1.

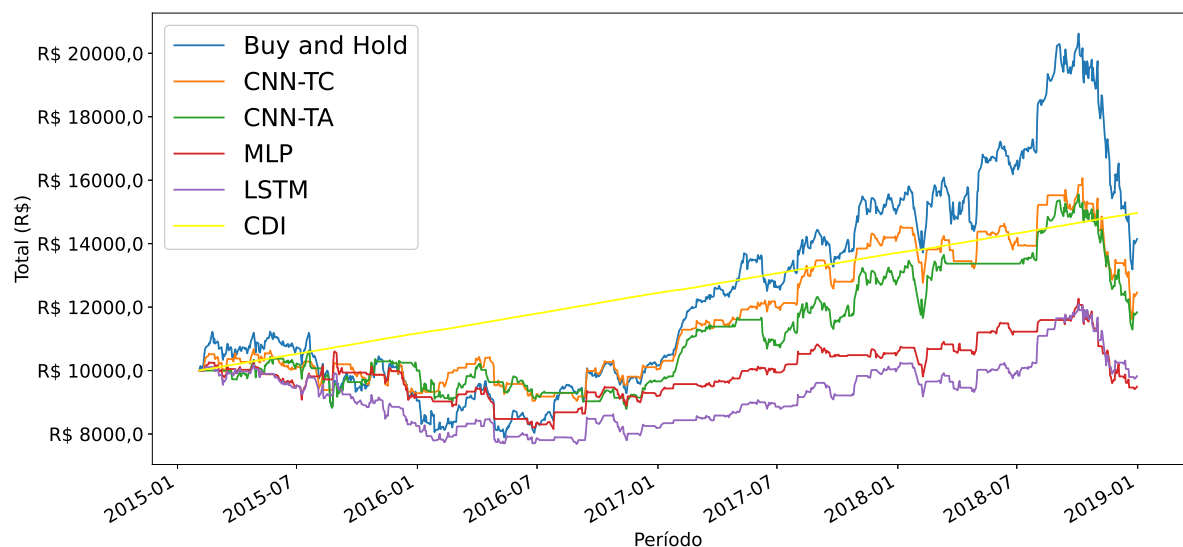
Tabela 8 – Métricas estatísticas para AAPL

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	53,69%	78,80%	53,69%	59,03%
CNN-TA	52,23%	72,56%	52,23%	58,74%
MLP	51,74%	76,44%	51,74%	57,87%
LSTM	50,41%	75,76%	50,41%	56,58%

Fonte: o autor

Como é possível observar na Figura 40, as abordagens usando CNN foram os que mais se aproximaram do retorno obtido pela estratégia BH. Entretanto os resultados não superaram esta estratégia nem o retorno do CDI ao longo do período. Assim como comentado na ação WEGE3, neste caso, os períodos de alta do preço não foram totalmente acompanhados pelos modelos, onde foi prevista venda antecipada impactando nos resultados.

Figura 40 – Retorno sobre investimento para AAPL



Fonte: o autor

Vale também ressaltar que tais movimentos no ativo aconteceram de forma exponencial, o que impacta diretamente nos cálculos dos indicadores levando-os rapidamente para região de sobrecompra, desta forma, o algoritmo é induzido a entender como um possível topo do preço.

6.8 WFC

A seguir são apresentadas as métricas para o ativo WFC. O comportamento do preço desta ação oscilou em uma faixa de valores sem uma tendência evidente. Entretanto, neste ativo a estratégia que se destacou nas métricas foi a LSTM. Apesar das CNNs terem apresentado resultados próximos, nota-se que o CNN-TC alcançou taxa de F1 superior.

Tabela 9 – Métricas estatísticas para WFC

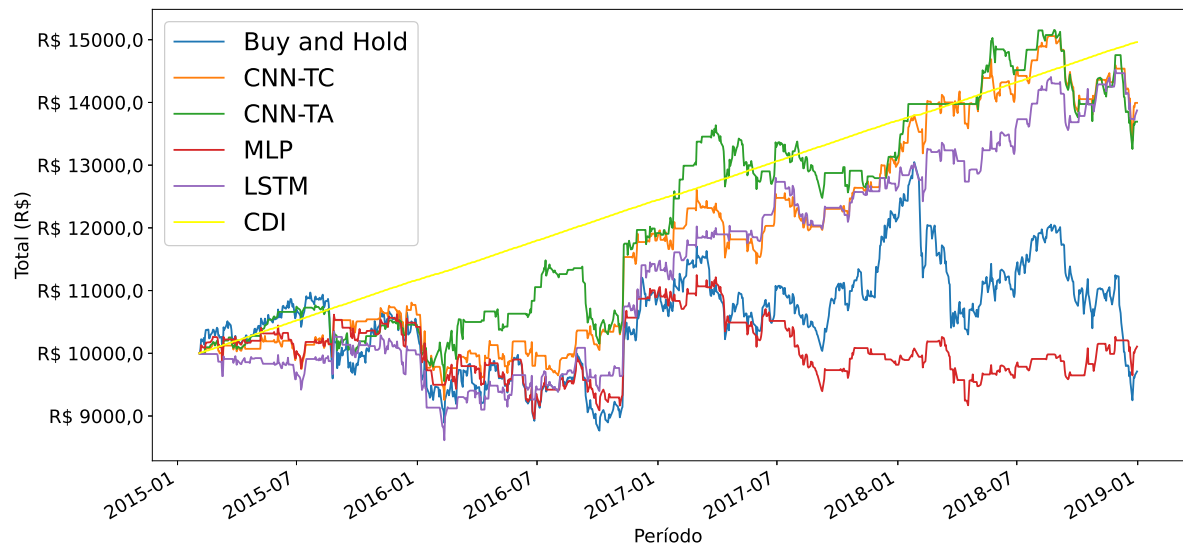
Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	63,32%	78,61%	63,32%	67,52%
CNN-TA	60,33%	79,10%	60,33%	65,78%
MLP	53,69%	75,72%	53,69%	60,05%
LSTM	66,19%	80,18%	66,19%	70,02%

Fonte: o autor

Apesar desta ação não ter apresentado uma tendência altista ou baixista, como se pode observar na Figura 41, todas as abordagens, exceto o MLP, obtiveram bons resultados financeiros. Assim como comentado sobre as métricas estatísticas, é possível notar um comportamento semelhante os sistemas CNN-TC, CNN-TA e LSTM. Já o MLP não foi capaz de acompanhar os períodos de alta no preço, impactando negativamente seu resultado, o que evidencia a superioridade dos demais modelos, principalmente CNN, na classificação dos pontos de operação neste ativo.

Pode-se também destacar que os 2 métodos de rotulagem utilizados demonstraram-se eficazes para o treinamento dos sistemas, o que fica evidenciado pelo comportamento CNN-TC, CNN-TA e LSTM. Nota-se que o CNN-TA chegou próximo do resultado financeiro dos demais, o que demonstra que o seu método e rotulagem é bem aplicável para o comportamento deste ativo. Por último, vale comentar que apesar do CNN-TA ter superado o CDI em 2 momentos, todas as estratégias obtiveram retornos abaixo do CDI no final.

Figura 41 – Retorno sobre investimento para WFC



Fonte: o autor

6.9 JPM

A Tabela 10 apresenta as métricas para o ativo JPM. E termos de acurácia e precisão, os sistemas CNN-TC e LSTM foram superiores frente aos demais. Quando analisa-se as taxas de F1, o MLP foi o mais inferior; enquanto o LSTM obteve resultado estatístico, no geral, superior, seguido pelo CNN-TC.

