UNIVERSIDADE ESTADUAL DE CAMPINAS Faculdade de Tecnologia

Bruno Nogueira Renzo

Redes LSTM Aplicadas à Predição de Valores do Mercado de Câmbio

Bruno Nogueira Renzo

Redes LSTM Aplicadas à Predição de Valores do Mercado de Câmb

Monografia apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. João Roberto Bertini Junior

Este trabalho corresponde à versão final da Monografia defendida por Bruno Nogueira Renzo e orientada pelo Prof. Dr. João Roberto Bertini Junior.

FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da monografia em defesa do Título de Bacharel em Sistemas de Informação, a que se submeteu o aluno Bruno Nogueira Renzo, em 29 de novembro de 2021 na Faculdade de Tecnologia – FT/UNICAMP, em Limeira/SP.

Prof. Dr. Nome do Orientador Presidente da Comissão Julgadora

Prof. Dr. Plínio Roberto Souza Vilela FT/UNICAMP

Mestre Arthur Emanuel de Oliveira Carosia IFSP São Paulo

Ata da defesa, assinada pelos membros da Comissão Examinadora, encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós Graduação da Faculdade de Tecnologia.

Agradecimentos

Gostaria de agradecer primeiramente a minha família pelo apoio e contribuição para minha formação pessoal e profissional, por terem me auxiliado durante todo o projeto.

Agradeço também aos professores da UNICAMP pelo embasamento teórico e prático que possibilitaram o desenvolvimento do presente trabalho. Em especial, um agradecimento ao meu orientador João Bertini, pela disponibilidade, confiança e divulgação de conhecimento.

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

Resumo

Tanto no ambiente corporativo quanto na comunidade acadêmica, a predição de valores do mercado financeiro sempre representou um grande desafio. A descrição matemática do comportamento do mercado é uma tarefa extremamente complexa, são diversos os fatores que influenciam a flutuação dos preços. Contudo, Redes Neurais Artificiais têm demonstrado grande potencial no reconhecimento das relações não lineares do mercado financeiro, alcançando bons resultados de predição. O presente trabalho estrutura uma rede neural LSTM para realizar predições da taxa de câmbio. É apresentado ao modelo ainda em etapa de treinamento um conjunto de dados contendo 5.760 registros históricos do par de moedas EUR/USD, com o acompanhamento de minuto em minuto das oscilações do preço de fechamento desse par. Após aprendizado e abstração das relações contidas nesse conjunto de treinamento, a real capacidade preditiva do modelo é avaliada em um conjunto de teste, contendo 1.440 novos registros. Os resultados são promissores e evidenciam que a rede LSTM foi capaz de reconhecer dependências de longo prazo na série temporal financeira, com taxa de erro 71.8% menor que a obtida com uma RNN padrão.

Abstract

Exchange rate forecasting has always been a challenge for governments, industries, investors, and the academic community. There are so many factors that can affect market behavior, in such a way that a mathematical description of these relationships is almost impossible. However, artificial neural networks have reached great prediction results in the last years, showing themselves capable of capturing the non-linearity of financial markets relationships. This paper examines the potential of LSTM for exchange rate forecasting. A total of 5.760 historical records were used during the training phase. Once the model has learned the relationships from the training dataset, 1.440 new records are used to test the model and evaluate its performance. For benchmarking purposes, standard RNN was applied to solve the same prediction problem. Results showed better performance of LSTM over standard RNN, achieving error rates lower by 71.8%.

Lista de Figuras

2.1	Grafo computacional de uma RNN	17
2.2	Grafo computacional de uma LSTM	18
4.1	Reestruturação do conjunto de dados original.	22
4.2	Conjunto de inicial dados após as transformações necessárias	22
4.3	Filtragem de dados.	23
4.4	Separação dos dados em conjunto de treino e teste.	24
4.5	Fluxo de atividades para normalização Min-Max	24
4.6	Reestruturação do conjunto de dados original	25
4.7	Definição do modelo LSTM	26
4.8	Definição de parâmetros de treinamento do modelo LSTM	26
5.1	Curva de aprendizado do modelo	27
5.2	Gráfico com visão ampla da performance do Modelo	29
5.3	Gráfico com visão ampla da performance do Modelo	30
5.4	Gráfico com visão delimitada a etapa de teste	30
5.5	Gráfico com visão delimitada a etapa de teste	31
5.6	Gráfico com visão aproximada da etapa de teste	31

Lista de Tabelas

5.1	Tabela com as métricas de	performance	28
~		Perrename	- `

Lista de Abreviaturas e Siglas

ANN Redes Neurais Artificiais

BIS Banco de Compensações Internacionais

EUR/USD euro para dólares americanos

FOREX Foreign Exchange market

LSTM Long Short-Term Memory

MSE erro quadrático médio

NYSE New York Stock Exchange

RMSE raiz do erro quadrático médio

RNN Redes Neurais Recorrentes

Sumário

1	Introdução				
2	Fun	damentação teórica	14		
	2.1	Introdução ao mercado de câmbio	14		
		2.1.1 Previsão da taxa de câmbio	15		
	2.2	Redes Neurais Artificiais	15		
		2.2.1 Redes Neurais Recorrentes	16		
		2.2.2 Redes Long Short-Term Memory	17		
3	Lev	antamento bibliográfico	19		
4	Des	envolvimento	21		
	4.1	Ingestão de dados	21		
	4.2	Tranformação	21		
	4.3	Limpeza de dados	22		
	4.4	Conjuntos de treino e teste	23		
	4.5	Normalização	23		
	4.6	Reestruturação dos dados	24		
	4.7	Criação do modelo LSTM	25		
	4.8	Treinamento do modelo LSTM	26		
5	Res	ultados	27		
	5.1	Curva de aprendizado	27		
	5.2	Métricas de perfornamce	28		
	5.3	Gráficos da predição	29		
6	Con	clusões	32		
R.	ferê	acias hibliográficas	34		

Capítulo 1

Introdução

O mercado de câmbio é o mercado financeiro destinado a transações de câmbio, operações onde ocorrem a troca da moeda de um país pela moeda de outro. O mercado FOREX é o maior mercado financeiro do mundo, onde trilhões de dólares são transacionados todos os dias. É o mercado com maior liquidez dentre todos os mercados do mundo financeiro (KUMAR, 2014a).

A taxa de câmbio é reconhecida com uma conexão decisiva entre a economia interna de um país e a economia Internacional. A valorização ou desvalorização da moeda de um país, impacta os mais diversos setores de sua economia, e por isso o compreendimento/predição da taxa de câmbio é peça chave para determinar a estabilidade macroeconômica de um país e sua participação no mercado internacional (WILLIAMSON, 2008).

