

CENTRO UNIVERSITÁRIO FEI
GUILHERME ALBERTINI DE OLIVEIRA

**PREDIÇÃO DO MERCADO FINANCEIRO COM UMA ARQUITETURA DE
EXTRAÇÃO DE CONTEXTO BASEADA EM DECOMPOSIÇÃO DE SERIES
TEMPORAIS**

São Bernardo do Campo

2022

GUILHERME ALBERTINI DE OLIVEIRA

**PREDIÇÃO DO MERCADO FINANCEIRO COM UMA ARQUITETURA DE
EXTRAÇÃO DE CONTEXTO BASEADA EM DECOMPOSIÇÃO DE SERIES
TEMPORAIS**

Dissertação de Mestrado apresentada ao Centro Universitário FEI, para obtenção do título de Mestre em Engenharia Elétrica. Orientado pelo Prof. Dr. Paulo Sérgio Silva Rodrigues.

São Bernardo do Campo

2022

Albertini de Oliveira, Guilherme.

Predição do mercado financeiro com uma arquitetura de extração de contexto baseada em decomposição de series temporais / Guilherme Albertini de Oliveira. São Bernardo do Campo, 2022.

112 p. : il.

Dissertação - Centro Universitário FEI.

Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues.

1. Machine Learning. 2. Bolsa de Valores. 3. Predição. 4. Redes Neurais. 5. Séries Temporais. I. Silva Rodrigues, Paulo Sérgio, orient. II. Título.

Aluno(a): Guilherme Albertini de Oliveira

Matrícula: 120121-9

Título do Trabalho: Predição do Mercado Financeiro com uma arquitetura de extração de contexto baseada em decomposição de series temporais

Área de Concentração: Processamento de Sinais e Imagens

Orientador(a): Prof. Dr. Paulo Sérgio Silva Rodrigues

Data da realização da defesa: 05/08/2022

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

O aluno foi arguido por todos os membros da banca e respondeu adequadamente todas as perguntas. Assim, a banca aprovou o referido aluno por unanimidade.

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

APROVADO

REPROVADO

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Paulo Sérgio Silva Rodrigues

Prof. Dr. Reinaldo Augusto da Costa Bianchi

Prof. Dr. Bruno Travençolo

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Dedico este trabalho aos meus pais que me apoiaram a vida toda e fizeram com que isto se tornasse possível.

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Paulo Sergio Rodrigues, pela confiança em mim e no meu trabalho, por toda a paciência e atenção, por todas as aulas e ensinamentos passados a mim, e por me ajudar a amadurecer gradualmente como um pesquisador profissional ao longo dos últimos anos.

A todos os amigos e colegas que interagi ao longo do curso, pelo compartilhamento de ideias e dúvidas, e pelos momentos de discussão e descontração.

A todos os meus familiares. Principalmente à minha mãe e meu pai que fizeram com que esta conquista fosse possível.

Ao Centro Universitário FEI, pela infraestrutura, disponibilidade de recursos para o desenvolvimento do trabalho e pelos funcionários que me auxiliaram durante esse período.

“O mistério gera curiosidade e a curiosidade é a base do desejo humano para compreender.”

Neil Armstrong

RESUMO

O mercado de ações do Brasil se encontra em uma grande alta. Durante a pandemia da COVID-19, a bolsa brasileira teve um crescimento na quantidade de pessoas físicas de 92,1% entre os anos 2019 e 2020. Este fato se deve a diversos fatores, dos quais os principais estão: a disseminação do mercado de ações via redes sociais, propagandas em comunidades de vídeos e a baixa da taxa de juros do país neste período. Deste modo, com a alta de negociadores na bolsa, a busca por algoritmos que reduzem o risco nas operações de ativos, visando maximizar o lucro obtido na compra e venda de papéis, também cresce. Muitos trabalhos sugerem a utilização de métodos de *machine learning* para a predição de valores na bolsa. Contudo, obter resultados assertivos com a alta volatilidade, e não-linearidade das séries temporais da bolsa, ainda é um desafio. Sendo assim, este trabalho propõe uma nova metodologia para modelar matematicamente os dados financeiros da bolsa de valores, utilizando redes neurais recorrentes, técnicas de decomposição de séries temporais, correlação entre bolsas e ativos e arquiteturas de extração de contextos. Com a arquitetura proposta também foi possível fazer a predição de pontos de reversão de queda para alta ou de alta para baixa, dos valores de uma determinada ação. Como resultado, os melhores algoritmos para a regressão foram os que fizeram a utilização de redes *Long Short-Term Memory* (LSTM) e *Discrete Wavelet Transform* (DWT), com e sem utilização de extração de contextos, ambos com 0.9 de *Root Mean Squared Error* (RMSE). Porém, estes algoritmos são, em média, 38% melhor quando treinado somente na bolsa brasileira. Para a classificação os melhores algoritmos foram os que fizeram a utilização de LSTM juntamente com DWT, e *Echo State Network* (ESN) com DWT, com 35% e 34.6% de *f1-score*, respectivamente. Por fim, foi possível observar que, em geral, a metodologia proposta tem melhores resultados em relação às métricas estatísticas, quando comparado com algoritmos tradicionais que não fazem o uso da técnica de extração de contexto para séries temporais.

Palavras-chave: LSTM, ESN, Correlação de Ativos e Bolsas, Séries Temporais, Decomposição de Series Temporais, Bolsa de Valores, Predição, *Machine Learning*, Extração de Contexto, Redes Neurais Recorrentes.

ABSTRACT

The Brazilian stock market has been growing in a fast pace. It has increased 92,1% from 2019 to 2020, in quantity of investors. This is happening due to multiple factors, however, two of the main reasons is related to the spread of information about the stock market in the social media, propaganda in video communities and the decrease of the interest rate of the country. These factors contribute to a higher number of investors in the stock market, and hence, the search for algorithms that reduce the risk of trading stocks also grows. Many papers suggest the use of machine learning to predict the stock market. However, it is still a challenge to obtain results with good accuracy, due to non-linearity and non-stationarity, of the stock data series. Therefore, this work proposes a brand new methodology to mathematically predict financial data from the stock market, using recurrent neural networks, time series decomposition techniques, correlation between different stock markets and stocks and a context extraction architecture. With the proposed architecture it is also possible to predict reversal points, from down trend to up trend, or from up trend to down trend, on the time series of a stock. As result, the best algorithm for the regression was the one using *Long Short-Term Memory* (LSTM) and *Discrete Wavelet Transform* (DWT), with and without context extraction, both evaluated with 0.9 of RMSE. However these algorithms are, in average, 38% better when trained using only brazilian stock market. For the classification the best algorithms were the ones using LSTM with DWT, e *Echo State Network* (ESN) with DWT , both using context extraction, with 35% and 34.6% of f1-score, respectively. In conclusion, it was possible to observe that, in general, the proposed methodology showed better results, relating to statistics metrics, when compared to tradicional algorithms that do not use time series context extraction technique.

Keywords: LSTM, ESN, Stock Correlation, Time Series, Time Series Decomposition, Stock Market, Prediction, Machine Learning, Context Extraction, Recurrent Neural Network

LISTA DE ILUSTRAÇÕES

Ilustração 1 – Quantidade de artigos publicados agrupados por ano (cor) e separados por classes (<i>regression</i> e <i>classification</i>) e categoria de dados utilizados (<i>daily</i> ou <i>intra-day</i>)	35
Ilustração 2 – Quantidade de artigos publicados separados por ano (cor) e modelo utilizado	36
Ilustração 3 – Volume de trabalhos publicados e separados por ano, agrupados por tipo dado utilizado no treinamento	36
Ilustração 4 – Representatividade de variáveis utilizadas para a predição de valores na bolsa.	38
Ilustração 5 – Quantidade de artigos agrupados por qual bolsa foi utilizada além dos modelos utilizados.	39
Ilustração 6 – Representatividade dos modelos na literatura, cada quadrado representa em tamanho a representatividade do modelo e o número de artigos publicados com este modelo.	39
Ilustração 7 – Representatividade dos modelos nos trabalhos analisados para a bibliografia.	40
Ilustração 8 – Representatividade das métricas utilizadas na revisão realizada.	41
Ilustração 9 – Comparação das duas classes de série temporal.	44
Ilustração 10 – Decomposição da série temporal do preço de fechamento do ativo da empresa Apple Inc. (AAPL) entre os anos de 2014 e 2021	45
Ilustração 11 – Máximas e mínimas de uma função senoidal.	48
Ilustração 12 – Tipos de rede neural	50
Ilustração 13 – <i>Unfold</i> de uma <i>Recurrent Neural Network</i> (RNN)	51
Ilustração 14 – Célula da <i>Long-Short Term Memory</i> (LSTM) que mitiga o risco de desaparecimento de gradiente	52
Ilustração 15 – Tipos de redes neurais	54
Ilustração 16 – Exemplo de modelo hierárquico proposto por Seyedhosseini, Sajjadi, and Tasdizen (2013)	57
Ilustração 17 – Representação <i>Candlestick</i> sobre os dados <i>Open High Low Close</i> (OHLC)	58
Ilustração 18 – Gráfico comparativo entre as janelas de Média Móvel	60
Ilustração 19 – Gráfico de demonstração de bandas de Bollinger	61
Ilustração 20 – Comparação de normalização	65
Ilustração 21 – Matrix de confusão	66
Ilustração 22 – K-Fold	67
Ilustração 23 – K-Fold para séries temporais	68
Ilustração 24 – K-Fold para séries temporais tipo <i>sliding window</i>	68
Ilustração 25 – Fluxograma geral da metodologia proposta	70
Ilustração 26 – Arquitetura <i>Cascaded Hierarchical Model</i> (CHM) proposta	75
Ilustração 27 – Processo de Sintetização	76

Ilustração 28 – Janela de Validação	78
Ilustração 29 – RMSE utilizando somente a bolsa brasileira	82
Ilustração 30 – RMSE utilizando bolsas globais	82
Ilustração 31 – Série temporal da ação B3SA3 utilizando somente a bolsa brasileira do método CHM-LSTM-DWT, período de teste 2013-2016T-2017V	83
Ilustração 32 – Série temporal da ação B3SA3 utilizando somente a bolsa brasileira do método LSTM-DWT, período de teste 2013-2016T-2017V	83
Ilustração 33 – Série temporal da ação B3SA3 utilizando bolsas globais do método CHM-LSTM-DWT, período de teste 2013-2016T-2017V	84
Ilustração 34 – Série temporal da ação B3SA3 utilizando bolsas globais do método LSTM-DWT, período de teste 2013-2016T-2017V	84
Ilustração 35 – RMSE utilizando bolsas globais avaliada em apenas 30 dias	84
Ilustração 36 – RMSE da ação ABEV utilizando bolsas globais por período avaliado	85
Ilustração 37 – RMSE da ação ABEV utilizando somente a bolsa brasileira por período avaliado	86
Ilustração 38 – RMSE da ação ITUB3 utilizando bolsas globais por período avaliado	86
Ilustração 39 – RMSE da ação ITUB3 utilizando somente a bolsa brasileira por período avaliado	87
Ilustração 40 – RMSE da ação B3SA utilizando bolsas globais por período avaliado	87
Ilustração 41 – RMSE da ação B3SA utilizando somente a bolsa brasileira por período avaliado	88
Ilustração 42 – RMSE média dos modelos	88
Ilustração 43 – Métricas de classificação para a ação B3SA utilizando somente bolsa brasileira	89
Ilustração 44 – Métricas de classificação para a ação B3SA utilizando as bolsas mundiais	90
Ilustração 45 – Métricas de classificação para a ação ITUB3 utilizando somente bolsa brasileira	90
Ilustração 46 – Métricas de classificação para a ação ITUB3 utilizando as bolsas mundiais	91
Ilustração 47 – Métricas de classificação para a ação ABEV utilizando somente bolsa brasileira	91
Ilustração 48 – Métricas de classificação para a ação ABEV utilizando as bolsas mundiais	92
Ilustração 49 – Série de classificação com modelo CHM-LSTM-DWT para a ação ITUB utilizando somente bolsa brasileira, no período 2013-2016T-2017V	92
Ilustração 50 – Série de classificação com modelo LSTM-DWT para a ação ITUB utilizando somente bolsa brasileira, no período 2013-2016T-2017V	92
Ilustração 51 – Série de classificação com modelo CHM-LSTM-DWT para a ação ITUB utilizando bolsas mundiais, no período 2013-2016T-2017V	93
Ilustração 52 – Série de classificação com modelo LSTM-DWT para a ação ITUB utilizando bolsas mundiais, no período 2013-2016T-2017V	93
Ilustração 53 – <i>f1-score</i> média dos modelos	93
Ilustração 54 – Exemplo de retreino utilizando uma janela de 5 dias	94

Ilustração 55 – Série de predição dos valores da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas brasileiras	95
Ilustração 56 – Comparação dos valores de RMSE da ação ITUB no período 2010-2013T-2014V para as bolsas brasileiras utilizando o modelo CHM-LSTM-DWT com retreino e sem retreino	95
Ilustração 57 – Série de predição dos valores da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas mundiais	95
Ilustração 58 – Comparação dos valores de RMSE da ação ITUB no período 2010-2013T-2014V para as bolsas mundiais utilizando o modelo CHM-LSTM-DWT com retreino e sem retreino	96
Ilustração 59 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas brasileiras sem retreino	96
Ilustração 60 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas brasileiras com retreino	96
Ilustração 61 – Comparação dos valores de <i>f1-score</i> , Acurácia e Precisão do modelo CHM-LSTM-DWT com retreino e sem retreino sobre ação ITUB no período 2010-2013T-2014V para as bolsas brasileiras	97
Ilustração 62 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas mundiais sem retreino	97
Ilustração 63 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas mundiais com retreino	98
Ilustração 64 – Comparação dos valores de <i>f1-score</i> , Acurácia e Precisão do modelo CHM-LSTM-DWT com retreino e sem retreino sobre ação ITUB no período 2010-2013T-2014V para as bolsas mundiais	98
Ilustração 65 – Comparação dos valores de RMSE entre modelos que utilizam frequências instantâneas e os que não utilizam.	99
Ilustração 66 – Comparação dos valores de <i>f1-score</i> entre modelos que utilizam frequências instantâneas e os que não utilizam.	100

LISTA DE TABELAS

Tabela 1 – Exemplo da base de dados	71
Tabela 2 – Variáveis do <i>dataset</i> final	71
Tabela 3 – Ativos do <i>dataset</i> final.	72
Tabela 4 – Variações do modelo	77

LISTA DE ABREVIATURAS

ADX	<i>Average Directional Index.</i>
ANN	<i>Artificial Neural Network.</i>
ARIMA	<i>AutoRegressive Integrated Moving Average.</i>
ARMA	<i>AutoRegressive Moving Average.</i>
B3	<i>Brasil Bolsa Balcão.</i>
BB	<i>Bollinger Bands.</i>
BP	<i>Backpropagation.</i>
BPTT	<i>Backpropagation Through Time.</i>
CEEMDAN	<i>Complete Ensemble Empirical Mode Decomposition with Adaptive Noise.</i>
CHM	<i>Cascaded Hierarchical Model.</i>
CNN	<i>Convolutional Neural Network.</i>
CoV	<i>Coefficient of Variation.</i>
CPCA	<i>Complex Principal Component Analysis.</i>
CSO	<i>Competitive Swarm Optimizer.</i>
DFA	<i>Detrended Fluctuation Analysis.</i>
DL	<i>Deep Learning.</i>
DNN	<i>Deep Neural Network.</i>
DRP	<i>Down Reversal Point.</i>
DWT	<i>Discrete Wavelet Transform.</i>
EHM	<i>Efficient Market Hypothesis.</i>
EMD	<i>Empirical Mode Decomposition.</i>
ESN	<i>Echo State Network.</i>
ETF	<i>Exchange-Traded Funds.</i>
EWT	<i>Empirical Wavelet Transform .</i>
FL	<i>Fuzzy Logic.</i>
FN	<i>False Negative.</i>
FNN	<i>Factorization Machine based Neural Network.</i>
FOA	<i>Fruit Fly Optimization Algorithm.</i>
FP	<i>False Positive.</i>
GA	<i>Genetic Algorithm.</i>
GAN	<i>Generative Adversarial Network.</i>
GARCH	<i>Generalized Autoregressive Conditional Heteroskedasticity.</i>
IFR	<i>Índice de Força Relativa.</i>
IMF	<i>Intrinsic Mode Function.</i>
LS	<i>Local Search.</i>
LSTM	<i>Long-Short Term Memory.</i>

LSTM-CNN	<i>Long Short-Term Memory-Convolutional Neural Network.</i>
MACD	<i>Moving Average Convergence Divergence.</i>
MAD	<i>Mean Absolute Deviation.</i>
MAE	<i>Mean Absolute Error.</i>
MAPE	<i>Mean Absolute Percentage Error.</i>
MF DFA	<i>Multifractional Detrended Fluctuation Analysis.</i>
MF-X-DFA	<i>Multifractal Detrended Cross-Correlation Analysis.</i>
MLP	<i>Multi Layer Perceptron.</i>
MODWT	<i>Maximal Overlap Discrete Wavelet Transform.</i>
MSE	<i>Mean Squared Error.</i>
MST	<i>Minimum Spanning Tree.</i>
NASDAQ	<i>National Association of Securities Dealers Automated Quotations.</i>
NLP	<i>Natural Language Processing.</i>
NRMSE	<i>Normalized Root Mean Squared Error.</i>
NSE	<i>National Stock Exchange of India.</i>
NYSE	<i>New York Stock Exchange.</i>
OHLC	<i>Open High Low Close.</i>
PCA	<i>Principal Component Analysis.</i>
RMAE	<i>Root Mean Absolute Error.</i>
RMSE	<i>Root Mean Squared Error.</i>
RMT	<i>Random Matrix Theory.</i>
RNN	<i>Recurrent Neural Network.</i>
ROI	<i>Return Over Investment.</i>
RR	<i>Ridge Regression.</i>
S&P 500	<i>Standard & Poor's 500 Composite Stock Price Index .</i>
SAE	<i>Stacked Autoencoders.</i>
SAR	<i>Spatial Autoregressive.</i>
SMA	<i>Simple Moving Average.</i>
SMAPE	<i>Symmetric Mean Absolute Percentage Error.</i>
SO	<i>Stochastic Oscillator.</i>
SSEC	<i>Shanghai Stock Exchange Composite.</i>
STFT	<i>Short-Time Fourier Transform.</i>
SVM	<i>Support Vector Machine.</i>
SVR	<i>Support Vector Regressor.</i>
TN	<i>True Negative.</i>
TP	<i>True Positive.</i>
TRIX	<i>Triple Exponential Average.</i>
URP	<i>Up Reversal Point.</i>
VMD	<i>Variational Mode Decomposition.</i>

WMC *Wavelet Multiple Correlation.*
WT *Wavelet Transform.*

SUMÁRIO

1	Introdução	17
1.1	Objetivo	19
1.2	Estrutura do trabalho	19
2	Trabalhos Relacionados	20
2.1	<i>Long-Short Term Memory</i>	20
2.2	Echo State Networks	24
2.3	Decomposição de Séries Temporais	28
2.4	Correlação de Ativos e Bolsas	31
2.5	Artigos de Revisão	34
2.6	Análise da Bibliografia	40
2.7	Conclusão da Bibliografia	41
3	Conceitos Fundamentais	43
3.1	Series Temporais	43
3.2	Decomposição de Series Temporais	44
3.2.1	<i>Discrete Wavelet Transform</i>	46
3.2.2	<i>Trasformada de Hilbert-Huang</i>	47
3.3	Redes Neurais Recorrentes	49
3.3.1	<i>Long Short-Term Memory</i>	51
3.3.2	<i>Echo State Network</i>	53
3.4	Análise Contextual	55
3.5	Modelo Hierárquico em Cascata	56
3.6	Parâmetros do Mercado Financeiro	57
3.6.1	<i>Open High Low Close (OHLC)</i>	58
3.6.2	Frequência de Oscilação	59
3.6.3	Variáveis de Análise Técnica	59
3.6.3.1	<i>Médias Móveis</i>	59
3.6.3.2	<i>Bandas de Bollinger</i>	60
3.6.3.3	<i>Stochastic Oscillator (SO)</i>	61
3.6.3.4	<i>Índice de Força Relativa (IFR)</i>	62
3.7	Definição de Pontos de Reversão	62
3.7.1	<i>Trend</i>	63
3.7.2	Candidato A Ponto de Reversão	63
3.7.3	Pontos de Reversão	64
3.8	Normalização de Dados	64
3.9	Métricas de Avaliação	64
3.9.1	MSE	64
3.9.2	RMSE	65

3.9.3	Matriz de Confusão	65
3.9.3.1	Acurácia	66
3.9.3.2	Precisão	66
3.9.3.3	F1-Score	67
3.9.4	Validação Cruzada K-Fold	67
3.9.4.1	Validação Cruzada K-Fold Para Séries Temporais	67
4	Metodologia	69
4.1	Base de Dados	69
4.2	Preparação dos Dados	71
4.3	Treinamento do Modelo	73
4.3.1	CHM	74
4.3.1.1	Etapa Bottom-Up	74
4.3.1.2	Etapa Top-Down	76
4.4	Avaliação do Modelo	77
4.5	Implementação	78
5	Resultados e Discussão	80
5.1	Avaliação da Previsão do Preço Futuro	81
5.2	Avaliação da Previsão de Pontos de Reversão	85
5.3	Análise do Método de Retreino	91
5.4	Análise do Uso de Frequências Instantâneas	98
5.5	Contribuições	99
6	Conclusão	101
	REFERÊNCIAS	102

1 INTRODUÇÃO

A bolsa de valores é alvo de muitas especulações, onde compradores e vendedores tentam aproveitar o momento atual dos preços dos ativos para lucrar. O volume de negociações na bolsa de valores brasileira tem crescido de forma acelerada ao longo dos anos. A quantidade de investidores brasileiros na bolsa aumentou 92,1% de 2019 para 2020, foram contabilizadas 3.229.318 pessoas físicas no fim do ano de 2020 (GOEKING, s.d.). Este crescimento pode ser atribuído, em partes, pela redução do juros básico no Brasil (TORAMAN; BAŞARIR, 2014), que se encontrava em 13% a.a no início do ano de 2017, e em Junho de 2020 chegou a 2% a.a. O decaimento desta taxa básica pode amenizar o interesse de investidores em ativos de renda fixa, indexados por esta taxa, e direcionar esses investidores para ativos de renda variável, como ações. Porém, existe um risco eminente de operar grandes valores na bolsa, indicando a necessidade de métodos automatizados que auxiliem na negociação de papéis, reduzindo o risco das operações efetuadas.

Assim, a predição de valores da bolsa é um tópico de grande interesse, tanto por pessoas físicas, quanto por empresas de grande porte. Muitas corporações estabelecem departamentos especializados na compra e venda de papéis na bolsa, com o objetivo de reduzir riscos e ampliar o seu capital. Sendo assim, é evidente o valor gerado por uma metodologia computacional que possibilita a previsão de valores da bolsa. Diversos trabalhos utilizam técnicas estatísticas com o objetivo de prever pontos de inversão no preço de um ativo, para que assim seja possível vendê-lo, caso o ponto de inversão seja negativo, ou comprá-lo, caso o ponto de inversão identificado seja positivo. Porém, as séries históricas dos preços dos ativos têm um comportamento não-linear e com alta presença de ruídos, o que faz com que muitos métodos estatísticos triviais tenham baixa performance na predição destes valores (KUMAR; MURUGAN, 2013; SI; YIN, 2013).

Prever inversões no preço de um ativo não é uma tarefa simples. Embora existam muitas técnicas para auxiliar na compra e venda de papéis, a identificação de padrões em gráficos e valores de séries temporais pode apresentar uma série de problemas derivados da não linearidade dos valores financeiros da bolsa. Existem também diversos fatores que influenciam diretamente ou indiretamente os preços dos ativos, o que torna essa tarefa ainda mais complexa, como por exemplo: eventos políticos, como a troca de um líder político de um país; macro-econômicos, como a alteração de juros básico do país; e fatores socio-econômicos.

A Hipótese dos Mercados Eficientes, *Efficient Market Hypothesis* (FAMA, 1970), relata que os preços dos ativos da bolsa não se dão por comportamentos que obedecem padrões, e sim por eventos estocásticos que ocorrem externamente à bolsa de valores mas a afeta diretamente. Um exemplo atual é a ocorrência de uma crise mundial movida pela pandemia da COVID-19. Esses eventos, por sua vez, são complexos demais para serem previstos por modelos matemáticos simples. Entretanto, diversos trabalhos mostram a possibilidade de modelar padrões em determinadas janelas de tempo na bolsa (MALKIEL, 2003; SMITH, 2002; NOFSINGER, 2005;

FILHO; ROCHA, 2020). O estado da arte também mostra como os fatores socio-econômicos podem auxiliar na previsibilidade dos valores da bolsa (PRECHTER; PARKER, 2007). Deste modo, esses fatores podem ser explorados para a predição de valores na bolsa, utilizando modelos de *machine learning*, como RNN, que realizam a identificação automática de padrões em quaisquer conjuntos de dados que sejam apresentados.

Os autores Minh et al. (2018) sugerem a classificação dos modelos de predição que trabalham sobre os dados da bolsa de valores, em dois tipos, são eles: modelos de análises técnicas e modelos de análises fundamentalistas. O primeiro se baseia em valores históricos do preço do ativo e procura formular modelos matemáticos que representem o comportamento da bolsa de valores. O segundo tipo de análise faz uma análise das características administrativas e contábeis da empresa analisada, visando modelar o preço do ativo se apoiando na expectativa do mercado sobre a empresa. A grande maioria dos trabalhos atuais foca na primeira técnica apresentada (KUMAR; JAIN; SINGH, 2020).

Como as séries da bolsa de valores são, de forma geral, não lineares e não estacionárias, para auxiliar os modelos de predição a terem melhores resultados, o estado da arte apresenta algumas metodologias que agem como suporte para os modelos de predição. Um tópico que apresenta relevância para auxiliar os modelos de predição de valores de ativos é a correlação entre ativos da mesma bolsa, ou entre bolsas. Existem diversos trabalhos que visam utilizar esta correlação para obter valores mais assertivos na predição de valores financeiros (TIWARI et al., 2013; KANTELHARDT et al., 2002; ASGHARIAN; HESS; LIU, 2013). Embora esta seja mais uma técnica para aperfeiçoar os modelos de predição, a não-linearidade dos valores ainda tem um impacto relevante na performance dos modelos.

Outra metodologia suporte, que diversos trabalhos recentes têm utilizado, é a utilização da técnica estatística de decomposição de séries temporais (STOCCHI; MARCHESI, 2018; ALTAN; KARASU; BEKIROS, 2019; NIU; XU; WANG, 2020; ZHOU, F. et al., 2019; CAO; LI; LI, 2019; CHANDAR; SUMATHI; SIVANANDAM, 2016; YU-JUN; YI-MEI; JIAN-HUA, 2020). Esta técnica visa decompor uma série em diversas sub-séries mais lineares que, juntas, compõem a série original. Ao utilizar as sub-séries extraídas é possível reduzir o impacto da alta volatilidade dos preços da bolsa de valores e permitir que os modelos trabalhem em dados que apresentam menos ruídos.

Uma outra metodologia suporte que pode ser explorada é a análise de contexto. Em visão computacional, diversos trabalhos utilizam esse conceito para a análise de imagens. O trabalho de Seyedhosseini, Sajjadi e Tasdizen (2013) propõe uma arquitetura para a extração de contextos em imagens. Assim, no trabalho presente, esta arquitetura será modificada para que sejam analisadas séries temporais ao invés de imagens.

Por fim, com o cenário aqui apresentado, este trabalho propõe uma nova metodologia baseada nos conceitos apresentados para a predição dos valores dos ativos listados na bolsa de valores brasileira. A arquitetura proposta é baseada na utilização de: correlação entre bolsas de valores e ativos, decomposição de séries temporais, RNNs e extração de contexto.

1.1 OBJETIVO

Este trabalho tem como objetivo encontrar padrões nas bolsas mundiais do mercado financeiro para realizar a previsão do preço de fechamento diário de ativos da bolsa *Brasil Bolsa Balcão* (B3), utilizando redes RNN e técnicas de decomposições de séries temporais.

1.2 ESTRUTURA DO TRABALHO

O restante do trabalho é definido da seguinte forma:

O Capítulo 2 apresenta diversos trabalhos que são relacionados com o tema aqui abordado. Esse capítulo tem como objetivo expor, e contextualizar, o cenário do estado da arte do tema.

No Capítulo 3 serão descritos os funcionamentos dos conceitos utilizados na metodologia deste trabalho, para que a leitura da metodologia seja clara para o leitor.

O Capítulo 4 faz o detalhamento da metodologia utilizada para o desenvolvimento do objetivo proposto, apresentando o funcionamento da arquitetura e detalhes da implementação.

O Capítulo 5 apresenta os resultados gerados a partir da metodologia proposta, assim como uma discussão sobre os dados finais analisando o desempenho dos algoritmos em diferentes cenários.

Por fim no capítulo 6 é feita a conclusão, trazendo um breve resumo do trabalho e também sobre os objetivos alcançados e próximos passos.

2 TRABALHOS RELACIONADOS

Existem diversos trabalhos publicados que têm como objetivo a predição de ativos na bolsa de valores. Nesta seção serão apresentadas algumas dessas metodologias propostas que apresentam relação com o trabalho presente. O capítulo será dividido nas seguintes seções: LSTM Seção 2.1, *Echo State Network* (ESN) Seção 2.2, Decomposição de Séries Temporais Seção 2.3, Correlação de Ativos e Bolsas Seção 2.4, Artigos de Revisão Seção 2.5, Análise da Bibliografia Seção 2.6 e Conclusão da Bibliografia Seção 2.7.

2.1 LONG-SHORT TERM MEMORY

As redes LSTM, propostas por S. Hochreiter e J. Schmidhuber (1997), são frequentemente utilizadas para a predição de valores no mercado financeiro devido à capacidade que a rede tem de utilizar valores anteriores para auxiliar na predição dos valores futuros, assim como as redes recorrentes. Porém, essas redes são preferidas pelo fato de conterem uma célula de memória de longo prazo que mitiga o problema de *vanishing gradient* que impede que a rede aprenda de uma forma efetiva.

Um dos trabalhos mais reconhecidos que utilizam LSTM é o de Bao, Yue e Rao (2017), onde os autores propõem uma metodologia que consiste na junção de *Wavelet Transform* (WT), *Stacked Autoencoders* (SAE) e LSTM. Primeiramente, as séries são decompostas pela WT visando a redução de ruídos nas séries. Em seguida os dados suavizados são passados pela SAE, que tem como objetivos aprender uma representação dos dados de uma forma não-supervisionada. Por fim, os dados transformados pela SAE são enviados para a LSTM, por sua vez, que tem como objetivo prever um ponto à frente dos que foram utilizados como treinamento. Para os testes, os autores fazem a análise de *Mean Absolute Percentage Error* (MAPE) e rentabilidade, geradas pelo modelo, e comparadas contra outros 3 modelos. Segundo os autores a metodologia proposta tem melhor performance tanto em MAPE quanto em rentabilidade.

Os autores Selvin et al. (2017) usam *Deep learning* para prever o preço de empresas listadas na bolsa da Índia. Os autores utilizaram a técnica de *sliding window* que consiste em treinar o modelo entre os valores do tempo t_n e t_{n+x} , realizar o *Backpropagation* (BP) e deslizar a janela de treinamento de t_n e t_{n+x} para t_{n+y} e t_{n+y+x} , onde y pode ser qualquer valor. Para determinar o valor de x , os autores realizaram vários testes utilizando o tamanho de x que gera o menor erro. Os autores compararam o uso de *deep learning* em três modelos, sendo eles *Convolutional Neural Network* (CNN), RNN e LSTM. Segundo os autores, as redes tiveram capacidade de aprender dinâmicas escondidas na série. A rede que teve mais sucesso foi a CNN, devido a sua capacidade de identificar mudanças bruscas na direção da série.

Utilizando *deep learning* e *Natural Language Processing* (NLP) os autores Akita et al. (2016) tentaram prever o preço de abertura dos ativos de diversas empresas. Como metodologia, os autores utilizaram técnicas de transformação de texto para vetores compreendidos pela LSTM,

e como entrada, além dos valores textuais, os autores incluem os preços dos ativos das empresas que serão utilizadas como variável de saída. Diferente dos demais artigos, os autores têm como saída um vetor composto pelo preço das empresas analisadas, onde geralmente seria somente um valor. Segundo os autores, a LSTM teve bons resultados em conseguir prever os valores futuros dos ativos.

No trabalho de Sachdeva et al. (2019), é proposta uma metodologia que utiliza redes recorrentes para a predição de valores dos ativos *NIFTY 50 index* e *INFYOSYS Ltd*. Para solucionar o problema de *vanishing gradient*, os autores utilizam a estrutura de uma LSTM. Como entrada da rede, os autores utilizam os preços de abertura, fechamento, preço máximo e preço mínimo dos ativos listados na bolsa no período de 11 de Dezembro de 2007 até 11 de Dezembro de 2017. Os autores normalizam os valores de entrada utilizando o método *MinMaxScaler* para que as funções de ativação *sigmoid* tenham melhor funcionamento. Como resultado, os autores apresentaram os valores de acurácia de 97,64%.

Diversos trabalhos demonstram que redes LSTM são uma boa alternativa para a modelagem de séries temporais, não apenas pela sua capacidade de encontrar padrões baseados em valores anteriores da série em questão, mas também pelo fato de possibilitar o uso de outras variáveis independentes em conjunto com os valores da série, como apresentado no trabalho de Li, Bu e Wu (2017), que combina os valores diários de ativos da bolsa de valores com valores de análises de sentimentos retirados de *posts* de fóruns da internet. Os valores de sentimentos são classificados em Neutro, Positivo ou Negativo, utilizando o método de classificação *Naive Bayes*. Os autores afirmam que o modelo pode chegar a atingir uma AUC (*Area Under the Curve*) de 94,69%, colocando em questionamento a *Efficient Market Hypothesis* (EHM), uma hipótese que sugere que o mercado financeiro é eficiente ao ponto de não ser possível prevê-lo (FAMA, 1970).

Um trabalho recente que utiliza a LSTM para a predição da tendência de um ativo da bolsa foi o trabalho de U et al. (2020). Os autores utilizaram uma combinação de informações de *candle sticks* com análise técnica, resultando num total de 27 *features* a serem utilizadas no modelo. O objetivo do trabalho visava encontrar pontos de reversão de tendência, para identificar se a queda ou a alta de um ativo, é temporária ou permanente. Para os testes, os autores utilizaram 10 ativos da bolsa de valores da China e 10 ativos da bolsa dos Estados Unidos. A média de score F1 para os ativos americanos foi de 68,8% enquanto para os ativos chineses a média foi de 55,2%. Os autores afirmam que a acurácia aumentou em 5,9%, 11,7% e 5,3%, respectivamente em relação aos métodos *Support Vector Machine* (SVM), *Multi Layer Perceptron* (MLP) e CNN.

Os autores Hiransha et al. (2018) tentam prever valores de cinco diferentes empresas listadas nas bolsas *New York Stock Exchange* (NYSE) e *National Stock Exchange of India* (NSE). Para isto, os autores utilizaram a técnica de *sliding window*, em quatro modelos diferentes de *deep learning*, são eles: RNN, CNN, LSTM e MLP. Como resultado, os autores sugeriram que todos os métodos performaram melhor que os algoritmos lineares como o *AutoRegressive*

Integrated Moving Average (ARIMA). E entre os modelos de *Deep Learning* (DL) as CNN tiveram os melhores resultados.

Um trabalho frequentemente citado, e do estado da arte, é o de Kim e Kim (2019), onde os autores propõem um modelo denominado de *Long Short-Term Memory-Convolutional Neural Network* (LSTM-CNN) que tem como objetivo aprender diversas representações dos mesmos dados combinando imagens de gráficos de ações e os dados da série temporal do ativo. Para os dados de imagens, os autores utilizam a rede CNN, e para os dados numéricos e temporais a rede LSTM. Logo depois, os autores juntam as duas em uma camada totalmente conectada que fará a predição dos valores. Segundo os autores, os resultados para o modelo LSTM-CNN foi de 0,098 para RMSE, 0,2291 para *Root Mean Absolute Error* (RMAE) e 0,0209 para MAPE, que são, respectivamente, 22,09%, 20,89%, e 38,17% a menos dos valores obtidos pelo modelo CNN, e 18,18%, 17,56%, e 32,87%, respectivamente, a menos do que o modelo LSTM.

No trabalho de X. Zhou et al. (2018) os autores fazem o uso de redes *Generative Adversarial Network* (GAN), onde as redes agem como adversárias. O modelo de rede GAN, são divididas em duas topologias: a topologia G, *generative*, e a topologia D, *discriminative*. A topologia G tenta aprender o comportamento de distribuição dos dados de entrada, tendo como processo de treinamento maximizar a probabilidade da topologia D cometer um erro. A topologia D tem como objetivo definir se os dados que recebeu vieram da topologia G ou dos dados de treinamento do modelo. As topologias podem ser qualquer tipo de rede neural. Para a arquitetura, os autores utilizaram como topologia G uma rede LSTM, e como topologia D uma rede CNN. Os autores usam treze indicadores técnicos simples para o treinamento das redes. Segundo os autores, a metodologia proposta supera outros modelos adversários do estado da arte, tanto em acurácia na predição do valor do ativo, quanto a acurácia na predição direção do preço do ativo.

As LSTM também são utilizadas em conjunto com outras técnicas como por exemplo a CNN. No trabalho de Lu et al. (2020), os autores utilizaram uma rede RNN como um extrator de *features* que são utilizadas como entrada de uma outra rede LSTM. Segundo os autores, tal arquitetura foi adotada pelo fato das CNN terem a capacidade de extrair *features* de bases de dados, enquanto as LSTM conseguem identificar padrões nas séries temporais, como é o caso da bolsa de valores. Diferente do trabalho de U et al. (2020) os autores trabalharam com o objetivo de prever o valor da ação em si, e não em identificar um ponto de reversão de tendência. Para realizar os experimentos, os autores utilizaram o índice de Changai, utilizando 7127 dias, entre 1 de Julho de 1991 e 31 de Agosto de 2020. Para o treinamento foram utilizados os primeiros 6627 dias e o restante, 500 dias, foram utilizados para o teste do modelo. Tanto em termos de R^2 , como *Mean Absolute Error* (MAE), e RMSE, a arquitetura CNN-LSTM teve os melhores resultados, embora que marginais. Para a CNN-LSTM foram encontrados os seguintes valores de métrica: 0,9646 para R^2 , 39,688 para RMSE, e 27,564 para MAE. É importante ressaltar que os métodos mais efetivos foram os que utilizavam alguma rede recorrente, como LSTM ou RNN.

Ainda no âmbito de combinar diversos modelos para atingir um único objetivo, o trabalho de Kelotra e Pandey (2020) faz a junção de uma LSTM com uma rede convolucional profunda para prever os valores de ativos. Porém, em contraste com o trabalho de Lu et al. (2020), a rede convolucional não é utilizada para a extração de *features*, e sim para auxiliar a LSTM na identificação de padrões. Como entrada para a rede, os autores utilizaram diversos indicadores técnicos. Em seguida, os indicadores são agrupados utilizando uma metodologia de clusterização. Como otimização de tempo de treinamento da rede profunda, os autores utilizaram um algoritmo bio-inspirado Rider-MBO para treinar os pesos da rede. Para treino, foram utilizados os papéis das empresas *Infosys, HCL Technologies, Wipro, Bharti Infratel, Larsen & Toubro Infotech, e Siemens*, de Janeiro de 2017 até Dezembro de 2018. Os autores afirmam terem reduzido a métrica RMSE em 2.6923 pontos, utilizando a metodologia proposta.