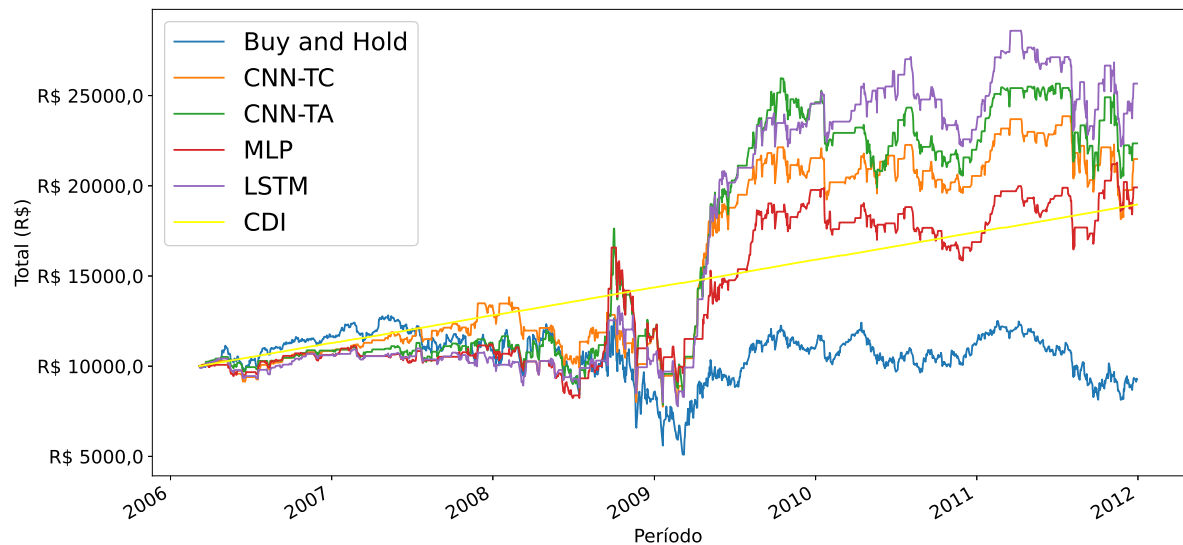
Tabela 10 – Métricas estatísticas para JPM

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	56,51%	79,17%	56,51%	61,28%
CNN-TA	53,41%	76,63%	53,41%	60,53%
MLP	53,14%	75,53%	53,14%	58,60%
LSTM	58,86%	80,49%	59,86%	63,48%

Fonte: o autor

Durante o período de avaliação ocorreu uma forte tendência baixista entre 2008/2009 conforme ser visto no gráfico da Figura 42 onde o BH apresentou um retorno de quase -50%. No entanto, os sistemas oscilaram entre perdas e ganhos próximos do valor inicial. No restante do período, a ação não apresentou uma tendência primaria clara. Apesar dos pontos comentados acima sobre as métricas e a crise financeira ocorrida durante o período de avaliação, todas as estratégias obtiveram bons resultados financeiros, tendo destaque para a LSTM e CNN-TA. Além disso, todos os sistemas superaram o CDI do período de avaliação.

Figura 42 – Retorno sobre investimento para JPM



Fonte: o autor

Apesar das taxas de acurácia não tenham sido totalmente satisfatórias, sendo próximas de 55%, os resultados financeiros não foram impactados negativamente. O CNN-TA alcançou um resultado financeiro próximo dos demais, mesmo tendo apresentado métricas próximas ao MLP. Embora possa-se visualizar um comportamento semelhante dos ativos durante certo período, evidenciando eficácia de ambos os métodos de rotulagem para treinar os modelos, pode-se concluir que o método de rotulagem do CNN-TA mostrou-se mais adequado para este ativo, pois mesmo com métricas inferiores este foi capaz de alcançar resultados próximos. Caso o sistema mostra-se mais preciso, poderia ter alcançado resultado melhor.

Acredita-se que a crise financeira ocorrida durante o período avaliado tenha impactado negativamente a capacidade dos modelos de identificarem tendências e padrões que pudessem fornecer pontos de operação mais rentáveis em alguns momentos. Após o período da crise é possível observar que os modelos foram capazes de acompanhar o movimento do preço com maior precisão.

6.10 TRV

A Tabela 11 apresenta as métricas para o ativo TRV. Neste ativo todos os sistemas apresentaram taxas de precisão elevadas, próximas de 77%, onde o CNN-TA obteve o melhor resultado. Entretanto a melhor acurácia contribuiu para um melhor resultado do CNN-TC em

termos de F1 dentre todos, com uma vantagem de 1,28% frente ao CNN-TA, evidenciando que o CNN-TC foi capaz de realizar mais classificações corretas, no geral.

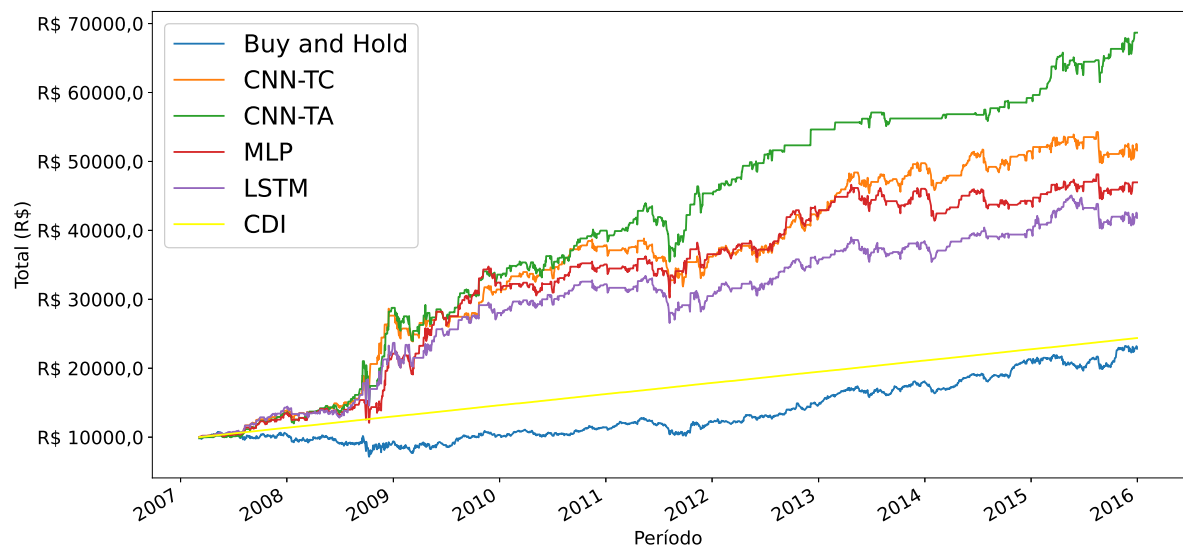
Tabela 11 – Métricas estatísticas para TRV

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	61,16%	77,25%	61,16%	65,73%
CNN-TA	56,77%	78,71%	56,77%	64,45%
MLP	59,35%	76,47%	59,35%	64,22%
LSTM	58,57%	75,84%	58,57%	63,65%

Fonte: o autor

Na Figura 43 é possível observar que todos as estratégias apresentaram resultados acima do BH e do CDI acumulado do período. Neste ativo o sistema CNN-TA apresentou o melhor retorno financeiro, seguido pelo CNN-TC. Devido às métricas de ambos não haverem apresentado valores muito distantes, pode-se concluir que o método de rotulagem utilizado para o CNN-TA auxiliou no retorno financeiro acima dos demais.

Figura 43 – Retorno sobre investimento para TRV



Fonte: o autor

6.11 XLE

Na Tabela 12 a seguir são apresentadas as métricas para a ETF XLE. Como se pode observar, os sistemas CNN-TC e LSTM apresentaram as melhores métricas, com taxas de acu-

rácia e precisão próximas de 55% e 80%, respectivamente. Na sequência o CNN-TA apresentou precisão próxima do LSTM, com 76,52% e resultado de 59,71% para F1.