Contudo, o mercado financeiro tem uma natureza complexa, o que dificulta muito a predição de taxas de câmbio futuras. A irregularidade, dinamismo, e alta volatilidade, fazem com que a estruturação do problema em um modelo matemático seja quase que impossível.

De fato, os primeiros estudos apresentados na literatura indicavam que a predição da taxa de câmbio não era possível. O pessimismo acerca do tema é em muito impulsionado pela publicação de Meese e Rogoff (1983), onde o autor aponta que um passeio aleatório (do inglês, *Random Walk*) tem performance tão boa quanto qualquer outro modelo de predição. Ainda que em Díaz (2008) a asserção do passeio aleatório tenha sido colocada à prova, uma vez que o trabalho foi conduzido com comparações unicamente entre modelos lineares, os trabalhos que se sucederam na área, com propostas não lineares, também não obtiveram melhores performances.

Pela importância tanto prática quanto teórica do problema, ao longo das últimas décadas diferentes técnicas e modelagens (lineares e não lineares) foram apresentadas buscando

ultrapassar os resultados obtidos com o passeio aleatório. Bons resultados têm sido alcançados nos últimos anos, especialmente por conta de avanços advindos da computação e aprendizado de máquina, onde modelos com base no uso Redes Neurais Artificiais têm demonstrado grande potencial na descoberta dos padrões não lineares do mercado financeiro (GURESEN; KAYAKUTLU; DAIM, 2011).

Em Yu, Shouyang Wang e Lai (2007) foram enumeradas algumas características das ANNs que as tornam ferramentas valiosas às tarefas de predições. Primeiramente, ANNs contrastam com os modelos estatísticos tradicionais, no sentido de que ANNs são métodos orientados por dados e auto adaptativos. Essa é uma característica única e muito desejada na predição financeira, onde em muitos dos casos o conjunto de dados é abundante, porém pouco se sabe sobre os mecanismos ou forças que influenciam o comportamento dessas variáveis (QI; ZHANG, G. P., 2001). Segundo, ANNs podem generalizar. Terceiro, redes neurais são aproximadoras universais (HORNIK; STINCHCOMBE; WHITE, 1989). E, por último, redes neurais artificiais são modelos não-lineares (ZHANG, G.; EDDY PATUWO; Y. HU, 1998).

Atualmente, uma das classes de redes neurais artificiais que mais tem se destacado na literatura são as Redes Neurais Recorrentes. RNNs são projetadas especificamente para reconhecer padrões em séries temporais, o que torna a arquitetura adequada para a predições financeira em escopo de análise técnica. Uma vez que definimos a sequência histórica de preços, volumes, ou outros indicadores técnicos como a entrada do modelo, podemos esperar que o modelo RNN abstraia as relações sequenciais entre esses indicadores, e efetue prospectos de valores futuros com efetividade.

Ainda, um dos modelos que pode elevar os resultados de predição alcançados com a estrutura padrão de uma RNN, são as RNNs com Memória de Curto e Longo Prazo (LSTM) apresentadas por Hochreiter e Schmidhuber (1997). As LSTMs propõe algumas modificações à estrutura padrao das RNNs para que seja possível o reconhecimento de dependências de longo prazo na série temporal.

O presente trabalho tem por objetivo analisar o desempenho da rede LSTM nesse cenário tão dinâmico que é o mercado de câmbio, e comparar seus resultados com uma arquitetura RNN padrão.

Desta forma, o restante do documento é organizado da seguinte forma: No Capítulo 2 é apresentada a fundamentação teórica deste trabalho, capítulo importante para estruturar os

conceitos que são elaborados no decorrer da monografia. No Capítulo 3 o levantamento bibliográfico, apresentando estudos que trazem contextualização geral sobre ANNs ou, mais especificamente, estudos que decorrem sobre modelagens LSTMs semelhantes à apresentada nesse trabalho. No Capítulo 4 são listadas as principais etapas do desenvolvimento do modelo, desde a extração de dados até a definição de parâmetros de treinamento. No Capítulo 5, os resultados obtidos, apresentando a curva de aprendizado do modelo e outras análises de performance. Por fim, no Capítulo 6, uma conclusão do trabalho.

Capítulo 2

Fundamentação teórica

2.1 Introdução ao mercado de câmbio

O mercado de câmbio (ou do inglês, *Foreign Exchange market* – FOREX) refere-se a troca de moedas de um país, por moedas de outro país. A necessidade dessas trocas surge do turismo, a compra e venda de bens internacionalmente, ou investimentos que ocorrem além das fronteiras de um país (MELVIN; NORRBIN, 2013).

Tradicionalmente o mercado de câmbio se referia a grandes bancos de centros financeiros como, Nova York ou Londres, que, entre si, trocavam depósitos em moeda estrangeira. Contudo, desde os anos 2000 os avanços tecnológicos têm tornado o processo cambial mais acessível, e cada vez mais, investidores individuais de diferentes tamanhos têm participado deste mercado (SOBOL; SZMELTER, 2020).

O mercado de moedas estrangeiras é o maior mercado financeiro do mundo. De acordo com os dados apresentado pelo Banco de Compensações Internacionais (BIS), em abril de 2019 o mercado FOREX teve um volume de transações em torno de \$6 trilhões por dia. Um aumento significativo dos \$5.3 trilhões calculados no mesmo período em 2016 (BIS, 2019).

O mercado FOREX tem diferentes sessões de negociação em diferentes fusos horários, sendo as três maiores os pregoes da Australasia, Europa, e America do Norte. Nenhuma das sessões fica aberta durante finais de semana, e também em dias uteis não estão abertas durante o dia todo. Porém, alternando entre as janelas de diferentes sessões é possível executar transações 24h por dia, 5 dias por semana (KUMAR, 2014b).

Quanto ao ativo negociado, os principais pares de moeda são EUR/USD, USD/JPY, GBP/USD e USD/CHF. Sendo que destes, o mais popular é o de euro para dólares americanos (EUR/USD).

2.1.1 Previsão da taxa de câmbio

O mercado de câmbio é um dos mercados de dinâmica mais complexo, sendo extremamente volátil, não linear, e irregular. As flutuações nesse mercado são altamente impactadas por fatores como taxa de juros, inflação, desenvolvimento comercial, e crescimento econômico, o que torna a previsão das taxas de câmbio uma tarefa extremamente difícil. Justamente por isso, tem surgido uma crescente área de pesquisa de interesse à indústria e comunidade acadêmica, que busca a descrição de modelos de alguma forma capazes de prever o comportamento e as flutuações desse mercado (YU; WANG, S.; LAI, 2007).