No estudo de Yu-jun, Yi-mei e Jian-hua (2020), os autores citam como é frequente os valores dos ativos nas bolsas de valores terem um comportamento não linear e de difícil compreensão, até mesmo para os algoritmos de previsão de series temporais. Sendo assim, os autores propuseram uma metodologia que visa decompor a série temporal da bolsa em sub-séries que compõem a série original. Para isso os autores utilizaram o algoritmo de EEMD (*Extended Empirical Mode Decomposition*), que pode ser comparado com técnicas como transformada de Fourier e Wavelets. A metodologia proposta pelos autores consiste em dividir a série temporal dos valores do ativo em estudo em n sub-séries, cada sub-séries teria um modelo de LSTM especializado e que são treinados individualmente. Ao final, uma etapa de fusão faz a junção das n previsões das n LSTM. Os experimentos dos autores mostram que o método proposto têm resultados significativos em termos de R^2 , MAE e RMSE.

O trabalho de Nelson, Pereira e Oliveira (2017) também mostra como as LSTM podem ser efetivas para prever valores da bolsa, desta vez os autores utilizaram informações do mercado da bolsa Brasileira, assim como este trabalho também o fará. É comum a utilização de análise técnica como entrada para os modelos. Cerca de 25% dos trabalhos, que têm como objetivo prever valores na bolsa, publicados entre 2017 e 2019, utilizaram análise técnica como *features* para algoritmos de machine learning (JIANG, 2020). Os autores Nelson, Pereira e Oliveira (2017) utilizaram uma rede LSTM, que tem como entrada os valores históricos de ativos e *Exchange-Traded Funds* (ETF) da bolsa de valores brasileira, juntamente com valores de análise técnica, para prever se o preço do ativo em análise irá ser maior ou menor que o atual em um período de 15 minutos no futuro. Assim, os autores esperam aprimorar os resultados da rede embasando seu treinamento em uma tipo de modelo fundamentado. Neste trabalho, as redes LSTM, foram implementadas com a característica de treino constante. Ao final de cada dia, a rede é retreinada utilizando os valores dos últimos 10 meses. Os testes foram baseados em quatro ações da bolsa e um ETF, usando os valores dos dez meses anteriores para o treino da rede, e os valores da semana anterior para os testes. Os autores utilizaram valores da bolsa de 2008 até 2015. Para o processamento, os autores normalizaram a série utilizando a seguinte expressão denominada log-return: $\log(p_i) - \log(p_{i-1})$, este processamento é utilizado para estabilizar a média e a va-

riância da série temporal tornando-a estacionária. Os autores apresentaram resultados de, em média, 55,9% de acurácia. Importante ressaltar que neste trabalho foram utilizadas somente informações baseadas no preço dos ativos.

Um modelo que traz resultados estatisticamente relevantes é o *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) (FU, 2009). Considerando isso, o trabalho de Kim e Won (2018) utiliza LSTM apoiadas no modelo GARCH para tentar melhorar ainda mais os resultados da predição de volatilidade da bolsa. A metodologia proposta baseia-se em utilizar como entrada da rede os valores de múltiplos modelos GARCH, sob a hipótese de que se cada variante do modelo GARCH, por exemplo o EGARCH, tem um potencial em previsão de séries temporais, então a junção dos modelos, com uma rede LSTM, tem a potência de todas as variantes do modelo GARCH. Como resultados, os autores apresentaram uma melhora de 10% em MAE sobre o modelo LSTM tradicional. Os testes foram executados sobre o ativo *KOSPI 200 index* da bolsa da Coreia. Para o treino foi utilizado o período de Janeiro 2001 até Setembro 2011, e para os testes os autores aplicaram os modelos sobre o período de Outubro de 2011 até Janeiro de 2017.

No trabalho de Borovkova e Tsiamas (2019) é proposta uma arquitetura de *ensemble* de LSTM; isto é, o modelo utiliza diversas instâncias da rede LSTM com o objetivo de melhorar a capacidade preditiva da arquitetura proposta para classificar momentos de compra e momentos de venda na bolsa. Os autores propõem uma metodologia que utiliza doze LSTM devido à quantidade de *features* utilizada. Como entrada da rede, os autores geraram diversas *features*, tais como: o preço histórico da ação, os competidores principais da empresa analisada, indicadores técnicos, auto regressões da série, etc., totalizando em 400 *features*. Para os treinamento, os autores utilizaram uma sequência de entrada para a LSTM de tamanho 5, e uma entrada de batch com 32 observações. Os autores fizeram testes para 22 ações americanas diferentes, prevendo a tomada de decisão a cada 5 minutos, afirmando que o modelo tem resultados melhores que escolhas aleatórias na maioria das vezes.

As pesquisas apresentadas nesta seção mostram a relevância das redes LSTM na área de predição de valores financeiros. A seção também evidencia os bons resultados obtidos com a rede, indicando ser uma forte candidata para aplicações de previsão de valores na bolsa de ações.

2.2 ECHO STATE NETWORKS

As redes ESN proposta por Jaeger (2007) não são uma arquitetura nova de redes, porém apenas recentemente ganharam mais atenção no mercado financeiro. Assim como as redes LSTM, as redes ESN têm como principal característica a recorrência. Dessa forma, este tipo de rede consegue levar em conta os valores passados para predizer um novo valor futuro, ideal para os valores da bolsa. Porém, em contraste com a LSTM as redes ESN têm um tempo de treinamento mais reduzido, e podem trazer bons resultados para a predição de valores temporais.

Entretanto, a quantidade de trabalhos publicados que utiliza este algoritmo para a predição de valores de ativo na bolsa ainda é pequena.

Ilies et al. (2007) utilizaram a rede ESN para prever dados de séries temporais, embora o objetivo do presente trabalho seja a aplicação de modelos em séries temporais de ativos da bolsa de valores da B3, o trabalho de Ilies et al. (2007) se assemelha em sua metodologia de decomposições de séries temporais e utilização de redes ESN, para aplicações em séries financeiras. Os autores optaram por utilizar um *ensemble* de ESN, onde toda ESN é treinada sob os mesmos dados e no fim é aplicada a média sobre as predição de todas as ESN. Para o treino, os autores utilizaram a técnica *holdout*, separando os últimos doze pontos da série como teste. Como resultado, é mostrado que as ESN tem um valor de *Symmetric Mean Absolute Percentage Error* (SMAPE) inferior a uma SVM de regressão.

Dois anos depois, Lin, Yang e Song (2009) utilizaram as redes ESN para prever valores na bolsa de valores americana. O objetivo dos autores era prever o valor de fechamento do dia da ação. Para isso utilizaram um modelo que considera apenas os valores prévios da ação, sem considerar nenhum tipo de indicador técnico. Segundo os autores, a predição dos valores da bolsa em um curto período de tempo não é diretamente afetada por fatores externos. Sendo assim, apenas os valores prévios seriam suficiente para a obtenção de bons resultados de curto prazo. Além disso, os autores utilizam a técnica de *Hurst Exponent* que identifica memórias de longo prazo em uma série temporal. Assim, os autores selecionaram qual seria o início da série temporal baseando-se em qual subsérie possui maior valor de *Hurst Exponent*. Para os experimentos foram considerados 491 ações da bolsa, com os valores entre Dezembro de 2001 e Novembro de 2005, somando um total de 1100 dias. Os primeiros 1000 dias foram utilizados para o treino da rede e os últimos 100 não foram incluídos no treinamento para serem utilizados como teste. Os autores concluíram que a rede ESN é apropriada para o uso de séries temporais financeiras, mas que, para cada ação da bolsa, é necessário achar parâmetros ótimos para a rede, sugerindo assim usar *Genetic Algorithm* (GA) para a seleção de parâmetros.

Um dos trabalhos que tentaram seguir por esta linha e utilizar GA para a otimização dos parâmetros da rede ESN, foi o trabalho de J. Liu et al. (2018). Nele, os autores também têm por objetivo a utilização da rede na bolsa de valores para prever valores futuros. Porém, propõem uma metodologia nova que utiliza o algoritmo bio inspirado *Fruit Fly Optimization Algorithm* (FOA). Este algoritmo se baseia no conceito de *swarm intelligence* que, por sua vez, se baseia no comportamento de enxames para otimizar um alvo. O algoritmo FOA, especificamente, baseia-se no comportamento de moscas de frutas. Os algoritmos bio-inspirados frequentemente têm como principal objetivo a otimização de algum parâmetro, ou parâmetros, específicos. No caso do trabalho de J. Liu et al. (2018), o algoritmo FOA foi utilizado para otimizar quatro parâmetros da rede ESN, sendo eles: a quantidade de neurônios N ; o *Sparse Degree* SD , que determina a proporção de ligação entre os neurônios da camada oculta; *Spectral Radius* SR , que demonstra o estado de eco da rede quando se apresenta menor que 1; e, por último, *Input Scale* IS , que representa um valor utilizado para transformar a escala dos valores da matriz de peso de entrada.

Os autores afirmam que o método proposto tem resultados mais satisfatórios que os métodos tradicionais como o *backpropagation*. Em uma tabela, foram apresentadas as métricas *Normalized Root Mean Squared Error* (NRMSE) e R^2 . Para os valores de NRMSE, o *backpropagation* apresentou um valor de 0,0713 contra 0,0379 da metodologia proposta, e para os valores de R^2 o *backpropagation* apresentou um valor de 0,9948 contra 0,9986 da metodologia proposta.

Um ano depois, seguindo a ideia de utilizar GA em ESN, os autores Long, Zhang e Li (2020) propuseram uma evolução para as redes ESN profundas. As redes contêm alguns hiperparâmetros a serem escolhidos, tais como spectral radius, connectivity rate. Dessa forma, os autores utilizam o algoritmo *Competitive Swarm Optimizer* (CSO) juntamente com *Local Search* (LS) para encontrar automaticamente os valores dos hiperparâmetros da rede, aumentando assim a sua acurácia final independente do problema em que será aplicada. Nos teste aplicados pelos autores a rede que utiliza GA apresentou sempre os melhores resultados.

Segundo Dan et al. (2014) um dos problemas que impede a otimização da rede é a presença de aleatoriedade na construção da camada oculta. Sendo assim, os autores propuseram um modelo denominado *Deterministic Echo State Network*, que tem como princípio básico retirar a característica de aleatoriedade na construção da camada oculta, impondo três topologias fixas, denominadas de *Delay line reservoir* (DLR), *Delay line reservoir with Backwards Connections* (DLRB) e *Simple Cycle Reservoir* (SCR). Para os testes, os autores utilizaram o Índice da bolsa de Changai e o índice da bolsa americana. Os resultados no índice de Changai mostraram que as redes com topologias fixas obtiveram 20% mais acurácia que as redes tradicionais, porém quase nenhuma melhora no índice da bolsa americana.

Um dos trabalhos mais recentes que utiliza ESN é o trabalho de Patel, Patel e Cuesta (2020). Os autores discorrem sobre a capacidade das ESN em prever séries temporais caóticas devido a presença do seu reservatório de neurônios. Para trazer mais resultados, os autores não fazem somente a predição dos valores de um único ativo, mas sim de diversos ativos. E em seguida fazem a seleção do portfólio de investimentos utilizando o algoritmo *K-Means* para a clusterização dos ativos. Após a clusterização dos ativos, os autores executam testes de Monte Carlo para otimizar a seleção dos ativos e, conseqüentemente, aumentar o retorno dos investimentos. Para calcular o teste dos retornos, os autores fazem as compras durante a semana e, ao fim dela, todos os ativos são vendidos, sem exceção, para assim descartar um possível viés de uma ou mais ações aumentarem o retorno do portfólio. Os autores afirmam que a rede consegue prever os valores com mais de 95% de confiança, porém nem sempre têm bons resultados ao prever retornos. Entretanto, é afirmado que apenas os dados históricos não foram suficientes para prever os ativos com acurácia.

Como um dos principais motivadores para a utilização das redes ESN é o seu tempo de treinamento reduzido, é possível modelar múltiplos ativos listados em uma bolsa em um período de tempo factível, como Bernal fez em seu trabalho em 2012 (BERNAL; FOK; PIDAPARTHI, 2012) selecionando 50 ativos aleatórios listados na S&P500, um indicador da bolsa americana que contem as maiores 500 empresas listadas em bolsa. Os autores coletaram dados entre 2004

e 2009, e, além dos principais valores OHLC, o trabalho utilizou médias móveis de 5, 10, 15 e 20 dias, o volume do ativo em bolsa e o próprio valor do índice S&P500. Como a rede ESN depende muito de seus estados iniciais, os autores propuseram passar os dados pela rede por um período de tempo antes de realizar o treinamento dos pesos da rede, para que assim o estado inicial dos pesos da rede sejam menos propensos a direcionar a rede para uma má performance. Para a validação do modelo, os autores conduziram os experimentos com a técnica de *Leave-One-Out*, que consiste em treinar a rede com o máximo de dados possível deixando apenas um único ponto de dados para o teste. Os autores compararam a rede com o filtro de Kalman, que consistem também em estimar valores mas baseando-se nas grandezas da série temporal. Como conclusão, é tomado que a rede ESN consegue fazer a modelagem de rápidas alterações no movimento da série, enquanto o filtro de Kalman não.

Barjaktarović, Barjaktarovic e Konjikuić (2020) discute sobre as vantagens do tempo computacional das ESN e também ressalta o fato da técnica ainda não ter muitos trabalhos publicados que a utilizam para a predição de valores em séries temporais da bolsa de valores. Este fato abre um espaço interessante para pesquisas futuras. Assim como já citado na Seção 1.1, o presente trabalho também tem como objetivo fazer uma comparação entre os métodos para verificar a eficácia da rede ESN em séries da bolsa de valores. Segundo o autor, as redes ESN possuem melhores resultados quando a série apresentada para seu treinamento tem uma característica caótica, e não tem comportamentos lineares ou estacionários, sendo este um dos principais motivadores para utilizar as redes ESN para tentar prever tendências na bolsa.

Porém, apesar das vantagens de utilizar redes ESN, alguns trabalhos afirmam ter problemas de *overfit*, como no caso de Zhengfu Liu et al. (2017); isto é, a rede não consegue generalizar ao predizer os valores, assim, as métricas para avaliação de performance de modelo são satisfatórias no domínio de treinamento, porém, ao sair deste conjunto a rede falha em predizer os valores com acurácia. Isto inviabiliza a utilização da mesma para uma aplicação real. Para tentar mitigar este problema, os autores de (LIU, Z. et al., 2017) tentam prever os valores da bolsa utilizando uma metodologia proposta denominada *Multi-objective Diversified Echo State Network* (MODESN).

Os autores afirmam que um dos principais motivos para a rede ESN tender ao *overfit* é a quantidade de neurônios localizados na camada escondida. Entretanto, quanto menor a quantidade de neurônios nessas camadas, mais é necessário que eles carreguem informações diferentes sobre a entrada do modelo, para que a rede tenha uma memória significativa da série temporal, melhorando, assim, a acurácia na predição dos próximos valores. Sendo assim, os autores propuseram otimizar a distância entre as séries temporais de entrada, para que seja obtida a menor quantidade de informações redundantes possível. Seguindo esta hipótese, os autores utilizaram algoritmos bio-inspirados, mais especificamente *Multi-objective evolutionary algorithm* (MO-EAs), para otimizar a distância entre as séries e evitar o *overfit*. Por fim, a metodologia gera 200 redes ESN com 70 neurônios apenas na camada escondida. Após o treinamento, as dezoito redes que apresentarem maior distância entre as séries, e maior acurácia, são selecionadas e in-

tegradadas por votos no conjunto de testes. A acurácia do modelo integrado na base de teste foi de 76,1%, mostrando que a metodologia proposta é promissora tendo mais acurácia que o modelo tradicional e menos propensão a *overfit*.

Uma arquitetura de redes ESN profundas é apresentada por Kim e King (2020). Com esta arquitetura, juntamente com a decomposição aditiva da série, os autores esperam melhorar o resultado da predição de valores em séries temporais. A rede ESN profunda tem a característica da presença de N camadas ocultas, interligadas por pesos que são treinados pelo algoritmo BP. O trabalho apresenta testes da rede ESN profunda na predição do índice estadunidense S&P500, utilizando dados de 14 de Agosto de 1999 até 01 de Julho de 2019, com a técnica de validação *holdout* com 80% dos dados para treino e o restante para testes. Assim os autores aplicam na série o teste de Dickey-Fuller, que indica a presença de uma série não estacionária, o que, normalmente, indica uma série de difícil previsão. Os autores afirmam que os resultados são extremamente positivos ao utilizar a rede ESN profunda para a previsão de séries não estacionárias, como a bolsa de valores.

Os trabalhos mostrados anteriormente evidenciam a capacidade que as redes ESN têm de se adaptar a séries temporais caóticas. O que é um indício que sua aplicação na bolsa de valores possa trazer resultados satisfatórios.

2.3 DECOMPOSIÇÃO DE SÉRIES TEMPORAIS

A decomposição de series temporais é utilizada para simplificar uma série temporal quebrando a mesma em diversas partes. Existem diversos algoritmos que fazem este processo, como o *Empirical Mode Decomposition* (EMD) (HUANG et al., 1998), *Empirical Wavelet Transform* (EWT) (GILLES, 2013), *Variational Mode Decomposition* (VMD) (DRAGOMIRETSKIY; ZOZZO, 2014) e *Discrete Wavelet Transform* (DWT). Uma série que antes poderia não apresentar resultados em um modelo de redes neurais como a ESN, ou a LSTM, pode apresentar bons resultados depois de ser processada por uma decomposição de séries temporais.

Os autores Stocchi e Marchesi (2018) utilizam a DWT para decompor séries de criptomoedas para tentar aprimorar a predição da série. Os autores afirmam que a técnica tem bons resultados, não somente no conjunto de dados utilizados como teste mas também em diversos outros que apresentam uma característica temporal, como a bolsa de valores.

Para prever valores de moedas digitais Altan, Karasu e Bekiros (2019) utilizaram a transformada EWT para decompor a série das moedas BTC, XRP, DASH and LTC, em n sub séries denominadas *Intrinsic Mode Function* (IMF), que foram utilizadas como entrada, posteriormente, para as n LSTM. Cada rede tem como entrada uma única sub série. Inicialmente, os autores utilizaram a função *MinMaxScaler* para manter os valores das séries entre 0 e 1. Isto é feito para manter a mesma importância entre todas as sub-séries, impedindo que o valor absoluto de uma sub-série tenha mais impacto apenas por ter uma magnitude maior. Logo após, é realizado o treino das LSTM e as predições dos valores das sub-séries. Então, os autores fazem

a junção das n previsões em um único valor utilizando uma arquitetura similar a de um neurônio em uma rede neural. Uma multiplicação com pesos otimizados pelo algoritmo *Cuckoo Search Optimization* (CSO) é feita sobre as previsões obtidas pelas redes LSTM. Logo após é feita a soma dos produtos obtidos. O algoritmo CSO otimiza os pesos utilizando como função objetivo a *Mean Squared Error* (MSE). Para os testes, os autores decompuseram as séries das moedas BTC, XRP, DASH and LTC, em 22, 12, 17, e 12 séries cada, respectivamente. Os dados foram coletados de 18 Julho 2010 até 28 Março 2019 para BTC, 22 Janeiro 2015 até 28 Março 2019 para XRP, 14 Fevereiro 2014 até 28 Março 2019 para DASH, e 24 Agosto 2016 até 28 Março 2019 para LTC. Para validação, foi utilizado o método *holdout* utilizando 85% dos primeiros pontos das séries para treino e os 15% restantes para testes. Os autores compararam os resultados com o trabalho de Lahmiri e Bekiros (2019) que, segundo os autores, a decomposição juntamente com o método de otimização, os valores de RMSE foram melhorados em 77,33%, 82,36% e 85,60% para a BTC, XRP e DASH, respectivamente.

Um dos trabalhos mais recentes que utilizam a decomposição de séries é o trabalho de Niu, Xu e Wang (2020). Os autores propõem uma arquitetura similar a de Altan, Karasu e Bekiros (2019). Primeiramente, a série é decomposta em n sub-séries que alimentam uma série de n LSTM, porém, ao invés de utilizarem a EWT para a decomposição das séries, os autores utilizam a VMD. É comum em trabalhos que fazem a decomposição de séries que, após a realização da decomposição, seja necessária uma junção das previsões. Neste caso, os autores não utilizaram um algoritmo de otimização para a junção assim como foi feito no trabalho de Niu, Xu e Wang (2020), mas fizeram uma simples soma das previsões obtidas. Por fim, os autores fazem a avaliação do modelo com métricas como MAPE, MAE e RMSE. Para fins de comparação, também foram utilizadas outras redes após a decomposição das séries, como a *Artificial Neural Network* (ANN) e a CNN. Para testes, os autores utilizaram quatro índices de bolsas de valores afirmando que a decomposição VMD tem melhores resultados em comparação com a EMD, e que a rede LSTM tem maior performance que as outras redes em todos os *datasets*.

Um trabalho que utiliza a decomposição EMD é o de F. Zhou et al. (2019). Os autores têm como objetivo a predição dos valores de fechamento dos índices de bolsa *Shanghai Stock Exchange Composite* (SSEC), *National Association of Securities Dealers Automated Quotations* (NASDAQ) e o *Standard & Poor's 500 Composite Stock Price Index* (S&P 500). A arquitetura se difere dos demais trabalhos principalmente na rede que é utilizada para a predição dos valores. Segundo esta revisão bibliográfica, a arquitetura mais comum para esta aplicação é a LSTM devido a sua capacidade recorrente. Porém, neste trabalho os autores optaram por utilizar uma arquitetura denominada *Factorization Machine based Neural Network* (FNN), que tem como característica a utilização de alguns neurônios com a função de ativação alterada para interações fatorizadas. Para o treino da rede é utilizado o BP. Para os experimentos, os autores utilizaram a técnica de *holdout* utilizando 80% dos dados para treino e os 20% restantes para teste. Segundo os autores, a utilização de decomposição de séries aumentou significativamente a performance do modelo.

Outro trabalho que mostra a importância da decomposição de séries é o trabalho de Cao, Li e Li (2019), onde os autores utilizam uma variação do modelo de decomposição EMD que faz uso da técnica de *ensemble*, denominado *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN). Assim como os demais trabalhos, após a decomposição da série, os autores enviam as séries decompostas para uma rede. Neste caso, a rede escolhida foi a LSTM. Para a junção das n previsões feitas pelas redes, os autores optaram por utilizar uma camada densamente conectado com as saídas das LSTM. Dessa forma a arquitetura aprende também, a unir os resultados através do algoritmo de BP. Para os testes, é realizada a previsão sobre os ativos S&P500, HSI, DAX e SSE, com dados coletados entre 13 de Dezembro de 2007 até 12 de Dezembro de 2017. Como técnica de avaliação foi utilizado o *holdout* com 90% dos dados para treinamento e os 10% restantes para teste. Como métricas os autores utilizaram MAE, RMSE, e MAPE, concluindo que a arquitetura performou melhor que outros modelos como SVM e a própria LSTM sem a decomposição de séries.

Embora as LSTM sejam comuns quando tem-se a decomposição de séries temporais, alguns autores utilizam técnicas mais solidificadas na literatura, como no caso do trabalho de Cheng e Wei (2014) que utiliza SVM, mais especificamente *Support Vector Regressor* (SVR), para a previsão dos valores decompostos ao invés de uma rede neural. Os autores focam na previsão de valores do índice da bolsa de Taiwan, TAIEX. Os autores justificam a utilização do algoritmo EMD pelo fato de alguns algoritmos estatísticos, como o próprio SVR, necessitarem de premissas sobre os dados a serem utilizados, como no caso da maioria dos algoritmos de previsão de séries temporais que assumem que a série é estacionária. A arquitetura proposta baseia-se na coleta dos dados da série do índice TAIEX do ano 2000 até o ano de 2006, resultando em 7 sub-séries, uma para cada ano. Os autores utilizam para treino os meses de Janeiro até Outubro de cada ano, e os meses restantes para teste. Em seguida, é aplicada a decomposição EMD com 7 sub-séries IMF e uma série residual. Diferente dos trabalhos apresentados anteriormente, os autores deste trabalho não utilizam um modelo para cada sub-série, sendo assim não necessitam de uma arquitetura de junção de previsões. Para as avaliações os autores utilizaram a métrica RMSE, onde concluem que o modelo proposto tem resultados mais eficientes quando comparados com modelos de auto-regressão tradicionais no mesmo conjunto de dados.

Dois anos antes Yu e Liu (2012) também utilizaram SVM para a previsão em combinação com o EMD. Os autores têm por objetivo prever os valores do índice da bolsa de Changai, utilizando dados de 27 de Novembro de 2008 até 22 de Fevereiro de 2010. Em contraste com o trabalho de Cheng e Wei (2014), os autores optaram por criar uma SVM para cada IMF decomposta pelo EMD, criando assim uma etapa de junção das previsões das n SVM. Os autores também ressaltam que a utilização da decomposição de séries incrementa a performance da previsão significativamente.

No trabalho de Chandar, Sumathi e Sivanandam (2016), a DWT é utilizada para fazer a previsão de ações da bolsa de valores. A metodologia dos autores é similar a dos demais trabalhos. Primeiramente, é feita a coleta dos dados seguida da decomposição das séries que, por sua vez,

alimentam uma rede. Neste caso, a rede escolhida foi a ANN. Os autores utilizaram os valores das ações das empresas Tata steel, Wipro, SBI, TCS e Infosys, entre 1 de Janeiro de 2010 até Junho de 2015, utilizando a técnica de validação *holdout* com 80% dos dados para treinamento e 20% para testes, e como métrica os valores de *Coefficient of Variation* (CoV), *Mean Absolute Deviation* (MAD), RMSE e MAE. Os autores afirmam que em todas os testes e métricas, os valores da arquitetura que utiliza a decomposição DWT são melhores.

O trabalho recente de Yu-jun, Yi-mei e Jian-hua (2020), já foi discutido na Seção 2.1, mas também demonstra a relevância da utilização das técnicas de decomposição. Neste caso os autores utilizaram a técnica de EMD.

Esta seção apresentou trabalhos que utilizam a decomposição de séries temporais. Os trabalhos evidenciam que a utilização da decomposição de séries temporais pode alavancar o resultado dos modelos que visam prever valores de uma série temporal caótica. Dessa forma, é possível concluir que utilizar a decomposição como pré-processamento das séries pode aumentar significativamente o resultado dos modelos de predição.

2.4 CORRELAÇÃO DE ATIVOS E BOLSAS

Uma das variáveis que tem um grande potencial de ser responsável pela melhora na predição de ativos são outros ativos de diferentes bolsas de diversos países. Nesta subseção, serão apresentados alguns trabalhos que fazem o estudo da influência de ativos de diferentes bolsas em outras bolsas, ou a influência de determinados ativos, como ações ou *commodities* sobre outros ativos. O fenômeno que descreve a intervenção, direta ou indireta, de uma bolsa sobre outra é denominado de *spillovers*.

Encontrar correlação entre bolsas e ativos não é um tópico recente, pois existem diversos trabalhos publicados que discorrem sobre o tema, além de diversas técnicas para estudar a hipótese de que uma bolsa, ou ativo, pode afetar diretamente, ou indiretamente os ativos de outras bolsas. O trabalho de Tiwari et al. (2013) faz o uso de transformadas baseadas em uma técnica variante da DWT, chamada de *Maximal Overlap Discrete Wavelet Transform* (MODWT), para identificar correlações entre nove diferentes bolsas asiáticas, sendo elas as bolsas do: Japão, Malásia, Indonésia, Hong Kong, Singapura, Coreia do Sul, Taiwan, China e Índia. A metodologia dos autores baseia-se na análise de variâncias e covariâncias entre diferentes escalas das *wavelets* obtidas por DWT, além de medidas pelo cálculo da *Wavelet Multiple Correlation* (WMC). Para a simplificação, os autores utilizaram as séries dos índices mais importantes de cada bolsa. Os dados utilizados para a análise estão compreendidos entre 4 de Janeiro de 2005 até 28 de Fevereiro de 2012. Os autores também analisaram a correlação em diferentes períodos, sendo eles de 2 a 4 dias *intra-weekly*, 4 a 8 dias *weekly*, 8 a 16 dias *fortnightly*, 16 a 32 dias *monthly*, 32 a 64 dias *monthly to quarterly*, 64 a 128 dias *quarterly to biannual*. Como conclusão, os autores mostraram que a correlação de *intra-weekly* é de 0.81, para *weekly* 0.85, *monthly* 0.89, e chegando a 0.95 nas menores frequências, indicando que um investidor com interesses de longo

prazo pode se basear nas frequências mais baixas, enquanto um investidor de curto prazo nas frequências mais altas. Assim, os autores concluem que a bolsa asiática é integrada, tendo como maior correlação as séries com menores frequências e com maior prazo.

Para a análise de memória de longo prazo em séries temporais, muitos trabalhos utilizam *Detrended Fluctuation Analysis* (DFA) e *Multifractional Detrended Fluctuation Analysis* (MFDFA) (KANTELHARDT et al., 2002), desde sua aplicação no trabalho de Peng et al. (1994) para análise de estruturas de DNA, segundo Ma, Wei e Huang (2013). Ainda no mercado asiático, os autores Ma, Wei e Huang (2013) investigam a correlação cruzada entre a bolsa da China contra as do Japão, Coreia do Sul, e Hong Kong, utilizando a técnica *Multifractal Detrended Cross-Correlation Analysis* (MF-X-DFA), proposta por Zhou (2008). A técnica de MF-X-DFA baseia-se no valor de *hurst exponent*, representado por H , utilizado para identificar memórias de longo prazo em uma série temporal. Sendo assim, quando os valores de H_{xy} , sendo x e y diferentes séries temporais, apresentam valores iguais a 0,5, não existe correlação cruzada entre as séries x e y ; quando o valor de H_{xy} é maior que 0,5, as séries têm uma correlação cruzada positiva de longo termo, o que significa que os valores futuros tendem a ter valores maiores que os anteriores; quando os valores de H_{xy} forem menores que 0,5, a série tem uma correlação cruzada positiva tendendo a ter valores futuros alternantes; isto é, um valor alto tende a ser seguido de um valor baixo e um valor baixo tende a ser seguido de um valor alto. Para os testes, os autores utilizaram valores compreendidos entre 1 de Janeiro de 1997 até 31 Dezembro de 2011. Dessa forma, como é afirmado que as bolsas têm correlações entre si, é possível inferir que uma bolsa pode conter informações relevantes para a predição de valores de uma bolsa adjacente.

O trabalho de Asgharian, Hess e Liu (2013) analisa a conexão entre diversas bolsas de valores utilizando técnicas de econometria. Os autores se apoiam nas análises de dependências espaciais que refletem o fenômeno que ocorre quando o valor de uma variável dependente em um local específico depende de valores observados em outros locais. A técnica mais comum, segundo os autores, é a *Spatial Autoregressive* (SAR). Para a análise, os autores utilizaram dados de Janeiro de 1995 até Dezembro de 2011. Dessa forma, os autores afirmaram ter encontrado que, independente da medida de distância considerada, a importância de proximidade reduz da primeira para a segunda metade das observações; isto é, antes e depois de Junho de 2003; enquanto o grau de dependência dos mercados crescem pouco a pouco ao longo do tempo. Segundo os autores, isto pode ser uma indicação de que o mercado financeiro global alcançou uma integração ao longo dos anos. Foi observado também o mesmo padrão quando foram comparadas as recessões globais de 2001 e entre 2007 e 2009, com os períodos restantes. A proximidade do espaço hipotético criado pelos autores tende a ter menos importância durante recessões, e a dependência total entre os mercados é maior. Este trabalho mostra como a integração dos mercados é relevante. Sendo assim, pode ter um grande potencial para a utilização de diversos mercados para a predição de um mercado em análise.

Um trabalho mais recente, escrito por Guo, Zhang e Tian (2018), tem como objetivo a análise de correlação entre ativos da bolsa de Changai. Os autores seguem por uma linha um

pouco diferente dos demais, e aplicam a correlação intra-bolsa; isto é, a análise é feita sobre uma única bolsa de valor, utilizando os ativos listados na mesma. Para a análise de conexão entre os ativos, os autores optaram pela utilização da estrutura de *Minimum Spanning Tree* (MST), onde cada nó da estrutura é um ativo, e o peso das suas conexões representa a afinidade de um ativo com o outro; ou seja, a distância hipotética entre os ativos. Os autores calculam esta distância utilizando a entropia de *Shannon* e o coeficiente de correlação de *Pearson*. Para a construção da MST, os autores utilizaram o algoritmo de *Prim*. Para os testes foram utilizados 280 ativos da bolsa de Changai com dados compreendidos entre os anos 2014 e 2016. Segundo os autores, as informações mútuas entre os ativos é mais eficaz para determinar uma correlação entre diferentes ativos.

No trabalho de Arai, Yoshikawa e Iyetomi (2015), os autores utilizam uma versão modificada do algoritmo *Principal Component Analysis* (PCA), denominado de *Complex Principal Component Analysis* (CPCA). Os autores utilizam dados de 2008 até 2011 de 483 ativos listados no índice americano S&P500. O CPCA permite aos autores extrair correlações entre os ativos da bolsa, e assim utilizar o *Random Matrix Theory* (RMT) para distinguir verdadeiras correlações de ruídos na série temporal. As informações extraídas são então projetadas em uma rede de correlações onde os pares de ativos que têm uma forte correlação contêm pesos relevantes na rede. Os autores afirmam ter encontrado o que foi denominado de "comunidade de ativos" que tem variações similares entre seus preços. Entretanto, os autores afirmam que a metodologia não encontrou relação significativa entre os grupos de ativos. Sendo assim, os autores sugerem ir além da correlação entre pares, e utilizar a correlação de dois ou mais ativos.

Além de analisar a correlação entre ações de empresas, também é possível analisar a correlação entre *commodities* e outros ativos da bolsa como no trabalho de Boldanov, Degiannakis e Filis (2015) que faz uma análise da influência do preço do petróleo na volatilidade dos ativos da bolsa. Mais especificamente, os autores analisam os seis maiores países importadores e exportadores de petróleo. Para a metodologia, os autores primeiramente estimaram a volatilidade dos ativos. Para isto, utilizaram o modelo GARCH para identificar *clusters* de volatilidade. Também consideraram a volatilidade realizada tanto para os ativos quanto para o preço do petróleo. Logo após fazem a análise de correlação condicional entre as volatilidades. Como conclusão, os autores mostram que a correlação entre o preço do petróleo e os ativos não é constante, e depende do período para ser verdadeira. Entretanto, é apontado que a correlação é sempre positiva, independente do período.

O mercado financeiro tem períodos de quedas bruscas, denominadas de crises. Alguns trabalhos levam esta característica do mercado em consideração enquanto outros preferem analisar o mercado ignorando este fato. O trabalho de Moldovan (2011) faz a análise da correlação entre as bolsas de Nova Iorque, Londres e Tóquio durante e fora o período de crise. Para o estudo, os autores utilizaram uma simples regressão aplicada com o método dos mínimos quadrados para identificar a interdependência dos mercados, onde a regressão utiliza como variáveis independentes as bolsas adjacentes à analisada. Embora não seja o melhor método para o es-

tudo, os autores encontraram que as correlações entre as bolsas são maiores em tempos de crise, sugerindo que a correlação venha do pânico gerado nos investidores durante este período, onde a liquidação dos investimentos tende a crescer significativamente, atrelada a uma aversão ao risco eminente dos ativos em tal período.

No trabalho de Hoseinzade e Haratizadeh (2019) os autores tentam utilizar o modelo CNN para extrair *features* de múltiplos mercados tentando aprimorar os resultados da predição. No trabalho são propostas dois tipos de arquiteturas, denominadas de 2D-CNNpred e 3D-CNNpred. Ambas utilizam dados de diversos mercados com diversas *features*. A diferença entre as arquiteturas está na topologia dos modelos. A 2D-CNNpred tem uma arquitetura onde um único sistema de predição é criado; isto é, para k bolsas de valores, os autores alimentam um único modelo de CNN com todos os valores. Em contraste com o modelo 3D-CNNpred, que assume que o comportamento dos mercados são muito diferentes e, portanto, é necessário criar um modelo para cada mercado. Os autores afirmam que, estatisticamente, os modelos superam os do estado da arte, sendo então uma evidência de que modelos que utilizam diversos mercados para a predição podem ter sucesso.

Os trabalhos apresentados mostram estudos sobre a correlação de ativos entre bolsas de valores. Assim, os trabalhos citados nesta seção suportam a ideia de utilizar diversas bolsas para prever o valor de um ativo na bolsa de valores brasileira.

2.5 ARTIGOS DE REVISÃO

Nesta seção serão apresentados alguns artigos de revisão que mostram o estado da arte no estudo de predição de ativos nas bolsas de valores.

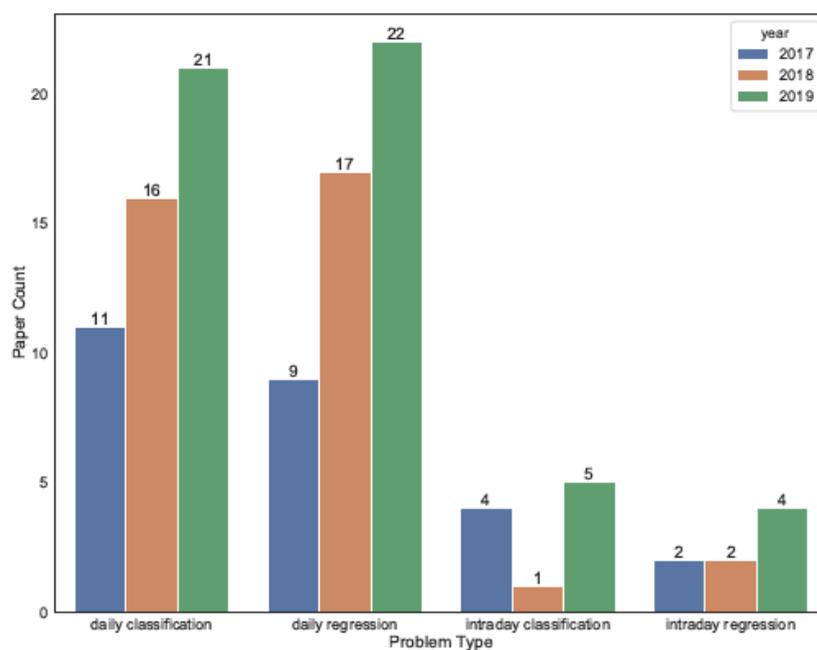
Na revisão de Jiang (2020), os autores fazem a análise de um total de 124 artigos, dos quais 115 têm data de publicação compreendida entre 2017 e 2019. Para obter os artigos, os autores usaram a plataforma *Google Scholar* utilizando palavras chaves como *deep learning*, *stock prediction*, *stock forecasting*, entre outras.

Assim os autores classificam os artigos em duas classes: *regression* e *classification*. Os artigos classificados em *regression* são aqueles que focam na predição de um valor específico do preço do ativo da bolsa. Por sua vez, os artigos classificados em *classification* têm como objetivo prever a direção do preço do ativo em tendências de alta ou tendências de baixa.

Os autores também separam os artigos por categoria de dados utilizados. As categorias são *daily*, e *intra-day*. A categoria *daily* se refere ao valor de fechamento diário do ativo; isto é, seus valores de OHLC se referem a um dia inteiro. Onde, o *O (Open)* é o valor com que o ativo abriu o mercado no dia; *H (High)* é o valor mais alto que o ativo atingiu no dia; *L (Low)* é o valor mais baixo que o ativo atingiu no dia; e *C (Close)* é o valor com que o ativo fechou o mercado no dia. Por sua vez, a categoria *intra-day* os valores de OHLC se referem a valores que são compreendidos no mesmo dia, como por exemplo o fechamento de 15 minutos dos ativos.

Em um gráfico de barras (Figura 1) os autores mostram a quantidade de trabalhos que focam na predição de valores *daily*, e trabalhos que focam na predição de valores *intra-day*. Os separando também em *regression* e *classification*. Segundo os autores, a Figura 1 indica que trabalhos que tentam prever valores diários são mais comuns do que artigos que focam na predição de valores que têm uma frequência maior, como por exemplo os fechamentos de 15 minutos, 30 minutos, etc. Os autores argumentam que este fato pode ser ocorrência da falta de informações *intra-day* na internet.