Tabela 12 – Métricas estatísticas para XLE

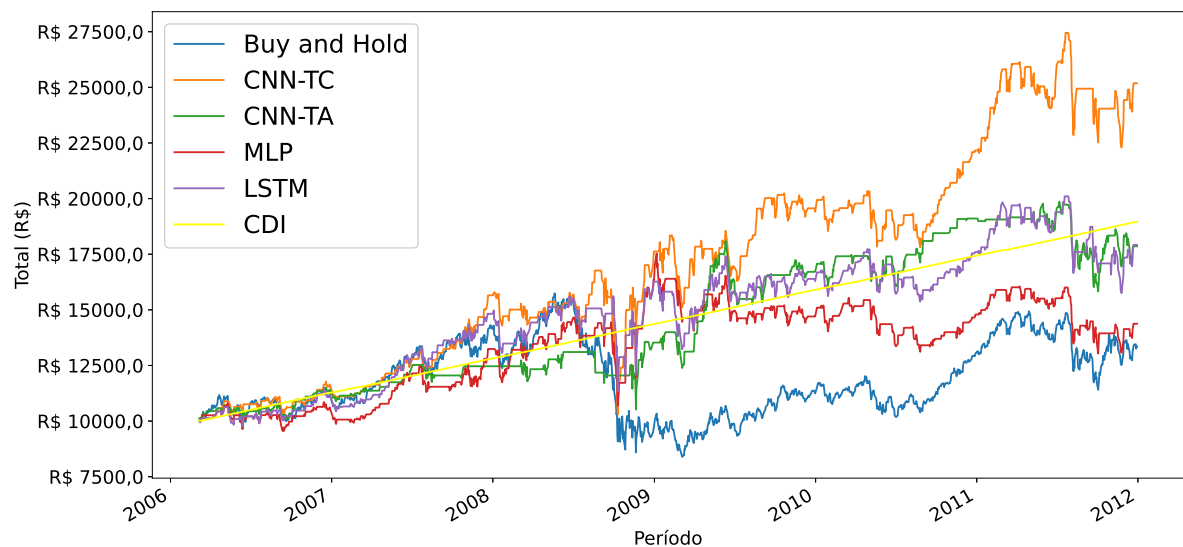
Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	56,53%	78,16%	56,53%	61,39%
CNN-TA	52,79%	76,52%	52,79%	59,71%
MLP	52,58%	75,65%	52,58%	58,10%
LSTM	54,45%	77,33%	54,45%	59,81%

Fonte: o autor

Nesta ETF o melhor resultado financeiro apresentado foi o do sistema CNN-TC, conforme ilustrado na Figura 44. É possível observar também que os sistemas CNN-TA e LSTM apresentaram retornos próximos, porém abaixo do retorno do CDI. Apesar de ter apresentado um retorno expressivo frente à estratégia BH, o MLP não se distanciou desta estratégia no final do período de avaliação.

Ademais, embora as métricas para o sistema CNN-TA tenham sido consideravelmente boas, isso não se refletiu em retorno financeiro. Desta forma, pode-se concluir que o método de rotulagem proposto para o CNN-TC bem como as métricas superiores foram responsáveis pelo elevado retorno financeiro frente às demais abordagens.

Figura 44 – Retorno sobre investimento para XLE



Fonte: o autor

6.12 XLF

Por fim, as métricas para a ETF XLF são demonstradas na Tabela 13. No período avaliado o preço deste ativo apresentou uma tendência de queda seguido de uma tendência de alta íngreme após o ano de 2012. Em termos de acurácia a maioria dos sistemas apresentaram resultados semelhantes (em torno de 55%). Em acurácia e precisão, o CNN-TC destacou-se frente aos demais, seguido pelo LSTM. Pode-se visualizar pela métrica F1 que o CNN-TC foi o sistema superior, o que é refletido também no retorno financeiro, principalmente quando comparado com o MLP e LSTM, treinados com o mesmo método de rotulagem, que apesar de métricas relativamente próximas, pode-se notar uma diferença no retorno deste entre 2009 e 2012.

Tabela 13 – Métricas estatísticas para XLF

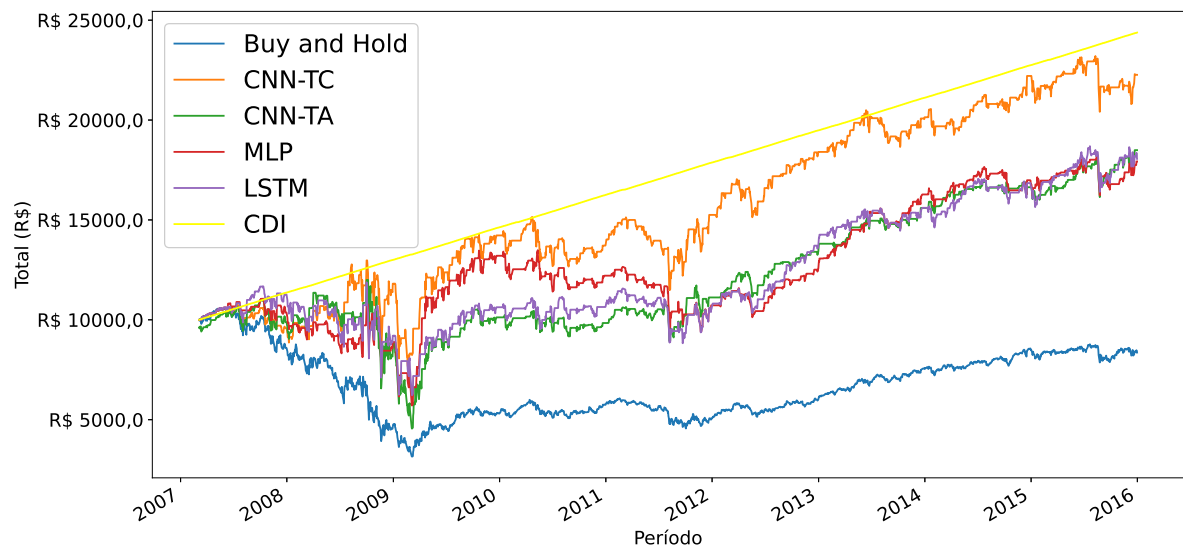
Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	60,81%	79,21%	60,81%	65,15%
CNN-TA	55,64%	71,41%	55,64%	61,01%
MLP	56,49%	75,84%	56,49%	61,55%
LSTM	55,89%	76,16%	55,89%	60,88%

Fonte: o autor

Na Figura 45 é apresentado o retorno o retorno financeiro deste ativo. Apesar de superarem a estratégia BH, nenhuma das estratégias superou o retorno do CDI. Isso pode ser explicado devido à tendência de baixa entre 2007 e 2009, que impactou no retorno dos sistemas durante parte do período de avaliação. Apesar disto, os sistemas oscilaram entre perdas e ganhos acima do retorno do BH.

O diferencial se deu no início da tendência de alta que se iniciou logo após o ano de 2009, onde o CNN-TC descolou-se de forma mais evidente dos demais. O MLP comportou-se de forma semelhante em parte deste período de recuperação. No restante do período as estratégias demonstraram retornos financeiros muito próximos, exceto pelo CNN-TC que foi capaz de identificar com melhor precisão os topos e fundos do preço. Por último, pode-se concluir que o método de rotulagem alinhado com as métricas estatísticas apresentadas pelo CNN-TC proporcionaram-lhe o melhor desempenho financeiro para este ativo.

Figura 45 – Retorno sobre investimento para XLF



Fonte: o autor

6.13 Custos e Emolumentos

A B3 (Bolsa Brasil Balcão) é a empresa responsável por gerenciar as operações do mercado financeiro brasileiro. Esta surgiu em 2017 como resultado da fusão entre a antiga BM&F Bovespa (Bolsa de Valores, Mercadorias e Futuros de São Paulo) e a Cetip (Central de Custódia e de Liquidação Financeira de Títulos) ¹.

Neste trabalho não foram consideradas taxas de corretagem, pois atualmente existem corretoras que oferecem planos sem corretagem, o que não impacta o resultado financeiro em nenhuma das estratégias.

Para as operações na bolsa brasileira, é cobrada uma taxa para todas as ordens executadas, chamada Emolumentos. Esta é subtraída de uma porcentagem do volume financeiro total negociado. A tabela 14 apresenta as taxas de operação para pessoas físicas e fundos de investimento ².

Para a estratégia BH, os Emolumentos são irrisórios. Mas para sistemas que realizam operações com maior frequência seria injusta uma comparação sem considerar tais taxas. Para isto foi calculado o desconto médio por execução para verificar se é cabível não desconsiderar tais taxas da B3.