As duas principais abordagens para a predição de mercados financeiros, são (i) a análise Técnica e (ii) a análise Fundamentalista. A abordagem mais fundamentalista, estuda os mais diversos fatores econômicos que podem vir a influenciar a movimentação do mercado. A análise técnica por outro lado, acredita que no preço do ativo em si já estão representados todos os fatores que o influenciam. Nesse sentido, abordagens técnicas trabalham a partir dos valores históricos do próprio mercado, na perspectiva de que nesses dados está contida a representação de futuras tendencias e comportamentos do mercado (YU; WANG, S.; LAI, 2007; MURPHY, 1999).

2.2 Redes Neurais Artificiais

O interesse no estudo de redes neurais vem de uma notável crescente durante os últimos anos. Esses esforços têm-se caracterizado de diversas formas: como a neurociência computacional, arquiteturas conexionistas, sistemas de processamento paralelo e distribuído, computação neuromórfica, e sistemas neurais artificiais. Todas essas áreas compartilham uma visão comum sobre cérebro, elas o observam como um dispositivo de computação paralela (RUMELHART; WIDROW; LEHR, 1994).

A estratégia tem sido desenvolver modelos matemáticos simplificados de sistemas cerebrais e então estudar esses modelos para entender como diferentes problemas computacionais podem ser resolvidos por esses dispositivos.

Embora os detalhes de cada proposta sejam diferentes, os modelos mais comuns têm o neurônio como a unidade básica de processamento. Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais (feed-forward). Com a associação de pesos para cada conexão, é possível armazenar o conhecimento representado no modelo e ponderar a entrada recebida por cada neurônio na rede. São as Redes Neurais Artificiais – ANN (BRAGA; LUDERMIR; CARVALHO, 2007).

Pode-se dizer que o problema de aprendizagem em uma rede neural é simplesmente o problema de encontrar um conjunto de pesos para essas conexões, os quais permitirão que a rede carregue a computação desejada.

2.2.1 Redes Neurais Recorrentes

Redes Neurais Recorrentes (RNN) são uma família de redes neurais para o processamento de dados sequenciais. Especializadas para processar uma sequência de valores, redes recorrentes podem escalar para sequencias muito maiores do que seria possível para redes sem essa especialização sequencial.

Diferente das redes unidirecionais (*feed-forward*), nas Redes Neurais Recorrentes existem conexões cíclicas entre os neurônios, o que representa um compartilhamento de parâmetros entre diferentes partes do modelo. Esse comportamento cíclico representa a influência do presente valor de uma variável, com o valor dessa mesma variável em um momento futuro (GOODFELLOW; BENGIO; COURVILLE, 2016a). A computação recorrente das RNNs pode ser representada como na Figura 2.1.

As implicações dessa diferente arquitetura são muitas. Redes neurais unidirecionais podem fazer apenas o mapeamento entre vetores de entrada e saída, enquanto RNNs a princípio podem mapear a sequência histórica de uma entrada para uma determinada saída. O ponto chave, é que essas conexões recorrentes permitem que uma "memória" das entradas passadas persista no estado interno da rede, e, desta forma, influenciem na saída da rede (GRAVES, 2012).

Contudo, conforme a profundidade e complexidade dessas redes recorrentes aumentam, o repetido compartilhamento de parâmetros e suas consequentes multiplicações matriciais impossibilitam o processo de aprendizado quando o reconhecimento de dependências de longo prazo são necessárias. O problema matemático que deriva desse cenário é conhecido na literatura como *vanishing and exploding gradient problem*, problema que foi apresentado

por trabalhos de diferentes pesquisadores (BENGIO; FRASCONI; SIMARD, 1993; BENGIO; SIMARD; FRASCONI, 1994; HOCHREITER, 1991; INFORMATIK et al., 2003).

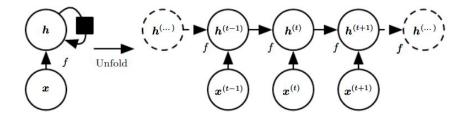


Figura 2.1: À esquerda, uma RNN sem valores de saída. Nessa rede a informação é extraída da variável de entrada \mathbf{x} e é incorporada ao estado \mathbf{h} , que é transmitido através do tempo. À direita, o desdobramento da computação recorrente dessa mesma rede em um grafo de estrutura sequencial. No grafo desdobrado cada um dos nós está associado a uma instancia de tempo t, e a função f é responsável pelo mapeamento do estado em t para o estado em t+1. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016b).

2.2.2 Redes Long Short-Term Memory

Redes Neurais Recorrentes com Memória de Curto e Longo Prazo (do inglês, *Long Short-Term Memory* – LSTM) se mostraram um modelo efetivo e escalável para diversos tipos de problemas no aprendizado de dados sequenciais, especialmente quanto àqueles que tratam da captura de dependências de longo prazo.

Apresentadas pela primeira vez em Hochreiter e Schmidhuber (1997), o objetivo era endereçar o problema relacionado ao desaparecimento de gradientes e, assim, possibilitar o armazenamento de informação durante intervalos estendidos de tempo.

A ideia central por traz da arquitetura das LSTMs é uma célula de memória capaz de preservar seu estado interno através do tempo, e unidades de portões (*gates*) não lineares que são responsáveis por controlar o fluxo de informação que entra ou sai desta célula (GREFF et al., 2015).

As LSTMs se mostraram eficientes para uma série de problemas que não eram passiveis de serem solucionados utilizando RNNs. Assim como foi apresentado em Greff et al. (2015), LSTMs têm sido utilizadas para o reconhecimento e reprodução de textos escritos à mão, geração de texto, tradução, reconhecimento de voz, predição de estrutura secundária de proteínas, analises de áudio, analises de vídeo, entre outras aplicações.

A estrutura LSTM mais comumente utilizada na literatura é descrita por Graves e Schmidhuber (2005). Como ilustrado na Figura 2.2, uma "LSTM simples" é composta por três

portões: esquecimento, entrada e saída. O fluxo de informação em cada um desses portões é definido por uma função de ativação sigmoide.

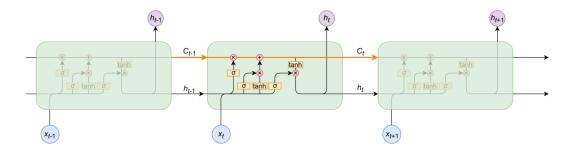


Figura 2.2: Estrutura básica de uma LSTM. Sendo ${\bf h}$ o estado interno do bloco LSTM, ${\bf c}$ a célula de memória, e ${\bf x}$ a variável de entrada do bloco.