Figura 1 – Quantidade de artigos publicados agrupados por ano (cor) e separados por classes (*regression* e *classification*) e categoria de dados utilizados (*daily* ou *intra-day*)

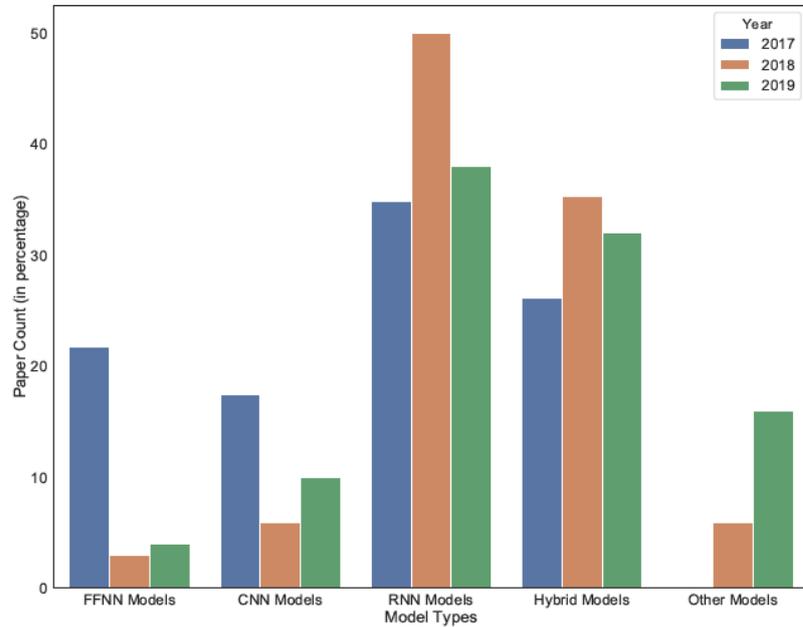


Fonte: Jiang (2020)

Os autores também analisaram os modelos que estão sendo utilizados para realizar a predição de valores na bolsa. A Figura 2 compara o volume de trabalhos agrupados por modelo de predição utilizado. Os autores apontam que a quantidade de trabalhos que utilizam redes recorrentes têm crescido ao longo dos anos. A hipótese para este comportamento é que isto se deva à habilidade das redes de levar em consideração os valores passados de uma série temporal como a bolsa de valores.

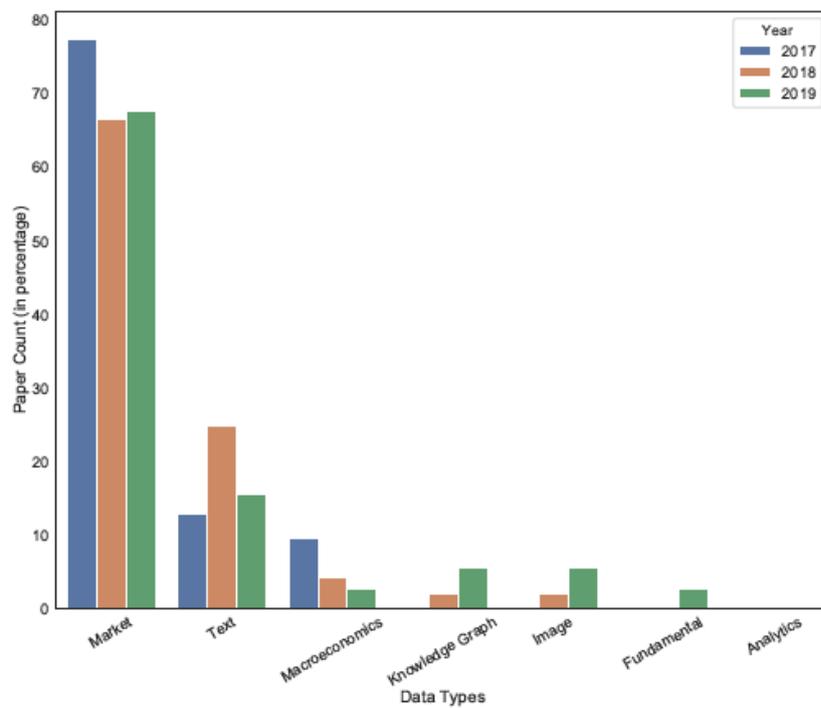
Outro dado interessante que Jiang (2020) estuda é o dado utilizado para treinamento da rede. Os autores apontam a maior representatividade dos trabalhos que utilizam os valores de mercado como entrada do modelo de *machine learning*, em relação aos que utilizam outros valores, como: texto de redes sociais para a captação de sentimentos com NLP; dados macroeconômicos do país, onde a bolsa analisada esta situada; e dados fundamentalistas, como por exemplo dados contábeis da empresa referente à ação analisada. A Figura 3 mostra o volume de trabalhos publicados e separados por ano, agrupado por tipo dado utilizado no treinamento.

Figura 2 – Quantidade de artigos publicados separados por ano (cor) e modelo utilizado



Fonte: Jiang (2020)

Figura 3 – Volume de trabalhos publicados e separados por ano, agrupados por tipo dado utilizado no treinamento



Fonte: Jiang (2020)

Outro trabalho que faz um estudo sobre o estado da arte é o de Shah, Isah e Zulkernine (2019). Neste artigo, os autores escrevem que teorias populares afirmam que a bolsa é uma

série randômica e não é possível prevê-la, mas demonstra que existem diversos trabalhos publicados que mostram o contrário, sendo possível ter sucesso na área. Os autores dividiram os artigos publicados em cinco macro categorias de técnicas para predição, são elas: Estatísticas, Reconhecimento de Padrões, Aprendizado de Máquina, Análise de Sentimentos e Híbridos.

Como técnicas estatísticas são citados os seguintes modelos: *AutoRegressive Moving Average* (ARMA), ARIMA e GARCH. Essas técnicas, normalmente, exigem algumas premissas sobre a série estudada, tais como: linearidade, estacionariedade, e normalidade das séries temporais. Porém, estas características dificilmente são encontradas em séries da bolsa de valores.

O Reconhecimento de Padrões foca em encontrar padrões nos valores da série OHLC do ativo analisado. As técnicas de Reconhecimento de Padrões podem ser utilizadas como entrada do Aprendizado de Máquina.

Por sua vez, as técnicas de Aprendizado de Máquina utilizam diversos dados para prever os valores OHLC da bolsa.

Outra técnica apresentada é a Análise de Sentimentos que, embora utilize muitas técnicas de Aprendizado de Máquina, usa principalmente valores textuais para realizar a predição do ativo, capturando sentimentos em *posts* de redes sociais, fóruns, entre outras.

Por fim, os modelos Híbridos são apresentados, os quais possuem como característica o uso de dois ou mais modelos apresentados, como por exemplo o Aprendizado de Máquina juntamente com Análise de Sentimentos. Normalmente, este tipo de modelo é empregado para tentar aprimorar um modelo anterior. Segundo os autores, os modelos híbridos de Aprendizado de Máquina e Estatística são os mais propensos a terem bons resultados. Os autores também afirmam que os algoritmos de redes como ANN têm crescido muito nos últimos anos.

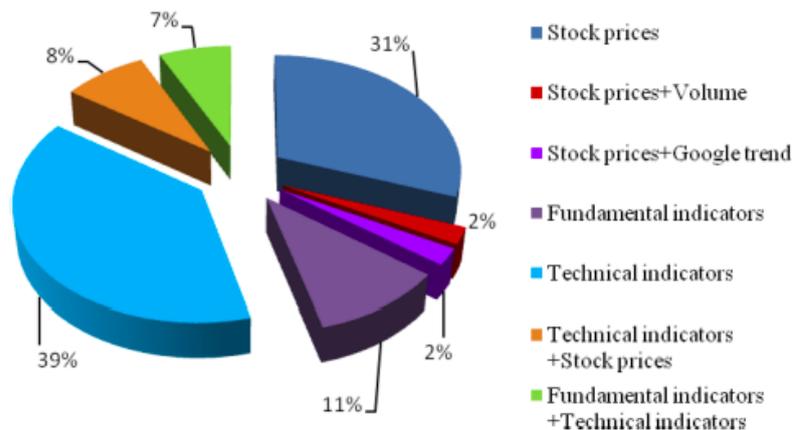
O recente trabalho de Kumar, Jain e Singh (2020) faz um estudo de seis diferentes aspectos que um trabalho que visa fazer a predição de valores na bolsa deve seguir. Os autores dividiram os aspectos da seguinte forma: estudo da bolsa analisada e seus dados, tipos de variáveis utilizadas, técnicas de pré-processamento utilizadas, técnicas de seleção de variáveis, modelos de *forecast* e métricas utilizadas para avaliar a performance do modelo.

Os autores colocam como primeiro passo, para realizar um trabalho sobre predição de valores na bolsa, a seleção da bolsa de valores. No trabalho é listado diversos trabalhos e quais bolsas foram utilizadas nestes trabalhos. É apontado que os trabalhos estudados escolhem ou um índice específico da bolsa analisada, ou um ativo específico bolsa.

No âmbito de variáveis utilizadas, os autores chegaram em uma conclusão similar ao estudo de Jiang (2020), indicando que os valores mais utilizados são os valores do ativo na bolsa, além de indicadores de análise técnica derivados dos valores desses ativos. Esta ideia pode ser observada na Figura 4, que mostra a representatividade, em porcentagem, das variáveis utilizadas.

Para o estudo do pré-processamento dos dados, os autores comentam a importância do procedimento para aumentar a confiabilidade e acurácia do modelo, caso tais técnicas sejam

Figura 4 – Representatividade de variáveis utilizadas para a predição de valores na bolsa.



Fonte: Kumar, Jain e Singh (2020)

utilizadas. Para a redução do viés de magnitude das variáveis, muitos trabalhos utilizam a técnica de normalização *Min-Max*, que consiste em deixar todas as variáveis no mesmo intervalo de valor, fazendo com que a magnitude das variáveis não seja levada em consideração pelo modelo.

Os autores também discorrem sobre a seleção e extração de variáveis. Tal seleção de variáveis consiste na eliminação de variáveis que podem ser tanto insignificantes para o modelo, quanto prejudiciais em questão de acurácia e tempo de treinamento dos modelos. Um exemplo de como realizar a seleção de variáveis é a utilização de técnicas estatísticas, como correlação, para determinar a importância das variáveis. Por sua vez, a extração de variáveis consiste em utilizar variáveis existentes para gerar novas variáveis que contenham informações mais relevantes e, consequentemente, trazem mais acurácia para o modelo.

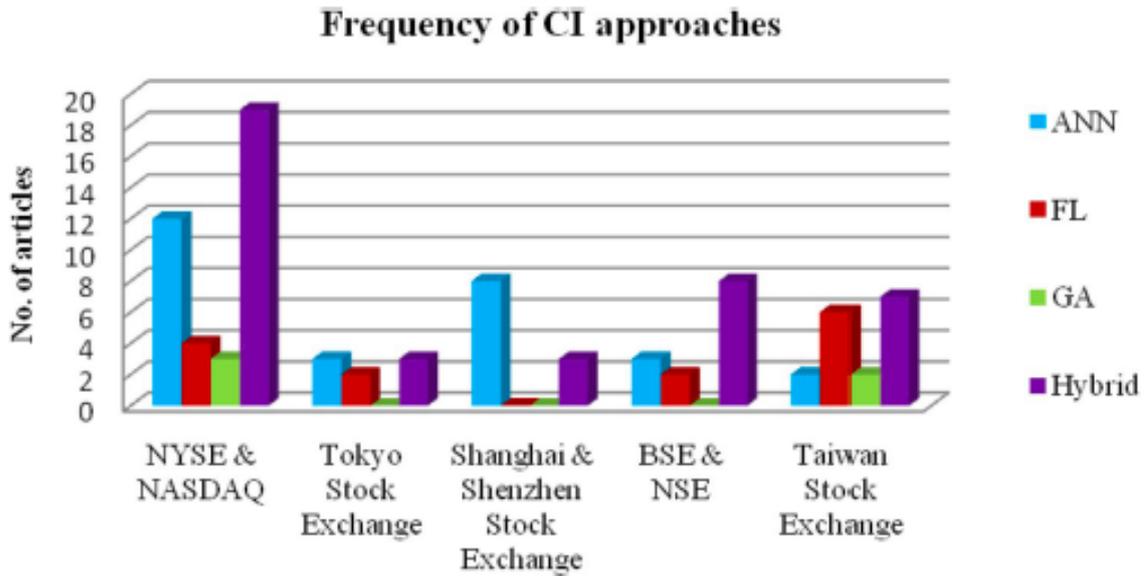
No próximo tópico, os autores mostram quais modelos que mais estão sendo utilizados para a predição de valores na bolsa. Entre os mais utilizados estão compreendidas as categorias de ANN, GA, *Fuzzy Logic* (FL) e híbridos; isto é, que utilizam dois ou mais dos apresentados anteriormente. O gráfico da Figura 5 mostra que o modelo mais utilizado é a ANN nas bolsas *NYSE* e *NASDAQ*.

Como tópico final são apresentadas as métricas utilizadas para a avaliação dos modelos utilizados. Os autores dividem as métricas em classificação e regressão, sendo que em classificação existem três divisões, são elas: Acurácia (Recall, Precision, F-Score, Accuracy e Hit Rate), Comparação (JPE), e Proximidade (MSE, RMSE, MAE, MAPE, entre outras.).

Em regressão, temos também três divisões, são elas: Acurácia Direcional (Hit Ratio, DS, POCID), Comparação (ARV), e Rentabilidade (*Return Over Investment* (ROI)). Assim, os autores concluem o artigo ressaltando a relevância das técnicas computacionais nos problemas de predição de valores na predição de valores da bolsa.

No artigo de revisão de Nosratabadi et al. (2020), os autores realizaram um estudo de 57 artigos relacionados à aplicação de algoritmos de *machine learning* na bolsa de valores, *e-commerce*, *cripto-moedas*, *marketing* e *corporate banking*. Os autores mostram uma tabela que

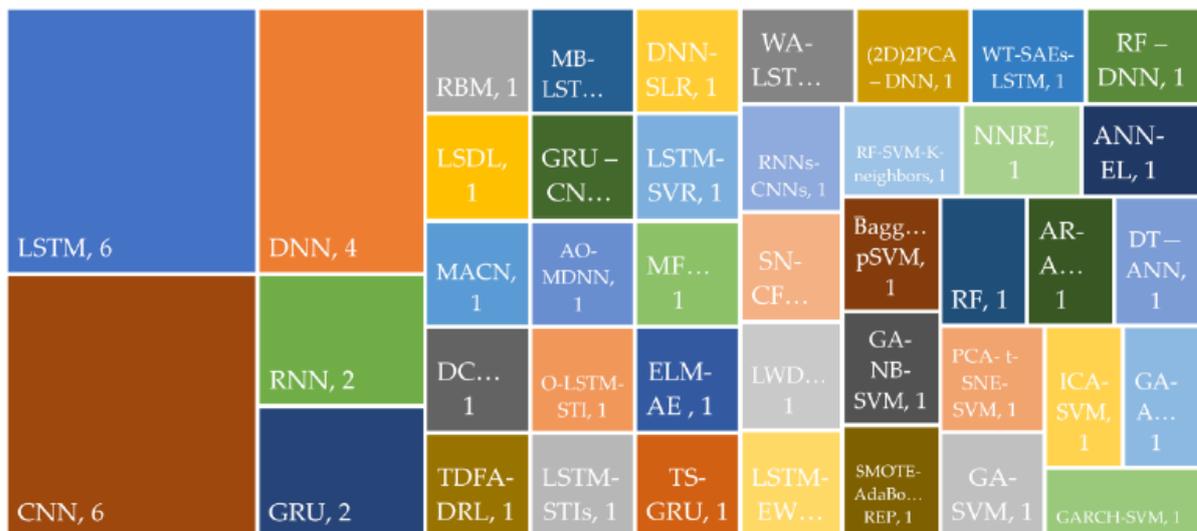
Figura 5 – Quantidade de artigos agrupados por qual bolsa foi utilizada além dos modelos utilizados.



Fonte: Kumar, Jain e Singh (2020)

demonstra a representatividade dos algoritmos dentro dos 57 artigos revisados. Novamente, é evidente que as LSTM têm uma presença mais forte nos trabalhos publicados, seja utilizando o modelo diretamente ou em conjunto com outros. A Figura 6 mostra a representatividade dos modelos na literatura, onde o tamanho dos retângulos é proporcional à quantidade de artigos que utilizam o modelo.

Figura 6 – Representatividade dos modelos na literatura, cada quadrado representa em tamanho a representatividade do modelo e o número de artigos publicados com este modelo.



Fonte: Nosratabadi et al. (2020)

Em seguida, os autores discorrem sobre os modelos utilizados, com foco em algoritmos de aprendizado de máquina. Conclui-se que os modelos LSTM, CNN e *Deep Neural Network* (DNN) são os mais frequentes para tentar prever os valores de série temporal da bolsa. Segundo os autores, muitos trabalhos que tiveram como metodologia um modelo híbrido também tiveram melhores resultados em relação à aqueles utilizaram apenas um modelo, sugerindo assim a utilização de mais de um modelo para aprimorar a acurácia da predição.

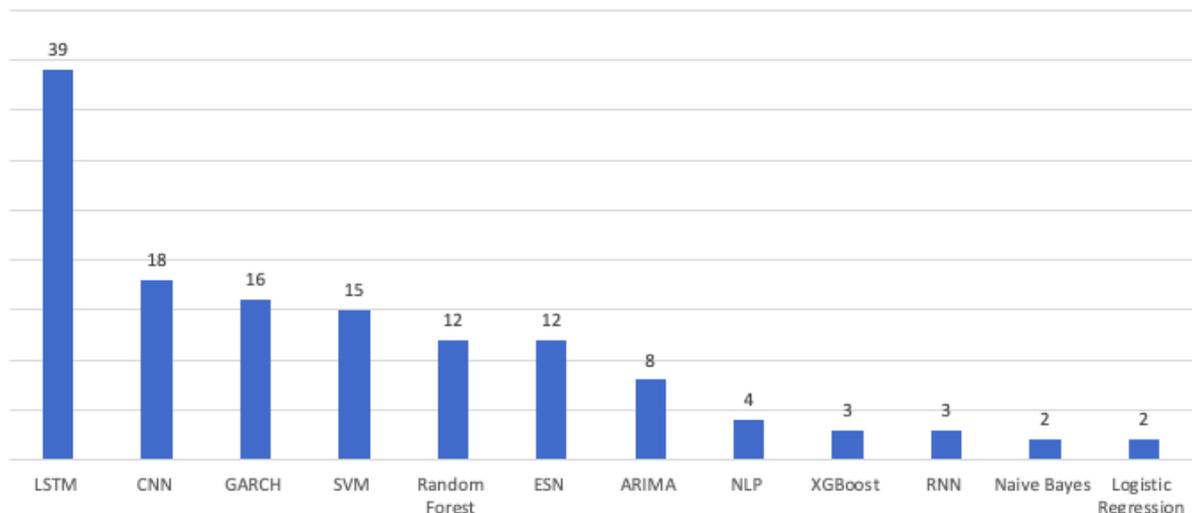
Os artigos citados nesta seção mostram que existem diversos trabalhos que tem sucesso na predição de valores de ativos na bolsa de ações. Além de evidenciar as metodologias, métricas e *features* mais utilizadas.

2.6 ANÁLISE DA BIBLIOGRAFIA

Para realizar a pesquisa sobre o estado da arte do tema deste trabalho, foram utilizados um total de 117 artigos, sendo 24 com ano de publicação entre 2010 e 2014, e 93 com o ano de publicação entre 2015 e 2020. As características relevantes, descritas a seguir, são aquelas que sustentam a metodologia proposta neste trabalho.

Um dos pontos que chamaram a atenção foi a quantidade de trabalhos publicados que utilizam a rede LSTM, seja como uma auxiliar, em um modelo híbrido, ou como a rede principal de predição. Assim, construímos um gráfico com base nos 117 artigos revisados, que traz uma visão dos modelos mais utilizados para a predição na bolsa de valores (Figura 7) que mostra a quantidade de trabalhos que utilizaram cada modelo.

Figura 7 – Representatividade dos modelos nos trabalhos analisados para a bibliografia.



Fonte: Autor

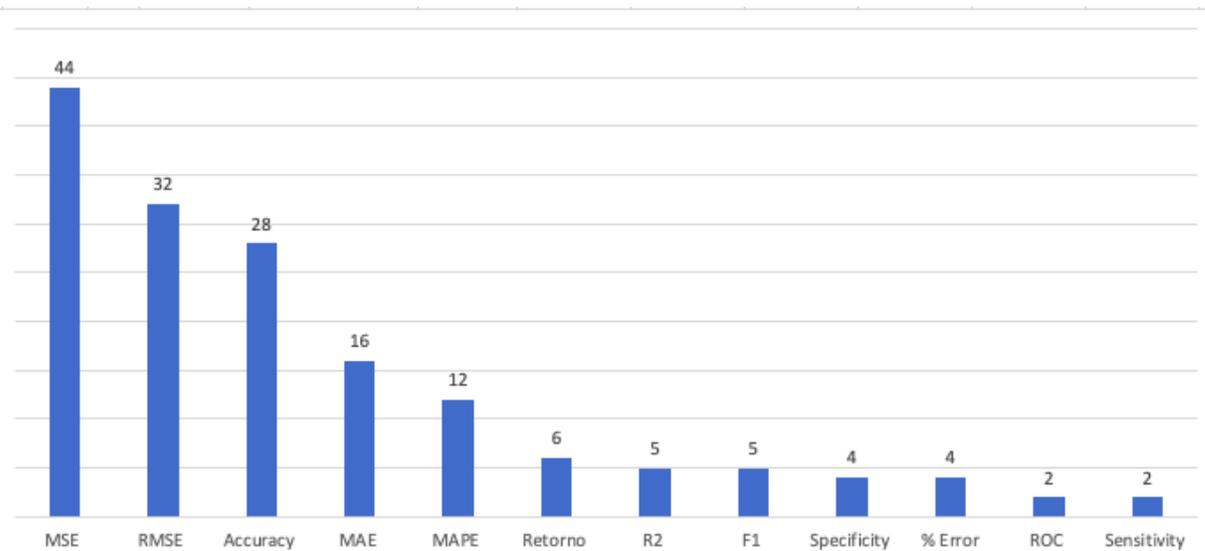
Baseado na Figura 7, as redes LSTM têm um espaço considerável nesta área, principalmente por se tratar de uma análise de uma série temporal, dados onde as LSTM conseguem

modelar de maneira satisfatória devido a sua capacidade de lembrar de valores passados como referência para tentar prever os valores futuros da série histórica dos ativos avaliados.

Um segundo ponto, que é possível notar ao analisar a Figura 7, é que o levantamento sugere uma preferência dos autores ao utilizar algoritmos de redes. Os algoritmos tradicionais de estatística vêm perdendo espaço entre os trabalhos mais recentes que têm como foco a predição de valores na bolsa. Entre os trabalhos analisados, apenas 35% utilizaram algoritmos estatísticos tradicionais.

Outro ponto que foi levantado é a respeito das métricas utilizadas para avaliar os modelos de predição da bolsa. Segundo o levantamento, a maior parte dos trabalhos utiliza a métrica MSE para avaliar a performance da predição, como mostra a Figura 8. De acordo com a figura, a representatividade da métrica MSE também sugere que a quantidade de artigos que preveem o valor do ativo, ao invés de realizar uma classificação de *compra* ou *venda*, é maior.

Figura 8 – Representatividade das métricas utilizadas na revisão realizada.



Fonte: Autor

2.7 CONCLUSÃO DA BIBLIOGRAFIA

Com a leitura dos artigos apresentados na seção de trabalhos relacionados, e a análise feita na Seção 2.6, foi possível encontrar os seguintes pontos importantes.

Primeiramente, de acordo com o resultado dos trabalhos estudados, é razoável sugerir que é possível ter bons resultados na área de predição de valores no mercado de ações utilizando algoritmos de *machine learning*, o que vai contra o estudo de Fama (1970), que afirma que o mercado é eficiente e sua predição não é factível. O fato da revisão apresentar diversos trabalhos que têm bons resultados, sustenta a metodologia proposta que se baseia principalmente na utilização de análise de dados para prever os valores do mercado financeiro.

A análise do estado da arte possibilitou o levantamento de outro ponto importante: é possível afirmar que os trabalhos mais recentes têm como foco a utilização de algoritmos de redes neurais, *deep learning*, NLP ou modelos híbridos. Assim, sugerindo uma fronteira de sabedoria na área. Considerando tal levantamento, o presente trabalho propõe um modelo híbrido utilizando técnicas de análise numérica para a decomposição de valores das séries temporais da bolsa, juntamente com algoritmos de *machine learning* e uma arquitetura de extração de contextos das séries temporais.

Também foi possível concluir que a predição de informações sobre os ativos da bolsa de valores é um tópico que chama a atenção de pesquisadores ao redor de todo o mundo, e demonstra uma atividade grande destes pesquisadores, onde a cada ano, novos algoritmos são propostos. Isto indica que a área tem um futuro promissor no mundo acadêmico devido à diversas metodologias possíveis capazes de resolver os problemas de ruídos, e não-linearidade da série temporal da bolsa.

Considerando a revisão bibliográfica realizada, pode-se notar que há uma tendência generalizada no uso de redes LSTM aplicadas à análise de mercados financeiros. Essas redes têm como característica principal o acompanhamento do comportamento temporal dos valores dos ativos individuais. No entanto, informações contextuais são geralmente negligenciadas nesses modelos. Até onde notamos, mesmo com estes modelos representando mais de 33% das metodologias propostas nos trabalhos aqui avaliados, nenhum trabalho considera esse tipo de informação para a análise de dados. Sabe-se, no entanto, que tais informações podem potencializar as previsões.

Na área de Visão Computacional, que trata também de análises temporais, o uso de modelos hierárquicos para captura de informações contextuais, já é uma realidade (SEYEDHOSSEINI; SAJJADI; TASDIZEN, 2013). Assim, neste trabalho propomos o uso de modelos hierárquicos, com submódulos RNN, para análise e previsibilidade de mercado de ações. Até onde sabemos, essa proposta ainda não foi abordada na literatura científica para esse tipo de aplicação.

3 CONCEITOS FUNDAMENTAIS

Neste capítulo serão apresentados os conceitos fundamentais da metodologia proposta neste trabalho. O capítulo foi dividido nas seguintes seções: Séries Temporais (Seção 3.1), Decomposição de Séries Temporais (Seção 3.2), Redes Neurais Recorrentes (Seção 3.3), Análise Contextual (Seção 3.4), Modelo Hierárquico em Cascata (Seção 3.5), Parâmetros do Mercado Financeiro (Seção 3.6), Normalização de Dados (Seção), e Métricas de Avaliação (Seção 3.9).

3.1 SERIES TEMPORAIS

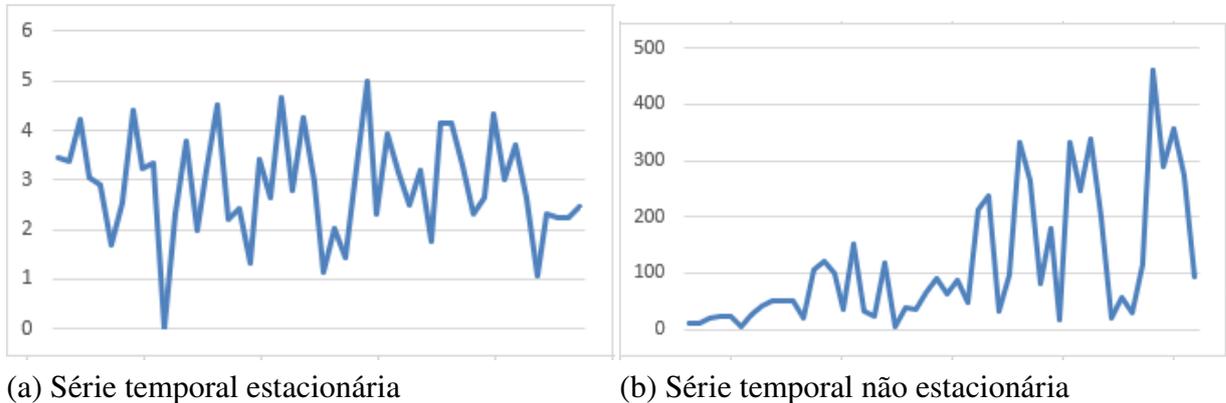
As séries temporais são utilizadas para representar dados que apresentam alguma característica relativa ao tempo em um sinal contínuo ou discreto. Normalmente, as séries temporais são apresentadas em formato de gráficos de linha, onde seu eixo x corresponde a um intervalo temporal e o eixo y corresponde a qualquer informação que desejamos analisar. O termo séries temporais é amplamente utilizado na literatura de diversas áreas de pesquisa devido à sua flexibilidade na representação de diversos tipos de dados. Apesar de não ser um termo novo, ainda é utilizado com frequência em diversos trabalhos atuais.

Uma série temporal pode ser definida como um conjunto de observações x_t , onde cada observação é gravada em um determinado tempo t . Estas observações podem ser medidas em tempos discretos ou contínuos. Para os tempos discretos, as observações devem ser medidas em intervalos temporais fixos, tais como: meses, dias, horas, minutos, etc. Sendo assim, t representa uma única medida temporal, $t_0 \in \mathbb{N}^*$ (BROCKWELL; DAVIS, 1991). Para os tempos contínuos, as observações são medidas continuamente sobre os intervalos; assim t representa um conjunto de medidas não-finito, $t_0 = [0,1]$.

As séries temporais apresentam uma grande relevância sobre a ordem em que os dados são apresentados. Ao levar em consideração a ordem dos dados, as séries temporais são capazes de expor comportamentos e padrões, antes não revelados, que auxiliam na predição de valores futuros. Em contraste, as séries de dados não-temporais são utilizadas para aplicações onde a ordem em que as observações são apresentadas não tem relevância para a capacidade preditiva do modelo que esta sendo aplicado sobre os dados.

De um modo geral, podemos categorizar as séries temporais em duas grandes classes: estacionárias e não estacionárias. As séries estacionárias são aquelas em que as distribuições não mudam conforme o tempo. Formalmente, considerando um processo estocástico $X = \{x_i; i \in \mathbb{Z}\}$, este processo é estacionário se a distribuição dos vetores $(x_{t_1}, \dots, x_{t_n})$ e $(x_{t_1+\tau}, \dots, x_{t_n+\tau})$ para $t \subset \mathbb{Z}$, $n \in \mathbb{N}$ e para qualquer $\tau \subset \mathbb{Z}$, forem iguais; isto é, suas médias e desvios padrão são iguais. As séries temporais são consideradas não estacionárias caso a condição descrita anteriormente não seja satisfeita (HAWKES; COX; MILLER, 1967). A Figura 9a apresenta uma série temporal estacionária, enquanto a Figura 9b apresenta uma série temporal não estacionária.

Figura 9 – Comparação das duas classes de série temporal.



Fonte: Autor

As séries temporais aparecem em diversas áreas de estudo, como por exemplo: engenharia, biologia, medicina, computação e economia. O uso de tais séries se faz relevante no presente trabalho devido ao tema explorado: bolsa de valores. Cada ativo da bolsa de valores apresenta um preço de mercado, e este preço, por sua vez, varia conforme o tempo. Dessa forma, temos uma série temporal em que cada observação x_t no tempo corresponde ao preço de uma ação. Ao levar em consideração o eixo temporal pode-se tirar vantagens dos comportamentos e padrões para inferir o preço dos ativos.

Neste trabalho serão analisadas principalmente séries temporais não estacionárias por se tratar de valores de ativos no mercado financeiro. Existem diversos fatores que podem impactar no preço de um ativo ofertado no mercado financeiro (FAMA, 1970). Devido a este fato, a série temporal analisada se torna mais volátil e instável. Muitos algoritmos estatísticos têm a habilidade de modelar uma série temporal mesmo que ela seja não estacionária, como o ARIMA. Este algoritmo se baseia no algoritmo ARMA, que por sua vez tem como principal base a média móvel da própria série temporal.

O presente trabalho utiliza técnicas de *machine learning*, como LSTM e ESN, que não necessitam que as séries analisadas sejam estacionárias. Assim, torna-se possível a modelagem de séries temporais como encontramos na bolsa de valores.

3.2 DECOMPOSIÇÃO DE SERIES TEMPORAIS

A decomposição de séries temporais são técnicas frequentemente utilizadas para eliminar ruídos de séries temporais (ALTAN; KARASU; BEKIROS, 2019). Alguns algoritmos de *machine learning* podem apresentar alta sensibilidade a ruídos na série temporal. Por este motivo, pode-se utilizar a decomposição de uma série para que tais ruídos não interfiram na assertividade do modelo. Porém, em alguns casos, tais ruídos podem apresentar informações relevantes e desconsiderá-los pode reduzir a performance do modelo ao invés de aprimorá-la.

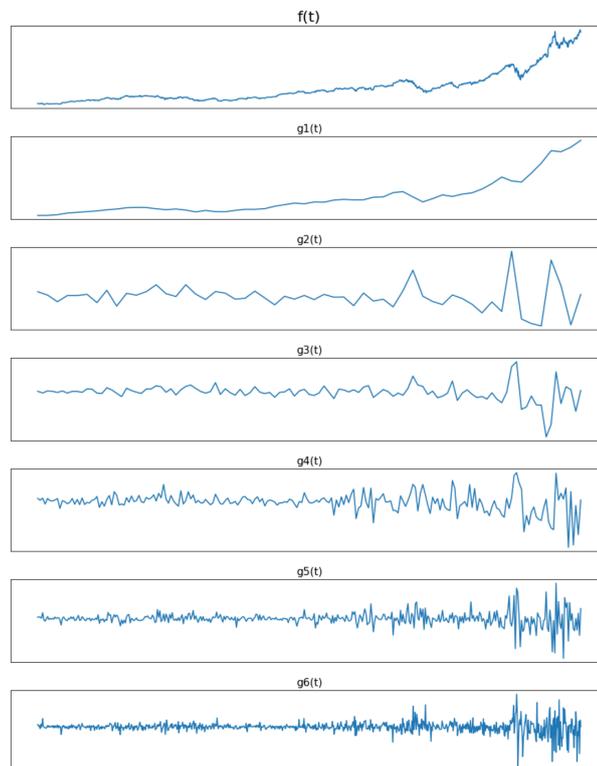
As técnicas de decomposição de séries temporais consistem em encontrar diversas funções que, juntas, compõem uma função original. Considerando uma série temporal $f(t)$, pode-se descrevê-la em sub-funções $g_i(t)$, como mostra a Equação (1).

$$f(t) = \sum_{i=1}^n g_i(t) \quad (1)$$

Esta representação da função original nos permite analisar a série temporal de diversas perspectivas, devido ao fato das representações exporem padrões e características antes encobertas pela desordem da série temporal. Assim, as funções $g_i(t)$ são utilizadas como entradas do modelo, ao invés da função $f(t)$. Cada função $g_i(t)$ representa um determinado nível de detalhe da função original, como mostra a Figura 10. Pode-se notar na figura como a função $g_6(t)$ possui uma frequência maior que as demais, e conseqüentemente contém mais detalhes sobre a série $f(t)$. Dessa forma, muitos trabalhos como o de Altan, Karasu e Bekiros (2019) utilizam esta decomposição e descartam as séries com mais detalhes com o intuito de realizar uma filtragem dos ruídos. Neste trabalho, a decomposição de séries não é utilizada como um filtro de ruídos. No presente trabalho as frequências não serão descartadas e as maiores frequências serão utilizadas na arquitetura de CHM para criar contextos a partir da série original, auxiliando o modelo a encontrar padrões de compra ou venda de ativos na bolsa de valores.

A seguir serão descritas as técnicas de decomposição utilizadas neste trabalho.

Figura 10 – Decomposição da série temporal do preço de fechamento do ativo da empresa Apple Inc. (AAPL) entre os anos de 2014 e 2021



Fonte: Autor

3.2.1 Discrete Wavelet Transform

A DWT é uma decomposição que baseia-se nas chamadas ondaletas, ou *wavelets*, para decompor sinais no tempo. As ondaletas são pequenas funções com média igual a zero, que se deslocam no tempo para encontrar mudanças bruscas na série analisada. Esta transformada tem características similares à transformada de Fourier.

A transformada de Fourier realiza a transformação da série multiplicando funções senoidais pela série original, formalmente descrito na Equação (2).

$$X(F) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi Ft} dt \quad (2)$$

onde F corresponde à frequência analisada, t corresponde ao tempo analisado e $x(t)$ à série temporal em análise.

A integral da Equação (2) é tomada no domínio da função para a obtenção de uma função no espaço da frequência e amplitude que define a ocorrência de qualquer frequência F na função $x(t)$. A função obtida é então utilizada para definir quais frequências têm maior relevância na série $x(t)$ (BRACEWELL; BRACEWELL, 1986). Contudo, a transformada não possui a capacidade de definir em quais tempos t as frequências obtidas são reveladas. Para contornar esta situação, pode-se limitar a transformada para atuar apenas em janelas de tempo t compreendidas em um intervalo discreto. Assim, para cada janela, tem-se uma análise de Fourier e conseqüentemente tem-se também uma noção de tempo nos quais as frequências aparecem. Esta aplicação consecutiva da transformação de Fourier em janelas de tempo t sobre a série temporal, é denominada de *Short-Time Fourier Transform* (STFT). Porém, devido à utilização de uma janela, pode-se perder a precisão temporal da ocorrência da frequência. Esta precisão é comumente chamada de resolução do tempo e resolução de frequência. Conforme aumenta-se a resolução do tempo, reduzindo a janela de análise, a resolução de frequência também se reduz, uma vez que existem menos pontos para serem analisados. Similarmente, se aumentada a janela de análise, tem-se mais resolução de frequência, mas a precisão temporal se reduz.

Em contrapartida com a transformada de Fourier, a DWT faz a substituição das funções senoidais por uma ondaleta, conforme a Equação (3).

$$X(a, b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}^*(t) dt \quad (3)$$

Onde ψ é a função ondaleta. Devido a esta substituição, a DWT possui uma maior precisão ao determinar quais frequências ocorrem em quais tempos t . Isto ocorre pois as ondaletas possuem a capacidade de se "esticar" para identificar as mudanças de frequência. A proporção em que a ondaleta será esticada é definida pelo parâmetro b na Equação (3), onde $b \in \mathbb{R}$. Assim como a STFT, as ondaletas se trasladam no tempo para obter uma resolução temporal das ocorrências das frequências. Esta translação é definida pelo parâmetro a na Equação (3), onde $a \in \mathbb{R}$. Os valores de a e b definem a precisão em que a ondaleta irá determinar o tempo de ocorrência de cada frequência.

Embora a STFT consiga determinar o tempo nas quais as frequências aparecem, a resolução de tempo e frequência são baixas. Entretanto, a DWT torna essa resolução maior com a aplicação de ondaletas.

3.2.2 Transformada de Hilbert-Huang

A transformada de Hilbert-Huang, proposta por Huang et al. (1998), tem como seu principal objetivo a análise de uma série temporal caótica. Esta transformada é composta de duas principais etapas, são elas: EMD e transformada de Hilbert. A junção dessas duas etapas resulta em uma transformada que apresenta informações instantâneas sobre a frequência de uma série temporal. Isto implica em uma maior resolução de frequência e tempo da função analisada. Dessa forma, a seguir serão apresentadas o funcionamento das duas etapas citadas.

A EMD é uma técnica utilizada para a decomposição de séries temporais de forma empírica; isto é, a técnica não assume premissas a priori e baseia-se inteiramente nos dados analisados. Esta técnica foi proposta por Huang et al. (1998) para a análise de sinais naturais. No princípio, a técnica foi utilizada na NASA para estudar as ondas do mar. Os sinais obtidos da natureza geralmente tem uma característica caótica e tendem a ser não-estacionários.

Em contraste com a transformada de Fourier, a EMD não realiza a projeção os dados originais em um domínio de frequência, se mantendo no espectro de tempo. Dessa forma, a EMD, extrai funções $g_i(t)$, denominadas de IMF, que compõem uma função original $f(t)$, porém, em contraste com a transformada de Fourier, mantêm informações instantâneas sobre as frequências.