¹<https://www.sunoresearch.com.br/artigos/b3/>

²http://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/listados-a-vista-e-derivativos/renda-variavel/tarifas-de-acoes-e-fundos-de-investimento/a-vista/

Tabela 14 – Taxas de emolumentos B3 para negociação e liquidação

Investidor	Negociação	Liquidação	Total
Pessoas físicas e demais investidores	0,003247%	0,0275%	0,030747%
Fundos e clubes de investimento locais	0,003247%	0,0200%	0,023247%

Os sistemas executam de 60 a 100 operações anualmente, sendo uma média de 80 operações e aproximadamente 50% para cada possível operação (compra e venda). Sendo assim, foram consideradas em média 40 negociações de compra e 40 de venda. Considerou-se para este trabalho as taxas para investidores de pessoa física, sendo esta 0.030747%.

O lucro médio entre os ativos brasileiros neste trabalho utilizando o sistema CNN-TC foi de 157.34%. Isto representa um lucro médio de R\$15.734,00. Com base neste valor pode-se calcular o desconto médio por operação calculando $R\$15.734,00 * 0,00030747 = R\$4,83$. E o valor médio total dos Emolumentos, sendo 4 anos nos períodos de análise, $80 * R\$4,83 * 4 = R\$1.545,60$. Este valor representa apenas 9,82% de desconto do lucro total, não impactando de forma significativa o retorno financeiro. Sendo assim não é necessário a consideração desta taxa para fins de comparação.

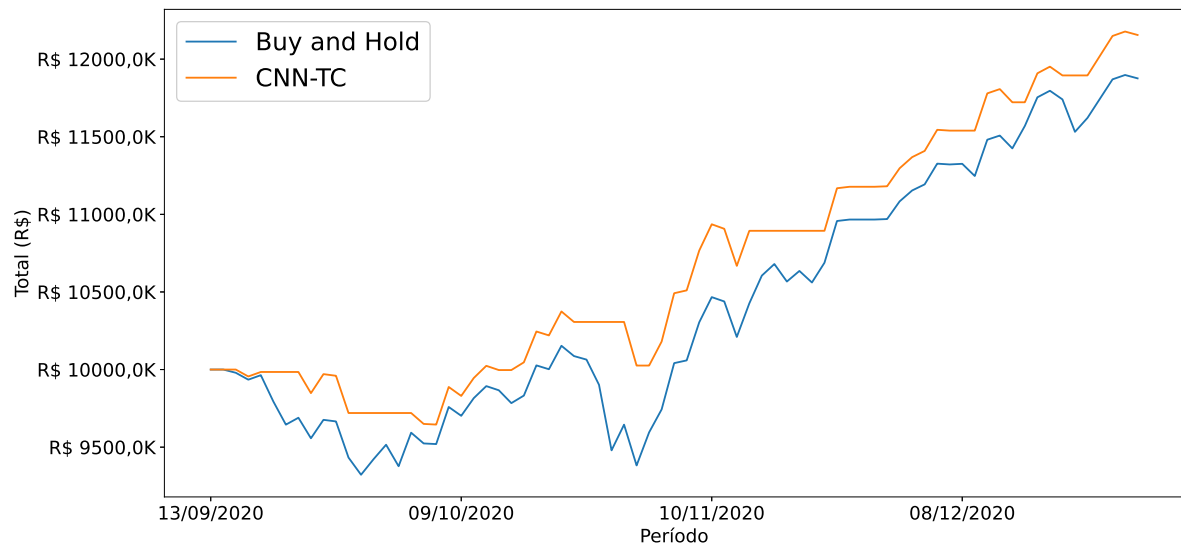
6.14 Teste em ambiente simulado

Para esta avaliação foi escolhido o ativo BOVA11 com o mesmo modelo utilizado na avaliação dos resultados neste mesmo capítulo.

A Figura 46 apresenta o retorno obtido com o sistema CNN-TC em comparação com a estratégia BH. Como se pode observar, apesar do retorno financeiro semelhante em alguns momentos e de um período não tão longo quanto os testes realizados anteriormente, que o modelo foi capaz de identificar topos e fundos durante o período analisado, alcançando assim um resultado financeiro superior.

Nota-se que nos momentos de queda, o modelo foi capaz de vender antecipadamente, preservando capital para comprar a um preço inferior. A Figura 47 apresenta o relatório financeiro gerado pelo software MetaTrader 5. Como se pode observar, o valor inicial investido foi de R\$ 10.000.000,00 e ao final do período avaliado, o resultado do CNN-TC foi um lucro de R\$ 2.155.593,10, enquanto a estratégia BH obteve um lucro de R\$ 1.876.092,40.

Figura 46 – Retorno sobre investimento no teste em ambiente simulado



Fonte: o autor

Figura 47 – Captura de tela do relatório do MetaTrader

Relatório da Conta de Negociação

Nome: DAVI GUARACHO NASCIMENTO

Conta: 62483510 (BRL, 1:1, XPMT5-Demo, demo, Netting)

Corretora: XP Investimentos CCTVM S/A

Data: 2021.01.03 19:02

Posições

Ativo	Position	Horário	Tipo	Volume	Preço	S / L	T / P	Preço de Mercado	Swap	Lucro	Comentário
									0.00	0.00	
		Saldo:		12 155 593.1			Margem Livre:		12 155 593.1		
		Facilidade do Crédito:		0.00			Margem:		0.00		
		Flutuante P/L:		0.00			Nível de Margem:		0.00%		
		Capital Líquido:		12 155 593.1							

Fonte: o autor, com uso do software MetaTrader

6.15 Discussões

A Tabela 15 sumariza os resultados financeiros das estratégias analisadas. Como se pode ver, o sistema CNN-TC obteve melhor resultado em 6 dos 12 ativos analisados frente à todas as estratégias, e quando analisado somente as estratégias utilizando RNAs, este foi melhor em 9 dos 12 ativos. Os resultados financeiros foram satisfatórios comparados às demais estratégias adotadas para comparação, apresentando grande diferença no valor final em alguns casos.

Tabela 15 – Resultados financeiros das estratégias analisadas

Ativo	Estratégias					
	BH	CDI	CNN-TC	CNN-TA	MLP	LSTM
BOVA11	R\$ 17.366,00	R\$ 14.962,62	R\$ 22.278,80	R\$ 15.2585,10	R\$ 16.425,90	R\$ 16.838,50
VALE3	R\$ 26.152,93	R\$ 14.962,62	R\$ 60.315,19	R\$ 20.753,56	R\$ 36.894,08	R\$ 28.927,61
PETR4	R\$ 21.677,08	R\$ 14.962,62	R\$ 14.928,37	R\$ 31.238,00	R\$ 5.142,22	R\$ 8.491,26
ITUB4	R\$ 21.090,62	R\$ 14.962,62	R\$ 24.337,77	R\$ 17.873,84	R\$ 21.196,86	R\$ 19.106,60
ABEV3	R\$ 10.019,29	R\$ 14.962,62	R\$ 15.555,81	R\$ 12.861,00	R\$ 11.508,70	R\$ 11.619,07
WEGE3	R\$ 15.517,57	R\$ 14.962,62	R\$ 18.283,69	R\$ 14.872,74	R\$ 13.392,80	R\$ 16.190,10
AAPL	R\$ 14.146,26	R\$ 14.962,62	R\$ 12.460,39	R\$ 11.836,75	R\$ 9.492,48	R\$ 9.823,27
WFC	R\$ 9.710,28	R\$ 14.962,62	R\$ 13.993,24	R\$ 13.693,00	R\$ 10.107,42	R\$ 13.873,28
JPM	R\$ 9.274,50	R\$ 18.961,86	R\$ 21.484,83	R\$ 22.351,96	R\$ 19.911,53	R\$ 25.668,57
TRV	R\$ 22.826,07	R\$ 24.379,59	R\$ 51.599,03	R\$ 68.681,20	R\$ 46.967,11	R\$ 41.808,71
XLE	R\$ 13.359,37	R\$ 18.961,86	R\$ 25.185,89	R\$ 17.857,00	R\$ 14.375,93	R\$ 17.918,82
XLF	R\$ 8.360,38	R\$ 24.379,59	R\$ 22.262,97	R\$ 18.492,84	R\$ 17.927,65	R\$ 18.057,11

Fonte: o autor

Para melhor visualização das métricas dos sistemas apresentados neste trabalho, os ativos de cada mercado alvo deste estudo foram agrupados e calculado média e desvios padrão das principais métricas. Nas Tabelas 16 e 17 são apresentadas as respectivas médias e desvios padrão para o mercado brasileiro, e na sequência, nas Tabelas 18 e 19, para o mercado americano.