Fonte: cutt.ly/zjAvSi2

Capítulo 3

Levantamento bibliográfico

A previsão de movimentos do mercado financeiro é um problema há muito conhecido pelo universo acadêmico e coorporativo. Durante os últimos anos foram desenvolvidos diversos artigos científicos investigando o uso da inteligência artificial, aprendizado de máquina, e automação nessa área. Entretanto são muitas as arquiteturas, técnicas, e abordagens possíveis de serem utilizadas. Nesta seção são levantadas algumas das soluções que já foram estudadas e apresentadas em trabalhos científicos.

Para a previsão de índices de ações, em Jian-Zhou Wang et al. (2011) é apresentada uma arquitetura hibrida de ANNs, a chamada *Wavelet De-noising-based back propagation*. Para diminuir o inerente ruido das séries temporais financeiras, primeiramente os dados de entrada são separados em múltiplas camadas, as chamadas *wavelets*, e só então, em seguida, são utilizadas as já conhecidas redes neurais com *back-propagation*.

Soluções envolvendo Sistemas especialistas, agentes e multi-agentes foram elaboradas em Barbosa e Belo (2008). No artigo é descrita uma estrutura para implementação de agentes inteligentes híbridos com a habilidade de negociar ativos financeiros sem a necessidade de supervisão humana.

Quanto à abordagem fundamentalista da previsão de ativos, tem-se o trabalho desenvolvido por Mittermayer (2004). Através da análise e processamento de texto, é implementado um sistema para prever tendencias nos preços de ações imediatamente após publicações de notícias.

Em Melo (2012), é feita primeiramente uma análise sucinta das principais técnicas que podem ser encontradas na literatura para modelagem e previsão de movimentos no mercado financeiro, e por final é proposta uma abordagem fora desta linha tradicional de modelos.

O artigo busca resgatar duas das técnicas anteriormente utilizadas nesta área: a inteligência aumentada e a inteligência coletiva.

Redes neurais do tipo LSTM são amplamente disseminadas no processamento de linguagem natural, onde já se provaram precisas e eficientes para a construção de modelos de linguagem e tarefas de reconhecimento de fala. Entretanto, recentemente, alguns artigos têm trabalhado com essa arquitetura também na predição de séries temporais financeiras.

Dzitkowski e Kozicki (2015) propõe um modelo LSTM para previsão multivariada da série temporal construída a partir de registros históricos de preços extraídos do repositório da *New York Stock Exchange* (NYSE).

Yıldırım, Toroslu e Fiore (2021) fazem predições do direcionamento dos movimentos no mercado FOREX. Para isso são trabalhados em paralelos dois conjuntos de dados e dois modelos LSTMs. Um modelo hibrido que endereça tanto a análise técnica quanto a análise fundamentalista do mercado.

Muitas das técnicas utilizadas nos trabalhos citados acima auxiliariam no desenvolvimento da solução apresentada neste projeto, em especial os últimos dois artigos apresentados, que assim como o presente trabalho, desenvolveram uma solução a partir de modelos LSTM e propuseram comparativos dos resultados obtidos com outras estruturas neurais.

Capítulo 4

Desenvolvimento

4.1 Ingestão de dados

O conjunto de dados utilizado nesse trabalho consiste de valores reais do mercado FOREX, todos os registros foram extraídos do repositório online oferecido pelo Dukascopy Swiss Bank Group. O intervalo selecionado foi do dia 07 ao dia 12 de abril de 2021, e contém 8.640 registros que acompanham as oscilações do par de moedas EUR/USD de minuto em minuto, o que permite uma análise detalhada e uma observação quase que instantânea das possíveis variações de preço.

Como ilustrado na Figura 4.1, os dados extraídos da Dukascopy Bank formam uma tabela com as seguintes colunas: o preço de fechamento, preço de abertura, preço mínimo, preço máximo, volume, e o horário ao que cada registro se refere. Todos esses valores, dizem respeito unicamente aos preços de oferta (bid prices).

4.2 Tranformação

Aplica-se as transformações necessárias aos dados brutos. Primeiro, como é necessário apenas o acompanhamento das variáveis de tempo e preço de fechamento, as colunas fora do escopo são removidas. Segundo, para facilitar a visualização e utilização da variável foi aplicado um tratamento na tipagem do dado na coluna de tempo. Terceiro, foi redefinido o índice da tabela, substituindo a ordenação numérica inicial pela própria data de cada registro. Ao final, temos uma outra estrutura de dados, como observa-se na Figura 4.2.

	Gmt time	Open	High	Low	Close	Volume
0	07.03.2021 00:00:00.000	1.19128	1.19128	1.19128	1.19128	0.0
1	07.03.2021 00:01:00.000	1.19128	1.19128	1.19128	1.19128	0.0
2	07.03.2021 00:02:00.000	1.19128	1.19128	1.19128	1.19128	0.0
3	07.03.2021 00:03:00.000	1.19128	1.19128	1.19128	1.19128	0.0
4	07.03.2021 00:04:00.000	1.19128	1.19128	1.19128	1.19128	0.0
		***			***	
8635	12.03.2021 23:55:00.000	1.19476	1.19476	1,19476	1.19476	0.0
8636	12.03.2021 23:56:00.000	1.19476	1.19476	1.19476	1.19476	0.0
8637	12.03.2021 23:57:00.000	1.19476	1.19476	1.19476	1.19476	0.0
8638	12.03.2021 23:58:00.000	1.19476	1.19476	1.19476	1.19476	0.0
8639	12.03.2021 23:59:00.000	1.19476	1.19476	1.19476	1.19476	0.0

8640 rows × 6 columns

Figura 4.1: Reestruturação do conjunto de dados original \mathbf{d} em múltiplas amostras, cada uma com suas respectivas variáveis de entrada \mathbf{x} e variáveis de saída \mathbf{y} .

Fonte: Elaborado pelo autor.

	bidclose
datetin	ie
2021-03-07 00:00:0	1.19128
2021-03-07 00:01:0	1.19128
2021-03-07 00:02:0	1.19128
2021-03-07 00:03:0	00 1.19128
2021-03-07 00:04:0	00 1.19128
2021-03-12 23:55:0	00 1.19476
2021-03-12 23:56:0	00 1.19476
2021-03-12 23:57:0	00 1.19476
2021-03-12 23:58:0	1.19476
	0 1.19476

Figura 4.2: Conjunto de inicial dados após as transformações necessárias. Fonte: Elaborado pelo autor.