As funções IMF são funções com média zero extraídas da série temporal original. Em contraste com a transformada de Fourier as funções extraídas não são necessariamente senoidais, e variam de acordo com a função original analisada. Para uma função extraída ser considerada IMF, a mesma deve ter duas características fundamentais apresentadas a seguir.

- a) **A diferença entre a quantidade picos e a quantidade de vezes que a função cruza o eixo X deve ser no máximo um.**
- b) **Em qualquer ponto, o valor médio da envoltória definida pelos valores de local máximo, e envoltória definida pelos valores de local mínimo deve ser zero.**

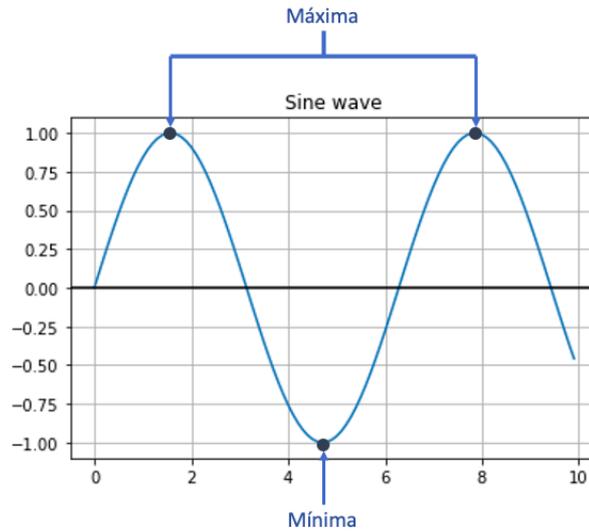
Considerando um caso onde os dados originais foram normalizados pela Equação 4,

$$z = (x - u)/s \quad (4)$$

o número de máximas é dado pela quantidade de pontos que assumem o maior valor antes da inversão da tendência. Já o número de mínimas é dado pela quantidade de pontos que assumem o menor valor antes da inversão da tendência, como mostra a Figura 11.

Cada IMF extraída representa uma frequência de oscilação de função analisada. Sendo assim, uma função $f(t)$ que apresente mais de uma frequência de oscilação também terá mais

Figura 11 – Máximas e mínimas de uma função senoidal.



Fonte: Autor

de uma função IMF. Esta é grande aplicabilidade da EMD, pois, dessa forma, podemos analisar uma série temporal caótica de forma mais organizada e mais estacionária.

As IMF são extraídas através de um processo denominado *sifting*, ou "peneiramento". Este processo é iterativo e pode ser executado diversas vezes até que se possa obter todas as IMF de uma função. O método define como etapa inicial a identificação dos pontos de máximos e mínimos da função analisada. Esses pontos serão utilizados para a composição de duas funções que envolvem a série analisada, E_{sup} e E_{inf} . A função, E_{sup} , que envolve série pelo lado superior é obtida performando uma interpolação cubica de *splines* sobre os pontos de máximo. A função, E_{inf} , que envolve série pelo lado inferior é obtida performando uma interpolação cubica de *splines* sobre os pontos de mínima. Logo após, é feita a média das duas funções utilizando a Equação 5

$$E_m = (E_{sup} + E_{inf})/2 \quad (5)$$

Após obter a função E_m , a mesma é subtraída da função original e a função obtida da subtração é avaliada nos dois critérios citados anteriormente que definem uma IMF. Caso os critérios sejam atingidos a função obtida é armazenada como uma IMF no tensor C_i . Logo após, a IMF encontrada é subtraída da função original, caso a função obtida de subtração seja um resíduo todas as IMF foram encontradas. Se esta função não for um resíduo, a função obtida substitui a função original e o processo se repete como descrito.

Dessa forma, pode-se obter a Equação (6)

$$X(t) = \sum_{k=1}^n C_k(t) + r_n \quad (6)$$

onde $X(t)$ a função original, $C_i(t)$ é o tensor que armazena as IMF e $r_n(t)$ a função resíduo.

Após obter as funções IMF, podemos aplicar a transformada de Hilbert para se obter informação de frequência instantânea do sinal analisado. A EMD antes da transformada de Hilbert se faz necessária pois em um determinado tempo t , a função original pode apresentar mais de uma frequência de oscilação. A EMD auxilia a transformada de Hilbert a encontrar todas as frequências oscilatórias presentes na série temporal.

Para se obter a transformada de Hilbert das IMF temos a Equação (7)

$$D_k(t) = \frac{1}{\pi} P \int \frac{C_i(\tau)}{t - \tau} d\tau \quad (7)$$

onde P é o valor principal de Cauchy, tem-se uma função analítica expressada pela Equação (8)

$$Z_k(t) = C_k(t) + jD_k(t) = A_k(t)e^{j\theta_k(t)} \quad (8)$$

Então a frequência instantânea de $X(t)$ pode ser expressa pela Equação (9)

$$X(t) = Re\left\{ \sum_{k=1}^n A_k(t)e^{j\theta_k(t)} \right\} + r_n(t) \quad (9)$$

3.3 REDES NEURAIS RECORRENTES

As RNN são redes neurais que possuem características que permitem a modelagem de predições sobre dados sequenciais. Os dados sequenciais são aqueles em que os valores anteriores da variável predita, têm um impacto relevante sobre os valores futuros da mesma. São exemplos de dados sequenciais as frases de qualquer linguagem natural. As frases de uma determinada linguagem são compostas de palavras que unidas formam informações relevantes para serem transmitidas. A importância da ordem em que as palavras aparecem na frase é a característica que torna esses dados sequenciais.

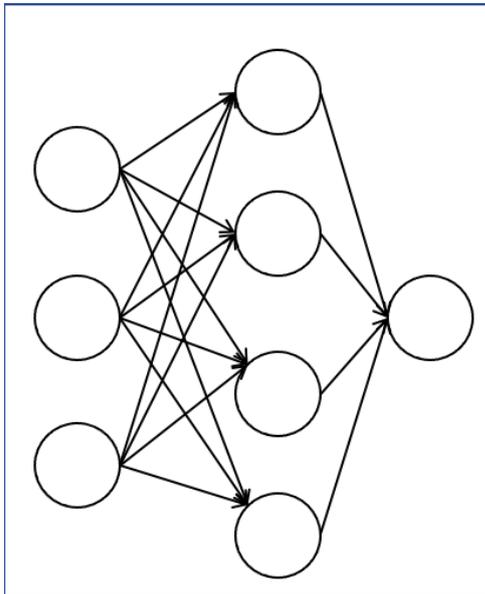
Outra característica importante das redes RNN é a capacidade de processar dados, multivariados, ou não, com comprimentos não fixos. Dessa forma, pode-se utilizar diferentes quantidades de pontos para o treinamento e para a predição dos valores. Este tipo de processamento não é possível ao utilizar redes *feed forward*.

Assim como uma rede neural *feed forward*, as RNN são compostas por neurônios que contêm uma função de ativação. O conjunto destes neurônios formam o que é denominado de "camada" na rede neural. As redes neurais mais comuns têm três tipos de camadas, são elas: camada de entrada, ou *input layer*; camada escondida, ou *hidden layer*; e a camada de saída, *output layer*.

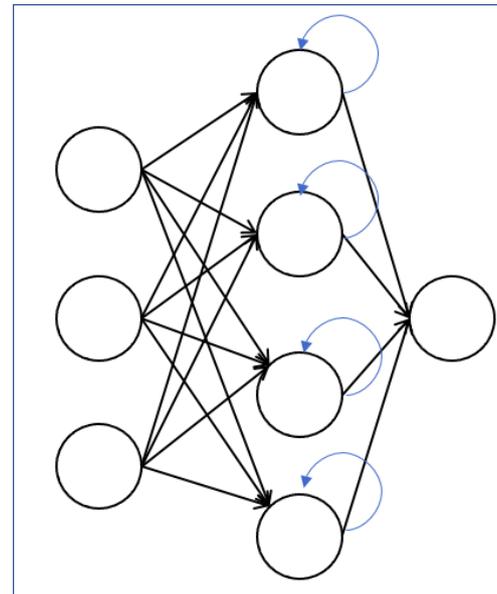
As RNN, em contraste com as redes *feed forward*, apresentam neurônios que se ligam a si mesmos. Esta característica é o que permite as redes RNN a considerarem a informação de sequência dos dados. A sequência dos dados é muito importante nas séries temporais, pois permite que padrões apresentados anteriormente sejam aprendidos pela RNN, para serem utilizados para prever valores futuros.

A Figura 12a apresenta uma rede neural *feed forward*. Esta rede somente envia informações em uma única direção, logo não considera informações sobre a sequência dos dados. Já a

Figura 12 – Tipos de rede neural



(a) Rede Neural Feed Forward



(b) Exemplo de Rede Neural Recorrente

Fonte: Autor

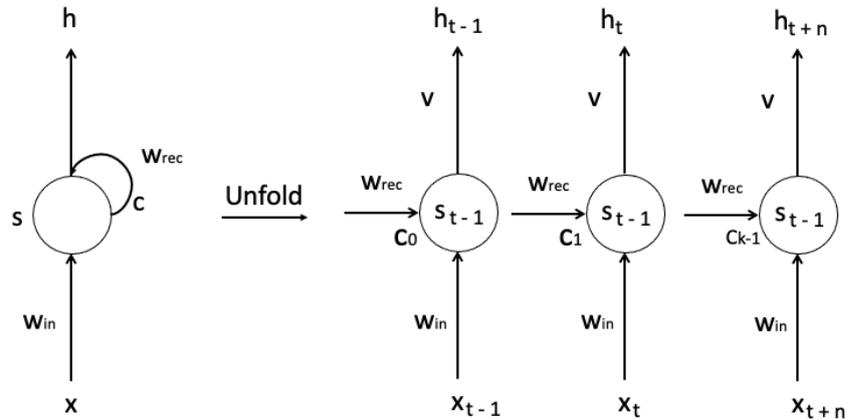
Figura 12b apresenta uma RNN, que leva em consideração informações da sequência dos dados. Pode-se notar na figura que nem todos os neurônios enviam informações em uma única direção, alguns neurônios se retroalimentam, e esta característica permite a captura de informação sobre a sequência dos dados.

O treinamento da rede RNN também utiliza o método de *Backpropagation Through Time* (BPTT) para atualizar os pesos da rede. Porém, assim como as redes *feed forward* com diversas camadas escondidas, as RNN apresentam uma explosão, ou, desaparecimento de gradiente. Isso se deve ao fato da RNN compartilhar os pesos ao longo do tempo. É possível visualizar na Figura 13 o *unfold* de uma RNN; isto é, como o neurônio é visto no tempo. Assim, quanto maior o tempo considerado para o treinamento da rede, maior será a probabilidade da explosão, ou, o desaparecimento de gradiente.

O desaparecimento, ou explosão, de gradiente acontece devido ao fato de ambos BP e BPTT serem computados através de derivadas parciais dos elementos da rede neural. Para computar o treinamento das redes utilizando BP ou BPTT, é utilizado a derivada parcial da função de erro em relação aos pesos da rede. Dessa forma obtemos a Equação (10)

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \sigma'(W_{rec} \cdot c_{t-1} + W_{in} \cdot x_t) \cdot W_{rec} \right) \frac{\partial c_1}{\partial W} \quad (10)$$

onde E é o erro da rede, u são os pesos de entrada, w são os pesos da recorrência, v o vetor de estados de cada tempo t , W todos os pesos da rede e σ a função de ativação tanh. Considerando

Figura 13 – *Unfold* de uma RNN

Fonte: Autor

que a derivada da função \tanh está contida no intervalo $[1, 0]$, o produtório da Equação (10) tende a 0 quando k assume maiores valores.

Porém, existem métodos para evitar o desaparecimento de gradiente nas redes RNN. A seguir serão apresentados dois destes métodos.

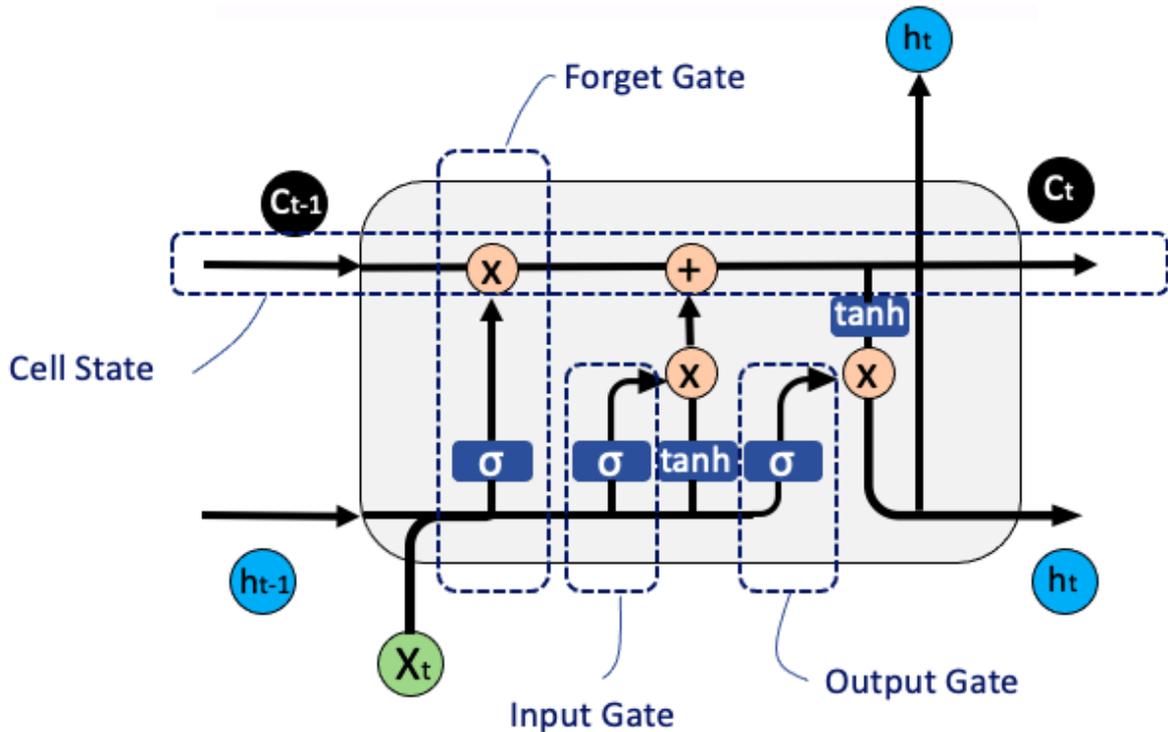
3.3.1 *Long Short-Term Memory*

As LSTM foram propostas por Sepp Hochreiter e Jürgen Schmidhuber (1997) para mitigar o problema de desaparecimento, ou explosão, de gradiente das RNN. Como dito anteriormente, tal problema ocorre devido a aplicação da derivada parcial na rede para a otimização dos pesos através dos métodos BP, ou BPTT. Dessa forma, para contornar este problema, as LSTM adicionam parâmetros na rede que compensam o produto obtido pelas derivadas parciais, evitando o desaparecimento do gradiente.

A LSTM apresenta três estruturas chamadas de *gates*, ou chaves, que comandam o fluxo de informações dentro da rede, são elas: *forget gate*, *input gate* e *output gate*. Essas são as estruturas responsáveis por impedir o desaparecimento de gradiente ocasionado pelo BP. A junção dessas estruturas formam uma célula de LSTM apresentada na Figura 14. As redes LSTM também contemplam um fluxo de informações denominado de *cell state* C_t , ou estado da célula. Este fluxo de informações define uma memória de longo prazo para a rede, e as chaves, anteriormente citadas, são responsáveis por alterar ou eliminar valores deste fluxo. Por fim, as células contêm um fluxo secundário denominado de *hidden state* h_t , ou estado escondido. Este estado envia informações adicionais que a célula seguinte deve utilizar para definir os valores do *cell state*.

A *forget gate* é responsável por definir quais informações devem permanecer na rede e quais devem ser esquecidas. Formalmente, essa chave é formada por uma rede composta por funções de ativação sigmoid e uma multiplicação ponto a ponto entre o vetor C_{t-1} e a concatena-

Figura 14 – Célula da LSTM que mitiga o risco de desaparecimento de gradiente



Fonte: Autor

ção dos vetores h_{t-1} e X_t . A rede sigmoide é utilizada para manter os valores da concatenação dos vetores no intervalo $[0, 1]$. A multiplicação ponto a ponto dos vetores obtidos é a característica que permite a LSTM esquecer valores anteriores, ou mantê-los, de acordo com o vetor obtido pela rede sigmoide. No momento um elemento do vetor assume um valor próximo a 0 as informações são esquecidas, e quando um elemento assume um valor próximo a 1 as informações são mantidas.

A segunda estrutura que compõe a célula de LSTM é denominada de *input gate*. Esta chave define quais valores do estado atual X_t devem ser concatenadas ao *cell state* atual. A estrutura é composta por uma rede sigmoide, uma rede com funções de ativação tanh, uma multiplicação ponto a ponto dos vetores obtidos pelas saídas das redes e uma operação de adição dos resultados da multiplicação ponto a ponto no *cell state*. A rede sigmoide, assim como na estrutura *forget gate*, é responsável por selecionar informações que devem ser adicionadas ao *cell state*. A rede tanh é utilizada para regularizar a rede e impedir que valores inseridos na rede tenham valores grandes devido às operações da rede LSTM.

A ultima estrutura é denominada de *output gate*. Esta chave tem como seu objetivo definir quais valores da *cell state* devem entrar para o estado escondido da próxima célula LSTM. A composição desta chave é uma rede sigmoide e uma multiplicação ponto a ponto entre o vetor de saída de sigmoide e o vetor da *cell state*. Esta chave é requerida para utilizar os valores de entrada do tempo t , X_t , como parâmetros para a decisão de quais valores do *cell state* seguirão para o próximo estado escondido h_t .

Matematicamente, as estruturas citadas acima, formam adições dentro do produtório da Equação (10). Se definirmos cada estrutura citada anteriormente de modo formal temos que

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) \quad (11)$$

onde f_t representa a *forget gate* no tempo t

$$i_t = \tanh(W_c \cdot [h_{t-1}, x_t]) \otimes \sigma(W_i \cdot [h_{t-1}, x_t]) \quad (12)$$

onde i_t representa a *input gate* no tempo t . Podemos separar a *input gate* em duas equações

$$i_{1_t} = \tanh(W_c \cdot [h_{t-1}, x_t]) \quad (13)$$

e

$$i_{2_t} = \sigma(W_i \cdot [h_{t-1}, x_t]) \quad (14)$$

então temos que

$$i_t = i_{1_t} \otimes i_{2_t} \quad (15)$$

assim podemos definir a *cell state* como

$$c_t = c_{t-1} \otimes f_t \oplus i_t \otimes i_{2_t} \quad (16)$$

derivando a Equação (16) pela regra da cadeia temos

$$\frac{\partial c_t}{\partial c_{t-1}} = \frac{\partial f_t}{\partial c_{t-1}} \cdot c_{t-1} + \frac{\partial c_{t-1}}{\partial c_{t-1}} \cdot f_t + \frac{\partial i_{2_t}}{\partial c_{t-1}} \cdot i_{1_t} + \frac{\partial i_{1_t}}{\partial c_{t-1}} \cdot i_{2_t} \quad (17)$$

pode-se notar que a derivada resultou em somas, então substituindo na Equação (10) temos

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \frac{\partial c_t}{\partial c_{t-1}} \right) \frac{\partial c_1}{\partial W} \quad (18)$$

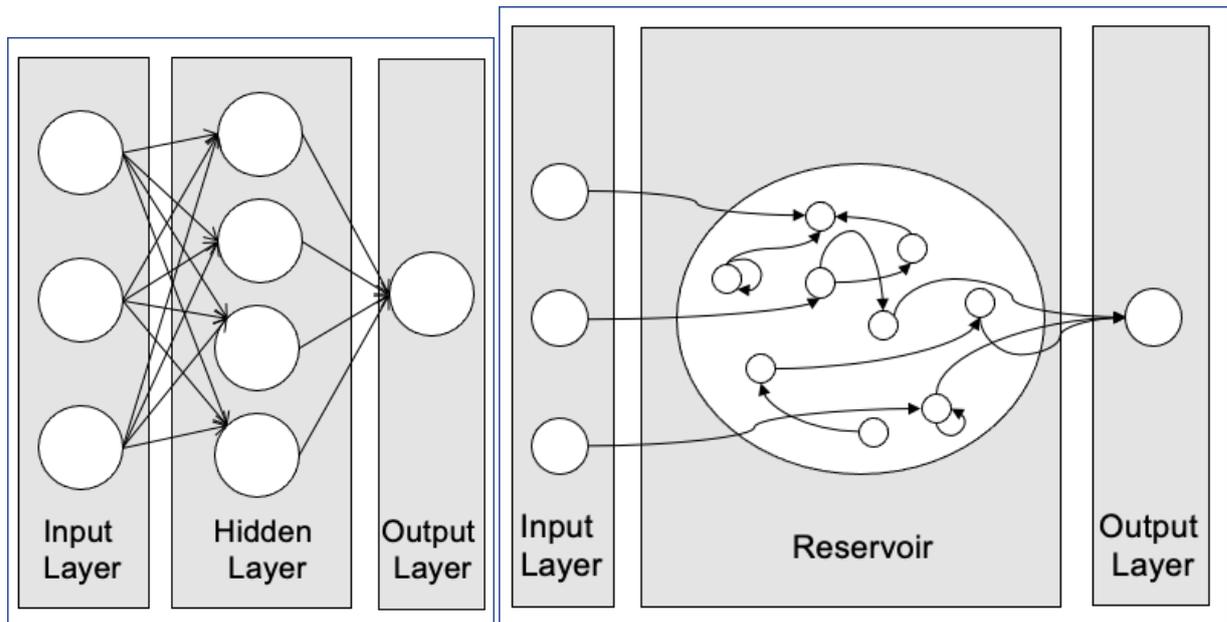
nesta forma as adições dentro do produtório impedem o desaparecimento do gradiente.

3.3.2 Echo State Network

As ESN, ou Computação de Reservatório, são uma variante das redes neurais recorrentes tradicionais propostas por Jaeger (2007). Estas redes possuem características especiais que permitem a modelagem de séries de dados caóticas devido a presença de uma estrutura denominada de *reservoir*. As ESN estão enquadradas em no espectro de modelos não-lineares.

De modo geral as ESN têm uma arquitetura similiar às redes neurais comuns. Apresentam uma camada de neurônios para a entrada de valores, camada de entrada (*input layer*), uma cama de neurônios para a saída dos valores preditos, camada de saída (*output layer*) e uma camada escondida (*hidden layer*) posicionada entre as duas anteriores. Porém, em contraste com redes neurais tradicionais, *feed forward*, as redes ESN são compostas de uma camada escondida com centenas, ou milhares, de neurônios não-densamente interligados, como mostra a Figura

Figura 15 – Tipos de redes neurais



(a) Rede Neural Feed Forward

(b) Echo State Network

Fonte: Autor

15b. Os neurônios desta camada tem conexões esparsas entre si; isto é, nem todo e qualquer neurônio é ligado com todos da mesma camada de *reservoir*. Outra característica que difere as redes ESN das tradicionais, é a interligação da camada de entrada com os neurônios do *reservoir*. Estas ligações são esparsas e geralmente aleatórias. Esta propriedade também é verdadeira para a camada de saída, muitas vezes denominada de camada de leitura nas ESN.

Embora ambas as redes ESN e *feed forward* tenham estrutura similares, o treinamento das redes ESN não é realizado por meio de métodos como BP ou BPTT. Em contraste com as redes *feed forward*, nem todos os pesos da ESN são modificados na etapa de treinamento do modelo. Os pesos da camada de *reservoir*, e os pesos que ligam a camada de entrada com o *reservoir*, são fixos e muitas vezes inicializados aleatoriamente. Estes pesos não se alteram na etapa de treinamento e permanecem sempre imutáveis. Entretanto, a camada de saída têm seus pesos treinados mas não pelos métodos tradicionais de treinamento de redes. O treinamento da ESN é realizado, geralmente, por modelos lineares como *Ridge Regression* (RR), que é uma regressão linear simples com penalidades de coeficientes. Esta característica permite a rede ESN terem o tempo de treinamento reduzido quando comparadas às redes *feed forward*.

A camada de *reservoir* permite às redes ESN projetar os dados originais em dimensões maiores. Este princípio toma vantagem do teorema de Cover (COVER, 1965), que implica que qualquer sistema não linearmente separável pode ser se projetado em dimensões maiores para que se torne linearmente separável. A metodologia de projeções dos dados em dimensões alternativas já foi exploradas em outros modelos como o SVM (CORTES; VAPNIK, 1995).

Os reservatórios das redes ESN devem respeitar três características essenciais sobre a conectividade e proporção do *reservoir*. A primeira delas é a ordem de grandeza do reservatório; isto é, a quantidade de neurônios. Tal característica permite que a rede crie representações diversas sobre os dados originais, para que se possa tomar vantagem dessa transformada. A segunda é a garantia que a conectividade dos neurônios é esparsa; isto é, não apresenta conexões densas entre os neurônios da camada do reservatório. É comum a conectividade ser inferior a 20% dos neurônios. Assim, a propagação de informação dentro do reservatório se torna lenta, impedindo a contaminação, ou entrelaçamento, de valores dentro reservatório. E por último, a inicialização dos pesos fixos devem, preferencialmente, ser inicializadas por uma distribuição uniforme com a média zero (LUKOSEVICIUS; JAEGER, 2009).

Segundo Jaeger (2007), a rede ESN deve apresentar uma propriedade denominada de *echo state property*, ou propriedade de estado de eco. Esta propriedade garante que um estado anterior, $x(n)$, e um conjunto de entradas anterior, $u(n)$, não tenha grandes influências no estado futuro $x(n + k)$. Assim, os efeitos dos estados anteriores, e das entradas anteriores, devem gradualmente desaparecer quando k tende a infinito. Essa propriedade pode ser assegurada se, a matriz de pesos W do *reservoir* for normalizada de modo que o maior valor absoluto do *eigenvalue* da matriz seja menor que 1.

Matematicamente, podemos definir a saída da rede ESN, pela Equação (19)

$$y(n) = f_{out}(W_{out}[u(n)|x(n)]) \quad (19)$$

onde $W_{out} \in \mathbb{R}^{N_y(N_u+N_x)}$, f_{out} corresponde a uma função não linear como a sigmoid, $u(n)$ é a matriz de entradas e $x(n)$ os valores de entrada da rede.

3.4 ANÁLISE CONTEXTUAL

Contexto é normalmente entendido como informações relevantes a se levar em consideração ao analisar uma cena. A definição de contexto em visão computacional é dada como "informação relevante para detectar um objeto, mas não necessariamente relativo a sua aparência"(WOLF; BILESCHI, 2006). Existem diversos trabalhos que sugerem que o ser humano tem a capacidade de interpretar uma imagem com base em contextos (COX; MEYERS; SINHA, 2004; BAR et al., 2005), conseguindo assim identificar, por exemplo, um objeto que tenha algumas de suas partes ocultas.

A análise contextual tem diversas aplicações relevantes em visão computacional, desde a área médica (TU; BAI, 2010), identificação de pessoas em álbuns de fotos (ANGUELOV et al., 2007), e identificação de objetos em geral (TORRALBA; MURPHY; FREEMAN, 2004). Essa técnica pode visar não somente automatizar a habilidade de identificar objetos, do ser humano, mas também a interpretação de cena em questão.

A ideia de análise contextual sugere que é possível inferir a identificação de objetos específicos se baseando nos objetos adjacentes, que não fazem parte do objeto foco da imagem apresentada. Essa identificação é realizada normalmente por características visuais que não

necessariamente pertencem ao objeto em questão. Na neurociência este processo é denominado de percepção visual baseada em contexto, ou percepção *top-down*. O trabalho de Seyedhosseini, Sajjadi e Tasdizen (2013) propõe uma metodologia que faz a extração automática do contexto de uma imagem utilizando técnicas de simplificação da imagem em diversos níveis.

O presente trabalho sugere que esta ideia pode ser transposta para séries temporais utilizando na metodologia uma técnica para simplificação das séries. Para implementar a simplificação das séries, é possível aplicar técnicas de decomposição de séries, como mostra a Seção 3.2. Sendo assim, a arquitetura proposta cria um contexto sobre as séries temporais da bolsa sem que sejam apresentados todos os detalhes da série temporal.

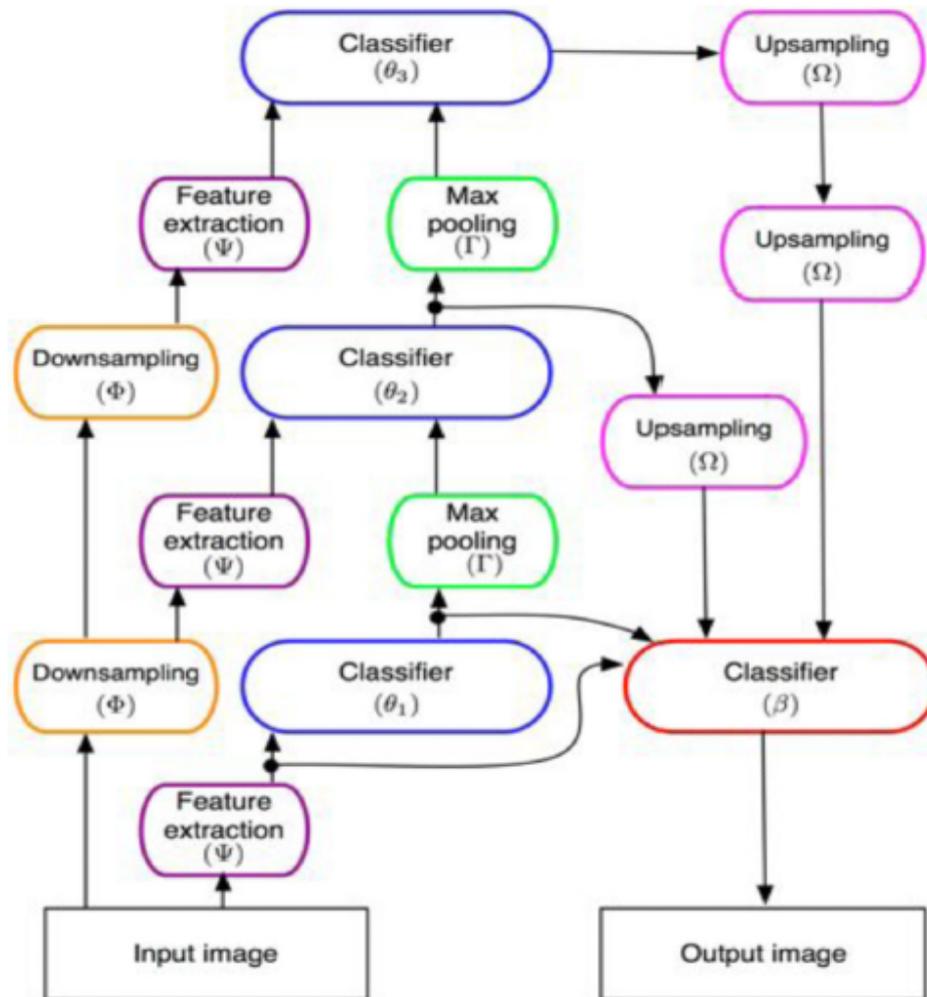
3.5 MODELO HIERÁRQUICO EM CASCATA

Até onde sabemos não foi encontrado na literatura relacionada ao mercado financeiro, modelos hierárquicos para predição de ativos. Considerando a ideia de captura de informações em cascata, pode ser uma boa estratégia para predição de ativos a partir de vários ativos correlatos, um dos objetivos desse projeto de pesquisa é a modelagem em cascata para análise e previsão do comportamento do mercado. Sendo assim, esta seção apresenta uma breve explicação mostrada a seguir do modelo hierárquico proposto por Seyedhosseini, Sajjadi e Tasdizen (2013) para um cenário de visão computacional, mas que será adaptado para o cenário aqui proposto, envolvendo mercado de ações.

No trabalho de Seyedhosseini, Sajjadi e Tasdizen (2013) foi apresentado um modelo hierárquico em cascata, chamado CHM. Trata-se de um *framework* utilizado em visão computacional para capturar informações contextuais de uma imagem, considerando ideias provenientes da neurociência de que este tipo de informação pode levar a uma maior assertividade destes sistemas. Esse *framework* é denominado de hierárquico em cascata uma vez que as informações de treinamento não passam por vários modelos que se comunicam mutuamente. A cada nível do modelo hierárquico a rede utiliza informações previstas pela rede do submodelo anterior juntamente com o *downsampling* da imagem, como mostra a Figura 16.

O trabalho de Seyedhosseini, Sajjadi e Tasdizen (2013) utiliza o modelo hierárquico em cascata com o objetivo de criar um modelo para a segmentação automática de uma imagem. Para isso, os autores separam a arquitetura CHM em três partes, são elas: etapa *bottom-up*, *top-down* e a cascata. A etapa *top-down* realiza o treino de vários classificadores, que por sua vez levam em consideração a imagem original e a imagem com um grau de redução de qualidade, chamado de *downsampling*. Por sua vez a etapa *bottom-up* é composta de n_i níveis, onde $i = \{1, 2, 3, \dots\}$, onde o nível n_i cria um contexto para o nível n_{i+1} . Ao alcançar o nível n_i , a etapa de *top-down* é inicializada. Esta etapa faz o treinamento de um único classificador que utiliza informação de saída dos n_i classificadores treinados na etapa *bottom-up*. Por fim, o modelo em cascata realiza os procedimentos descritos m vezes, a cada iteração m_i o modelo utiliza a saída da iteração m_{i-1} como entrada do nível 1 da etapa de *bottom-up*. Este modelo visa a extração de contexto das

Figura 16 – Exemplo de modelo hierárquico proposto por Seyedhosseini, Sajjadi e Tasdizen (2013)



Fonte: Seyedhosseini, Sajjadi e Tasdizen (2013)

variáveis, e segundo os autores tem resultados mais satisfatórios quando comparados com outros da categoria. Os autores ainda afirmam que embora tenham utilizado um modelo específico para os modelos classificadores intermediários, é possível utilizar qualquer modelo que seja robusto contra *overfit* e rápido no treinamento.

3.6 PARÂMETROS DO MERCADO FINANCEIRO

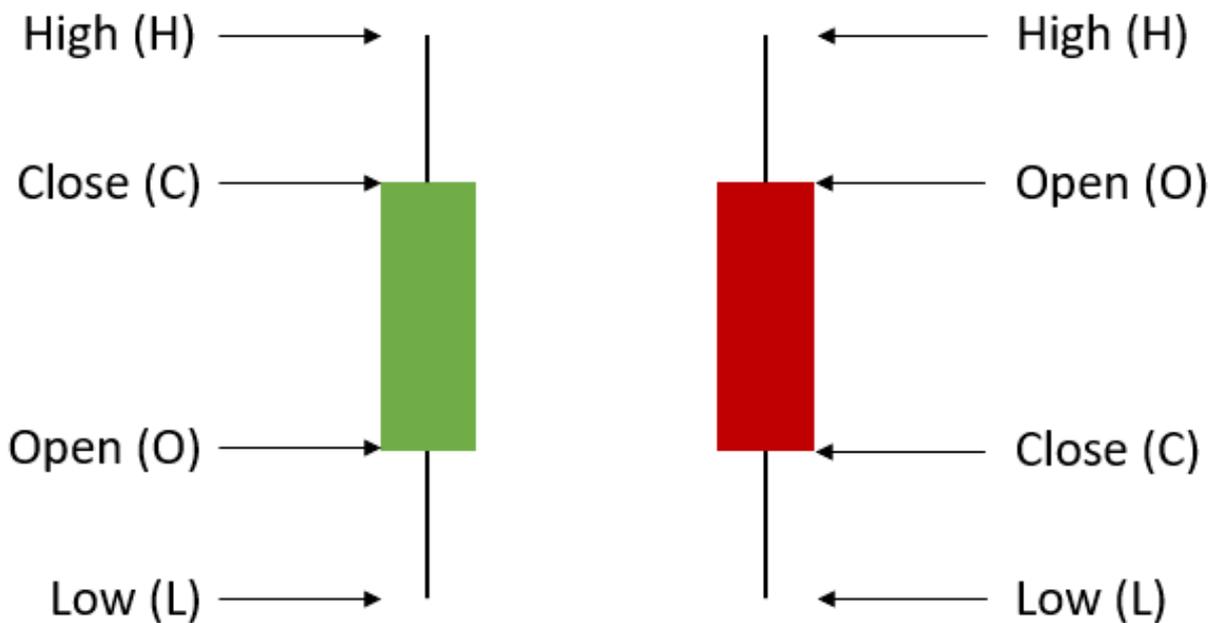
Ao longo dos anos a análise da bolsa de valores foi se tornando cada vez mais robusta e diversificada, tanto em diversidade de ferramentas utilizadas para a análise, quanto na assertividade delas. Contudo, todas as ferramentas utilizadas para auxiliar os negociadores da bolsa a lucrar mais na compra e venda de papéis utilizam um, ou mais, valores numéricos para quantificar preços, tendências e volatilidade.

Nesta seção serão apresentados os conceitos dos valores utilizados pelo modelo proposto neste trabalho.

3.6.1 *Open High Low Close (OHLC)*

Esta categoria é a mais básica e mais utilizada nos trabalhos que visam prever o valor de um ativo na bolsa de valores (KUMAR; JAIN; SINGH, 2020; JIANG, 2020). Como o próprio nome diz, os valores OHLC são compostos por 4 informações, são elas: o valor de abertura da ação (O), o valor mais alto que a ação atingiu (H), o valor mais baixo que a ação atingiu (L), e o valor de fechamento da ação (C). A Figura 17 mostra os dados OHLC através de um gráfico *candlestick*.

Figura 17 – Representação *Candlestick* sobre os dados OHLC



Fonte: Autor

Os gráficos de *candlestick* são exclusivamente formados pelas informações de OHLC do ativo. Estes gráficos são formados pela composição de diversos *candlestick* seguidos. Cada *candlestick* representa o período em que a informação OHLC está compreendida. Um *candlestick* pode representar 1 dia, 15 minutos, 30 minutos, 1 hora, e assim por diante. Sendo assim, as informações de OHLC são referentes ao período do *candlestick* analisado. Neste trabalho, serão utilizados exclusivamente os *candlesticks* de 1 dia. Este período foi selecionado devido à escassez de dados mais curtos como 15 minutos ou 1 hora.

3.6.2 Frequência de Oscilação

Os decompositores de séries temporais, apresentados na Seção 3.2, são capazes de extrair valores de frequência de oscilação no domínio do tempo, além de decompor uma série em diversas outras mais simples. Desta forma, esse valor pode ser utilizado para identificar em quais tempos t a volatilidade do preço da ação foi alta. Assim, revelando uma informação que pode ser utilizada no modelo proposto como variável explicativa das variações dos preços futuros.

3.6.3 Variáveis de Análise Técnica

As *features* de análise técnica são informações geradas a partir de técnicas estatísticas aplicadas ao valor OHLC da ação. São exemplos de *features* de análise técnica: Média Móvel Simples - *Simple Moving Average* (SMA), Média Móvel de Convergente e Divergente - *Moving Average Convergence Divergence* (MACD), Oscilador Estocástico - *Stochastic Oscillator* (SO), Média Exponencial Tripla - *Triple Exponential Average* (TRIX) e Índice Médio Direcional - *Average Directional Index* (ADX), Bandas de Bollinger - *Bollinger Bands* (BB), Índice de Força Relativa (IFR), entre outros. Neste trabalho, serão utilizadas as técnicas SMA, BB, IFR, e SO.

Estas informações são frequentemente utilizadas na área de predição de valores de ativos no mercado financeiro (KUMAR; JAIN; SINGH, 2020). No trabalho de Pimenta et al. (2018), os autores dividem as *features* de análise técnica em quatro sub-categorias, são elas: Seguidores de Tendência (*Trend Followers*), Osciladores (*Oscillators*), Sistemas de Banda (*Band Systems*) e Identificadores de Divergência (*Divergence Identifier*).