Tabela 16 – Médias das métricas estatísticas para mercado brasileiro

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	63,47%	78,30%	63,47%	67,66%
CNN-TA	56,74%	78,88%	56,74%	63,15%
MLP	56,44%	74,64%	56,44%	61,88%
LSTM	56,65%	74,60%	56,65%	61,98%

Fonte: o autor

Tabela 17 – Desvio padrão das métricas estatísticas para mercado brasileiro

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	±4,03%	±2,03%	±4,03%	±3,40%
CNN-TA	±6,58%	±4,05%	±6,58%	±6,00%
MLP	±4,53%	±2,79%	±4,53%	±3,89%
LSTM	±3,97%	±2,08%	±3,97%	±3,63%

Fonte: o autor

Tabela 18 – Médias das métricas estatísticas para mercado americano

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	58,63%	78,53%	58,63%	63,35%
CNN-TA	55,20%	75,82%	55,20%	61,70%
MLP	54,50%	75,94%	54,50%	60,07%
LSTM	56,53%	74,57%	56,53%	61,30%

Fonte: o autor

Tabela 19 – Desvio padrão das métricas estatísticas para mercado americano

Sistema	Acurácia	Precisão	Revocação	F1
CNN-TC	$\pm 3,35\%$	$\pm 0,67\%$	$\pm 3,35\%$	$\pm 2,97\%$
CNN-TA	$\pm 2,47\%$	$\pm 3,10\%$	$\pm 2,47\%$	$\pm 2,54\%$
MLP	$\pm 2,62\%$	$\pm 0,37\%$	$\pm 2,62\%$	$\pm 2,25\%$
LSTM	$\pm 3,99\%$	$\pm 2,06\%$	$\pm 3,99\%$	$\pm 3,64\%$

Fonte: o autor

Para o mercado brasileiro, nota-se que o sistema CNN-TC apresentou resultados mais consistentes que os demais. Apesar de desvios padrão próximos do sistema LSTM em algumas métricas, as médias das métricas ficaram acima de quase todas as estratégias comparadas. Na sequência tem-se o sistema CNN-TA, que apesar de um desvio padrão mais elevado que os demais, apresentou precisão levemente superior ao CNN-TC, com diferença de 0,58%, e F1 consideravelmente aceitável e acima do MLP e LSTM. Apesar disto, o único ativo no qual o CNN-TA mostrou resultado financeiro superior foi para PETR4, ativo o qual pode ter sido o grande influenciador para desvios padrão elevados entre os sistemas. Os sistemas utilizando MLP e LSTM apresentaram médias próximas e abaixo do CNN-TA.

No mercado de ações americano, embora em termos de acurácia os sistemas CNN-TC e LSTM se comportaram de forma semelhante, note-se um desvio padrão superior em termos de precisão para o LSTM, evidenciando resultados mais consistentes por parte do CNN-TC. É possível observar também que os sistemas CNN-TC, CNN-TA e LSTM apresentaram os melhores valores de F1, com uma diferença de 1,65% entre o CNN-TC e o CNN-TA, enquanto a diferença entre o CNN-TC e o sistema LSTM foi de 2,05%, o que ressalta melhor aplicabilidade do CNN-TC para este mercado. Vale ressaltar que os sistemas CNN-TC e LSTM apresentaram as melhores médias de acurácia, evidenciando melhor capacidade de classificação neste mercado frente ao CNN-TA, no geral. Entretanto, dos 6 ativos utilizados neste trabalho, o CNN-TA apresentou melhor retorno financeiro em apenas 1 ativo e praticamente empatou em retorno financeiro com os demais sistemas em 4 ativos. O sistema LSTM apresentou resultado

financeiro claramente melhor apenas no ativo JPM. Já o CNN-TC alcançou melhores resultados financeiros nos ativos XLE e XLF e empatou com o CNN-TA em 2 ativos.

Uma ferramenta de avaliação de tarefas de classificação é a matriz de confusão, que mostra quantas amostras foram classificadas em cada uma das classes e quais foram corretamente ou erroneamente classificadas. Como exemplo, a Tabela 20 apresenta as matrizes de confusão dos sistemas analisados no ativo BOVA11.

Como se pode observar, devido às classes Compra e Venda serem menos comuns, existe uma dificuldade inerente do algoritmo de capturar características específicas, podendo levar a uma sobre-classificação da classe Neutro. Desta forma, muitos pontos falsos de entrada e saída de operação são gerados. A confusão com a classe Neutro não necessariamente prejudica o retorno financeiro, pois em caso de classificações repetidas, nenhuma ação é tomada. Entretanto confundir a classe Venda com Compra (falso positivo) ou Compra com Venda (falso negativo) pode comprometer o retorno financeiro.

Nota-se que o sistema CNN-TC foi capaz de identificar com maior precisão as classes quando comparado com os demais algoritmos treinados com o mesmo método de rotulagem (MLP e LSTM). A quantidade de vendas classificadas pelo CNN-TC foi menor, permitindo que este sistema acompanhasse os períodos de alta no preço, além disso a quantidade de classificações erradas da classe Neutro foi menor. O sistema MLP apresentou maior confusão nas classes Compra e Venda quando comparado com os demais que utilizaram o mesmo método de rotulagem, onde o desempenho financeiro ficou próximo da estratégia BH. Semelhantemente o LSTM apresentou retorno financeiro próximo. Este foi capaz de identificar Vendas e Compras com maior precisão, entretanto a confusão com a classe Neutro foi mais frequente, o que pode ter executado vendas de forma antecipada, prejudicando o seu retorno financeiro.

O sistema CNN-TA apresentou maior confusão com a classe Neutro. As amostras rotuladas como Compra foram identificadas em menor quantidade que as de Venda. Estes podem ter sido o motivo de retorno financeiro inferior frente aos demais algoritmos comparados, onde é identificada venda de forma antecipada. Além disso, a quantidade de amostras de compra e venda geradas pelo método de rotulagem deste sistema é menor, e isto pode ter prejudicado o desempenho do modelo em abstrair características destas classes.

Tabela 20 – Matrizes de confusão para o ativo BOVA11

CNN-TC				
Verdadeiro				
		Venda	Compra	Neutro
Predito	Venda	55	0	31
	Compra	1	65	21
	Neutro	154	104	528
MLP				
Verdadeiro				
		Venda	Compra	Neutro
Predito	Venda	38	9	39
	Compra	10	43	34
	Neutro	182	165	439

CNN-TA				
Verdadeiro				
		Venda	Compra	Neutro
Predito	Venda	40	1	12
	Compra	6	26	27
	Neutro	252	227	368
LSTM				
Verdadeiro				
		Venda	Compra	Neutro
Predito	Venda	45	1	40
	Compra	2	48	37
	Neutro	192	182	412

Fonte: o autor

Apesar de algumas métricas, no geral, não terem ficado tão distantes, nota-se uma diferença no retorno financeiro entre os algoritmos. Por fim, com base nas avaliações realizadas, pode-se concluir que os sistemas baseados em CNN apresentaram métricas, no geral, melhores que os demais para classificação, isto evidencia a capacidade dos mesmos na identificação de padrões em séries temporais para classificação. Os sistemas que utilizaram LSTM demonstraram resultados estatísticos melhores em 2 ativos, demonstrando que este modelo de RNA também pode ser eficiente para tarefa de classificação em alguns casos.

Embora o sistema CNN-TA não tenha superado o retorno financeiro do CNN-TC na maioria das avaliações, pode-se associar este fato ao método de rotulagem utilizado em sua proposta, pois como mencionado nos resultados, no mercado americano este sistema apresentou métricas próximas do CNN-TC e retorno financeiro superior em alguns casos.

