

Redes Neurais Aplicadas Na Predição do Preço da Criptomoeda Ether: MLP vs bi-LSTM

Francisco Leonésio Carneiro Duarte¹, Náthalee Cavalcanti de Almeida Lima¹

¹Departamento de Engenharias e Tecnologia – Universidade Federal Rural do Semi-árido (UFERSA)

leonesiorf@hotmail.com, nathalee.almeida@ufersa.edu.br

Abstract. *The Ether has been gaining a lot of space in the world of virtual currencies, already being the second largest in the industry, in relation to capitalization. Given the attractiveness of this currency for investments, then the need to create models of prediction for the same. An efficient way to create such models is the use of artificial neural networks. In this paper, this work aims to compare the performance of the Multilayer Perceptrons (MLP) and Bidirectional Long Short Term Memory Units (bi-LSTM) networks in the prediction of the Ether time series. It was verified that the bi-LSTM presented a performance somewhat better than the MLP in the accomplishment of such task. Matlab software was used to perform the simulations.*

Resumo. *O Ether vem ganhando bastante espaço no mundo das moedas virtuais, já sendo a segunda maior do ramo, com relação à capitalização. Tendo em vista a atratividade desta moeda para investimentos, surge então a necessidade de criar modelos de predição para a mesma. Uma maneira eficiente de criar tais modelos é o emprego das redes neurais artificiais. Desta forma, este trabalho tem por objetivo comparar o desempenho das redes Multilayer Perceptrons (MLP) e Bidirectional Long Short Term Memory Units (bi-LSTM) na predição da série temporal do Ether. Verificou-se que a bi-LSTM apresentou um desempenho um pouco melhor que a MLP na realização de tal tarefa. Foi utilizado o software Matlab nas simulações.*

1. Introdução

A humanidade constantemente vem evoluindo suas relações econômicas. Elas iniciaram por meio do escambo, evoluíram para o uso de moedas e cédulas impressas, sendo que mais recentemente, a moeda perdeu sua forma física tornando-se virtual.

Ulrich (2014) apud Almeida e Cunha (2018) ressalta que as moedas digitais, surgiram durante a crise financeira de 2008 com o objetivo de apresentar uma possível solução para a instabilidade do sistema financeiro mundial. Neste sentido, as novidades trazidas pelas moedas virtuais instigaram a curiosidade dos investidores, trazendo assim um crescimento no mercado de criptomoedas nos últimos anos.

A primeira criptomoeda descentralizada denominada Bitcoin foi criada em 2009 por Satoshi Nakamoto, com o objetivo de permitir a transferência de dinheiro sem o intermédio de instituições financeiras ao mesmo tempo em que se garantia alta segurança nas transações.

De acordo com Bitcoin (2018), estas funcionalidades do Bitcoin são implementadas por meio do Blockchain, o qual constitui uma grande rede de computadores *peer-to-peer* responsável por validar as transações. A segurança do sistema é garantida, na medida em que uma tentativa de modificação maliciosa de um registro é rapidamente identificada pelo resto da rede, a qual descarta esta informação maliciosa. Desta forma, para que uma modificação maliciosa tenha sucesso é necessário um grande poder computacional, o qual é inviável de se conseguir.

Inspirando-se no Bitcoin, surgiram diversas outras criptomoedas denominadas de altcoins, cujo objetivo era superar algumas das limitações do Bitcoin, bem como trazer novas funcionalidades. Dentre as altcoins é possível destacar o Ether.

De acordo com Ethereum (2019), o Ether é uma moeda virtual utilizada para rodar os contratos inteligentes, pagar aos mineradores e também rodar os serviços dos aplicativos desenvolvidos para a plataforma Ethereum. Esta moeda teve a sua pré-venda realizada em 2014, de modo a arrecadar fundos para o lançamento da plataforma Ethereum. Uma das grandes diferenças do Ether para o Bitcoin, diz respeito ao fato de que o Ether não possui limite máximo de moedas geradas.

Coinmarketcap (2019) destaca que o Ether é a segunda maior criptomoeda do mercado, possuindo uma capitalização de mais de 53 bilhões de reais. Verifica-se, portanto, a importância desta moeda bem como sua atratividade para investimento, trazendo assim, a necessidade de empregar modelos de predição do preço do Ether de forma a orientar os investimentos.

Sabendo-se que o preço do Ether é muito variável, e possui comportamento complexo e não linear, evidencia-se a dificuldade em realizar a tarefa de predição do mesmo. Observou-se em trabalhos como os de Soares e Teixeira (2018) e de Mallqui e Fernandes (2017) que as redes neurais artificiais, constituem uma ferramenta de boa qualidade para realizar este tipo de tarefa. Estas redes podem ser compreendidas como uma forma de simular o cérebro humano, trazendo assim uma capacidade de aprendizado para o sistema computacional.