4.3 Limpeza de dados

Além do processo de transformação, também é necessário fazer uma limpeza dos dados.Como o mercado FOREX abre às 9PM/10PM GMT de domingo e fecha as 9PM/10PM na sexta-feira, é preciso fazer a exclusão de períodos onde o mercado não estava aberto. Com a redução, a base de dados passa a ter 7.200 registros, Figura 4.3.

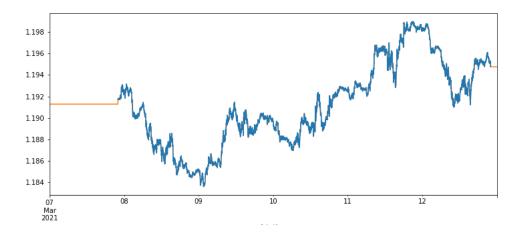


Figura 4.3: Em **laranja**, os períodos que foram retirados da base de dados. Fonte: Elaborado pelo autor.

4.4 Conjuntos de treino e teste

A separação dos dados entre conjuntos diferentes para treino e teste é um procedimento simples, e importante para a validação do modelo de *machine learning* em problemas de predição de séries temporais.

Nesta etapa, o conjunto de dados é subdividido em dois, um para treino, e um para teste. O subconjunto de treino, utilizado para ajustar o modelo de aprendizado de máquina, e o subconjunto de teste, utilizado para avaliar os resultados preditos. O objetivo é estimar a qualidade de predição do modelo em cenários que não fizeram parte do treinamento, ou seja, estimar a qualidade real de predição.

Para modelos de aprendizado de máquina essa separação tem uma proporção padrão de 80/20 ou 70/30, como apresentado em Dangeti (2017). A maior parte do conjunto é utilizada para o treinamento, e o restante utilizado para validação. Essa separação é ainda mais eficiente quando o volume de dados é grande, e ambos subconjuntos têm uma boa representatividade do cenário original.

Aplica-se então uma taxa de separação de 0.8. Para treino, 80% dos dados, e para teste, 20%. A linha que divide os conjuntos de treino e teste é ilustrada em Figura 4.4.

4.5 Normalização

Com o pré-processamento das entradas e saídas da rede neural, pode-se alcançar um treinamento mais eficiente. Funções de processamento das variáveis de entrada transformam

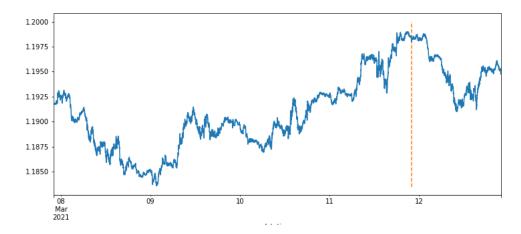


Figura 4.4: Separação dos dados em conjunto de treino e teste. À esquerda da faixa **laranja**, o conjunto de treino, à direita o conjunto de teste.

Fonte: Elaborado pelo autor.

os dados brutos em um melhor formato para rede neural. O processo de normalização para os dados brutos tem grandes efeitos na preparação do dado para se tornar adequado ao treinamento (JAYALAKSHMI; A., 2011).

Existem muitos tipos de normalização, para este trabalho em específico a arquitetura proposta utiliza a normalização de dados Min-Max com intervalo em [0, 1], como ilustrado na Figura 4.5.

```
scaler=MinMaxScaler(feature_range=(0,1))

train_data=scaler.fit_transform(np.array(train_data).reshape(-1,1))
test_data=scaler.transform(np.array(test_data).reshape(-1,1))|
```

Figura 4.5: Fluxo de atividades para normalização Min-Max Fonte: Elaborado pelo autor.

A função de normalização foi ajustada de acordo com os dados disponíveis no subconjunto de treinamento. Uma vez a escala definida, ambos subconjuntos tiveram seus dados redimensionados.

4.6 Reestruturação dos dados

Séries temporais univariadas são uma classe de especificações onde procura-se modelar e predizer variáveis utilizando-se apenas de informações contidas nos valores passados da própria variável. Modelos de séries temporais univariadas são normalmente ateóricos,

implicando que sua construção e uso não são embasadas em nenhum pressuposto teórico do comportamento da variável - são os chamados 'Modelos de séries temporais puras' (BROOKS, 2008).

O presente trabalho propõe o desenvolvimento de um modelo de rede neural LSTM para séries temporais univariadas, de passo único. Espera-se que, com base nas observações históricas da variável (*bidclose*), o modelo seja capaz de realizar a predição do próximo valor nessa sequência.

A partir do conjunto inicial de dados, é feita uma reestruturação da sequência original única em múltiplas amostras, cada uma com suas respectivas variáveis de entrada e saída. De modo que, temos sequências de observações históricas como variáveis de entrada, e o próximo valor observado nessa sequência como variáveis de saída. A transformação ocorre como ilustrado na Figura 4.6. Cada amostra com 100 observações históricas e 1 valor a ser predito.

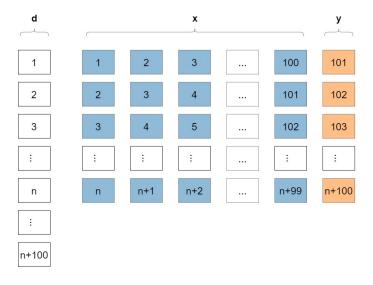


Figura 4.6: Reestruturação do conjunto de dados original \mathbf{d} em múltiplas amostras, cada uma com suas respectivas variáveis de entrada \mathbf{x} e variáveis de saída \mathbf{y} . Sendo (\mathbf{n} +100) o número total de registros no conjunto de dados original.

Fonte: Elaborado pelo autor.

4.7 Criação do modelo LSTM

A arquitetura proposta é uma *Stacked* LSTM, ou seja, um modelo composto de múltiplas camadas LSTM empilhadas. São implementadas três camadas ocultas do tipo LSTM com 15 neurônios cada uma, e uma camada Dense ao final. A rede foi treinada utilizando o método de otimização *Adam*, e a função de perda *Mean Squared Error*.