Ao analisar os gráficos dos resultados financeiros pode-se observar que o modelo é capaz de antecipar os maiores movimentos de queda, preservando capital para comprar na baixa uma maior quantidade de ações e assim aumentar o retorno ao longo do tempo. Embora o período de teste em ambiente simulado tenha sido curto, pode-se observar comportamento semelhante no mesmo.

Sendo assim, com base nas avaliações realizadas e tendo como base o retorno esperado dos métodos de rotulagem, pode-se concluir que boas métricas de classificação em conjunto com o método de rotulagem proposto neste trabalho proporcionaram resultados melhores para o sistema CNN-TC na maioria dos ativos analisados.

7 CONCLUSÃO

Neste trabalho, foi proposto um sistema que utiliza um modelo de CNN com o objetivo de prever os melhores momentos de compra e venda para o mercado financeiro. O sistema foi avaliado nos mercados de ações brasileiro e americano, e para isso foram utilizadas algumas das principais ações que compõe o IBOVESPA bem como alguns ativos selecionados do mercado americano, sendo 4 deles os mesmos utilizados na proposta do sistema CNN-TA para fins de comparação.

Como comparação, foram implementados 3 modelos de RNAs, uma MLP, uma LSTM e outro sistema baseado em CNN, chamado de CNN-TA (SEZER; OZBAYOGLU, 2018). Foram realizadas avaliações estatística e financeira a fim de mensurar a eficácia destes sistemas na identificação dos momentos de operação e avaliar se os mesmos superariam a estratégia BH, sendo esta a estratégia mais simples para o investimento em bolsa de valores.

Embora nos períodos analisados tenham ocorrido fatos atenuantes, tais como crise financeira e notícias relacionadas às empresas em específico, os sistemas apresentaram resultados considerados satisfatórios, tanto em termos financeiros como em métricas estatísticas. Em períodos menores o resultado final pode ser impactado por notícias e acontecimentos inesperados, mas quando analisado em períodos de tempo maiores, as estratégias apresentadas mostraram-se capazes de gerar lucro frente aos *benchmarks* de mercado utilizados, podendo assim reduzir as incertezas de investimento em algumas situações e apresentando bons indicadores financeiros de risco e retorno.

Como *benchmarks* de mercado para comparação foram utilizados a estratégia BH e o CDI, ambos indicadores comumente utilizados para mensurar o desempenho de investimentos. Na maioria dos casos, as estratégias utilizando modelos de RNAs mostraram-se capazes de superar tais indicadores, principalmente o BH quando os preços passam por forte tendência de queda.

O sistema CNN-TA mostrou resultados melhores quando avaliado nos ativos do mercado americano, sendo 4 destes os mesmos utilizados na proposta original. Pode-se notar uma característica em comum dentre os 4 ativos utilizados para comparação, todos apresentaram uma tendência de queda no início do período de avaliação seguida de uma recuperação nos preços que corroborou para o resultado financeiro final. Quando analisado no mercado brasileiro, nos ativos onde houve uma tendência de alta durante o período analisado este sistema não se mostrou tão eficaz. Nas ações da PETR4 e ABEV3, onde o preço oscilou em uma faixa de

valores na maior parte do tempo, este sistema apresentou resultados melhores, ganhando da proposta deste trabalho na PETR4.

Os sistemas MLP e LSTM apresentaram comportamentos variados. O MLP não apresentou os melhores resultados em nenhum dos ativos avaliados. Entretanto em alguns casos este aproximou-se do comportamento do CNN-TC. Já o LSTM apresentou o melhor retorno financeiro na ação JPM e se aproximou dos sistemas baseados em CNN nas ações WEGE3 e WFC. Apesar de ambos os sistemas terem apresentado boas métricas, pode-se notar menor capacidade de generalização para classificação dos momentos de operação quando comparado com o sistema CNN-TC.

O sistema CNN-TC foi capaz de analisar dados de séries temporais e executar classificação de sinal de negociação, obtendo retornos expressivos quando comparado às demais estratégias, onde das 12 avaliações realizadas, em 6 apresentou resultados financeiros superiores, e considerando somente as estratégias utilizando RNAs, este foi melhor em 9 ativos. Embora a vantagem nas métricas tenha sido por pequena diferença em alguns casos, este demonstrou maior capacidade para identificação e classificação dos padrões nas séries analisadas.

Nota-se ainda que o método de rotulagem proposto neste trabalho apresenta resultados melhores quando comparado ao da proposta do CNN-TA. Pode-se concluir que este método, em conjunto com métricas melhores proporcionadas pelo CNN-TC (ou próximas do sistema CNN-TA em alguns casos), permitiram retornos financeiros superiores para o CNN-TC.

Vale ressaltar que os resultados financeiros dos sistemas mostraram-se distantes do retorno total possível com 100% de precisão para os rótulos gerados por ambos os métodos, e que em alguns casos, uma pequena diferença de métricas mostrou resultados financeiros evidentemente diferentes. Isto reforça a necessidade de que um sistema de negociação deve ser preciso a fim de garantir retornos vantajosos frente às estratégias mais básicas de investimento, e evidencia que pequenas diferenças para este tipo de classificação podem influenciar os resultados a longo prazo.

O teste em ambiente simulado, apesar de realizado em um período relativamente curto, também apresentou bons resultados, alcançando retorno financeiro superior à estratégia BH, o que valida a aplicabilidade desta pesquisa em cenários reais.

O estudo nesta área de pesquisa exige um grande conhecimento interdisciplinar em tópicos como estatística, econometria, finanças, e inteligência artificial. Apesar de existirem variáveis não consideradas neste estudo, como notícias, fatores econômicos e fatos relevantes relacionados às empresas em específico, o retorno financeiro das avaliações mostrou-se vantajoso

quando comparado com a estratégia BH, a mais básica do mercado de renda variável. Sendo assim, os resultados mostraram-se promissores para um sistema de negociação automatizado.

Este trabalho teve sua contribuição e resultados iniciais publicados no evento XVII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2020) (NASCIMENTO; COSTA; BIANCHI, 2020).

7.1 Trabalhos Futuros

A quantidade e a qualidade dos dados de entrada é um fator que afeta a qualidade da previsão das RNAs. Neste trabalho, para o CNN-TC foram considerados apenas alguns recursos de entrada de forma empírica. É possível analisar e incluir outros elementos que possam contribuir positivamente para melhorar a classificação. Acredita-se que possam existir elementos, como dados econômicos que impactem diretamente as empresas, que podem melhorar a capacidade de previsão. Como exemplo, pode-se citar a empresa Petrobras S.A. que sofre influência da cotação do barril do petróleo, e a Vale S.A. que tem influência da cotação do minério de ferro no seu lucro operacional. Tais dados podem complementar a predição da tendência destes ativos. Além disso, existem indicadores que não foram explorados nesta pesquisa e que podem fornecer informações adicionais sobre o movimento dos preços, tais como Taxa de Variação (*Rate of Change - RSI*), RSI estocástico e Volume de equilíbrio (*On-balance volume - OBV*).

Analisar dados financeiros das empresas pode ser um recurso importante para modelos de previsão e classificação, visto que o mercado tende a reagir a tais informações. Sendo assim, pode-se complementar o modelo com dados fundamentalistas, a fim de aprimorar a tomada de decisão.

Para definição dos dados de entrada, pode-se aplicar análises matemáticas das variáveis disponíveis a fim de encontrar correlações que corroborem para melhor qualidade de classificação, tais técnicas como Análise de Componentes Principais (Principal Component Analysis) (PCA).

O processo de otimização dos parâmetros das RNAs é vital para a melhoria dos resultados. Pode-se utilizar de ferramentas como algoritmos de força bruta, ou mesmo algoritmos genéticos, a fim de encontrar melhores combinações para os modelos.

Realizar experimentos em períodos menores, inclusive em operações de alta frequência. Neste trabalho considerou-se apenas o tempo gráfico diário, mas pode ser possível obter ganhos maiores realizando operações intradiárias. Independentemente do período considerado, a leitura

dos indicadores permanece a mesma, o que indica que pode-se treinar um modelo em um tempo gráfico e utilizá-lo em outro.