Os Seguidores de Tendência são destinados a identificar o principal movimento do preço de um ativo. Os Osciladores têm o objetivo de identificar pontos de reversão no movimento de um ativo. Os Sistemas de Banda são caracterizados pela adição de três curvas em torno da curva do preço do ativo, uma a cima, uma abaixo e uma intermediária. Por fim, os Identificadores de Divergência fazem a comparação entre o indicador e o movimento do preço, identificando pontos de correção do preço do ativo.

3.6.3.1 Médias Móveis

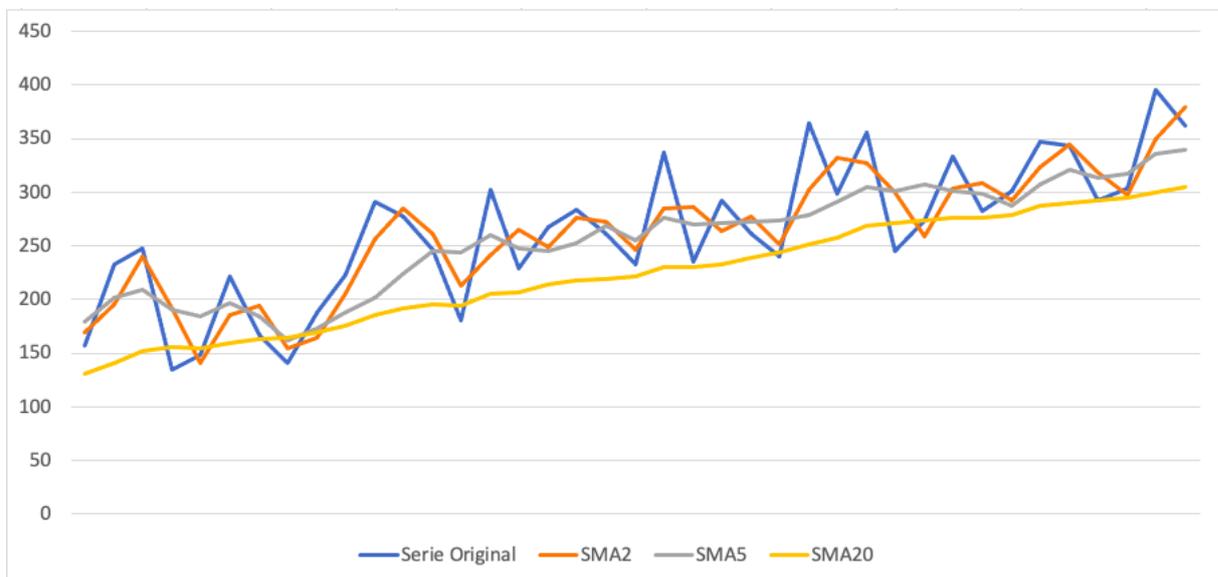
A média móvel, ou SMA, é uma técnica estatística, frequentemente utilizada em diversas áreas de análises de dados, que tem como principal objetivo suavizar os valores de um vetor V , onde $V \in \mathbb{R}$. Essa técnica pertence ao grupo de Seguidores de Tendência, sendo assim é majoritariamente aplicada para obter noções das movimentações futuras das séries temporais (KOUAISSAH et al., 2020). A média é calculada a partir da Equação (20)

$$M_{\tau}(t) = \frac{\sum_{i=0}^{\tau-1} x_{t-i}}{\tau} \quad (20)$$

onde $M_\tau(t)$ representa o média móvel da série temporal no tempo t , x representa a série temporal, τ é a janela de tempo utilizada para realizar a média e t é o instante observado. O valor de τ representa a janela em que a média móvel irá operar. Na bolsa de valores, por exemplo, é comum a utilização de uma média dos últimos τ preços de fechamento de um ativo para identificar sua tendência. O tamanho da janela τ determina o prazo da tendência, assim, valores altos para τ identificam tendências de longo prazo, e números baixos identificam tendências de curto prazo.

A Figura 18 mostra um exemplo de médias móveis, comparando janelas diferentes. É possível notar como a média com janela de tamanho 2 (SMA2) é pouco suavizada em relação à janela de tamanho 20 (SMA20). Sendo assim, quanto maior a janela utilizada para o cálculo da média móvel, maior será a suavização da curva.

Figura 18 – Gráfico comparativo entre as janelas de Média Móvel



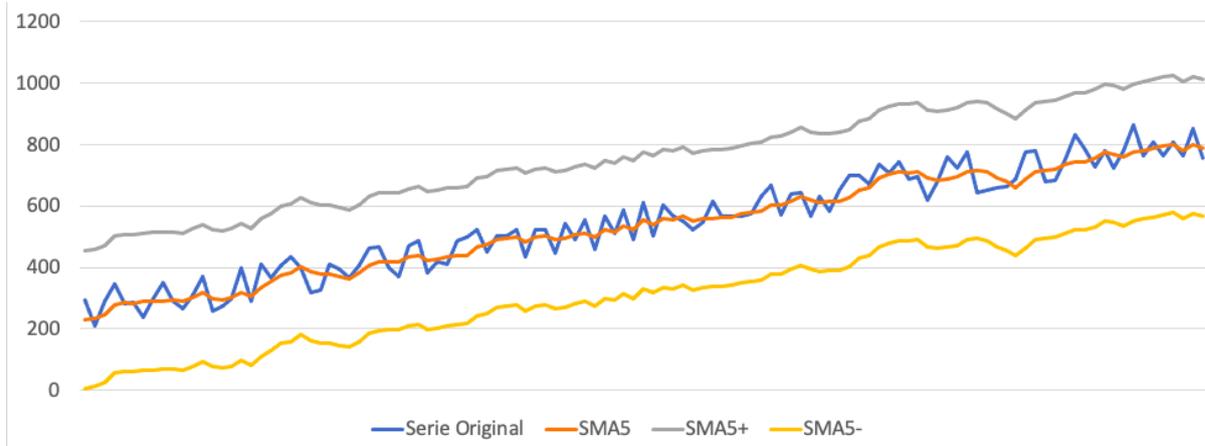
Fonte: Autor

3.6.3.2 *Bandas de Bollinger*

As bandas de Bollinger, ou BB, foram propostas por Bollinger (2001) para gerar indicadores de volatilidade dos ativos da bolsa. Essa técnica está contida no grupo de Sistemas de Bandas. Os Sistemas de Bandas são compostos por três curvas que, juntas, formam um invólucro na série temporal. As curvas são compostas por uma média móvel simples, comumente utilizando o período de vinte valores anteriores, e duas médias móveis que são derivadas da média citada anteriormente porém deslocadas por dois desvios padrão da série. A Figura 19 mostra uma série temporal e sua correspondente banda de Bollinger.

Este indicador técnico é frequentemente utilizado para identificar pontos de reversão de tendência na série temporal. É definido, então, uma regra para identificar tais pontos. O ponto de reversão de alta ocorre quando o preço do ativo inverte sua tendência de queda para uma

Figura 19 – Gráfico de demonstração de bandas de Bollinger



Fonte: Autor

tendência de alta. Essa reversão é identificada pelo cruzamento do preço do ativo com a curva de média inferior do invólucro da série temporal. Já o ponto de reversão de baixa é identificado pelo comportamento inverso ao descrito anteriormente; isto é, quando o valor do preço do ativo cruza a curva de média superior.

3.6.3.3 Stochastic Oscillator (SO)

Proposto por Lane (1950), os osciladores estocásticos, ou SO, pertencem a categoria de osciladores das técnicas de análise técnica. Os osciladores são utilizados para identificar o *momentum* da tendência de uma série temporal; isto é, o impulso dos valores na alta, ou baixa. Assim, é possível ter noções sobre o comportamento futuro da série. Os osciladores se baseiam principalmente na diferença entre o preço de fechamento do ativo hoje em comparação com o preço de fechamento de n dias anteriores. Assim como as médias móveis os osciladores estocásticos consideram uma janela de tempo para calcular dois valores que, juntos, compõem este indicador, são eles: %K e %D.

O valor de %K é definido pela Equação (21)

$$\%K = 100 \frac{Price - Low_n}{High_n - Low_n} \quad (21)$$

onde $Price$ se refere ao preço do ativo no tempo t , Low_n é o valor mínimo do preço do ativo nos últimos n períodos, $High_n$ é o valor máximo do preço do ativo nos últimos n períodos.

O valor de %D é dado pela média móvel dos últimos n valores de %K. Como mostra a Equação (22)

$$\%D = \frac{\%K_1 + \%K_2 + \%K_3 \dots \%K_n}{n} \quad (22)$$

A interpretação dos valores %D e %K é relativa a valores constantes utilizadas na literatura. Usualmente, quando os valores de %D e %K estão acima de 80, é considerado um ponto de

inversão negativo; isto é, o valor do ativo está alto indicando uma reversão na tendência de alta para uma tendência de baixa. Quando os valores de %D e %K estão abaixo de 20, o preço do ativo é considerado um ponto de inversão positivo; isto é, o valor do ativo está baixo indicando uma reversão na tendência de baixa para uma tendência de alta.

3.6.3.4 Índice de Força Relativa (IFR)

O índice de força relativa é pertencente ao grupo de Identificadores de Divergência, e foi proposto por Wilder (1978). Este indicador mede a força de um determinado ativo na bolsa, se baseando nos preços de fechamento dos períodos anteriores. O IFR tem características similares com o oscilador estocástico, pois também utiliza conceitos de *momentum* do preço do ativo, mas tem seu foco em identificar correções de preços ao invés de pontos de reversão.

O cálculo do índice se baseia no lucro, ou prejuízo, que o ativo apresentou ao longo de um determinado período. Primeiramente é identificada a movimentação do preço do ativo. Caso o preço do ativo no instante t for maior que um instante anterior $t - n$, então é computado a Equação (23)

$$U = close_t - close_{t-n} \quad (23)$$

caso o preço do ativo no instante t for menor que um instante anterior $t - n$, então é computada a Equação (24)

$$D = close_{t-n} - close_t \quad (24)$$

em seguida é calculado o valor de força relativa, expressada pela relação da Equação (25)

$$RS = \frac{MA(U, n)}{MA(D, n)} \quad (25)$$

onde MA representa uma média móvel com janela n . E por fim é calculado o índice da força relativa, pela Equação (26)

$$RSI = 100 - \frac{100}{1 + RS} \quad (26)$$

Assim como os osciladores estocásticos, o IFR é interpretado pelo cruzamento dos seus valores com duas constantes: 70 para alta e 30 para baixa. Quando os valores de IFR estão acima de 70, o preço do ativo é considerado como *overbought*; isto é, o valor do ativo está sobre-comprado e indica que seu preço não está "justo", e necessitará de correções. Quando os valores de IFR estão abaixo de 30, o preço do ativo é considerado como *oversold*; isto é, o valor do ativo está sobre-vendido e indica que seu preço não está "justo", e também necessitará de correções.

3.7 DEFINIÇÃO DE PONTOS DE REVERSÃO

Pontos de reversão são valores na série temporal de uma ação na bolsa que indicam que o preço está caminhando para uma alta, ou uma baixa (CAGINALP; LAURENT, 1998). Para

identificar um ponto de reversão primeiramente é necessário definir o que é uma *trend*, para avaliar se os indícios de reversão são definitivos ou apenas flutuações naturais no preço do ativo. Após esta definição é possível identificar um candidato a ponto e reversão, e por fim o ponto de reversão em si. Este trabalho utilizou a definição feita no trabalho de U et al. (2020). A seguir será detalhada as definições de *trend*, candidato a ponto de reversão e ponto de reversão.

3.7.1 *Trend*

Uma *trend*, no preço de um ativo, é um indicio de que os valores estão se ordenando de forma crescente, ou decrescente, assim, o estudo de Caginalp e Laurent (1998) define uma *trend* de alta e de baixa como as Equações (27) e (28), respectivamente.

$$MA_3(t-6) < MA_3(t-5) < \dots < MA_3(t-1) < MA_3(t) \quad (27)$$

$$MA_3(t-6) > MA_3(t-5) > \dots > MA_3(t-1) > MA_3(t) \quad (28)$$

onde $MA_3(t)$ representa a média móvel de 3 dias no dia t . Desta forma podemos definir *Trend* como a Equação (29)

$$Trend(t) = \begin{cases} 1 & \text{se a trend for de alta no tempo } t \\ -1 & \text{se a trend for de baixa no tempo } t \\ 0 & \text{caso contrário} \end{cases} \quad (29)$$

3.7.2 Candidato A Ponto de Reversão

Assim como no trabalho de U et al. (2020), este trabalho define um candidato a ponto de reversão (CRP), quando um ponto que tem como valor de alta para *Trend* tem um valor de fechamento maior do que o dia anterior, como mostra a Equação (30)

$$CRP(t) = \begin{cases} 1 & \text{se } Trend(t) = 1 \text{ e } Close(t) \leq Close(t-1) \\ -1 & \text{se } Trend(t) = -1 \text{ e } Close(t) \geq Close(t-1) \\ 0 & \text{caso contrário} \end{cases} \quad (30)$$

3.7.3 Pontos de Reversão

Utilizando as *trends*, e a definição de Candidatos a Ponto de Reversão podemos definir um ponto de reversão com a Equação (31)

$$RP(t) = \begin{cases} 2 & \text{se } CRP(t) = 1 \text{ e } \frac{MA_5(t) - \min_{i=1, a} MA_5(t+i)}{MA_5(t)} > \theta \\ 1 & \text{se } CRP(t) = -1 \text{ e } \frac{\max_{i=1, a} MA_5(t+i) - MA_5(t)}{MA_5(t)} > \theta \\ 0 & \text{caso contrário} \end{cases} \quad (31)$$

onde $RP(t) = 2$ representa um ponto de reversão de baixa, $RP(t) = 1$ representa um ponto de reversão de alta, e $RP(t) = 0$ não representa um ponto de reversão, em um período de tempo a e garantindo um retorno maior que θ .

3.8 NORMALIZAÇÃO DE DADOS

A principal ideia da normalização é prevenir que o modelo seja enviesado por um, ou mais, valores apenas por sua magnitude. Para isto, é subtraído do valor a sua média do período analisado e feita uma escala para a unidade de variância utilizando a Equação (32).

$$z = (x - u)/s \quad (32)$$

onde z é o novo valor que o *data point* assumirá, x é o valor a ser normalizado, u é a média de todos os *data points* analisados, e s é o desvio padrão dos *data points* analisados. A Figura 20a mostra os dados de volume de ações no seu estado natural, sem a normalização. É possível notar como as variações estão distorcidas, uma vez que uma ação tem uma magnitude de volume muito maior que as demais.

Na Figura 20b a normalização foi aplicada sendo possível notar a variância entre os ativos. Esta característica é importante para que a arquitetura proposta consiga prever o valor de um ativo baseando-se na variação temporal de outro.

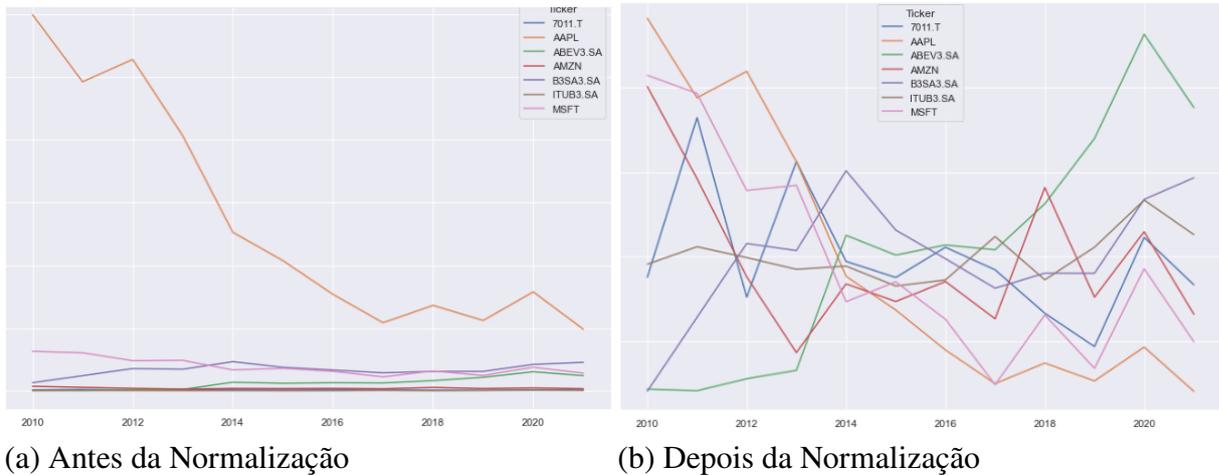
3.9 MÉTRICAS DE AVALIAÇÃO

Nesta seção serão apresentadas as métricas utilizadas neste trabalho. Essas métricas são utilizadas para avaliar a performance do modelo sobre os valores do mercado de ações.

3.9.1 MSE

O MSE é uma métrica utilizada para avaliar modelos que têm como principal objetivo prever valores contínuos. Intuitivamente, pode-se ler os valores de MSE como a distância entre os valores verdadeiros e os valores preditos pelo modelo em questão. O calculo é realizado da seguinte forma: primeiramente a diferença entre o valor verdadeiro para os valores predito pelo

Figura 20 – Comparação de normalização



Fonte: Autor

modelo, é calculada; logo após, o resultado desta diferença é elevada ao quadrado. É realizado estes dois primeiros passos para as n observações utilizadas como métrica, e por fim os valores obtidos são somados e divididos por n , como mostra a Equação (33)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (33)$$

3.9.2 RMSE

O RMSE é uma métrica utilizada para avaliar modelos que têm como principal objetivo prever valores contínuos, assim como o MSE (Seção 3.9.1). Os valores de RMSE são calculados da mesma forma que o MSE, com a adição da raiz, como mostra a Equação (34).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2} \quad (34)$$

Em contraste com o MSE, o RMSE assume valores que são diretamente comparáveis, em questões de escala, com os valores reais e preditos.

3.9.3 Matriz de Confusão

A matriz de confusão é uma métrica que avalia a performance de modelos de classificação (TING, 2010). Considerando um modelo com um objetivo de classificação binário, ou seja, um modelo que faz a classificação de uma observação em Positiva, ou Negativa, a matriz de confusão apresenta quatro tipos de resultado sobre as classificações realizadas sobre as observações, são elas: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) e *False Negative* (FN). A seguir serão descritos esses tipos:

- a) TP: Corresponde a quantidade de vezes em que o classificador classificou como positivo uma observação que de fato é positiva.
- b) TN: Corresponde a quantidade de vezes em que o classificador classificou como negativo uma observação que de fato é negativa.
- c) FP: Corresponde a quantidade de vezes em que o classificador classificou como positivo uma observação que na realidade é negativa.
- d) FN: Corresponde a quantidade de vezes em que o classificador classificou como negativo uma observação que na realidade é positiva.

Frequentemente a matriz de confusão é representada visualmente como a Figura 21 mostra.

Figura 21 – Matrix de confusão

		Valor Real	
		Positivo	Negativo
Valor Predito	Positivo	TP	FP
	Negativo	FN	TN

Fonte: Autor

A partir das informações que compõem a matriz de confusão, pode-se extrair algumas outras métricas. A seguir será descrito como obter essas métricas baseadas na matriz de confusão.

3.9.3.1 Acurácia

A acurácia é um valor que explicita a porcentagem de valores totais, tanto positivo quanto negativos, que foram corretamente preditas pelo modelo de classificação proposto, sendo calculado pela Equação (35).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (35)$$

Um ponto negativo da acurácia é devido a sua baixa sensibilidade a valores de classes desbalanceadas. Caso os dados apresentem poucos valores positivos, a acurácia ainda será alta mesmo que o modelo não tenha uma performance alta sobre as classes positivas.

3.9.3.2 Precisão

Em contraste com a acurácia, a precisão se baseia somente nos valores positivos da classificação. A precisão é um valor em porcentagem que explicita a quantidade de valores positivos preditos pelo classificador que foram classificados corretamente.

$$PR = \frac{TP}{TP + FP} \quad (36)$$

3.9.3.3 F1-Score

O F1-Score é dado pela média harmônica entre os valores de precisão e sensibilidade, como mostra a Equação (37)

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (37)$$

Essa métrica tem como vantagem levar em conta tanto valores positivos, quanto valores negativos, sendo recomendada para séries que tem valores desbalanceados entre positivos e negativos.

3.9.4 Validação Cruzada K-Fold

A validação cruzada *K-Fold* é uma técnica utilizada em *machine learning* para avaliar a generalização de um modelo sobre um conjunto de dados; isto é, verificar se o modelo não está condicionado a boa performance exclusivamente sobre uma partição específica de treino.

Para realizar a validação cruzada primeiramente criam-se diversas partições da base de treinamento denominadas de *Folds*. E logo após, o modelo é treinado nas diversas partições, obtendo, assim, diversos valores de uma métrica, como mostra a Figura (22). Podemos então tirar uma média das métricas obtidas para identificar se o modelo proposto tem uma generalização satisfatória.

Figura 22 – K-Fold

1 Fold	Teste	Treino	Treino	Treino	Treino
2 Fold	Treino	Teste	Treino	Treino	Treino
3 Fold	Treino	Treino	Teste	Treino	Treino
4 Fold	Treino	Treino	Treino	Teste	Treino
...					
K-Fold	Treino	Treino	Treino	Treino	Teste

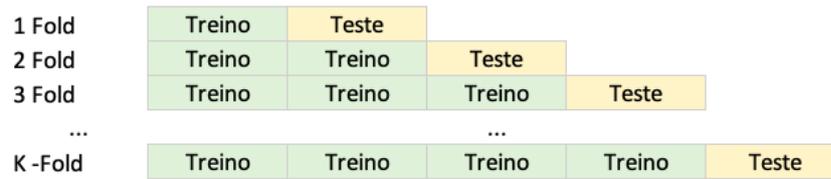
Fonte: Autor

3.9.4.1 Validação Cruzada K-Fold Para Séries Temporais

O método de validação cruzada para séries temporais utiliza os mesmos princípios, e tem os mesmos objetivos, da validação cruzada tradicional. Porém, para as séries temporais os valores futuros não são utilizados para criar as partições, pois, como as séries temporais tem dependência temporal, os valores futuros não são utilizados para prever valores passados.

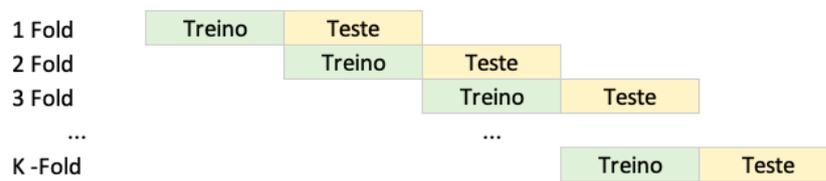
Existem dois tipos de validação *K-Fold* para séries temporais. O primeiro tipo aumenta a quantidade de valores de pontos observados a cada *Fold*, como mostra a Figura 23. O segundo tipo também é denominado de *sliding window*, pois não adiciona novas observações a cada *Fold*, apenas avança no tempo com uma janela de intervalo temporal fixo, como mostra a Figura 24.

Figura 23 – K-Fold para séries temporais



Fonte: Autor

Figura 24 – K-Fold para séries temporais tipo *sliding window*



Fonte: Autor

4 METODOLOGIA

Este capítulo apresenta a metodologia proposta neste trabalho para atingir o objetivo descrito na Seção 1.1. A arquitetura proposta neste trabalho é baseada nos conceitos apresentados no Capítulo 3, e também nos trabalhos relacionados descritos no Capítulo 2. A metodologia proposta utiliza modelos hierárquicos em cascata (Seção 3.5), juntamente com RNN (Seção 3.3) e decomposição de séries temporais (Seção 3.2), com o intuito de prever valores financeiros futuros da bolsa, para auxiliar na negociação de papéis. A estratégia para atingir o objetivo é extrair valores contextuais da série temporal da bolsa de valores e utiliza-los como dimensões de entrada para modelos de predição baseados em *machine learning*.

A motivação para utilizar modelos hierárquicos em cascata, juntamente com decomposição de séries temporais da bolsa de valores, provém da revisão feita no Capítulo 2. A Figura 7 da Seção 2.6 quantifica os trabalhos que utilizaram redes RNN acompanhadas, ou não, de decomposição de séries temporais. Como foi notado, essas técnicas são frequentemente utilizadas na área de predição de valores financeiros. Até onde foi possível notar, ainda que a revisão do estado da arte mostre diversos métodos que utilizam decomposição de séries temporais e RNN, os trabalhos apresentados não consideram valores contextuais (Seção 3.4) das séries temporais em questão. Esta observação indica então uma oportunidade para se obter melhores resultados na predição de ativos da bolsa, extraindo valores contextuais com modelos hierárquicos em cascata.

A Figura 25 mostra o fluxograma da metodologia proposta. A metodologia é dividida em três etapas principais, são elas: (1) Preparação dos Dados (Seção 4.2), (2) Treinamento do Modelo (Seção 4.3) e (3) Avaliação do Modelo (Seção 4.4).

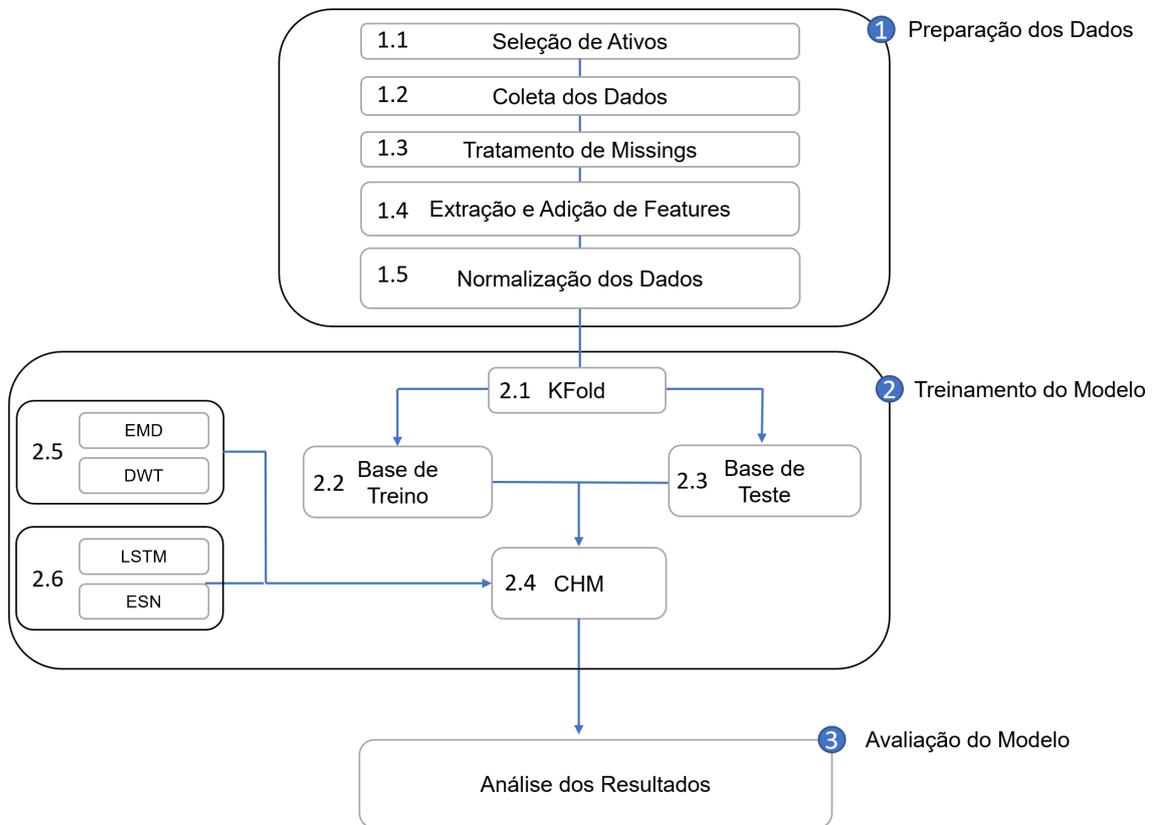
Nas seções a seguir cada etapa da metodologia será descrita em mais detalhes.

4.1 BASE DE DADOS

A base de dados utilizada neste trabalho, tanto para o treinamento do modelo quanto para a avaliação do mesmo, foi retirada do pacote *yfinance* do *python* (YFINANCE..., s.d.). Esta biblioteca oferece informações de OHLC sobre a maioria dos principais ativos de todo mundo. Para acessar as informações dos ativos das bolsas é necessário utilizar o *Ticker* (sigla da empresa na bolsa) da empresa listada na bolsa. Como neste trabalho serão utilizadas diversas bolsas e diversos ativos, foi construído um *script* para obter um único *dataset* que contenha informações de diversas bolsas de valores.

Para este trabalho, foram selecionadas as oito principais bolsas de valores do mundo, e as três empresas mais representativas de cada uma para o treinamento do modelo. Para compor as séries temporais das bolsas, o período de dados utilizado foi o de 1 dia por valor OHLC, devido à escassez das informações sobre os demais períodos. Os valores obtidos na base de dados estão compreendidos entre o período Janeiro de 2010 até Janeiro de 2021, totalizando 4019 *data points*

Figura 25 – Fluxograma geral da metodologia proposta



Fonte: Autor

para cada ação. Devido ao fato de um mesmo ativo estar presente em múltiplas bolsas de um único país (com exceção da China) apenas uma bolsa de cada país é considerada, para evitar a duplicidade dos dados. Caso o país tenha mais de uma bolsa, foi selecionada a maior bolsa em quantidade de ativos listados. Apesar disso, foram consideradas duas bolsas da China devido à autonomia de Hong Kong em relação ao país. As bolsas utilizadas neste trabalho estão listadas a seguir:

- a) Brasil Bolsa Balcão (Brasil)
- b) New York Stock Exchange (Estados Unidos)
- c) Tokyo Stock Exchange (Japão)
- d) Shanghai Stock Exchange (China)
- e) Hong Kong Stock Exchange (China)
- f) London Stock Exchange (Reino Unido)
- g) National Stock Exchange of India Ltd (Índia)
- h) Frankfurt Stock Exchange (Alemanha)

A Base de Dados gerada terá 4 colunas para cada ação obtida, uma coluna para cada valor de OHLC. A Tabela 1 mostra como ficaria a base de dados para a ação da Apple, código *ticker* AAPL.

Tabela 1 – Exemplo da base de dados

Date	AAPL-Open	AAPL-High	AAPL-Low	AAPL-Close
2010-01-04	6.535266	6.568029	6.503114	6.553025
2010-01-05	6.571091	6.601405	6.529754	6.564355
2010-01-06	6.564353	6.590381	6.453202	6.459939
2010-01-07	6.483824	6.491479	6.401149	6.447998
2010-01-08	6.439424	6.491479	6.401456	6.490866
...
2021-01-25	142.193417	144.201029	135.703412	142.044327
2021-01-26	142.720159	143.415867	140.503811	142.282852
2021-01-27	142.551192	143.415871	139.549706	141.189590
2021-01-28	138.665165	141.120033	135.862436	136.250046
2021-01-29	134.997773	135.902201	129.412211	131.151489

Tabela 2 – Variáveis do *dataset* final

OHLC	Open
	High
	Low
	Close
Análise Técnica	Médias Móveis
	Bandas de Bollinger
	Stochastic Oscillator
	Índice de Força Relativa
Frequências	Frequências de Oscilação

As *features* utilizadas na metodologia proposta podem ser divididas em três categorias: OHLC, Análise Técnica e Frequência de Oscilação. Todas estas informações são contínuas e possuem valor temporal. Para as *feature* de análise técnica foi utilizada uma técnica para cada sub-categoria definida por Pimenta et al. (2018), e descritas na Seção 3.6, são elas: SMA, para Seguidores de Tendência; BB, para sistemas de banda; SO, para osciladores; e IFR, para Identificadores de Divergência. A Tabela 2 mostra de uma forma mais organizada os tipos de variáveis utilizadas no trabalho. No total a base apresentará 216 séries temporais, compostas de 9 variáveis de 3 ativos das 8 bolsas listadas.

4.2 PREPARAÇÃO DOS DADOS

A etapa de Preparação dos Dados (Figura 25, bloco 1) é dividida em cinco sub-etapas, são elas: Seleção de Ativos, Coleta dos Dados, Tratamento de *Missings*, Extração e Adição de *Features*, e Normalização dos Dados. Cada sub-etapa utiliza a base de dados tratada pela sub-etapa anterior.

A sub-etapa de Seleção de Ativos (Figura 25, bloco 1.1) consiste na identificação dos ativos que contenham informações constantes ao longo dos anos analisados na base de dados;

isto é, a base deve contemplar a maioria das informações dos ativos. Esta seleção é feita para evitar que ativos que contenham poucas informações afetem a performance do modelo. Foi definido um critério de consistência de dados do ativo ao longo dos anos. O critério criado leva em consideração a representatividade do ativo listado na bolsa, e também se a maioria dos dados estão presentes durante todo o período analisado (Janeiro de 2010 a Janeiro de 2021). Caso um ativo não apresente dados durante um longo período de tempo, ou tenha menos informações do que os demais ativos, este ativo será descartado. Esta é uma etapa manual e é executada apenas uma vez. Ao final foram selecionados os ativos apresentados na Tabela 3.

Tabela 3 – Ativos do *dataset* final.

Bolsa	Ativo
Brasil Bolsa Balcao (B3)	Itaú Unibanco (ITUB3.SA)
	Ambev (ABEV3.SA)
	B3 (B3SA3.SA)
New York Stock Exchange (NYSE)	Apple Inc. (AAPL)
	Microsoft Corp. (MSFT)
	Google (GOOG)
Tokyo Stock Exchange (TSE)	Toyota (7203.T)
	Keyence (6861.T)
	Sony (6758.T)
Shanghai Stock Exchange (SSE)	Kweichow Moutai Co., Ltd. (600519.SS)
	China Construction Bank Corporation (601939.SS)
	Agricultural Bank of China (601288.SS)
Hong Kong Stock Exchange (HKSE)	Tencent Holdings Limited (0700.HK)
	Industrial and Commercial Bank of China Limited (1398.HK)
	China Merchants Bank Co., Ltd. (3968.HK)
London Stock Exchange (LSE)	Unilever (ULVR.L)
	AstraZeneca (AZN.L)
	HSBC (HSBA.L)
National Stock Exchange of India Ltd (NSE)	Reliance Industries Limited (RELIANCE.NS)
	Tata Consultancy Services (TCS.NS)
	HDFC Bank Limited (HDFCBANK.NS)
Frankfurt Stock Exchange (FSE)	SAP (SAP.DE)
	Siemens Aktiengesellschaft (SIE.DE)
	Volkswagen (VOW3.DE)

A Coleta dos Dados (Figura 25, bloco 1.2) consiste na obtenção da base de dados descrita na Seção 4.1 sobre os ativos selecionados na sub-etapa anterior, ao total serão 24 ativos selecionados. Estes dados serão utilizados nas etapas seguintes.

Em seguida, na sub-etapa de Tratamento de *Missings* (Figura 25, bloco 1.3), é realizado um tratamento de valores faltantes, ou *missings*, na base de dados tratada pela sub-etapa anterior. Em contraste com a sub-etapa de Seleção de Ativos, esta sub-etapa não elimina ativos a serem analisados, mas preenche os valores faltantes que são provenientes da diferença de dias úteis de cada bolsa de valor. Para este tratamento, caso o valor de um ativo de uma bolsa não esteja presente o valor do dia anterior será utilizado. Esta regra foi definida baseada no próprio comportamento da bolsa B3, que replica os valores do último dia útil para os dias não úteis.

A próxima sub-etapa é a Extração e Adição de *Features* (Figura 25, bloco 1.4) (Seção 3.6.3). Nesta etapa, são incluídas as *features* de análise técnica, calculadas a partir das variáveis OHLC.

Por fim é aplicada a sub-etapa de Normalização dos Dados (Figura 25, bloco 1.5) (Seção 3.8), os dados para evitar vieses provenientes da magnitude dos valores absolutos dos dados.

Após esta sub-etapa a base de dados esta pronta para ser direcionada para o modelo.

4.3 TREINAMENTO DO MODELO

Esta etapa consiste em realizar o treinamento do modelo com os dados obtidos da etapa anterior (Seção 4.2). Nesta etapa, a base dados é dividida em duas sub-bases: treinamento e validação. A sub-base de treinamento será utilizada na arquitetura CHM que, por sua vez, fará o treinamento das RNN sobre os dados. A predição, que será feita a partir do modelo CHM, será utilizada em conjunto com a sub-base de validação para avaliar o desempenho do modelo CHM na etapa 3 (Seção 4.4).

A divisão dos dados em treinamento e teste é realizada pelo algoritmo *k-fold* (Figura 25, bloco 2.1) (Seção 3.9.4). Esta divisão é feita para assegurar que o modelo treinado não apresentará um estado de *overfit*; isto é, o modelo generalizará ao fazer predições sobre valores futuros. O *k-fold* (Figura 25-2.1) criará duas bases a partir da base obtida pela etapa de Preparação dos Dados (Seção 4.2), a Base de Treino (Figura 25, bloco 2.2) e a Base de Teste (Figura 25, bloco 2.3). Após esta quebra, os dados da Base de Treino são enviados para o modelo CHM (Figura 25, bloco 2.4) descrito na Seção 3.5, que fará o treinamento das redes neurais RNN (Figura 25, bloco 2.6). Neste trabalho serão testados dois tipos de RNN, são elas: ESN e LSTM. Esta é a etapa onde o modelo proposto realiza a extração de valores contextuais da bolsa e os utiliza para predizer os valores futuros da bolsa. Para a decomposição dos valores de série temporal também serão utilizados dois tipos de decompositores (Figura 25, bloco 2.5), são eles: DWT e EMD.

Devido ao fato do modelo CHM permitir selecionarmos qual será o tipo de rede e tipo de decompositor, serão realizados algumas variações dos modelos, apresentados na Tabela 4. As variações são compostas por: uma RNN, um decompositor de série temporal e a inclusão ou não das variáveis de frequência de oscilação das ondas no tempo. Todos os modelos são treinados com a mesma quantidade de dados e sobre os mesmos ativos. Serão testados também modelos balizadores que não apresentam a arquitetura de CHM. Estes modelos serão utilizados para verificar o real efeito da extração de contexto das séries temporais. Ao total, serão avaliadas 12 variações do modelo.

Para o treino das redes, as variáveis de entrada, ou variáveis explicativas, serão as descritas na Seção 3.6. Para os modelos que utilizam EMD, também serão utilizados os valores de frequência de oscilação, descritos na Seção 3.6.1. E como variável de saída, ou variável resposta, serão utilizados os valores de fechamento do ativo para todos os modelos.

4.3.1 CHM

Nesta seção será apresentado o funcionamento da arquitetura de extração de contexto proposta neste trabalho. Primeiramente será apresentada uma ideia geral do modelo, e em seguida será feita em detalhes a descrição de cada etapa da arquitetura.

A Figura 26 apresenta a arquitetura geral do modelo CHM proposto. Assim como o trabalho de Seyedhosseini, Sajjadi e Tasdizen (2013) o processo tem um formato hierárquico, porém neste trabalho não serão utilizadas as estruturas de *stages*. Primeiramente as variáveis selecionadas passam pelo algoritmo de decomposição de séries temporais selecionado, DWT ou EMD (Seção 3.2). Logo após as séries decompostas são direcionadas para a etapa *bottom-up* do modelo. Os dados passarão primeiramente pela etapa *bottom-up* do modelo para criar um contexto sobre as séries temporais; isto é, treinar as redes para realizar previsões que serão utilizadas em outros nível como dados de contexto da série temporal. Por fim, os dados passarão pela etapa *top-down* para que a previsão dos valores do mercado financeiro seja realizada utilizando dados de contexto gerados pela etapa *bottom-up*.

A seguir, será descrito como o modelo realiza o aprendizado do contexto de uma série temporal, o processo é dividido em duas etapas principais, são elas: *bottom-up* e *top-down*.