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LSTM

model = Sequential()
model.add(LSTM(15,return_sequences=True, input_shape=(x_train.shape[1],x_train.shape[2])))
model.add(LSTM(15,return_sequences=True))
model.add(LSTM(15))
model.add(Dense(x_train.shape[2]))
model.add(Dense(x_train.shape[2]))
model.compile(loss='mean_squared_error',optimizer='adam')
```

Figura 4.7: Definição do modelo LSTM Fonte: Elaborado pelo autor.

4.8 Treinamento do modelo LSTM

Ainda, antes de iniciar de fato o treinamento, alguns hiperparâmetros são definidos para melhor regular esta etapa.

Como ilustrado na Figura 4.8, os seguintes parâmetros são definidos ao modelo: *epochs* como o número de vezes que o modelo irá repassar o treinamento sobre todas as amostras, *batch-size* como número de amostras que devem ser analisadas pelo modelo antes que ele possa atualizar seu estado interno, e *validation-split* referente à fração do conjunto de treino que será utilizada como dados de validação.

Figura 4.8: Definição do modelo LSTM Fonte: Elaborado pelo autor.

Capítulo 5

Resultados

Após a preparação do conjunto de dados e definição da arquitetura do modelo, os seguintes resultados foram obtidos.

5.1 Curva de aprendizado

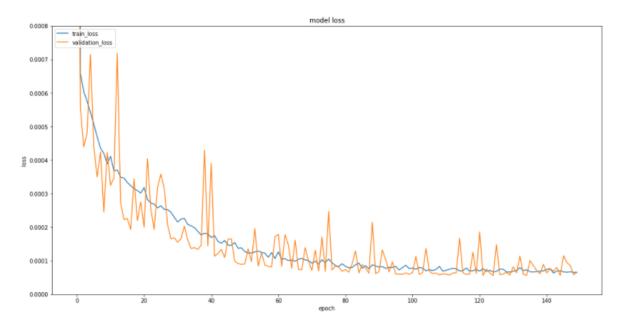


Figura 5.1: Curva de aprendizado do modelo Fonte: Elaborado pelo autor.

Na Figura 5.1, a curva de aprendizado da rede LSTM, representação matemática do desenvolvimento que ocorre durante o treinamento, onde de época em época os parâmetros internos da rede são atualizados. Para ambos os contextos, tanto de treino quanto de

validação, busca-se a minimização da função de perda, métrica que acompanha a diferença entre os valores preditos pelo modelo e os valores observados.

Da dinâmica e comportamento apresentados, tem-se um bom diagnostico do modelo. Ambas as curvas têm uma boa evolução ao longo das épocas e acabam por encontrar um ponto de estabilidade, o que representa que de fato houve um processo de aprendizado consistente. Por vezes, a curva de aprendizado poderia evidenciar cenários de *overfitting* ou *underfitting*, não foi o caso. Ainda, a pequena diferença entre os resultados em ambiente de treino e validação evidenciam uma boa capacidade de generalização do modelo.

5.2 Métricas de perfornamce

Para de fato mensurar a capacidade de predição do modelo, é necessária uma análise com dados novos, em ambiente de teste. As métricas utilizadas nessa etapa são as chamadas métricas de performance. São apresentadas na Tabela 5.1, o erro quadrático médio (MSE), e também, a raiz do erro quadrático médio (RMSE).

Para agregar na análise, também são demonstrados os resultados de uma rede neural recorrente clássica. Vale ressaltar que ambos modelos, LSTM e RNN, foram modelados e treinados em mesmo contexto e plataforma, a fim de garantir a integridade dos resultados.

Tr 1 1 F 4 Tr 1 1	/ 1	C	11 TOTAL D	TATE
Tabela 5.1: Tabela com	as metricas de	performance pa	ra os modelo LSTM e R	ININ

Métrica	Modelo	Treino	Teste
MSE	StackedLSTM	1.488111e-08	6.317745e-08
MSE	StackedRNN	2.236754e-08	7.989822e-07
RMSE	StackedLSTM	1.219881e-04	2.513512e-04
RMSE	StackedRNN	1.495578e-04	8.938580e-04

A estruturação do modelo LSTM proposta e a validação da arquitetura foram bem sucedidas. A rede teve bons desempenhos durante o aprendizado, minimizando o RMSE a 1.21e-04, e também em fase de teste, onde o RMSE foi de 2.51e-04. São boas métricas de performance, uma vez validada no mesmo cenário, a RNN padrão obteve RMSE mais elevados, com 1.49e-04 em treino, e 8.93e-04 em teste. Comparada ao modelo RNN, a rede LSTM apresenta um redução de 18.4% na taxa de erro em treino, e 71.8% na taxa de erro em teste.

5.3 Gráficos da predição

Observando Figura 5.2, Figura 5.4, e Figura 5.6 podemos ver de forma gráfica o resultado das predições feitas pelo modelo LSTM. Quanto mais próximas as linhas de predição estão dos dados reais, melhor a performance preditiva. Os resultados observados demonstram boa capacidade do modelo, tanto em etapa de treino, quanto em etapa de teste. Para melhor comparação, em paralelo, são apresentados os resultados da RNN clássica, Figura 5.3, Figura 5.5.

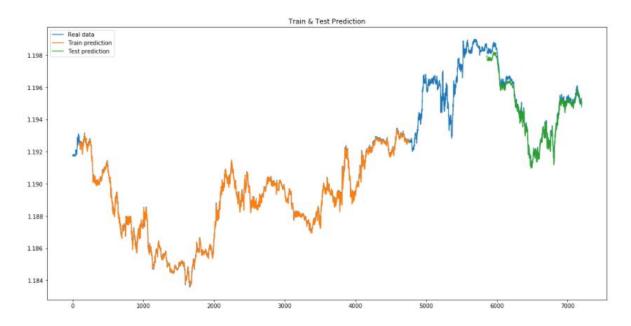


Figura 5.2: Gráfico com visão ampla da performance do modelo LSTM. Valores de predição de treino e teste nas cores **laranja** e **verde** respectivamente, e em **azul** os dados reais.

Fonte: Elaborado pelo autor.

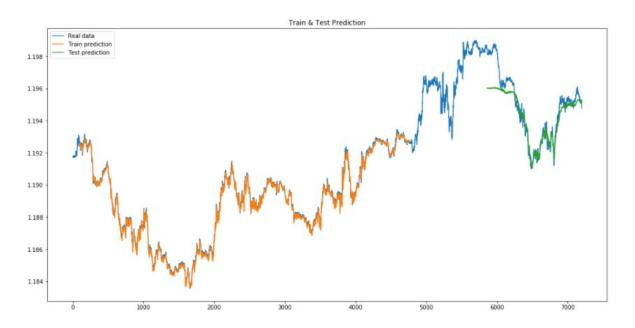


Figura 5.3: Gráfico com visão ampla da performance do modelo RNN. Valores de predição de treino e teste nas cores **laranja** e **verde** respectivamente, e em **azul** os dados reais.

Fonte: Elaborado pelo autor.

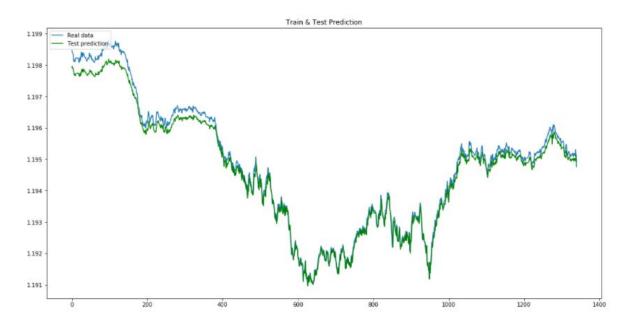


Figura 5.4: Gráfico com visão delimitada à performance em conjunto de teste. Em **verde** os valores preditos pelo modelo LSTM, e em **azul** os dados reais.

Fonte: Elaborado pelo autor.

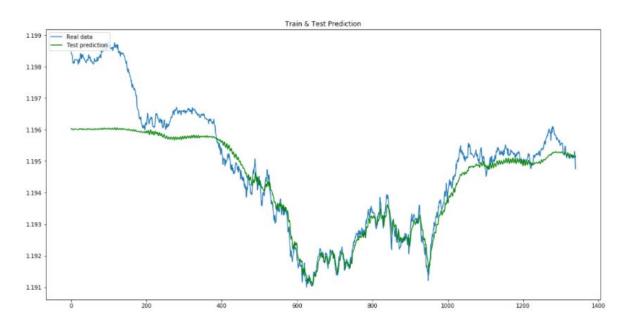


Figura 5.5: Gráfico com visão delimitada à performance em conjunto de teste. Em **verde** os valores preditos pelo modelo RNN, e em **azul** os dados reais.