Os métodos de rotulagem apresentados neste trabalho são baseados apenas no histórico do preço de fechamento. É cabível um estudo a fim de determinar com base em fatores mais concretos os melhores momentos de compra e venda a fim de encontrar padrões que proporcionem melhores métricas estatísticas e maior rentabilidade financeira.

Da mesma forma, o modo de operação considerou apenas o preço de fechamento. Pode-se analisar a tomada de decisão em outros momentos, como na abertura do mercado, a fim de proporcionar maiores ganhos financeiros bem como evitar perdas em momentos que o mercado é impactado por notícias no decorrer do dia.

É possível avaliar a implementação de *stop loss*, uma forma de delimitar uma perda máxima por operação, para situações em que o mercado é impactado por notícias, a fim de garantir uma rentabilidade melhor a longo prazo.

Por último, pode-se explorar modelos mais complexos de RNAs. Neste trabalho foram apresentadas comparações com redes LSTM. Estas têm como característica um tipo de memória de longo prazo, e acredita-se que possam fornecer recursos adicionais para a classificação. Sendo assim, sugere-se o estudo de modelos de arquiteturas mistas a fim de complementar os resultados finais.

REFERÊNCIAS

- ALONSO-MONSALVE, Saúl et al. Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. **Expert Systems with Applications**, Elsevier, v. 149, p. 113250, 2020.
- BAO, Wei; YUE, Jun; RAO, Yulei. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. **PLoS one**, Public Library of Science, v. 12, n. 7, 2017.
- BOUREAU, Y-Lan; PONCE, Jean; LECUN, Yann. A theoretical analysis of feature pooling in visual recognition. In: PROCEEDINGS of the 27th international conference on machine learning (ICML-10). [S.l.: s.n.], 2010. P. 111–118.
- BRACEWELL, Ronald Newbold; BRACEWELL, Ronald N. **The Fourier transform and its applications**. [S.l.]: McGraw-Hill New York, 1986. v. 31999.
- BRAGA, A de P. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: Livros Técnicos e Científicos, 2000.
- BULKOWSKI, Thomas N. **Encyclopedia of chart patterns**. [S.l.]: John Wiley & Sons, 2011. v. 225.
- CANZIANI, Alfredo; PASZKE, Adam; CULURCIELLO, Eugenio. An analysis of deep neural network models for practical applications. **arXiv preprint arXiv:1605.07678**, 2016.
- CAO, Jiasheng; WANG, Jinghan. Stock price forecasting model based on modified convolution neural network and financial time series analysis. **International Journal of Communication Systems**, Wiley Online Library, v. 32, n. 12, e3987, 2019.
- CAVALCANTE, Rodolfo C et al. Computational intelligence and financial markets: A survey and future directions. **Expert Systems with Applications**, Elsevier, v. 55, p. 194–211, 2016.
- CHOLLET, François et al. **Keras**. [S.l.: s.n.], 2015. <https://keras.io>.
- CHUNG, Hyejung; SHIN, Kyung-shik. Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. **Neural Computing and Applications**, Springer, v. 32, n. 12, p. 7897–7914, 2020.
- CIRESAN, Dan Claudiu et al. Convolutional neural network committees for handwritten character classification. In: IEEE. 2011 International Conference on Document Analysis and Recognition. [S.l.: s.n.], 2011. P. 1135–1139.
- COLBY, Robert W; MEYERS, Thomas A. **The encyclopedia of technical market indicators**. [S.l.]: Dow Jones-Irwin Homewood, IL, 1988.
- DAMODARAN, Aswath; CUNHA SERRA, Afonso Celso da. **Valuation: Como avaliar empresas e escolher as melhores ações**. [S.l.]: Grupo Gen-LTC, 2000.

DUCHI, John; HAZAN, Elad; SINGER, Yoram. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. **Journal of Machine Learning Research**, v. 12, p. 2121–2159, jul. 2011.

ECONOMATICA. [S.l.: s.n.], 2020. acessado em 14/07/2020. Disponível em: <<https://insight.economatica.com/volume-financeiro-negociado-na-bolsa-b3/>>.

EDWARDS, Robert D; BASSETTI, WHC; MAGEE, John. **Technical analysis of stock trends**. [S.l.]: CRC press, 2012.

ELDER, Alexander. **Como se transformar em um operador e investidor de sucesso**. [S.l.]: Alta Books Editora, 2018.

FAMA, Eugene F. Random walks in stock market prices. **Financial analysts journal**, Taylor & Francis, v. 51, n. 1, p. 75–80, 1995.

FATH, Aref Hashemi; MADANIFAR, Farshid; ABBASI, Masood. Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. **Petroleum**, Elsevier, v. 6, n. 1, p. 80–91, 2020.

FISCHER, Thomas; KRAUSS, Christopher. Deep learning with long short-term memory networks for financial market predictions. **European Journal of Operational Research**, Elsevier, v. 270, n. 2, p. 654–669, 2018.

GLOT, Xavier; BENGIO, Yoshua. Understanding the difficulty of training deep feedforward neural networks. In: PROCEEDINGS of the thirteenth international conference on artificial intelligence and statistics. [S.l.: s.n.], 2010. P. 249–256.

GLOT, Xavier; BORDES, Antoine; BENGIO, Yoshua. Deep sparse rectifier neural networks. In: PROCEEDINGS of the fourteenth international conference on artificial intelligence and statistics. [S.l.: s.n.], 2011. P. 315–323.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. [S.l.]: MIT press, 2016.

GRAVES, Alex; JAITLY, Navdeep; MOHAMED, Abdel-rahman. Hybrid speech recognition with deep bidirectional LSTM. In: IEEE. 2013 IEEE workshop on automatic speech recognition and understanding. [S.l.: s.n.], 2013. P. 273–278.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. [S.l.]: Artmed, 2007. ISBN 9788577800865. Disponível em: <<https://books.google.com.br/books?id=bhMwDwAAQBAJ>>.

HAYKIN, S.S. **Neural Networks: A Comprehensive Foundation**. [S.l.]: Prentice Hall, 1999. (International edition). ISBN 9780139083853. Disponível em: <<https://books.google.com.br/books?id=M5abQgAACAAJ>>.

HE, Kaiming et al. Deep residual learning for image recognition. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. P. 770–778.

HE, Kaiming et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: PROCEEDINGS of the IEEE international conference on computer vision. [S.l.: s.n.], 2015. P. 1026–1034.

HEBB, Donald Olding. **The organization of behavior: a neuropsychological theory**. [S.l.]: J. Wiley; Chapman & Hall, 1949.

HINTON, Geoffrey E et al. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012.

HIRANSHA, M et al. NSE stock market prediction using deep-learning models. **Procedia computer science**, Elsevier, v. 132, p. 1351–1362, 2018.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.

HOSEINZADE, Ehsan; HARATIZADEH, Saman. CNNpred: CNN-based stock market prediction using a diverse set of variables. **Expert Systems with Applications**, Elsevier, v. 129, p. 273–285, 2019.

HUANG, Jian; CHAI, Junyi; CHO, Stella. Deep learning in finance and banking: A literature review and classification. **Frontiers of Business Research in China**, Springer, v. 14, p. 1–24, 2020.

JAIN, Sneha; GUPTA, Roopam; MOGHE, Asmita A. Stock Price Prediction on Daily Stock Data using Deep Neural Networks. In: IEEE. 2018 International Conference on Advanced Computation and Telecommunication (ICACAT). [S.l.: s.n.], 2018. P. 1–13.

KAMENCAY, Patrik et al. A new method for face recognition using convolutional neural network. Vysoká škola báňská-Technická univerzita Ostrava, 2017.

KARIM, Fazle et al. LSTM fully convolutional networks for time series classification. **IEEE access**, IEEE, v. 6, p. 1662–1669, 2017.

KARPATHY, Andrej et al. Large-scale video classification with convolutional neural networks. In: PROCEEDINGS of the IEEE conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2014. P. 1725–1732.

KINGMA, Diederik P; BA, Jimmy. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KRAUSS, Christopher; DO, Xuan Anh; HUCK, Nicolas. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. **European Journal of Operational Research**, Elsevier, v. 259, n. 2, p. 689–702, 2017.