4.3.1.1 Etapa Bottom-Up

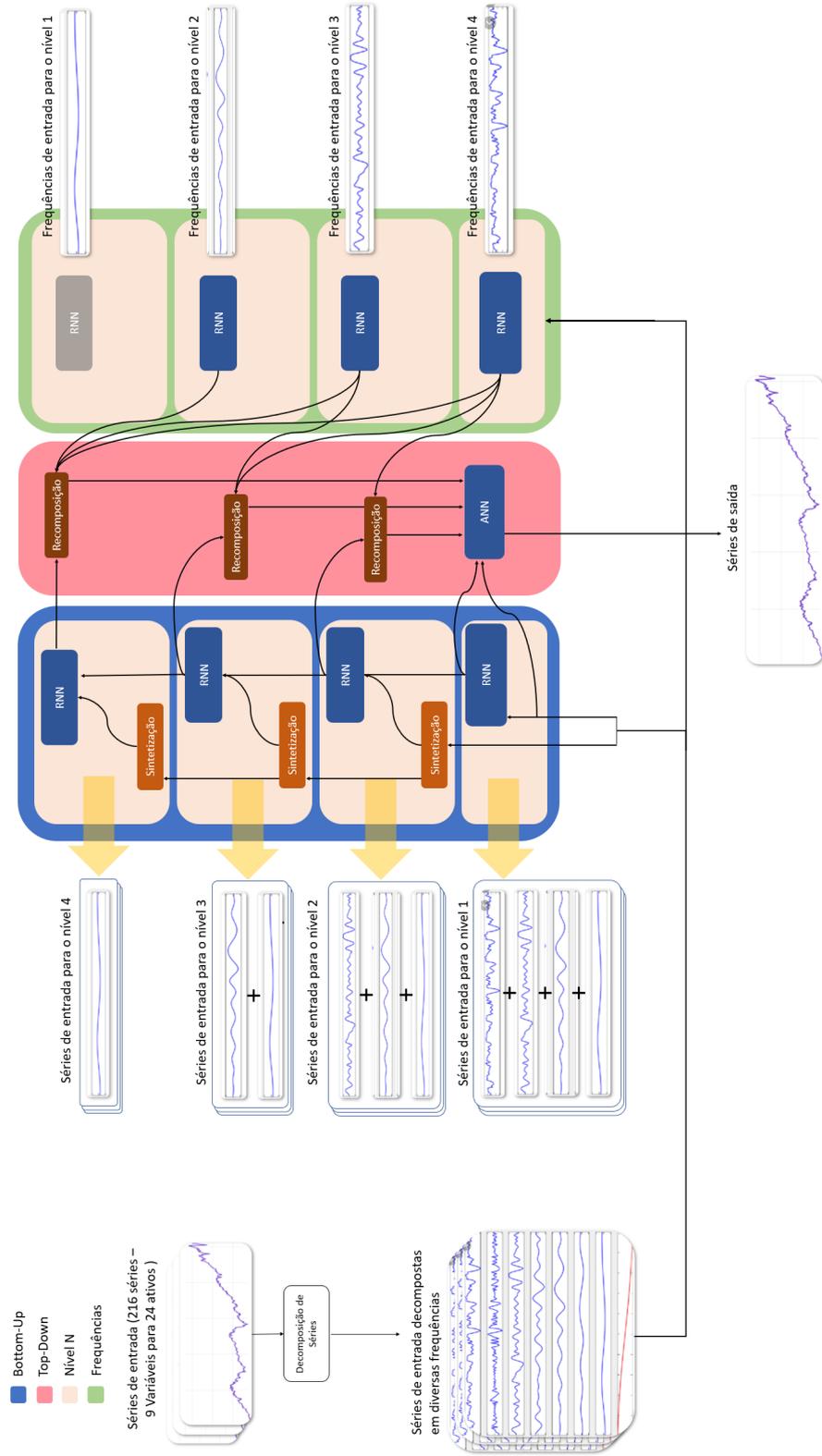
A etapa *bottom-up* é apresentada no quadrado azul da Figura 26. Esta etapa se refere ao processo que aprenderá o contexto das séries temporais. A arquitetura *bottom-up* é formada por diversos níveis hierárquicos onde cada nível da hierarquia é denominado N_i . A etapa contempla N níveis, sendo que, em cada nível N_i , uma rede é treinada sobre sub-séries simplificadas criadas a partir da série original. Os níveis na arquitetura se referem ao grau de simplificação da série temporal, quanto maior o nível menos frequências e ruídos estarão presentes na série temporal. Na Figura 26 também é possível notar que a cada nível N_i uma predição é feita para ser utilizada como contexto no próximo nível hierárquico. No último nível da etapa *bottom-up* a série apresenta o mínimo de detalhes possível, e este nível utiliza informações retiradas dos outros níveis para fazer a predição da sua série temporal.

Primeiramente, as variáveis obtidas na Seção 4.2 são decompostas utilizando uma das técnicas apresentadas na Seção 3.2. Após a decomposição, serão obtidas k sub-séries, SS , para cada variável de cada uma das ações. A soma de todas as sub-séries SS extraídas pela decomposição resultam na série original, S_o , que representa uma das variáveis decompostas, como mostra a Equação (38).

$$S_o = \sum_{i=0}^k SS_i \quad (38)$$

Desta forma, para cada nível é retirada uma, ou mais, sub-séries, obtidas da decomposição da série original, da série original resultando em uma nova série simplificada Sn que será utilizada

Figura 26 – Arquitetura CHM proposta



Fonte: Autor

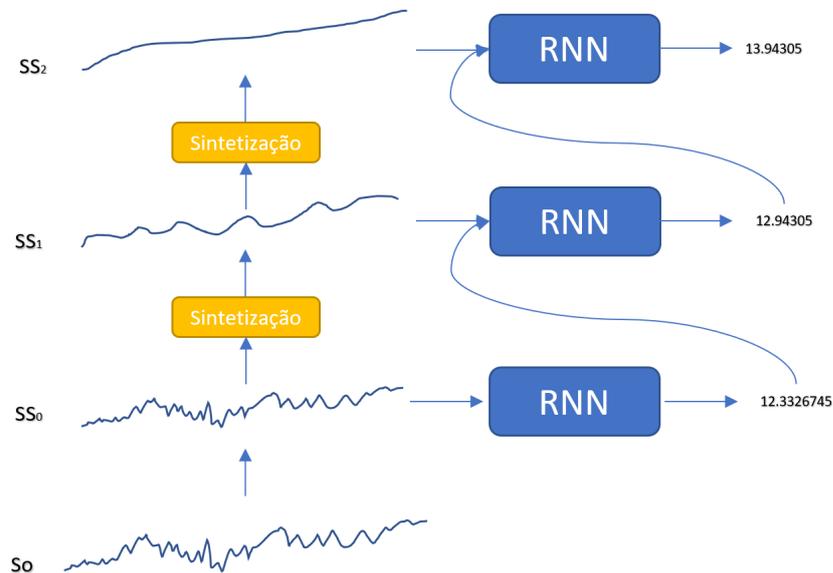
no nível atual, N_j . Esta operação é denominada de **sintetização**. Podemos definir as séries de cada nível utilizando a Equação (39)

$$Sn_j = S_o - \sum_{i=0}^j SS_i \quad (39)$$

, onde j é o nível atual. As séries subtraídas serão sempre as que apresentarem maior frequência.

Sendo assim, para o primeiro nível N_1 as variáveis são utilizadas como entrada de uma rede de regressão RNN sem aplicar a sintetização das séries. Essa rede será treinada e os resultados que forem originados de sua predição serão utilizados como entrada para o próximo nível. Para o nível N_2 , a série das variáveis de entrada serão sintetizadas, utilizando a Equação (39). As séries obtidas da sintetização serão utilizadas como entrada para o treino da rede de predição do nível N_2 , juntamente com a saída da rede de regressão do nível anterior (N_1). Para os seguintes níveis, assim como o nível N_2 , a saída da rede de regressão do nível anterior ao nível atual é utilizada para treinar a rede de predição do nível atual, como mostra a Figura 27. A utilização dos valores preditos pela rede de regressão do nível N_{i-1} cria um contexto da série temporal para o treinamento da próxima rede na etapa *top-down*. Os níveis apresentados nesta etapa é a característica que leva o modelo a receber a categoria de hierárquico, devido à ideia de níveis que se posicionam de forma hierárquica.

Figura 27 – Processo de Sintetização



Fonte: Autor

4.3.1.2 Etapa Top-Down

Em contraste com a etapa *bottom-up*, a etapa *top-down* realiza o treino de uma única rede ANN. Esta rede utiliza a saída de todas as redes treinadas na etapa *bottom-up* como *features*

Tabela 4 – Variações do modelo

Modelo	CHM	RNN	Decompositor	Features de Frequência
CHM-LSTM-DWT	Sim	LSTM	DWT	Não
CHM-ESN-DWT	Sim	ESN	DWT	Não
CHM-LSTM-EMD	Sim	LSTM	EMD	Não
CHM-ESN-EMD	Sim	ESN	EMD	Não
CHM-LSTM-EMD-Freq	Sim	LSTM	EMD	Sim
CHM-ESN-EMD-Freq	Sim	ESN	EMD	Sim
LSTM-EMD	Não	LSTM	EMD	Não
LSTM-DWT	Não	LSTM	DWT	Não
ESN-EMD	Não	ESN	EMD	Não
ESN-DWT	Não	ESN	DWT	Não
ESN-EMD-Freq	Não	ESN	EMD	Sim
LSTM-EMD-Freq	Não	LSTM	EMD	Sim

de treinamento. Porém, para cada sub-série excluída na etapa *bottom-up* é necessário antes recompor a série para a frequência inicial, somando as demais frequências retiradas em cada nível. Isto é necessário devido ao fato das séries se encontrarem em níveis de detalhes diferentes. Esta operação é denominada de "recomposição". Porém, não é possível realizar esta tarefa de forma simples pois os valores das frequências faltantes ainda não existem. Desta forma foi criada uma rede neural para prever os valores de cada uma das frequências. Este conjunto de redes foi denominado de redes preditoras de frequência. Quando uma rede da etapa *bottom-up* faz a predição de um valor, este valor se encontra na frequência do nível em que a rede se encontra, para retornar à frequência da série original é necessário somar com a predição das redes preditoras de frequência. A etapa *top-down* é a etapa responsável por realizar a junção dos contextos criados pela etapa *bottom-up*, e fará a predição do valor da ação do dia seguinte, com base nas predições das redes *bottom-up*, fazendo um julgamento de qual rede se aproxima mais do resultado real naquela situação.

4.4 AVALIAÇÃO DO MODELO

A última etapa da metodologia consiste na Avaliação do Modelo (Figura 25, bloco 3). Este processo faz uso da métrica RMSE (Seção 3.9.2) para avaliar a performance do modelo. Esta validação será feita através do método *K-Fold* (Seção 3.9.4.1).

Primeiramente, será utilizada a técnica de RMSE para quantificar o desvio do modelo em relação aos valores reais. Também será utilizada uma métrica de classificação baseada na predição de pontos de reversão identificados pelo modelo. A métrica de classificação será aplicada da seguinte forma: caso o modelo faça a predição no tempo $t + 1$ de um valor maior do que um valor no tempo t , este tempo t será marcado como *compra*. Caso o modelo aponte uma

baixa no tempo $t + 1$, em relação ao tempo t , este tempo será marcado como *venda*. Desta forma, também será gerada uma matriz de confusão (Seção 3.9.3).

Para definir os valores das métricas, o *dataset* será dividido em treino e teste (Seção 4.3). O objetivo desta divisão é validar se os resultados da arquitetura proposta são independentes da janela utilizada para treino. A divisão será feita como mostra a Figura 28. A célula azul refere-se aos dados de treino e a amarela à parte de teste que irá ser utilizada para validar o desempenho do modelo.

Figura 28 – Janela de Validação



Fonte: Autor

Esta etapa será realizada para cada variação de modelo apresentada na Seção 4.3, com o intuito de identificar a variação de modelo que apresenta a melhor performance na predição dos valores. Os modelos avaliados serão os listados na Tabela 4.

4.5 IMPLEMENTAÇÃO

A arquitetura proposta e os experimentos serão implementados utilizando a linguagem de programação Python 3.7.0. Serão utilizadas as bibliotecas: *yfinance* (YFINANCE..., s.d.), para obter os dados de treinamento; a biblioteca Tensorflow, proposta por Martín Abadi et al. (2015) para a construção das redes LSTM, juntamente com a API do Keras (CHOLLET et al., 2015); a biblioteca SKLearn proposta por Pedregosa et al. (2011) foi utilizada para a normalização dos

dados; a biblioteca Pandas (TEAM, 2020), para o tratamento dos dados; e a biblioteca Numpy (HARRIS et al., 2020), também para o tratamento dos dados. Por último, para a apresentação de gráficos de gráficos será utilizada a biblioteca Matplotlib (HUNTER, 2007).

5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os experimentos e testes realizados para validar a eficácia da metodologia proposta. Foi realizada uma comparação entre as diferentes variações da proposta, com o intuito de verificar a contribuição da arquitetura de extração de contexto sobre séries temporais para a performance de predição dos valores da bolsa. Para esta comparação, e validação, serão utilizadas métricas estatísticas para avaliação dos modelos apresentados na Tabela 4, apresentada na Seção 4. Os testes foram feitos sobre a predição dos valores de 3 ações brasileiras, são elas: Ambev ABEV3, Itaú ITUB3 e B3 B3SA3. Também serão apresentados testes com o retreino contínuo da metodologia proposta ao longo do tempo, e a comparação deste com o treino convencional.

Os experimentos foram realizados utilizando dados entre Janeiro de 2010 e Janeiro de 2021. Para garantir a eficácia do modelo em diversos cenários, foram utilizados 4 períodos distintos dentro destas duas datas. O primeiro período utiliza dados entre 2010 e 2013 para treino, realizando os teste em 2014. Para convenção neste trabalho, este período foi denominado de 2010-2013T-2014V. O segundo período utiliza informações entre 2013 e 2016 para treino e dados de 2017 para teste, denominado de 2013-2016T-2017V. O terceiro período utiliza dados entre 2016 e 2020 para treino, e Janeiro de 2021 para teste, denominado de 2016-2020T-2021V. E o último período contempla 66% dos dados totais para treino e 33% para teste, denominado de *33p_train_test*. Os dados foram retirados do banco de dados do Yahoo Finance (YFINANCE..., s.d.) e tratados utilizando a linguagem de Python.

A metodologia proposta foi comparada com modelos que não utilizam a arquitetura de extração de contexto proposta, redes ESN e LSTM, em combinação com EMD ou DWT para a decomposição de séries. Todas as variações foram testadas utilizando dois conjuntos de *features*, um conjunto que utiliza somente as 3 ações brasileiras citadas anteriormente, e outro conjunto utilizando as 3 ações brasileiras mais 3 ações de mais 8 bolsas citadas na Seção 4.1, totalizando 24 ações. Ambas utilizando informações de análise técnica sobre os valores. Neste trabalho foram utilizados os seguintes indicadores de análise técnica: SMA, BB, SO e IFR. Estas informações foram calculadas a partir dos dados obtidos da biblioteca Yahoo Finance, e utilizadas como entrada dos modelos propostos. Também foi testada a utilização de *features* de frequência de oscilação (Seção 3.6.2).

Para a normalização dos dados foi utilizada a função `MinMaxScaler` da biblioteca `SKLearn` do Python (PEDREGOSA et al., 2011), que transforma os dados em números normalizados entre 1 e 0. Esta normalização garante que a magnitude absoluta dos dados não crie vieses no modelo, como por exemplo a divergência de câmbio entre as diferentes moedas dos países em que as bolsas utilizadas se encontram.

Todos os modelos utilizaram 300 épocas para o treinamento das redes. Para a arquitetura das redes foram utilizados os seguintes parâmetros para as redes LSTM: 3 camadas para o treinamento, sendo compostas de 100 células LSTM na camada oculta; um *look back* de 7

dias; otimizador Adam; *loss* de MSE; função de ativação *tanh*; sem o uso de *dropout*. Para as redes ESN foi utilizado: 600 neurônios no reservatório; *spectral radius* de 0.9; *connectivity* de 0.1; otimizador Adam; *loss* de MSE; e função de ativação *tanh*. Para as redes ANN, utilizadas na etapa *top-down* da arquitetura proposta, foi utilizado: 100 neurônios na camada oculta; otimizador Adam; *loss* de MSE; e função de ativação *tanh*.

Para cada modelo, foram testados dois tipos de predição. O primeiro tipo de predição refere-se ao valor de fechamento do preço de um ativo no dia seguinte à aplicação do modelo. Para a avaliação foi utilizada a métrica de RMSE, apresentada na Seção 3.9.2. O segundo tipo de predição utiliza os valores preditos pelo modelo para identificar momentos de compra ou venda de ativos, utilizando o conceito de pontos de reversão (Seção 3.7). Para esta predição foram utilizadas as métricas de classificação, *f1-score*, acurácia e precisão, apresentadas na Seção 3.9.3.

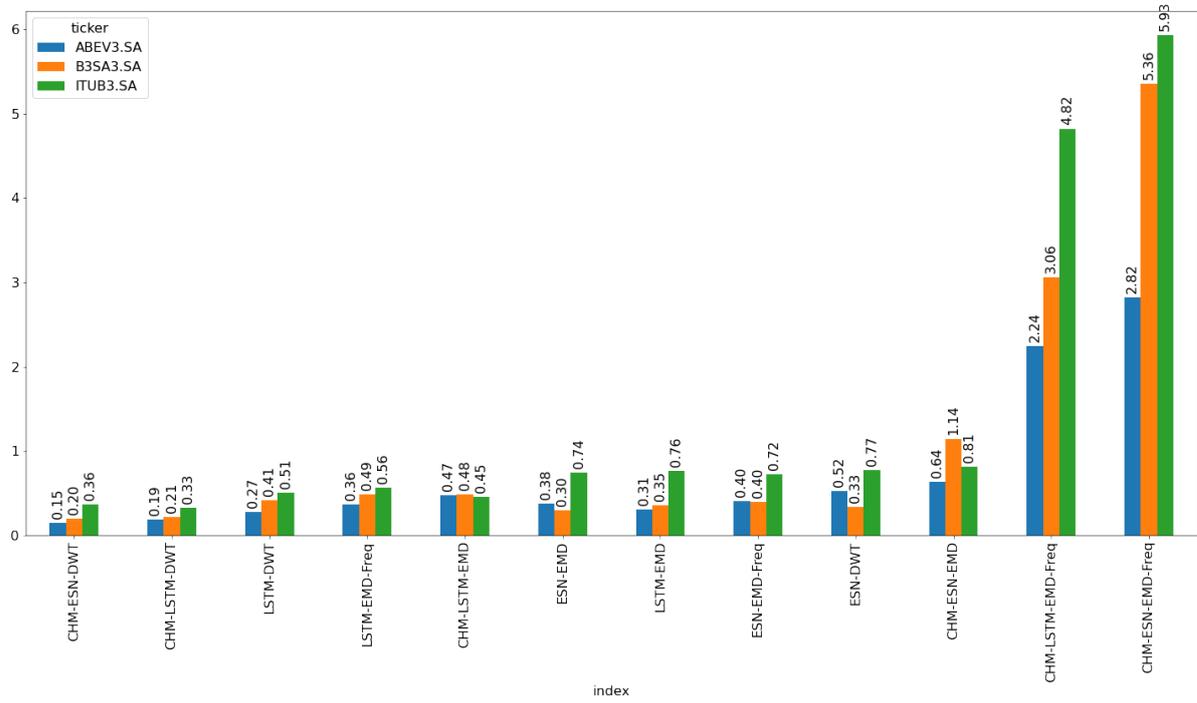
Nas seções a seguir será discutido em mais detalhes o desempenho do modelo proposto em previsões do preço futuro do ativo (Seção 5.1), o desempenho do modelo proposto em previsões de pontos de reversão (Seção 5.2), a avaliação da utilização de um retreino periódico (Seção 5.3), a avaliação do uso de *features* de frequência instantânea (Seção 5.4), e por fim serão discutidas as contribuições do presente trabalho para o estado da arte (Seção 5.5).

5.1 AVALIAÇÃO DA PREVISÃO DO PREÇO FUTURO

Nesta seção é apresentada a performance do modelo proposto para a previsão do preço do ativo no dia seguinte em que o modelo foi utilizado. Para realizar a comparação da performance do modelo proposto com outros modelos que não utilizam a extração de contexto, foi utilizada a métrica de RMSE (Seção 3.9.2). Assim, foi gerado um gráfico de barras que contém a avaliação para cada um dos modelos utilizados. Cada gráfico contém os valores de RMSE para as 3 ações utilizadas para avaliação (Ambev ABEV3, Itaú ITUB3 e B3 B3SA3), e também para cada variação de modelo apresentados na Tabela 4. Foram gerados 2 gráficos semelhantes a este, um para a avaliação utilizando somente a bolsa brasileira (Figura 29), e um utilizando as bolsas mundiais (Figura 30). Todos os gráficos foram gerados utilizando uma média aritmética simples dos resultados obtidos entre os 4 períodos avaliados.

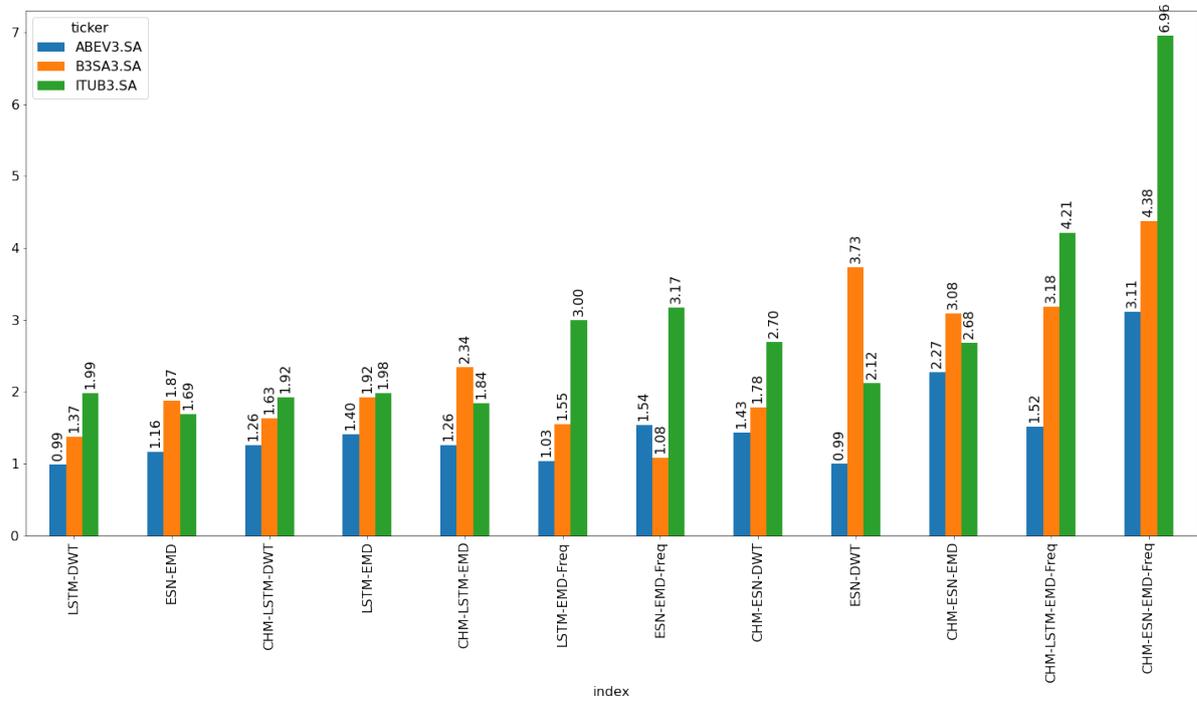
Ao observar os valores de RMSE para os modelos aplicados à bolsa de valores brasileira, Figura 29, podemos notar que os modelos que utilizam a extração de contexto, juntamente com a técnica de DWT, têm uma performance maior do que os modelos que não utilizam a arquitetura proposta. Entre os modelos mais similares CHM-LSTM-DWT, e LSTM-DWT, podemos notar uma redução de, em média, 38% para o modelo que utiliza a técnica de CHM proposta. Também foi possível notar que as variações que utilizaram os valores de frequência de oscilação, apresentada na Seção 3.6.2, não convergiram. Isto indica que, possivelmente, o número de épocas utilizado não foi o suficiente para a convergência da rede neste caso, ou as variáveis de frequência não são recomendadas para a previsão dos valores futuros. As Figuras 31 e 32 mostram a performance de predição dos valores da ação B3SA3 ao longo do tempo dos modelos

Figura 29 – RMSE utilizando somente a bolsa brasileira



Fonte: Autor

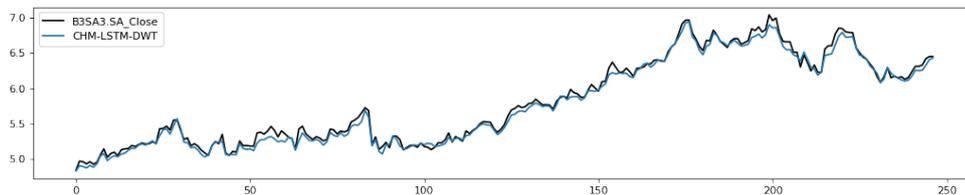
Figura 30 – RMSE utilizando bolsas globais



Fonte: Autor

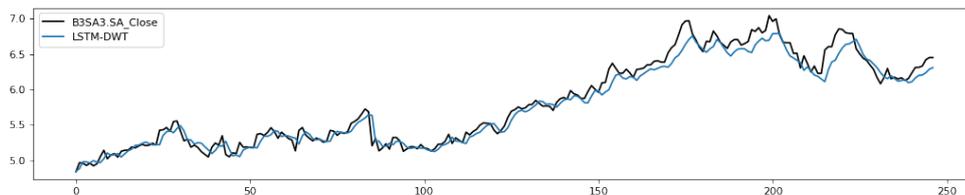
CHM-LSTM-DWT, e LSTM-DWT, respectivamente, como exemplo da série temporal predita para as ações.

Figura 31 – Série temporal da ação B3SA3 utilizando somente a bolsa brasileira do método CHM-LSTM-DWT, período de teste 2013-2016T-2017V



Fonte: Autor

Figura 32 – Série temporal da ação B3SA3 utilizando somente a bolsa brasileira do método LSTM-DWT, período de teste 2013-2016T-2017V

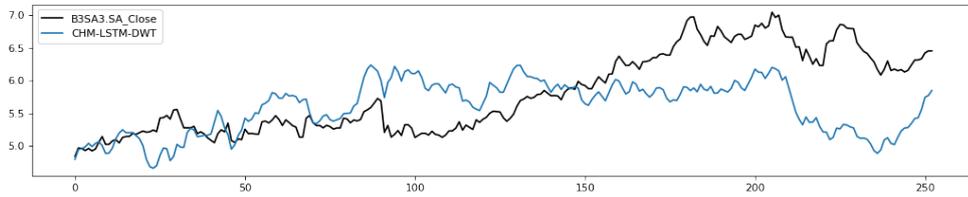


Fonte: Autor

Apesar do resultado positivo na bolsa brasileira, os resultados se invertem quando são utilizadas as bolsas mundiais, como mostra a Figura 30. É possível notar que os modelos que utilizam variáveis de frequência oscilatórias continuam com performances baixas. Contudo, o modelo tradicional LSTM-DWT, que não faz o uso da arquitetura proposta, agora apresenta resultados melhores que os modelos que utilizam em média 14%. Em geral, os modelos possuem uma performance inferior quando as variáveis das bolsas globais são utilizadas, indicando uma possível rejeição à hipótese inicial que a correlação entre bolsas poderia maximizar o resultado dos modelos de forma geral.

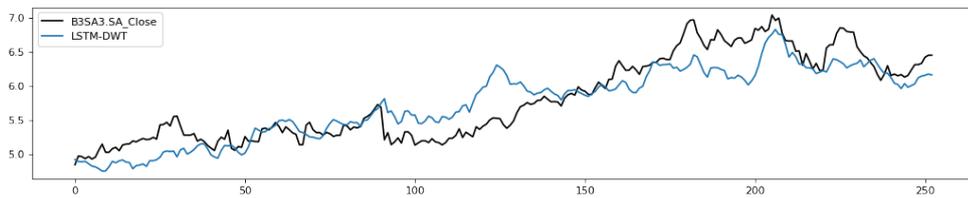
Apesar disto, foi observado que, mesmo com a inversão da performance entre os modelos CHM-LSTM-DWT e LSTM-DWT ao utilizar bolsas mundiais, os modelos que utilizaram a arquitetura proposta tendem a ter um resultado mais satisfatório nos primeiros pontos preditos pelo modelo. As Figuras 33 e 34, mostram a série temporal da predição dos modelos CHM-LSTM-DWT e LSTM-DWT, respectivamente. É possível notar que, nos primeiros 30 pontos de predição, a arquitetura CHM tem melhores resultados. Sendo assim, foi construído um gráfico semelhante aos anteriores, realizando a comparação de RMSE entre os modelos, porém desta vez apenas considerando os primeiros 30 pontos da predição, Figura 35. Considerando apenas os primeiros 30 dias de predição do modelo, assim o método proposto volta a ter uma performance melhor que o tradicional em média 33%.

Figura 33 – Série temporal da ação B3SA3 utilizando bolsas globais do método CHM-LSTM-DWT, período de teste 2013-2016T-2017V



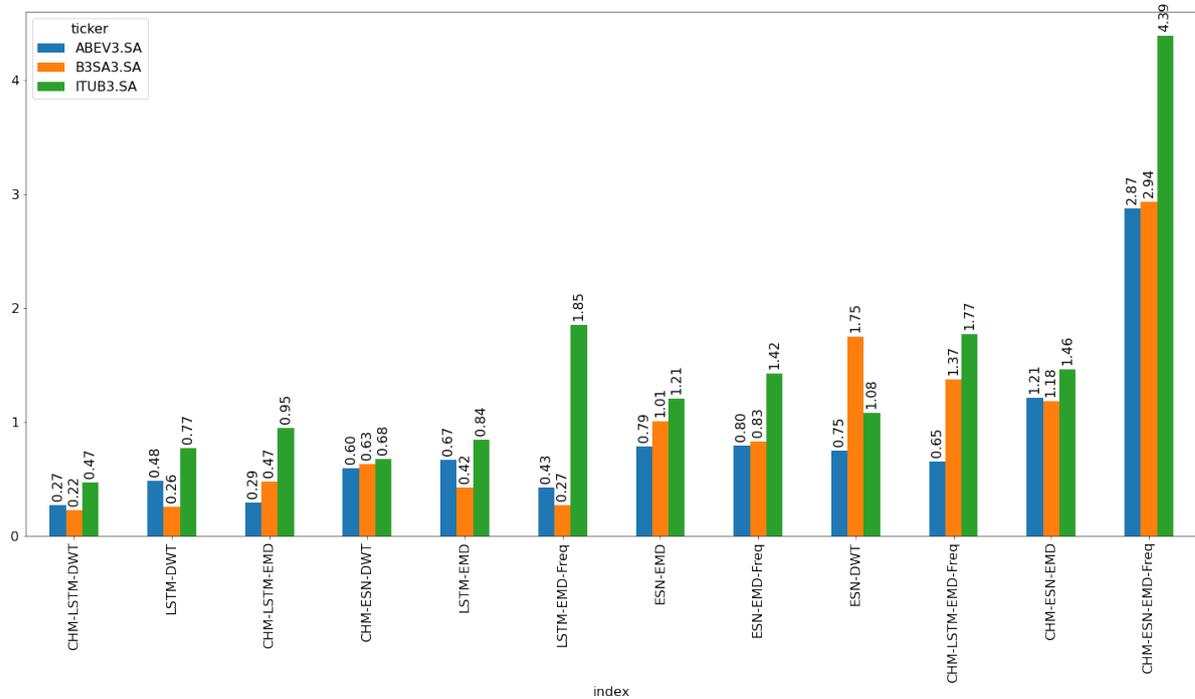
Fonte: Autor

Figura 34 – Série temporal da ação B3SA3 utilizando bolsas globais do método LSTM-DWT, período de teste 2013-2016T-2017V



Fonte: Autor

Figura 35 – RMSE utilizando bolsas globais avaliada em apenas 30 dias

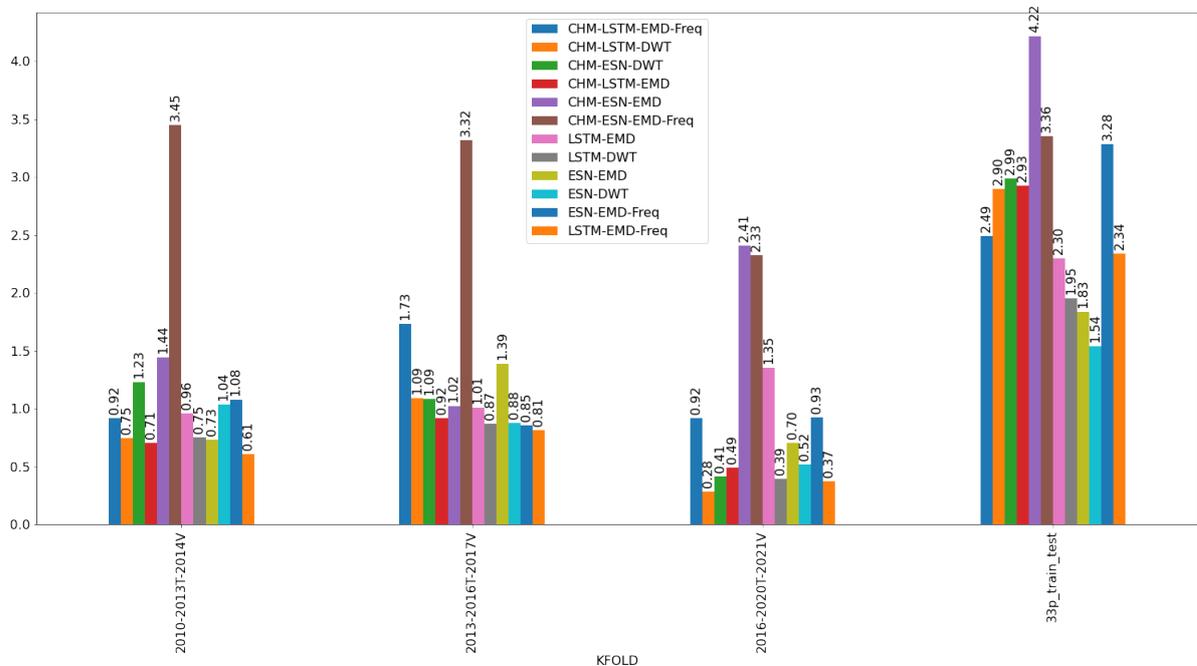


Fonte: Autor

As Figuras 36, 37, 38, 39, 40, 41, mostram a performance dos modelos por período avaliado, sem realizar a média aritmética utilizada anteriormente. Desta forma é possível avaliar

como os modelos se comportam no período de pandemia, utilizando o KFOLD 2016-2020T-2021V. Em média, o modelo CHM-LSTM-DWT é 35% mais performático, em termos de RMSE, do que o tradicional LSTM-DWT, quando os modelos são testados no período da pandemia.

Figura 36 – RMSE da ação ABEV utilizando bolsas globais por período avaliado



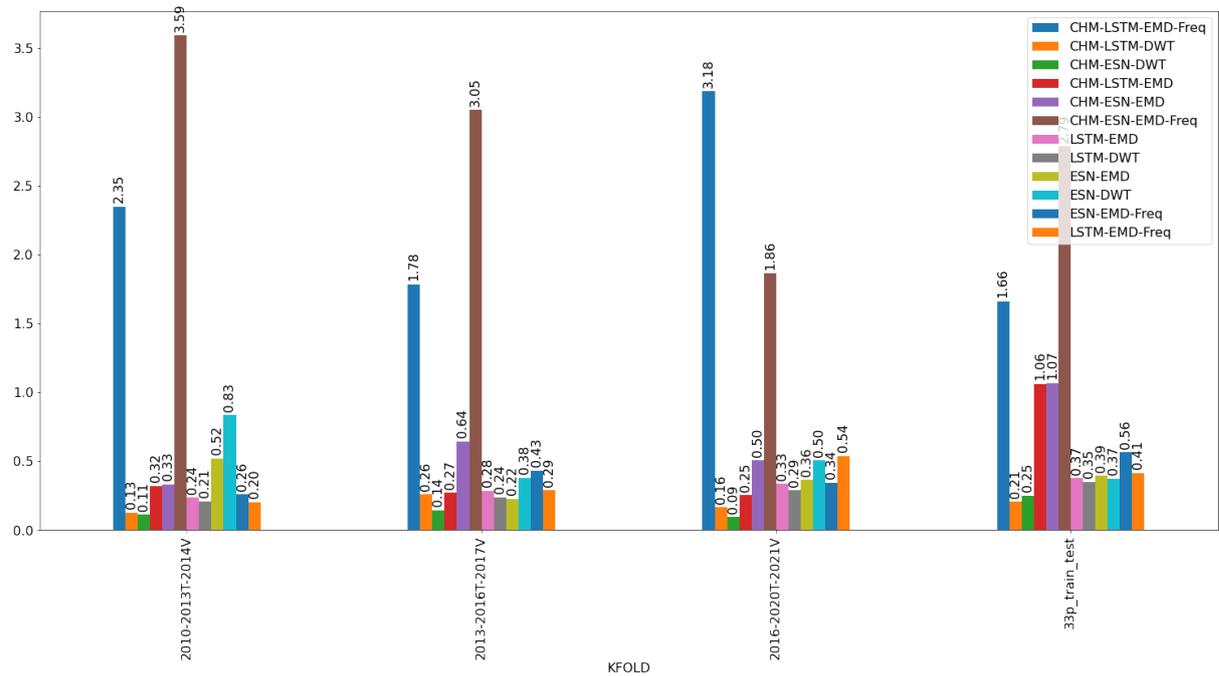
Fonte: Autor

Por fim, a Figura 42 mostra uma média geral entre, ações previstas, base utilizada e períodos. Assim, podemos concluir que os modelos com maior performance são CHM-LSTM-DWT e LSTM-DWT. Foi possível concluir que o método de CHM proposto tem resultados relevantes quando comparado a um modelo que utiliza a mesma rede sem a arquitetura proposta. Apesar disto, os valores finais, mostrados na Figura 42, são similares entre os modelos que utilizam a arquitetura proposta e o que não a utilizam. Este resultado foi impactado principalmente pela baixa performance do CHM-LSTM-DWT quando utilizado nas bolsas mundiais.

5.2 AVALIAÇÃO DA PREVISÃO DE PONTOS DE REVERSÃO

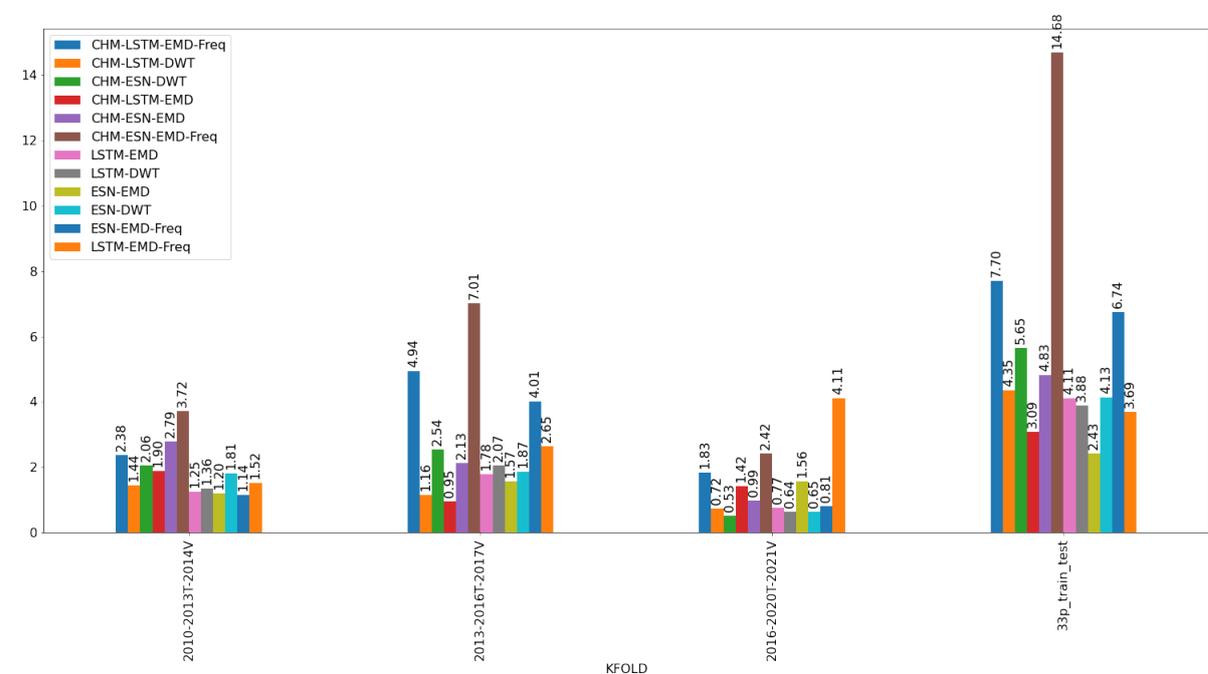
Esta seção apresenta uma comparação do modelo proposto com os demais para a previsão de pontos de reversão. Para realizar a comparação, foram utilizadas as métricas *f1-score* (Seção 3.9.3.3), *accuracy* (Seção 3.9.3.1) e *precision* (Seção 3.9.3.2). Os pontos foram classificados em *Down Reversal Point* (DRP) e *Up Reversal Point* (URP). Para um ponto ser considerado de reversão (Seção 3.7) foram utilizados os seguintes parâmetros: para o *trend* foram considerados 4 pontos anteriores, para o candidato a ponto de reversão foi utilizado o ponto previsto pelos algoritmos, para a verificação do ponto de reversão foi utilizado um retorno mínimo de 2% nos 10 dias seguintes.