Fonte: Elaborado pelo autor.

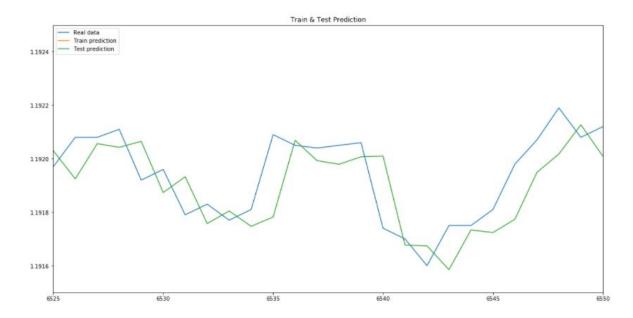


Figura 5.6: Gráfico com visão aproximada em período de 25 minutos de predição. Em **verde** os valores preditos pelo modelo LSTM, e em **azul** os dados reais.

Fonte: Elaborado pelo autor.

Capítulo 6

Conclusões

A estruturação do modelo LSTM proposta e a validação da arquitetura foram bem sucedidas. O modelo apresentou resultados com taxa de erro menores que as obtidas com a implementação de uma arquitetura RNN padrão, principalmente quando validado com dados de teste, onde observou-se uma redução de 71.8% no RMSE.

Dos resultados finais e a diferença de performance entre os dois modelos, entende-se que a rede LSTM foi capaz de abstrair dependências de longo prazo na série temporal, o que é um resultado surpreendente, uma vez que, como foi discutido durante o decorrer do trabalho, o mercado de cambio é extremamente complexo e dinâmico.

De modo geral, a pesquisa apresentada se mostrou efetiva e contribuiu com a literatura científica em torno do tema apresentado, no sentido que se alinha com outros trabalhos e contribui com a crescente evidenciação do valor/poder do uso de redes neurais aplicadas às mais diversas áreas.

Entretanto, algumas alternativas não abordadas poderiam ter enriquecido ainda mais o projeto. O uso de diferentes modelos para *benchmarking*, teria agregado muito à analise final dos resultados, inclusive uma comparação com o próprio modelo de passeio aleatório. Como foi comentado durante o trabalho, este modelo sempre serviu de base para as validações de modelos preditivos no mercado financeiro, seria interessante ter trazido comparações nesse sentido também.

Ainda, algumas alterações simples também poderiam ter sido feitas e trariam grandes benefícios. Se possível, a afinação de hiper parâmetros (*hyperparameter tuning*), o que poderia aumentar significante a performance do modelo.

Com certeza o desenvolvimento da presente pesquisa iluminou a possibilidade de estudo de diversas outras aplicações, que inclusive poderiam ser abordadas em futuros trabalhos (ainda com relação a predições financeiras). Para citar algumas delas, aplicações com modelos multivariados, onde múltiplos atributos seriam considerados como entrada do modelo, o que ampliaria o leque de possíveis relações a serem abstraídas pelo modelo. Modelos multistep também valeriam a análise, uma vez que a predição não seria feita apenas sobre um valor único, mas sobre uma sequência de valores futuros, o que expandiria a utilidade prática do modelo.

Referências bibliográficas

BARBOSA, R. P.; BELO, O. Algorithmic Trading Using Intelligent Agents. In: ARABNIA, H. R.; MUN, Y. (Ed.). Proceedings of the 2008 International Conference on Artificial Intelligence, ICAI 2008, July 14-17, 2008, Las Vegas, Nevada, USA, 2 Volumes (includes the 2008 International Conference on Machine Learning; Models, Technologies and Applications). [S.l.]: CSREA Press, 2008. p. 136–142.

BENGIO, Y.; FRASCONI, P.; SIMARD, P. Y. The problem of learning long-term dependencies in recurrent networks. **IEEE International Conference on Neural Networks**, 1183–1188 vol.3, 1993.

BENGIO, Y.; SIMARD, P. Y.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE transactions on neural networks**, v. 5 2, p. 157–66, 1994.

BIS. Triennial Central Bank Survey, Foreign exchange turnover in April 2019. [S.l.], set. 2019.

BRAGA, A. d. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. d. L. F. Redes neurais artificiais: teoria e aplicações. [S.l.]: LTC, 2007.

BROOKS, C. Univariate time series modelling and forecasting. In: INTRODUCTORY Econometrics for Finance. 2. ed. [S.l.]: Cambridge University Press, 2008. p. 206–264. DOI: 10.1017/CBO9780511841644.006.

DANGETI, P. **Statistics for Machine Learning**: Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R. [S.l.: s.n.], jul. 2017. v. 1, p. 34–35.

DÍAZ, M. Exchange rates forecasting: Local or global methods? **Applied Economics**, v. 40, p. 1969–1984, ago. 2008. DOI: 10.1080/00036840600905308.

DZITKOWSKI, K.; KOZICKI, T. **High Frequency Trading Price Prediction using LSTM Recursive Neural Networks**. 2015. Warsaw University of Technology. Disponível em: <ht tps://github.com/dzitkowskik/StockPredictionRNN/blob/master/docs/project.pdf>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. ht tp://www.deeplearningbook.org.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. p. 376. http://www.deeplearningbook.org.

GRAVES, A. **Supervised Sequence Labelling with Recurrent Neural Networks**. [S.l.: s.n.], jan. 2012. v. 385. ISBN 978-3-642-24796-5. DOI: 10.1007/978-3-642-24797-2.

GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. **Neural Networks**, v. 18, n. 5, p. 602-610, 2005. IJCNN 2005. ISSN 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2005.06.042. Disponível em: https://www.sciencedirect.com/science/article/pii/S0893608005001206.