KRIESEL, David. **A brief Introduction on Neural Networks**. [S.l.: s.n.], 2007.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: ADVANCES in neural information processing systems. [S.l.: s.n.], 2012. P. 1097–1105.

LAMBERT, Donald R. Commodity channel index: Tool for trading cyclic trends. **Technical Analysis of Stocks & Commodities**, v. 1, p. 47, 1983.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Yann et al. Handwritten digit recognition with a back-propagation network. In: ADVANCES in neural information processing systems. [S.l.: s.n.], 1990. P. 396–404.

LEE, Chien-Chiang; LEE, Jun-De; LEE, Chi-Chuan. Stock prices and the efficient market hypothesis: Evidence from a panel stationary test with structural breaks. **Japan and the world economy**, Elsevier, v. 22, n. 1, p. 49–58, 2010.

LEMO, FLAVIO ALEXANDRE CALDAS DE ALMEIDA. **Análise técnica dos mercados financeiros**. [S.l.]: Saraiva Educação SA, 2017.

LI, Yun. **80% of the stock market is now on autopilot**. 2019. Disponível em: <<https://www.cnn.com/2019/06/28/80percent-of-the-stock-market-is-now-on-autopilot.html>>. Acesso em:

MALKIEL, Burton G; FAMA, Eugene F. Efficient capital markets: A review of theory and empirical work. **The journal of Finance**, Wiley Online Library, v. 25, n. 2, p. 383–417, 1970.

MARTÍN ABADI et al. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. [S.l.: s.n.], 2015. Software available from tensorflow.org. Disponível em: <<http://tensorflow.org/>>.

MCCULLOCH, Warren S; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.

MEHTAB, Sidra; SEN, Jaydip. Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Timeseries. **arXiv preprint arXiv:2001.09769**, 2020.

NASCIMENTO, Davi; COSTA, Anna; BIANCHI, Reinaldo. Stock Trading Classifier with Multichannel Convolutional Neural Network. In: ANAIS do XVII Encontro Nacional de Inteligência Artificial e Computacional. Evento Online: SBC, 2020. P. 282–293. DOI: 10.5753/eniac.2020.12136. Disponível em: <<https://sol.sbc.org.br/index.php/eniac/article/view/12136>>.

NELSON, David MQ; PEREIRA, Adriano CM; OLIVEIRA, Renato A de. Stock market's price movement prediction with LSTM neural networks. In: IEEE. 2017 International joint conference on neural networks (IJCNN). [S.l.: s.n.], 2017. P. 1419–1426.

NG, Andrew et al. [S.l.: s.n.], 2015. acessado em 12/11/2020. Disponível em: <<http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/>>.

NIELSEN, Michael A. **Neural networks and deep learning**. [S.l.]: Determination press San Francisco, CA, 2015. v. 2018.

NIGRIN, Albert. **Neural networks for pattern recognition**. [S.l.]: MIT press, 1993.

NOBRE, João; NEVES, Rui Ferreira. Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. **Expert Systems with Applications**, Elsevier, v. 125, p. 181–194, 2019.

PADIAL, Darío López. **Technical Analysis Library in Python**. [S.l.], acessado em 01/03/2020. <https://technical-analysis-library-in-python.readthedocs.io/en/latest/>.

PIMENTA, Alexandre et al. An automated investing method for stock market based on multiobjective genetic programming. **Computational Economics**, Springer, v. 52, n. 1, p. 125–144, 2018.

POWERS, David MW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. **arXiv preprint arXiv:2010.16061**, 2020.

QUONG, Gene Soudack; SOUDACK, Avrum. A., Volume-weighted rsi: Money flow. **Technical Analysis of Stocks & Commodities ()**, 1989.

ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, David E; HINTON, Geoffrey E; WILLIAMS, Ronald J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.

RUSSELL, Stuart; NORVIG, Peter. Artificial intelligence: a modern approach, 2002.

SACHETIM, Henrique Martins. Análise técnica: estudo da confiabilidade dos principais indicadores de análise técnica, aplicados as ações mais negociadas na Bovespa no período de 1995 a 2005, 2006.

SAMARAWICKRAMA, AJP; FERNANDO, TGI. A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. In: IEEE. 2017 IEEE International Conference on Industrial and Information Systems (ICIIS). [S.l.: s.n.], 2017. P. 1–6.

SAYAVONG, Lounnapha; WU, Zhongdong; CHALITA, Sookasame. Research on Stock Price Prediction Method Based on Convolutional Neural Network. In: IEEE. 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS). [S.l.: s.n.], 2019. P. 173–176.

SELVIN, Sreelekshmy et al. Stock price prediction using LSTM, RNN and CNN-sliding window model. In: IEEE. 2017 international conference on advances in computing, communications and informatics (icacci). [S.l.: s.n.], 2017. P. 1643–1647.

SEZER, Omer Berat; OZBAYOGLU, A Murat; DOGDU, Erdogan. An artificial neural network-based stock trading system using technical analysis and big data framework. In: PROCEEDINGS of the southeast conference. [S.l.: s.n.], 2017. P. 223–226.

SEZER, Omer Berat; OZBAYOGLU, Ahmet Murat. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. **Applied Soft Computing**, Elsevier, v. 70, p. 525–538, 2018.

SHAN, Ketan. The Dow Theory. **Aweshkar Research Journal**, v. 13, n. 1, 2012.

SILVA, André Luiz Carvalhal da. Introdução à análise de dados. **Rio de Janeiro: E-papers**, 2009.

SILVEIRA BUENO, Rodrigo de Losso da. **Econometria de séries temporais**. [S.l.]: Cengage Learning, 2008.

SOLAI, Pavithra. [S.l.: s.n.], 2018. acessado em 10/11/2020. Disponível em: <<https://medium.com/@pavisj/convolutions-and-backpropagations-46026a8f5d2c>>.

SRIVASTAVA, Nitish et al. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

SZEGEDY, Christian et al. Going deeper with convolutions. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015. P. 1–9.

SZEGEDY, Christian et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: THIRTY-FIRST AAAI conference on artificial intelligence. [S.l.: s.n.], 2017.

SZEGEDY, Christian et al. Rethinking the inception architecture for computer vision. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. P. 2818–2826.

TSAI, Yun-Cheng; CHEN, Jun-Hao; WANG, Chun-Chieh. Encoding candlesticks as images for patterns classification using convolutional neural networks. **arXiv preprint arXiv:1901.05237**, 2019.

TSANTEKIDIS, Avraam et al. Forecasting stock prices from the limit order book using convolutional neural networks. In: IEEE. 2017 IEEE 19th Conference on Business Informatics (CBI). [S.l.: s.n.], 2017. v. 1, p. 7–12.

VELAY, Marc; DANIEL, Fabrice. Stock chart pattern recognition with deep learning. **arXiv preprint arXiv:1808.00418**, 2018.

WANG, Jia et al. Financial markets prediction with deep learning. In: IEEE. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). [S.l.: s.n.], 2018. P. 97–104.

WILDER, J.W. **New Concepts in Technical Trading Systems**. [S.l.]: Trend Research, 1978. ISBN 9780894590276. Disponível em: <<https://books.google.com.br/books?id=WesJAQAAMAAJ>>.

WILLIAMS, B.M. **New Trading Dimensions: How to Profit from Chaos in Stocks, Bonds, and Commodities**. [S.l.]: Wiley, 1998. (A Marketplace Book). ISBN 9780471295419. Disponível em: <<https://books.google.com.br/books?id=AFySgHdkxeMC>>.

WU, Meiyin; CHEN, Li. Image recognition based on deep learning. In: IEEE. 2015 Chinese Automation Congress (CAC). [S.l.: s.n.], 2015. P. 542–546.

YACIM, Joseph; BOSHOF, Douw. Impact of Artificial Neural Networks Training Algorithms on Accurate Prediction of Property Values. **Journal of Real Estate Research**, v. 40, p. 375–418, nov. 2018.

YAMASHITA, Rikiya et al. Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, Springer, v. 9, n. 4, p. 611–629, 2018.

ZHANG, Kang et al. Stock market prediction based on generative adversarial network. **Procedia computer science**, Elsevier, v. 147, p. 400–406, 2019.