Figura 37 – RMSE da ação ABEV utilizando somente a bolsa brasileira por período avaliado



Fonte: Autor

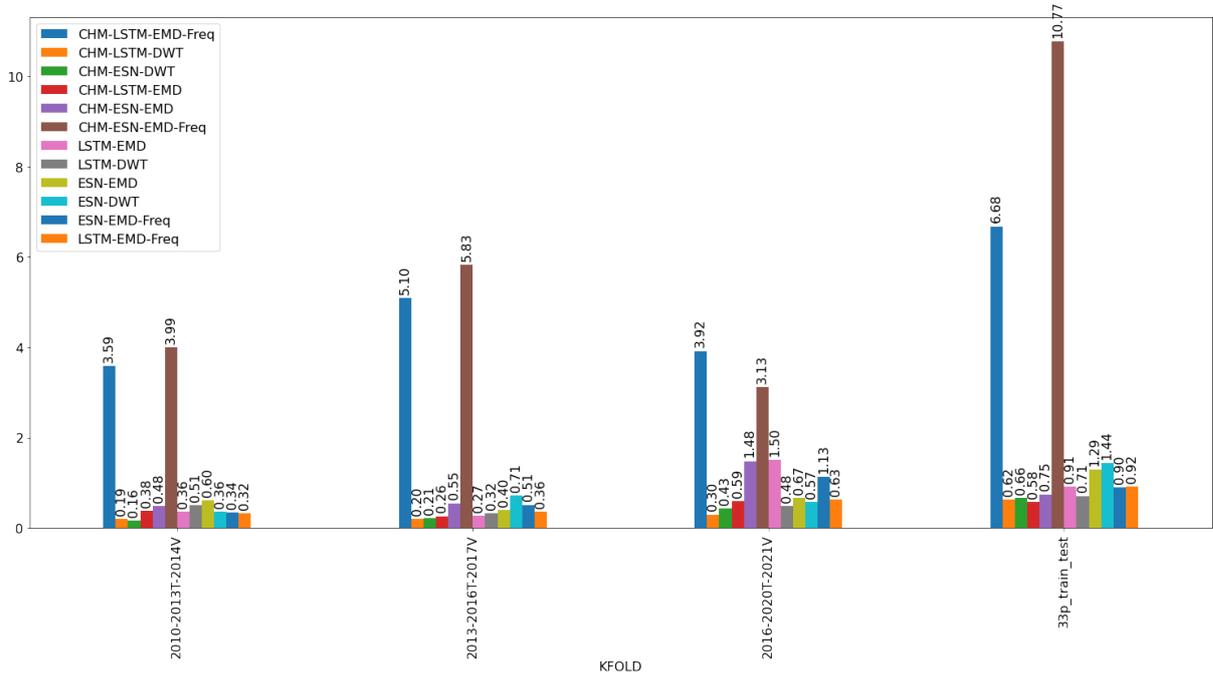
Figura 38 – RMSE da ação ITUB3 utilizando bolsas globais por período avaliado



Fonte: Autor

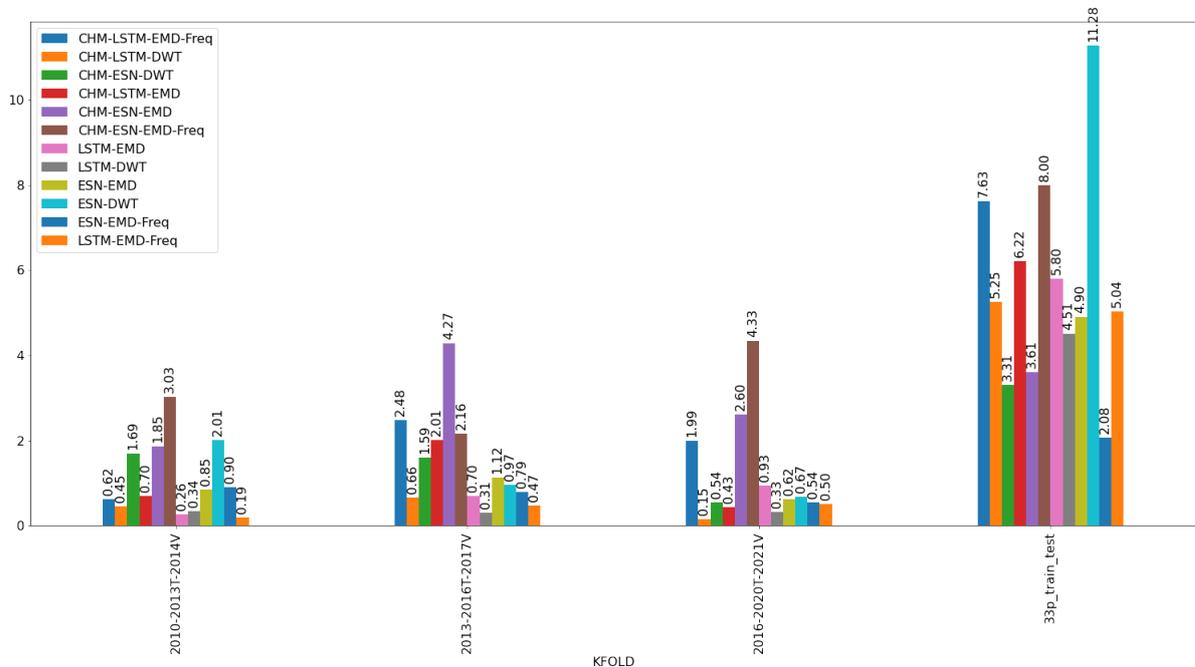
Assim, foram gerados seis gráficos de barras que apresentam a avaliação dos modelos. Cada gráfico contém as três métricas de classificação citadas anteriormente para cada variação de modelo, mais um valor de referência *no skill*, que determina a probabilidade de realizar uma

Figura 39 – RMSE da ação ITUB3 utilizando somente a bolsa brasileira por período avaliado



Fonte: Autor

Figura 40 – RMSE da ação B3SA utilizando bolsas globais por período avaliado

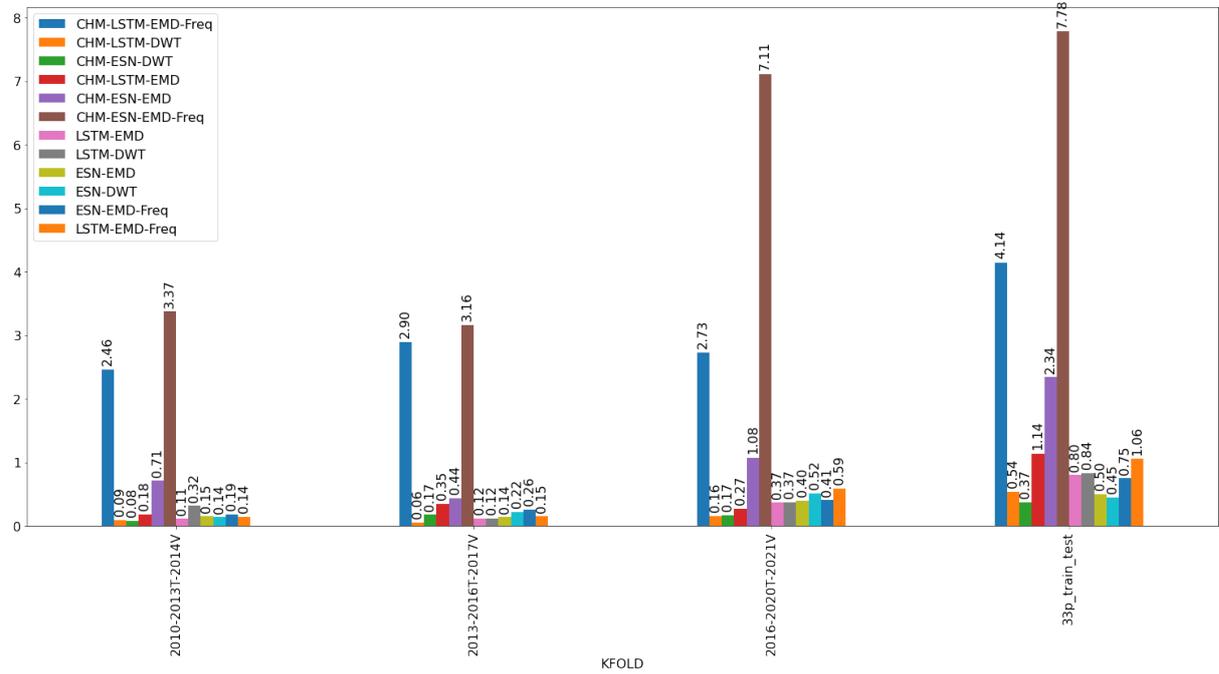


Fonte: Autor

compra ou venda correta sem utilizar qualquer tipo de modelo. O valor de *no skill* é calculado a partir da Equação (40).

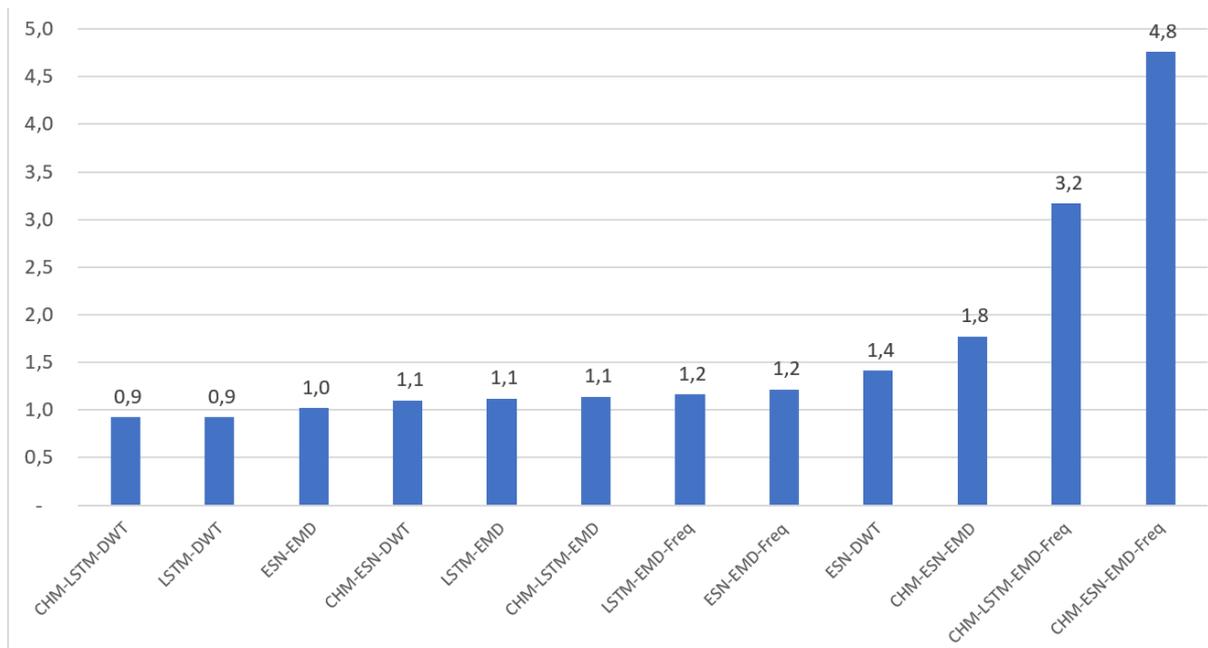
$$NoSkill = RP/N \tag{40}$$

Figura 41 – RMSE da ação B3SA utilizando somente a bolsa brasileira por período avaliado



Fonte: Autor

Figura 42 – RMSE média dos modelos



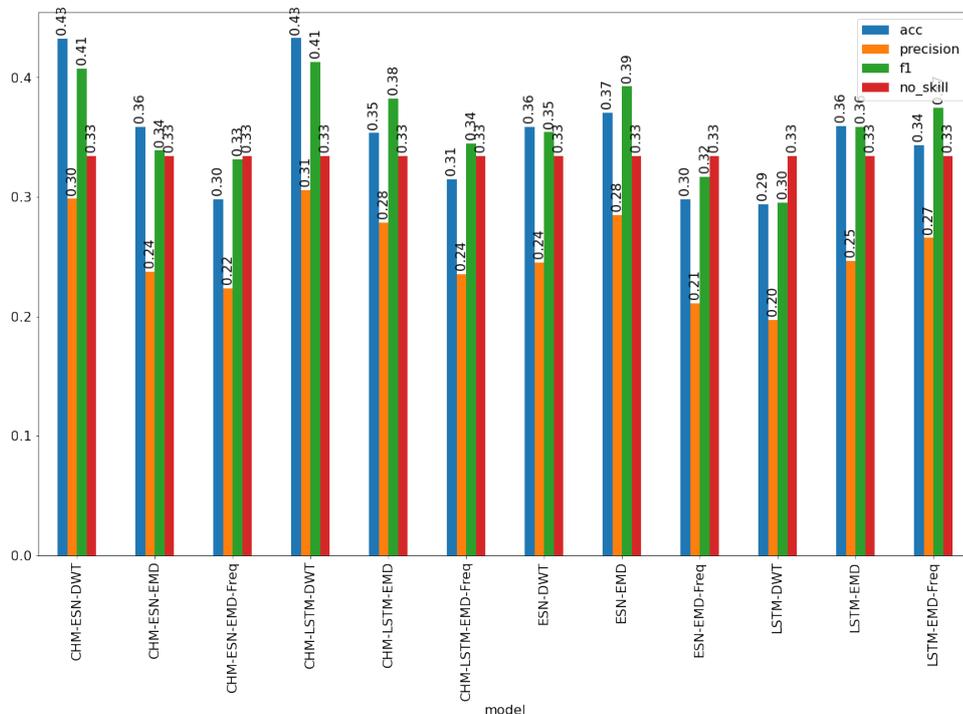
Fonte: Autor

Onde *RP* é a quantidade de *reversal points*, e *N* a quantidade de pontos da série temporal. Sendo assim, foram gerados dois gráficos para cada ação, um utilizando as bolsas mundiais e outro utilizando somente a bolsa brasileira, Figuras 43, 44, 45, 46, 47, 48. Todos os gráficos foram

gerados utilizando uma média aritmética simples dos resultados obtidos entre os 4 períodos avaliados.

Analisando os gráficos obtidos é possível notar uma melhora média de 35% no *f1-score*, utilizando a bolsa brasileira, e 10% do *f1-score* para bolsas mundiais, entre o modelo LSTM-DWT para o CHM-LSTM-DWT, o que é um indício da eficácia da metodologia proposta. Em uma média simples, podemos deduzir um aumento médio no *f1-score* de 22% ao utilizar a metodologia CHM proposta juntamente com LSTM e DWT. Na comparação entre os modelos ESN-EMD e CHM-ESN-EMD, é possível notar uma melhora de 21% na média do *f1-score*, utilizando a bolsa brasileira, e 5% do *f1-score* para bolsas mundiais. Em uma média simples, podemos deduzir um aumento médio no *f1-score* de 13% ao utilizar a metodologia CHM proposta, juntamente com ESN e EMD.

Figura 43 – Métricas de classificação para a ação B3SA utilizando somente bolsa brasileira

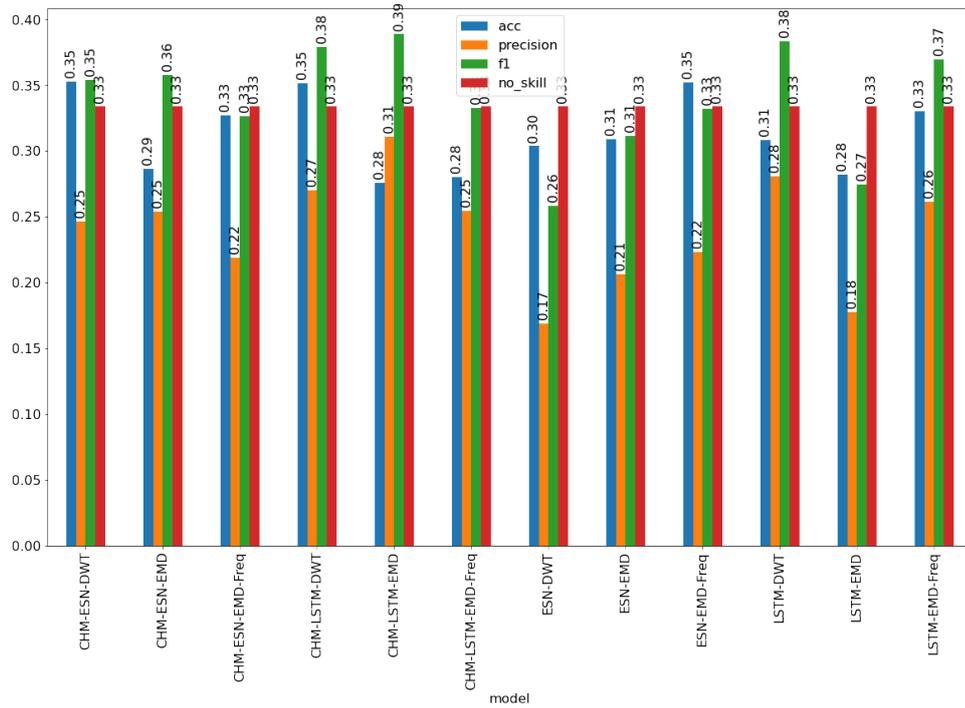


Fonte: Autor

Os gráficos das Figuras 49, 50, 51, 52 apresentam exemplos de série temporal com a identificação de pontos de reversão preditos: verde para o URP, e vermelho para o DRP. Os pontos preenchidos em triângulos na cor preta são considerados pontos de reversão verdadeiros, enquanto os outros serão apenas candidatos.

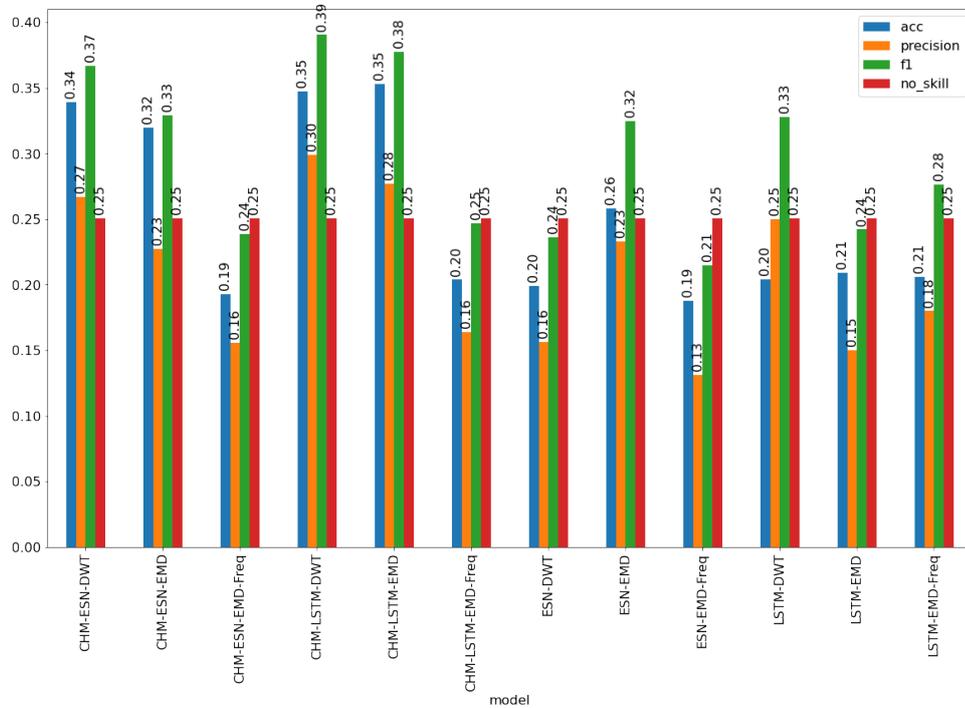
Por fim, a Figura 53 mostra uma média geral entre, ações preditas, base utilizada e períodos. Assim, podemos concluir que, apesar dos valores de *f1-score* não serem altos, em diversos casos os melhores modelos são os que utilizam a arquitetura de CHM sem a utilização de variáveis de frequência, e na maioria dos casos possuem valores de *f1-score* superiores aos valores de *no skill*.

Figura 44 – Métricas de classificação para a ação B3SA utilizando as bolsas mundiais



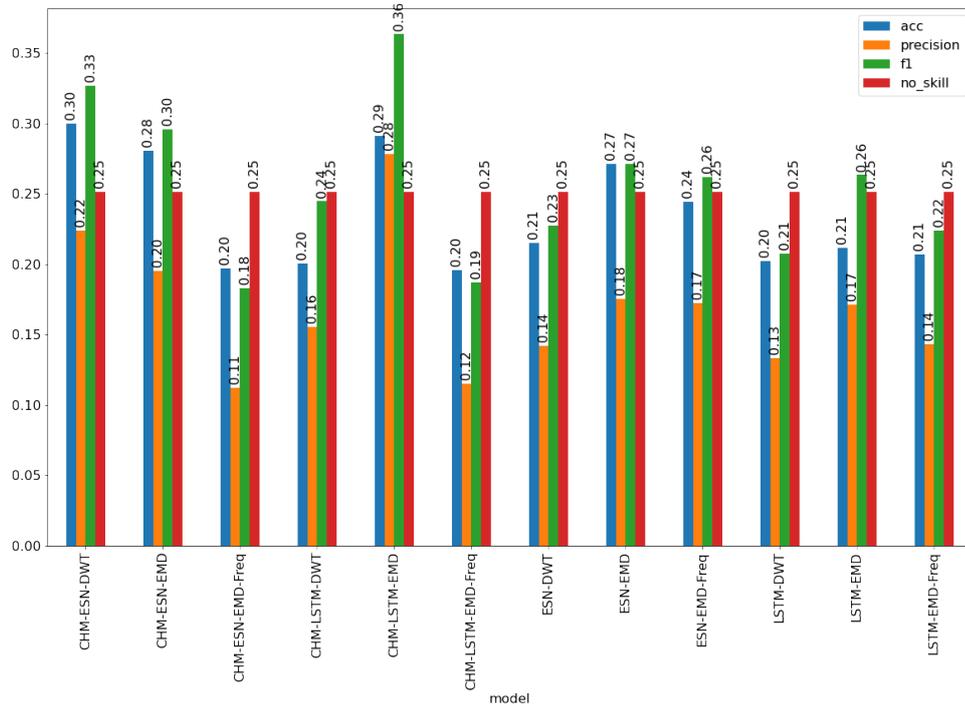
Fonte: Autor

Figura 45 – Métricas de classificação para a ação ITUB3 utilizando somente bolsa brasileira



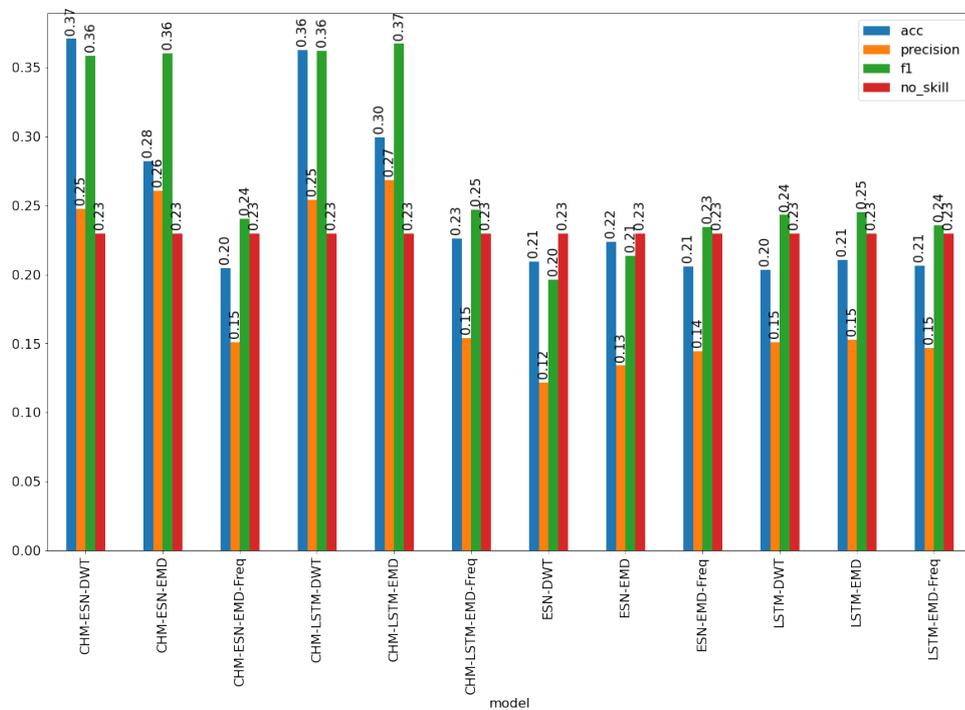
Fonte: Autor

Figura 46 – Métricas de classificação para a ação ITUB3 utilizando as bolsas mundiais



Fonte: Autor

Figura 47 – Métricas de classificação para a ação ABEV utilizando somente bolsa brasileira

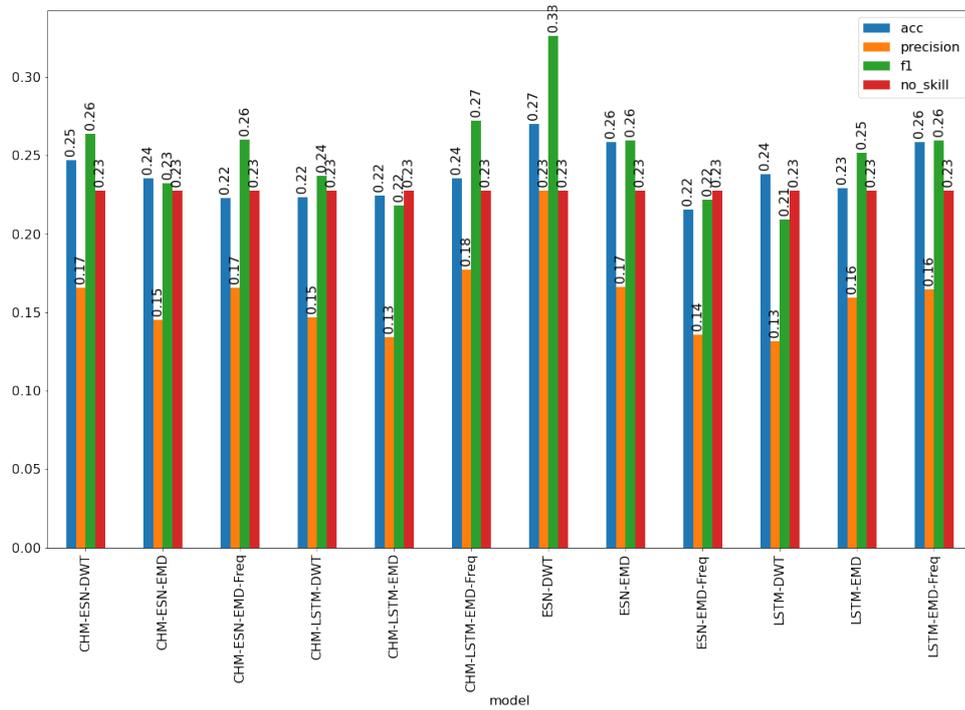


Fonte: Autor

5.3 ANÁLISE DO MÉTODO DE RETREINO

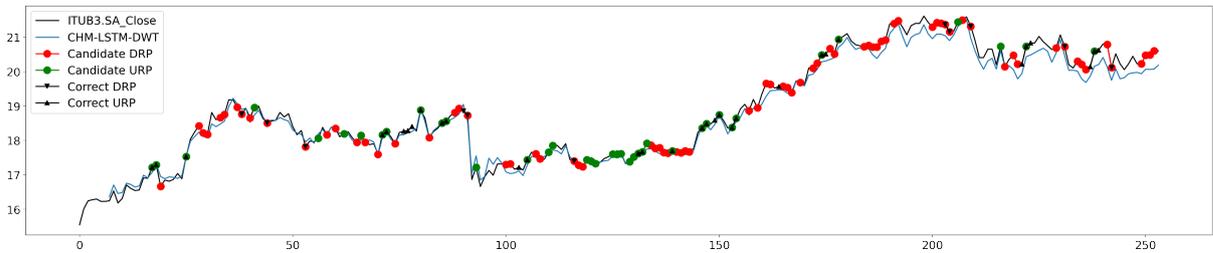
Esta seção apresenta os resultados da avaliação do método de retreino sobre a metodologia proposta. Como notado nas avaliações anteriores, os modelos tendem a perder performance

Figura 48 – Métricas de classificação para a ação ABEV utilizando as bolsas mundiais



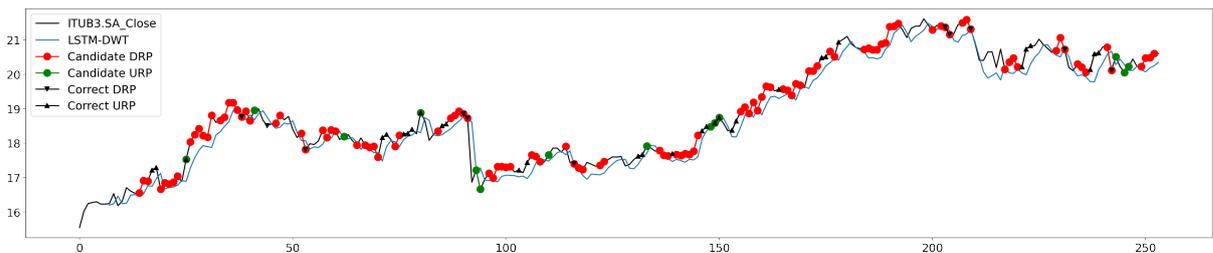
Fonte: Autor

Figura 49 – Série de classificação com modelo CHM-LSTM-DWT para a ação ITUB utilizando somente bolsa brasileira, no período 2013-2016T-2017V



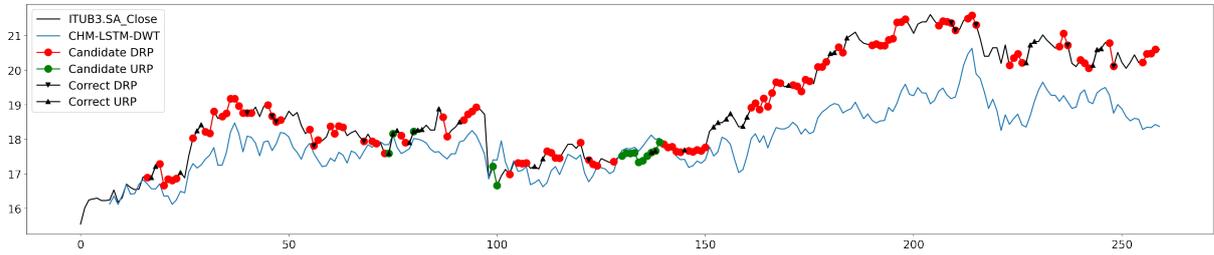
Fonte: Autor

Figura 50 – Série de classificação com modelo LSTM-DWT para a ação ITUB utilizando somente bolsa brasileira, no período 2013-2016T-2017V



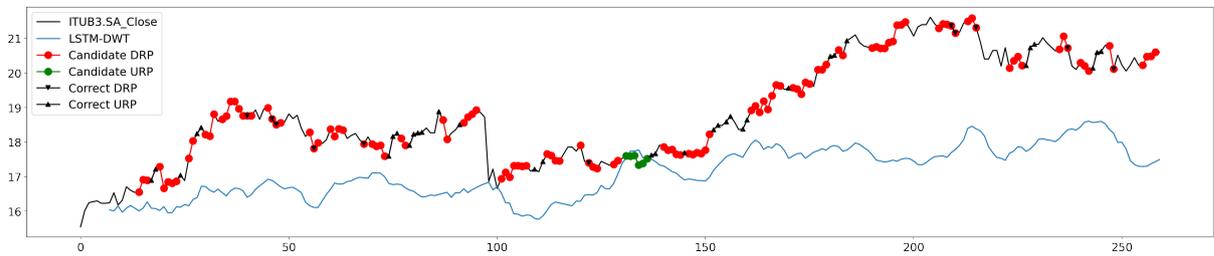
Fonte: Autor

Figura 51 – Série de classificação com modelo CHM-LSTM-DWT para a ação ITUB utilizando bolsas mundiais, no período 2013-2016T-2017V



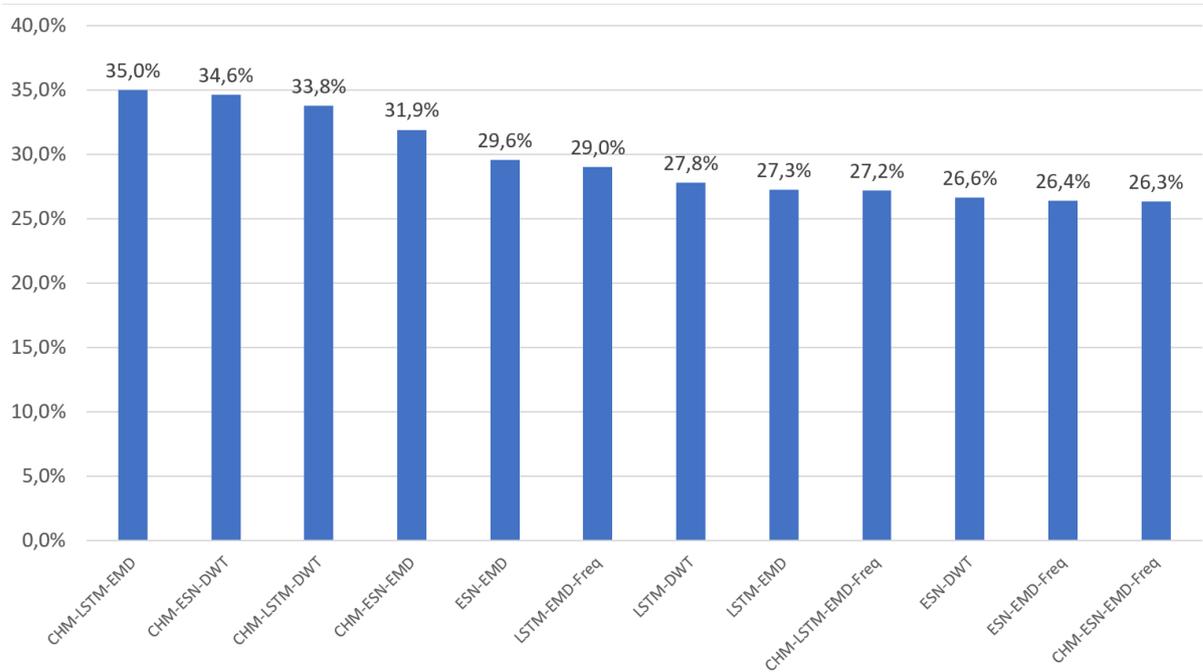
Fonte: Autor

Figura 52 – Série de classificação com modelo LSTM-DWT para a ação ITUB utilizando bolsas mundiais, no período 2013-2016T-2017V



Fonte: Autor

Figura 53 – *f1-score* média dos modelos

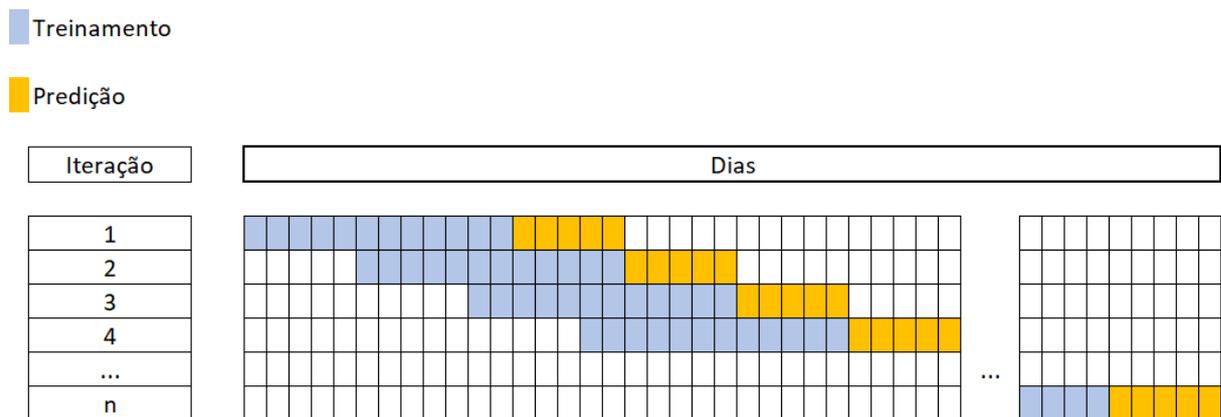


Fonte: Autor

ao longo do tempo, sugerindo a necessidade de um retreino do modelo para ajusta-lo à nova distribuição de dados. Devido à complexidade deste teste, em relação ao tempo de treinamento, não foi possível gerar o teste para todas as variações da metodologia proposta. Sendo assim, para realizar este teste foram utilizados os seguintes parâmetros: período 2010-2013T-2014V dos 4 períodos avaliados; somente a ação do Itaú (ITUB3); utilizando o modelo CHM-LSTM-DWT, tanto para a base que contém somente a bolsa brasileira, quanto para a base que utiliza as bolsas mundiais. O teste foi feito utilizando uma janela de retreino a cada 15 dias.

A técnica de retreino funciona da seguinte forma: primeiramente é selecionada uma janela de treino, onde o modelo faz o treinamento sobre esta janela. Em seguida, o modelo faz a previsão dos k dias que sucedem o último dia de treino, onde k é o comprimento da janela de treino. Então a janela de treino é deslocada para adiante em k dias, e o processo de treinamento é novamente calculado sobre os novos dados de treino, descartando as informações aprendidas anteriormente. Por fim, é feita uma nova previsão dos k dias que sucedem o último ponto da nova janela de treino. Este processo é reproduzido diversas vezes até terminar os pontos da série utilizada. A Figura 54 apresenta um exemplo deste processo para k igual a 5.