GREFF, K. et al. LSTM: A Search Space Odyssey. **CoRR**, abs/1503.04069, 2015. arXiv: 1503.04069. Disponível em: http://arxiv.org/abs/1503.04069.

GURESEN, E.; KAYAKUTLU, G.; DAIM, T. U. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, v. 38, n. 8, p. 10389–10397, 2011. ISSN 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2011.02.068. Disponível em: https://www.sciencedirect.com/science/article/pii/S0957417411002740>.

HOCHREITER, S. Untersuchungen zu dynamischen neuronalen Netzen, abr. 1991.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-term Memory. **Neural computation**, v. 9, p. 1735–80, dez. 1997. DOI: 10.1162/neco.1997.9.8.1735.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, v. 2, n. 5, p. 359–366, 1989. ISSN 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8. Disponível em: https://www.sciencedirect.com/science/article/pii/0893608089900208.

INFORMATIK, F.; BENGIO, Y.; FRASCONI, P.; SCHMIDHUBER, J. Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies. A Field Guide to Dynamical Recurrent Neural Networks, mar. 2003.

JAYALAKSHMI, T.; A., S. Statistical normalization and back propagation for classification. **International Journal Computer Theory Engineering (IJCTE)**, v. 3, p. 89–93, jan. 2011.

KUMAR, R. Chapter 5 - Stock Markets, Derivatives Markets, and Foreign Exchange Markets. In: KUMAR, R. (Ed.). **Strategies of Banks and Other Financial Institutions**. San Diego: Academic Press, 2014. p. 125–164. ISBN 978-0-12-416997-5. DOI: https://doi.org/10.1016/B978-0-12-416997-5.00005-1. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780124169975000051.

KUMAR, R. Chapter 5 - Stock Markets, Derivatives Markets, and Foreign Exchange Markets. In: KUMAR, R. (Ed.). **Strategies of Banks and Other Financial Institutions**. San Diego: Academic Press, 2014. p. 125–164. ISBN 978-0-12-416997-5. DOI: https://doi.org/10.1016/B978-0-12-416997-5.00005-1. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780124169975000051.

MEESE, R. A.; ROGOFF, K. Empirical exchange rate models of the seventies: Do they fit out of sample? **Journal of International Economics**, v. 14, n. 1, p. 3–24, 1983. ISSN 0022-1996. DOI: https://doi.org/10.1016/0022-1996(83)90017-X. Disponível em: https://www.sciencedirect.com/science/article/pii/002219968390017X.

MELO, B. Considerações cognitivas nas técnicas de previsão no mercado financeiro. 2012. Universidade Estadual de Campinas. Disponível em: https://www.dca.fee.unicamp.br/~gudwin/courses/IA889/2011/IA889-09.pdf.

MELVIN, M.; NORRBIN, S. C. Chapter 1 - The Foreign Exchange Market. In: MELVIN, M.; NORRBIN, S. C. (Ed.). **International Money and Finance (Eighth Edition)**. Eighth Edition. [S.l.]: Academic Press, 2013. p. 3-24. ISBN 978-0-12-385247-2. DOI: https://doi.org/10.1016/B978-0-12-385247-2.00001-9. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780123852472000019.

MITTERMAYER, M.-A. Forecasting Intraday stock price trends with text mining techniques. In: 37TH Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the. [S.l.: s.n.], 2004. p. 10–. DOI: 10.1109/HICSS.2004.1265201.

MURPHY, J. J. Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. [S.l.: s.n.], jan. 1999.

QI, M.; ZHANG, G. P. An investigation of model selection criteria for neural network time series forecasting. **European Journal of Operational Research**, v. 132, n. 3, p. 666-680, 2001. ISSN 0377-2217. DOI: https://doi.org/10.1016/S0377-2217(00)00171-5. Disponível em: https://www.sciencedirect.com/science/article/pii/S0377221700001715.

RUMELHART, D. E.; WIDROW, B.; LEHR, M. A. The basic ideas in neural networks. **Commun. ACM**, v. 37, p. 87–92, 1994.

SOBOL, I.; SZMELTER, M. Retail investors in the foreign exchange market. **Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu**, v. 64, nov. 2020. DOI: 10 . 15611 / pn . 2020.6.13.

WANG, J.-Z.; WANG, J.-J.; ZHANG, Z.-G.; GUO, S.-P. Forecasting stock indices with back propagation neural network. **Expert Systems with Applications**, v. 38, n. 11, p. 14346-14355, 2011. ISSN 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2011.04.222. Disponível em: https://www.sciencedirect.com/science/article/pii/S0957417411007494.

WILLIAMSON, J. Exchange Rate Economics. Peterson Institute for International Economics Working Paper No. 08-3, fev. 2008. DOI: http://dx.doi.org/10.2139/ssrn.1091612. Disponível em: https://ssrn.com/abstract=1091612.

YILDIRIM, D. C.; TOROSLU, I. H.; FIORE, U. Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. **Financial Innovation**, v. 7, n. 1, p. 1, jan. 2021. ISSN 2199-4730. DOI: 10.1186/s40854-020-00220-2.

YU, L.; WANG, S.; LAI, K. K. Foreign-Exchange-Rate Forecasting With Artificial Neural Networks. [S.l.: s.n.], jan. 2007. v. 107. ISBN 978-0-387-71719-7. DOI: 10.1007/978-0-387-71720-3.

ZHANG, G.; EDDY PATUWO, B.; Y. HU, M. Forecasting with artificial neural networks:: The state of the art. **International Journal of Forecasting**, v. 14, n. 1, p. 35–62, 1998. ISSN 0169-2070. DOI: https://doi.org/10.1016/S0169-2070(97)00044-7. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S016920709700044
7>.