Figura 54 – Exemplo de retreino utilizando uma janela de 5 dias



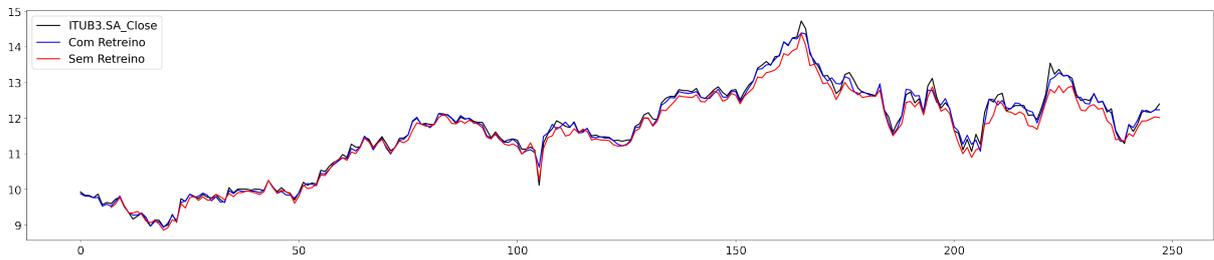
Fonte: Autor

A Figura 55 apresenta a série temporal dos valores da ação ITUB3 predita com o método CHM-LSTM-DWT proposto para a bolsa brasileira. Esta figura apresenta 3 linhas que fazem a comparação entre a predição utilizando o método de retreino e a predição que não faz o uso do método de retreino. A Figura 56 mostra uma redução de 47% no valor do RMSE quando a técnica faz o uso do método de retreino.

As Figuras 57 e 58 fazem a mesma análise descrita anteriormente, porém agora utilizando as bolsas mundiais. É possível notar uma redução ainda maior em relação aos valores que utilizam somente a bolsa brasileira. Desta vez, o método de retreino apresenta uma redução de 85% em relação ao método tradicional.

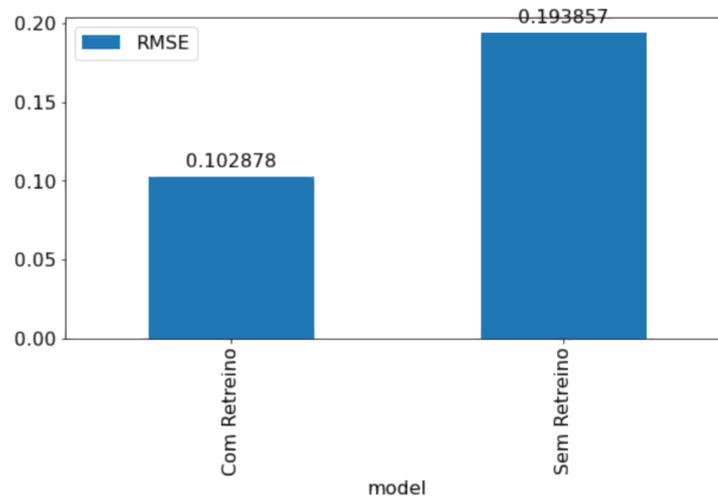
Também foi avaliada a performance de classificação de pontos de compra e venda dos ativos quando utilizada a técnica de retreino. A Figura 59 mostra a série temporal com os pontos

Figura 55 – Série de predição dos valores da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas brasileiras



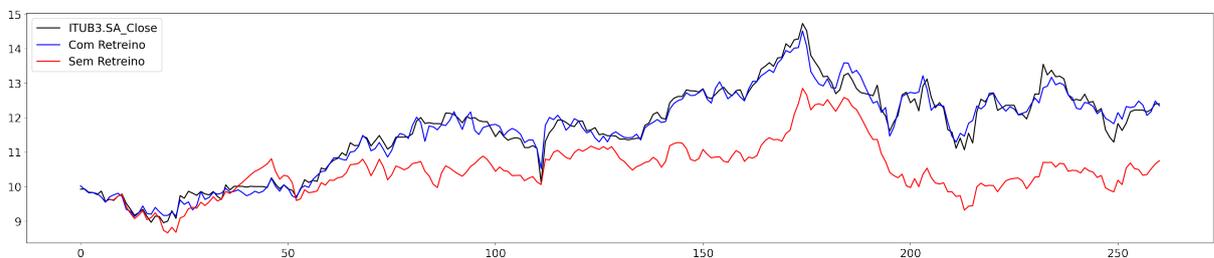
Fonte: Autor

Figura 56 – Comparação dos valores de RMSE da ação ITUB no período 2010-2013T-2014V para as bolsas brasileiras utilizando o modelo CHM-LSTM-DWT com retreino e sem retreino



Fonte: Autor

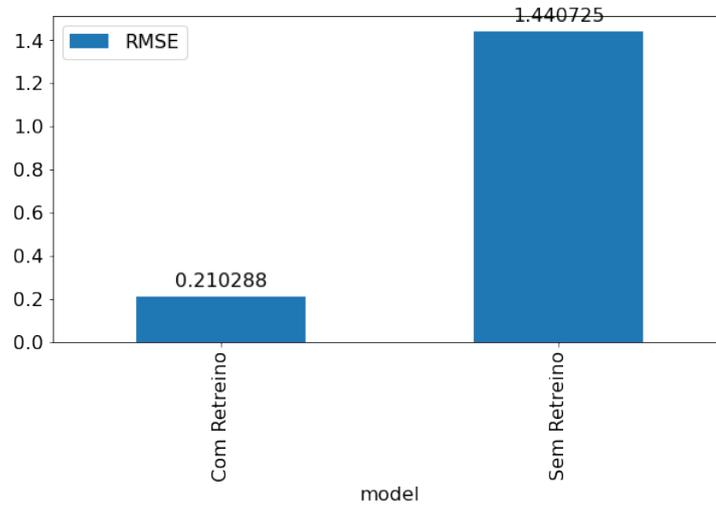
Figura 57 – Série de predição dos valores da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas mundiais



Fonte: Autor

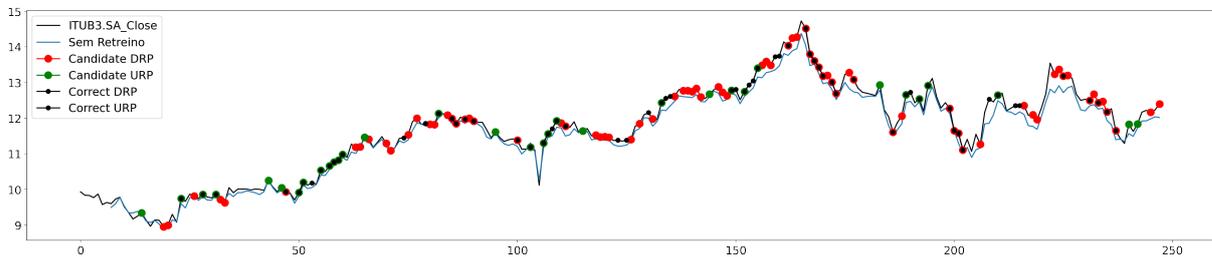
de DRP em vermelho, e os pontos de URP em verde da ação ITUB3 predita com o método CHM-LSTM-DWT proposto para a bolsa brasileira, sem o uso da técnica de retreino. A Figura 60 mostra a mesma série, porém fazendo o uso a técnica de retreino.

Figura 58 – Comparação dos valores de RMSE da ação ITUB no período 2010-2013T-2014V para as bolsas mundiais utilizando o modelo CHM-LSTM-DWT com retreino e sem retreino



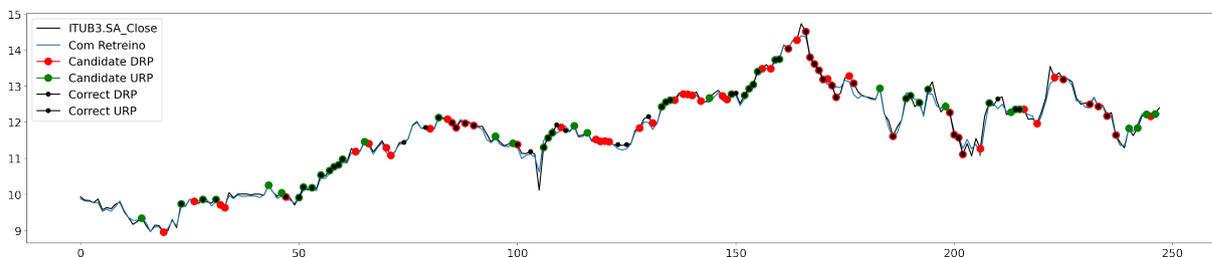
Fonte: Autor

Figura 59 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas brasileiras sem retreino



Fonte: Autor

Figura 60 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas brasileiras com retreino

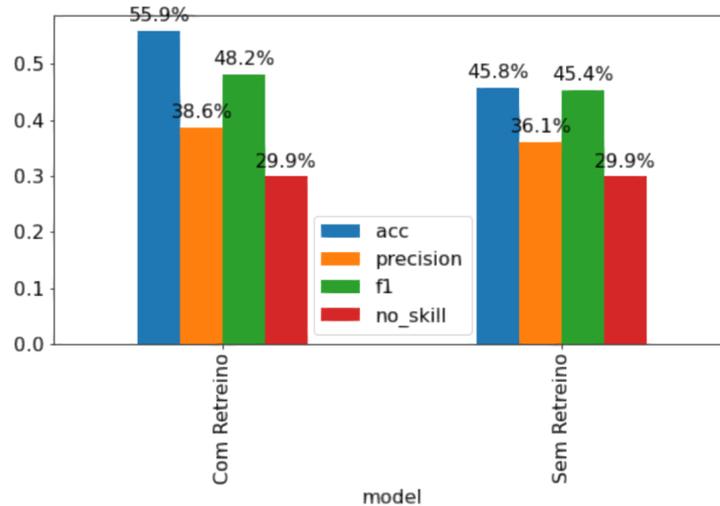


Fonte: Autor

A Figura 61 apresenta a performance do modelo que utiliza o método de retreino para realizar as classificações em relação ao modelo que não utiliza. Não foi possível notar uma

melhora significativa nos resultados quando é utilizada a técnica de retreino, dado que a métrica *f1-score* apresenta um acréscimo de 6% em relação ao método sem retreino.

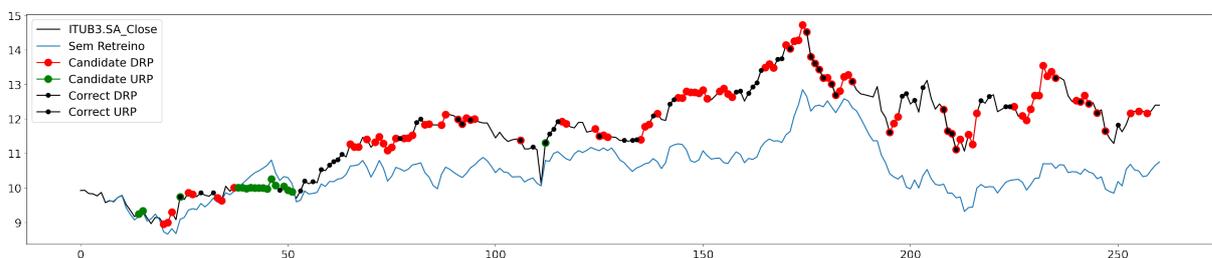
Figura 61 – Comparação dos valores de *f1-score*, Acurácia e Precisão do modelo CHM-LSTM-DWT com retreino e sem retreino sobre ação ITUB no período 2010-2013T-2014V para as bolsas brasileiras



Fonte: Autor

As Figuras 62 e 63 fazem a mesma análise descrita anteriormente, porém agora utilizando as bolsas mundiais. Desta vez, é possível notar uma redução significativa em relação aos valores que utilizam somente a bolsa brasileira, dado que o método de retreino apresenta um acréscimo de 95% de *f1-score* em relação ao método tradicional.

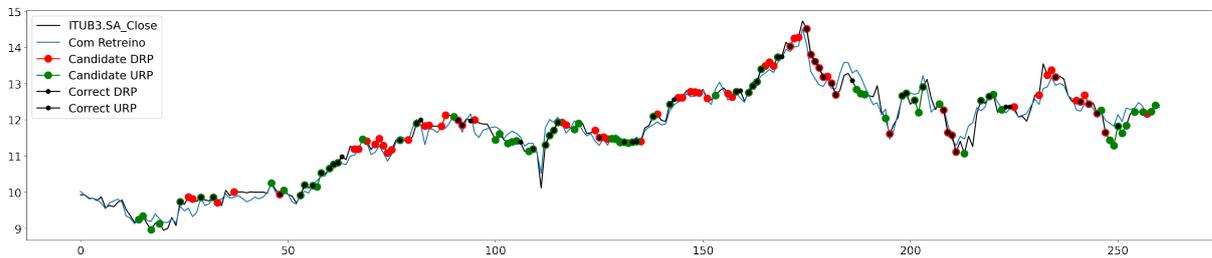
Figura 62 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas mundiais sem retreino



Fonte: Autor

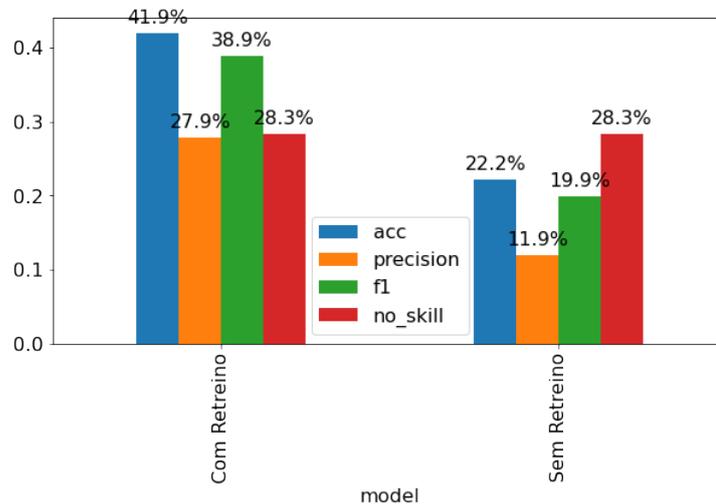
Por fim, foi possível concluir que o método de retreino traz melhoras significativas na performance do modelo proposto, dado que o mesmo tem performances altas nos primeiros pontos de predição, o retreino tende a manter esta performance inicial.

Figura 63 – Série de classificação da ação ITUB no período 2010-2013T-2014V utilizando o modelo CHM-LSTM-DWT para as bolsas mundiais com retreino



Fonte: Autor

Figura 64 – Comparação dos valores de *f1-score*, Acurácia e Precisão do modelo CHM-LSTM-DWT com retreino e sem retreino sobre ação ITUB no período 2010-2013T-2014V para as bolsas mundiais

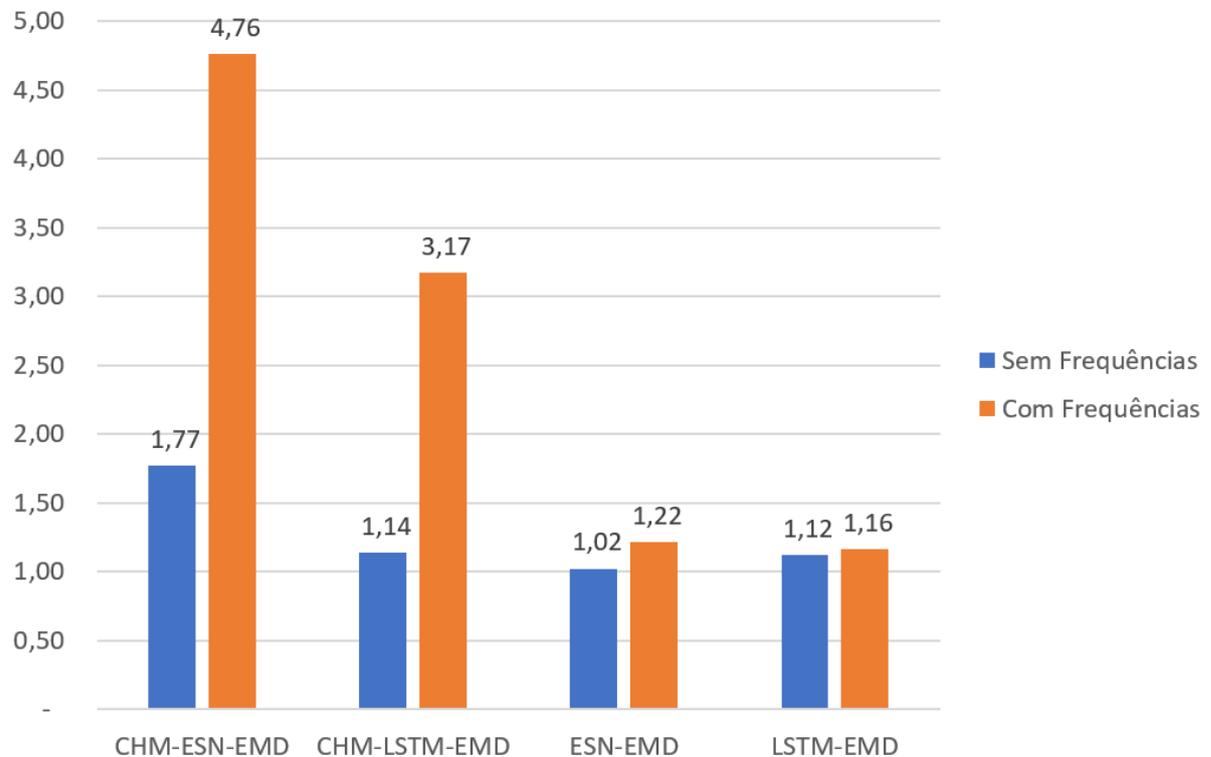


Fonte: Autor

5.4 ANÁLISE DO USO DE FREQUÊNCIAS INSTANTÂNEAS

Esta seção faz a análise do impacto das variáveis de frequência instantânea na performance dos algoritmos. A Figura 65 faz a análise utilizando a métrica RMSE para a predição do valor de fechamento da ação no dia seguinte. A Figura 66 faz a análise utilizando a métrica *f1-score* para a predição de pontos DRP e URP. Em ambos os casos, foi possível notar uma redução na performance dos algoritmos quando a variável de frequência instantânea é utilizada. Isto sugere que a variável intervêm negativamente sobre o treinamento das redes neurais. Também é possível notar que os modelos que utilizam a extração de contexto, proposta neste trabalho, são mais impactadas que os modelos que não a utilizam.

Figura 65 – Comparação dos valores de RMSE entre modelos que utilizam frequências instantâneas e os que não utilizam.



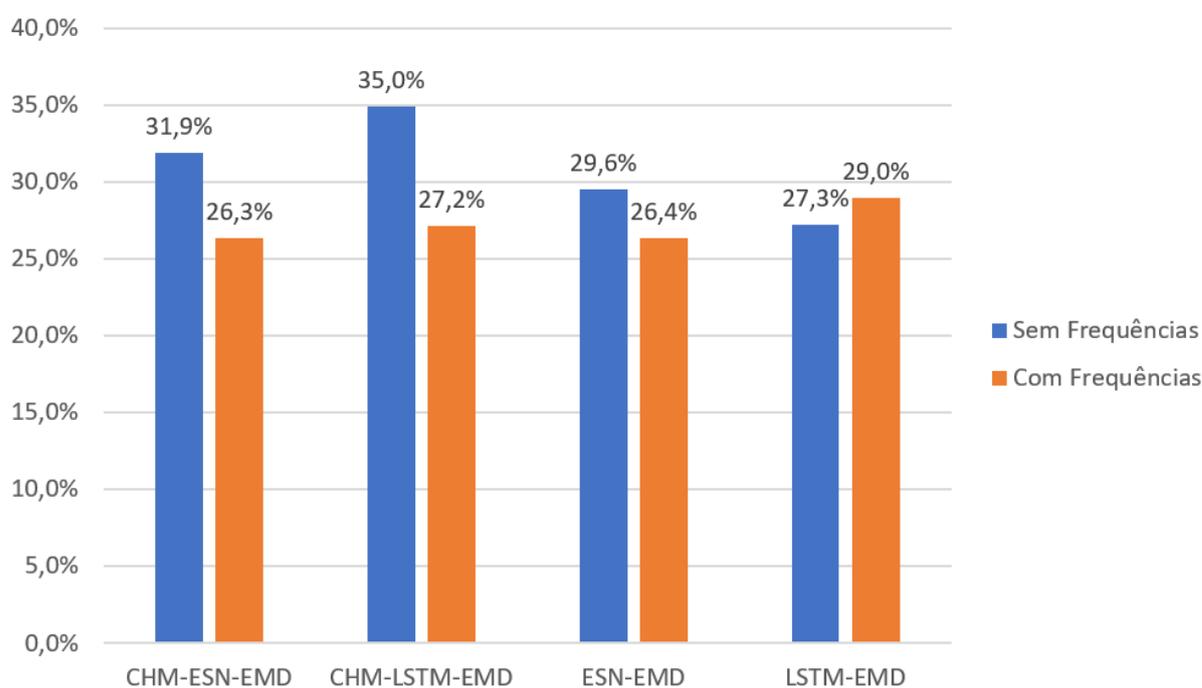
Fonte: Autor

5.5 CONTRIBUIÇÕES

Após a análise dos resultados apresentados, é possível destacar alguns pontos relevantes que o presente trabalho agrega na área de predição de valores do mercado de ações, sendo eles:

- A proposta de uma nova metodologia para a previsão de séries temporais. Apesar da metodologia de extração de contexto, proposta neste trabalho, ter sido utilizada para a previsão da bolsa de valores, o algoritmo proposto pode ser utilizado nas diversas áreas que utilizam séries temporais que apresentem distribuições que mudam constantemente.
- A comparação de metodologias que fazem a utilização de extração de contexto sobre as séries temporais com algoritmos que não levam em consideração esta variável. Os resultados mostraram que as metodologias que fazem o uso de extração de contexto têm performances iguais ou maiores do que os que não utilizam, principalmente quando utilizadas em conjunto com o algoritmo de DWT para a decomposição das séries.
- O impacto de utilizar diversas bolsas para realizar a previsão dos valores da bolsa brasileira. Ao contrário dos resultados esperados da hipótese levantada no início do trabalho, a utilização das bolsas mundiais em conjunto com a bolsa brasileira,

Figura 66 – Comparação dos valores de *f1-score* entre modelos que utilizam frequências instantâneas e os que não utilizam.



Fonte: Autor

não aumentou a performance dos modelos, na maioria dos casos teve impactos negativos.

- d) O impacto da utilização de frequências oscilatórias, ou frequências instantâneas, na predição dos valores futuros da série temporal. Como foi possível notar com a análise dos resultados, os algoritmos que fazem a utilização tiveram resultados inferiores aos que não fazem o uso destas variáveis. Rejeitando, assim, a hipótese levantada no início do trabalho.
- e) O impacto da técnica de retreino constante dos algoritmo proposto. O trabalho apresentou os resultados gerados pela metodologia proposta com o retreino, e sem o retreino constante. Foi possível notar uma melhora significativa na performance do modelo quando aplicada a técnica de retreino.

6 CONCLUSÃO

Neste trabalho, foi proposta uma nova metodologia para a predição das séries temporais da bolsa de valores brasileira. A metodologia proposta faz uma adaptação da técnica de extração de contexto para imagens, proposta por Seyedhosseini, Sajjadi e Tasdizen (2013), para a sua aplicação em séries temporais. Foram utilizados também algoritmos que não fazem o uso da técnica proposta, LSTM e ESN, para validar a eficácia da metodologia proposta.

Para a comparação, os algoritmos foram aplicados tanto na regressão, para prever o valor de fechamento do dia seguinte da ação, quanto para classificação, utilizando o valor previsto para indicar pontos de reversão na série. Assim, foram utilizadas técnicas estatísticas para avaliação das regressões e classificações, sendo elas: RMSE, acurácia, precisão e *F1-score*. Então, foram comparados 12 variações de modelo, sendo que 6 utilizam a metodologia proposta e 6 não utilizam, para 3 ações brasileiras, com 2 conjuntos de *features* diferentes. Como resultado, os melhores algoritmos para regressão foram o CHM-LSTM-DWT e LSTM-DWT, ambos com 0.9 de RMSE. Porém, o algoritmo CHM-LSTM-DWT é, em média, 38% melhor quando treinado somente na bolsa brasileira. Para a classificação, os melhores algoritmos foram o CHM-LSTM-DWT e CHM-ESN-DWT, com 35% e 34.6% de *f1-score*, respectivamente.

Além disso, foram testadas duas hipóteses adicionais: o impacto da utilização de variáveis de frequência instantânea na performance dos modelos e o impacto da utilização de diversas bolsas de valores na performance dos modelos. Como resultado, foi observado que tanto a frequência instantânea quanto a utilização de diversas bolsas têm impacto negativo na predição dos valores da série temporal.

Também foi testada a técnica de retreino constante do algoritmo ao longo do tempo. Ao analisar os resultados dos algoritmos, quando treinados com as bolsas mundiais, foi possível notar que o algoritmo proposto tinha uma boa performance nos primeiros 15 pontos de predições após o treinamento. Sendo assim, foi testado um retreino a cada 15 pontos. Os resultados mostraram uma redução significativa nos valores de RMSE, em 85%, e também um aumento nos valores de *f1-score*, em 95%.

Por fim, foi possível observar que, em geral, a metodologia proposta tem melhores resultados em relação às métricas estatísticas, quando comparado com algoritmos tradicionais que não fazem o uso da técnica de extração de contexto para séries temporais. A metodologia proposta é promissora tanto na predição de ações do mercado financeiro quanto em qualquer série temporal que tenha uma distribuição que varia ao longo do tempo.

No entanto, ainda existem tópicos que podem ser explorados em trabalhos futuros, alguns exemplos são: a avaliação de outras topologias de redes neurais; comparação com metodologias que descartam as maiores frequências obtidas pelos algoritmos de decomposição de séries; utilizar mais ações e mais variáveis no treinamento da rede; explorar as diversas funções de ativação das redes neurais da metodologia proposta; e utilizar diferentes períodos de valores como por exemplo 1 minuto, ao invés de 1 dia entre os pontos da série temporal.

REFERÊNCIAS

- AKITA, Ryo et al. Deep learning for stock prediction using numerical and textual information. **2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)**, p. 1–6, 2016.
- ALTAN, Aytac; KARASU, Seçkin; BEKIROĞLU, S. Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. **Chaos Solitons & Fractals**, v. 126, p. 325–336, 2019.
- ANGUELOV, Dragomir et al. Contextual Identity Recognition in Personal Photo Albums. **2007 IEEE Conference on Computer Vision and Pattern Recognition**, p. 1–7, 2007.
- ARAI, Y.; YOSHIKAWA, Takeo; IYETOMI, H. Dynamic Stock Correlation Network. In: KES. [S.l.: s.n.], 2015.
- ASGHARIAN, Hossein; HESS, W.; LIU, L. A spatial analysis of international stock market linkages. **Journal of Banking and Finance**, v. 37, p. 4738–4754, 2013.
- BAO, Wei; YUE, Jun; RAO, Yulei. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. **PLoS ONE**, v. 12, 2017.
- BAR, M. et al. The contribution of context to visual object recognition. **Journal of Vision**, v. 5, p. 88–88, 2005.
- BARJAKTAROVIĆ, Lidija; BARJAKTAROVIC, M.; KONJIKUIĆ, Sneana. ECHO STATE NETWORKS USAGE FOR STOCK PRICE PREDICTIONS. In:
- BERNAL, Armando; FOK, Sam; PIDAPARTHI, R. Market Time Series Prediction with Recurrent Neural Networks. In:
- BOLDANOV, Rustam; DEGIANNAKIS, Stavros; FILIS, G. Time-Varying Correlation Between Oil and Stock Market Volatilities: Evidence From Oil-Importing and Oil-Exporting Countries. **Econometric Modeling: Commodity Markets eJournal**, 2015.
- BOLLINGER, J. Bollinger on Bollinger Bands. In:
- BOROVKOVA, S.; TSIAMAS, Ioannis. An ensemble of LSTM neural networks for high-frequency stock market classification. **Journal of Forecasting**, v. 38, p. 600–619, 2019.
- BRACEWELL, Ronald Newbold; BRACEWELL, Ronald N. **The Fourier transform and its applications**. [S.l.]: McGraw-Hill New York, 1986. v. 31999.

BROCKWELL, P.; DAVIS, Richard A. Time Series: Theory and Methods. In:

CAGINALP, Gunduz; LAURENT, Henry. The Predictive Power of Price Patterns. **FEN: Behavioral Finance (Topic)**, 1998.

CAO, J.; LI, Z.; LI, J. Financial time series forecasting model based on CEEMDAN and LSTM. **Physica A-statistical Mechanics and Its Applications**, v. 519, p. 127–139, 2019.

CHANDAR, S.; SUMATHI, M.; SIVANANDAM, S. Prediction of Stock Market Price using Hybrid of Wavelet Transform and Artificial Neural Network. **Indian journal of science and technology**, v. 9, 2016.

CHENG, C.; WEI, Liang-Ying. A novel time-series model based on empirical mode decomposition for forecasting TAIEX. In:

CHOLLET, François et al. **Keras**. [S.l.: s.n.], 2015. <https://keras.io>.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

COVER, T. Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. **IEEE Trans. Electron. Comput.**, v. 14, p. 326–334, 1965.

COX, David; MEYERS, Ethan M; SINHA, P. Contextually Evoked Object-Specific Responses in Human Visual Cortex. **Science**, v. 304, p. 115–117, 2004.

DAN, Jingpei et al. Deterministic Echo State Networks Based Stock Price Forecasting. **Abstract and Applied Analysis**, v. 2014, p. 1–6, 2014.

DRAGOMIRETSKIY, Konstantin; ZOSSO, Dominique. Variational Mode Decomposition. **IEEE Transactions on Signal Processing**, v. 62, p. 531–544, 2014.

FAMA, Eugene. Efficient Capital Markets: A Review of Theory and Empirical Work. **Journal of Finance**, v. 25, p. 383–417, 1970.

FILHO, T.; ROCHA, Paulo M. M. Evidence of inefficiency of the Brazilian stock market: The IBOVESPA future contracts. **Physica A-statistical Mechanics and Its Applications**, v. 543, p. 123200, 2020.

FU, Fangjian. Idiosyncratic Risk and the Cross-Section of Expected Stock Returns. **Risk Management**, 2009.

GILLES, J. Empirical Wavelet Transform. **IEEE Transactions on Signal Processing**, v. 61, p. 3999–4010, 2013.

GOEKING, Weruska. **Apenas 3% dos brasileiros investiram em ações em 2020 e média aplicada caiu 31%**. [S.l.: s.n.]. <https://valorinveste.globo.com/mercados/renda-variavel/bolsas-e-indices/noticia/2021/01/15/apenas-3percent-dos-brasileiros-investiram-em-aco-es-em-2020-e-media-aplicada-caiu-31percent.ghtml>. Accessed: 2021-06-05.

GUO, X.; ZHANG, Hu; TIAN, T. Development of stock correlation networks using mutual information and financial big data. **PLoS ONE**, v. 13, 2018.

HARRIS, Charles R. et al. Array programming with NumPy. **Nature**, Springer Science e Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <<https://doi.org/10.1038/s41586-020-2649-2>>.

HAWKES, A.; COX, D.; MILLER, H. D. The Theory of Stochastic Processes. **The Mathematical Gazette**, v. 51, p. 261, 1967.

HIRANSHA, M. et al. NSE Stock Market Prediction Using Deep-Learning Models. **Procedia Computer Science**, v. 132, p. 1351–1362, 2018.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, p. 1735–1780, 1997.

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 Syst. Appl.**, v. 129, p. 273–285, 2019.

HUANG, N. et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. **Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences**, v. 454, p. 903–995, 1998.

HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007.

ILIES, I. et al. Stepping forward through echoes of the past : forecasting with Echo State Networks. In:

JAEGER, H. Echo state network. **Scholarpedia**, v. 2, p. 2330, 2007.

JIANG, WeiWei. Applications of deep learning in stock market prediction: recent progress. **ArXiv**, abs/2003.01859, 2020.

YU-JUN, Yang; YI-MEI, Yang; JIAN-HUA, Xiao. A Hybrid Prediction Method for Stock Price Using LSTM and Ensemble EMD. **Complexity**, v. 2020, p. 1–16, 2020.

KANTELHARDT, J. et al. Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. **Physica A-statistical Mechanics and Its Applications**, v. 316, p. 87–114, 2002.

KELOTRA, Amit; PANDEY, Prateek. Stock Market Prediction Using Optimized Deep-ConvLSTM Model. **Big data**, v. 8 1, p. 5–24, 2020.

KIM, H.; WON, C. H. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. **Expert Syst. Appl.**, v. 103, p. 25–37, 2018.

KIM, T.; KING, Brian R. Time series prediction using deep echo state networks. **Neural Computing and Applications**, p. 1–19, 2020.

KIM, Taewook; KIM, H. Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. **PLoS ONE**, v. 14, 2019.

KOUAISSAH, Nouredine et al. Theoretical and practical motivations for the use of the moving average rule in the stock market. **Ima Journal of Management Mathematics**, v. 31, p. 117–138, 2020.

KUMAR, D. A.; MURUGAN, S. Performance analysis of Indian stock market index using neural network time series model. **2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering**, p. 72–78, 2013.

KUMAR, Gourav; JAIN, Sanjeev; SINGH, U. P. Stock Market Forecasting Using Computational Intelligence: A Survey. **Archives of Computational Methods in Engineering**, p. 1–33, 2020.

LAHMIRI, Salim; BEKIROU, Stelios. Cryptocurrency forecasting with deep learning chaotic neural networks. **Chaos, Solitons & Fractals**, v. 118, p. 35–40, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960077918310233>>.

LANE, George C. Stochastic Oscillator. In:

LI, Jiahong; BU, Hui; WU, J. Sentiment-aware stock market prediction: A deep learning method. **2017 International Conference on Service Systems and Service Management**, p. 1–6, 2017.

LIN, Xiaowei; YANG, Zehong; SONG, Yixu. Short-term stock price prediction based on echo state networks. **Expert Syst. Appl.**, v. 36, p. 7313–7317, 2009.

LIU, J. et al. Financial Data Forecasting Using Optimized Echo State Network. In: ICONIP. [S.l.: s.n.], 2018.

LIU, Zhengfu et al. Predicting stock trend using multi-objective diversified Echo State Network. **2017 Seventh International Conference on Information Science and Technology (ICIST)**, p. 181–186, 2017.

LONG, Jianyu; ZHANG, Shaohui; LI, C. Evolving Deep Echo State Networks for Intelligent Fault Diagnosis. **IEEE Transactions on Industrial Informatics**, v. 16, p. 4928–4937, 2020.

LU, W. et al. A CNN-LSTM-Based Model to Forecast Stock Prices. **Complex.**, v. 2020, 6622927:1–6622927:10, 2020.

LUKOSEVICIUS, M.; JAEGER, H. Reservoir computing approaches to recurrent neural network training. **Comput. Sci. Rev.**, v. 3, p. 127–149, 2009.

MA, Feng; WEI, Y.; HUANG, Dengshi. Multifractal detrended cross-correlation analysis between the Chinese stock market and surrounding stock markets. **Physica A-statistical Mechanics and Its Applications**, v. 392, p. 1659–1670, 2013.

MALKIEL, B. The Efficient Market Hypothesis and Its Critics. **Journal of Economic Perspectives**, v. 17, p. 59–82, 2003.

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

MINH, Dang Lien et al. Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. **IEEE Access**, v. 6, p. 55392–55404, 2018.

MOLDOVAN, I. Stock Markets Correlation: before and during the Crisis Analysis. **Theoretical and Applied Economics**, p. 111–122, 2011.

NELSON, D.; PEREIRA, Adriano M.; OLIVEIRA, Renato A. de. Stock market's price movement prediction with LSTM neural networks. **2017 International Joint Conference on Neural Networks (IJCNN)**, p. 1419–1426, 2017.

NIU, Hongli; XU, Kunliang; WANG, W. A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network. **Applied Intelligence**, p. 1–14, 2020.

NOFSINGER, John. Social Mood and Financial Economics. **Journal of Behavioral Finance**, v. 6, p. 144–160, 2005.

NOSRATABADI, S. et al. Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. In:

PATEL, Dev; PATEL, K.; CUESTA, Charles Dela. Approaching Stock Market Trading with Echo State Networks. **Journal of Student Research**, v. 9, 2020.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PENG, C. et al. Mosaic organization of DNA nucleotides. **Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics**, v. 49 2, p. 1685–9, 1994.

PIMENTA, A. et al. An Automated Investing Method for Stock Market Based on Multiobjective Genetic Programming. **Computational Economics**, v. 52, p. 125–144, 2018.

PRECHTER, R. R.; PARKER, W. The Financial/Economic Dichotomy in Social Behavioral Dynamics: The Socionomic Perspective. **Journal of Behavioral Finance**, v. 8, p. 108–84, 2007.

SACHDEVA, Akshay et al. An Effective Time Series Analysis for Equity Market Prediction Using Deep Learning Model. **2019 International Conference on Data Science and Communication (IconDSC)**, p. 1–5, 2019.

SELVIN, Sreelekshmy et al. Stock price prediction using LSTM, RNN and CNN-sliding window model. **2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**, p. 1643–1647, 2017.

SEYEDHOSSEINI, M.; SAJJADI, Mehdi S. M.; TASDIZEN, T. Image Segmentation with Cascaded Hierarchical Models and Logistic Disjunctive Normal Networks. **2013 IEEE International Conference on Computer Vision**, p. 2168–2175, 2013.

SHAH, Dev; ISAH, H.; ZULKERNINE, F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. **International Journal of Financial Studies**, v. 7, p. 26, 2019.

SI, Yain-Whar; YIN, Jiangling. OBST-based segmentation approach to financial time series. **Eng. Appl. Artif. Intell.**, v. 26, p. 2581–2596, 2013.

SMITH, V. Constructivist and Ecological Rationality in Economics. **The American Economic Review**, v. 93, p. 465–508, 2002.

STOCCHI, M.; MARCHESI, M. Fast wavelet transform assisted predictors of streaming time series. **Digit. Signal Process.**, v. 77, p. 5–12, 2018.

TEAM, The pandas development. **pandas-dev/pandas: Pandas**. [S.l.]: Zenodo, fev. 2020. Disponível em: <<https://doi.org/10.5281/zenodo.3509134>>.

TING, Kai Ming. Confusion Matrix. In: **Encyclopedia of Machine Learning**. Edição: Claude Sammut e Geoffrey I. Webb. Boston, MA: Springer US, 2010. P. 209–209. Disponível em: <https://doi.org/10.1007/978-0-387-30164-8_157>.

TIWARI, A. et al. Stock Market Integration in Asian Countries: evidence from Wavelet multiple correlations. **Journal of Economic Integration**, v. 28, p. 441–456, 2013.

TORAMAN, Cengiz; BAŞARIR, Ç. The Long Run Relationship Between Stock Market Capitalization Rate and Interest Rate: Co-integration Approach. **Procedia - Social and Behavioral Sciences**, v. 143, p. 1070–1073, 2014.

TORRALBA, A.; MURPHY, K.; FREEMAN, W. Contextual Models for Object Detection Using Boosted Random Fields. In: NIPS. [S.l.: s.n.], 2004.

TU, Zhuowen; BAI, X. Auto-Context and Its Application to High-Level Vision Tasks and 3D Brain Image Segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, p. 1744–1757, 2010.

U, JuHyok et al. A new LSTM based reversal point prediction method using upward/downward reversal point feature sets. **Chaos, Solitons & Fractals**, v. 132, p. 109559, 2020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960077919305168>>.

WILDER, J. New Concepts in Technical Trading Systems. In:

WOLF, Lior; BILESCHI, S. A Critical View of Context. **International Journal of Computer Vision**, v. 69, p. 251–261, 2006.

YFINANCE. [S.l.: s.n.]. <https://pypi.org/project/yfinance/>. Accessed: 2021-06-05.

YU, Honghai; LIU, H. Improved Stock Market Prediction by Combining Support Vector Machine and Empirical Mode Decomposition. **2012 Fifth International Symposium on Computational Intelligence and Design**, v. 1, p. 531–534, 2012.

ZHOU, F. et al. EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction. **Expert Syst. Appl.**, v. 115, p. 136–151, 2019.

ZHOU, W. Multifractal detrended cross-correlation analysis for two nonstationary signals. **Physical review. E, Statistical, nonlinear, and soft matter physics**, 77 6 Pt 2, p. 066211, 2008.

ZHOU, X. et al. Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets. **Mathematical Problems in Engineering**, v. 2018, p. 1–11, 2018.