Tendo em vista esta possibilidade de emprego de técnica de redes neurais para resolução de problemas relacionados a predição de séries temporais, este trabalho tem por objetivo comparar o desempenho da rede neural MLP com o da Rede Neural bi-LSTM, sendo ambas aplicadas na predição do preço da criptomoeda Ether.

O Ether foi escolhido como alvo da pesquisa, por ser a segunda maior criptomoeda. Além disso, mesmo tendo tanta importância ele possui poucos trabalhos na literatura relacionados ao desenvolvimento de modelos preditivos para o mesmo. Por sua vez, o Bitcoin que é a maior criptomoeda, já possui diversos estudos relacionados a esta temática.

O presente trabalho é dividido em 6 Seções. Na Seção 2 tem-se a Fundamentação Teórica, onde será discutido os conceitos da rede neural MLP com entradas atrasadas no tempo, da célula LSTM, bem como da rede bi-LSTM. A Seção 3 apresenta a metodologia da pesquisa. Para a quarta Seção será apresentado os resultados da pesquisa, sendo que a quinta Seção apresenta as conclusões obtidas neste trabalho. Na sexta Seção as referências utilizadas neste trabalho são listadas.

2. Fundamentação Teórica

2.1. Rede MLP com entradas atrasadas no tempo

Silva, Spatti e Flauzino (2010) ressaltam que as redes MLP com entradas atrasadas, pertencem a arquitetura *feedforward* de múltiplas camadas. A predição é realizada a partir de um instante t , com base nos valores anteriores da série, de acordo com a equação abaixo.

$$x(t) = f(x(t-1), x(t-2), \dots, x(t-n_p)) \quad (1)$$

Onde n_p é a quantidade de medidas passadas necessárias para estimação do valor. A rede apresenta topologia apresentada na Figura 1. Verifica-se por meio da figura que a rede possui atrasos temporais na camada de entrada. Esta linha de atrasos corresponde a uma memória, de forma a garantir que amostras anteriores sejam sempre consideradas pela rede na estimação do valor futuro.

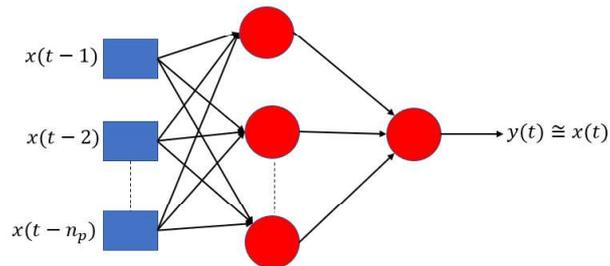


Figura 1. Topologia Para a MLP com Entradas Atrasadas no Tempo

Fonte: Elaborado Pelos Autores

Durante o processo de treinamento a rede deve minimizar o erro entre a saída real e a saída estimada, sendo este erro dado pela expressão: $E = x(t) - y(t)$, para o intervalo $(n_p + 1) \leq t \leq N$, onde N é a quantidade total de amostras da série temporal. O processo de treinamento e aprendizado da rede com atrasos é similar ao da MLP convencional, sendo necessário ter o cuidado de organizar os dados de entrada e das saídas esperadas, considerando os atrasos aplicados no processo.

2.2. LSTM

Segundo MathWorks (2019), a rede LSTM é um tipo de rede neural recorrente capaz de aprender, processar e classificar dados sequenciais. Além disso, ela é aplicada em várias áreas como: análise de sentimentos, modelagem de linguagem, reconhecimento da fala, bem como em análise de vídeo.

As redes neurais treinadas por meio de retropropagação apresentam problemas relacionados ao gradiente de fuga, fazendo com que a rede consiga capturar apenas dependências de curto prazo. O problema do gradiente de fuga, ocorre devido a diminuição dos gradientes associados aos pesos, até que os mesmos se tornem zero ou fiquem próximos a este valor. Quando isto ocorre, a multiplicação do gradiente pela taxa de aprendizado faz com que os pesos se alterem lentamente, o que dificulta o treinamento.

Para resolver este tipo de problema a LSTM foi criada. Ela supera o problema do gradiente de fuga por meio de portões que têm o papel de reter informações importantes e esquecer informações que não são importantes.

Na Figura 2, tem-se o esquema da arquitetura de um bloco LSTM. Percebe-se nesta figura a existência de uma célula de memória, um portão de entrada, um portão de saída, um portão de esquecimento, além do estado escondido.

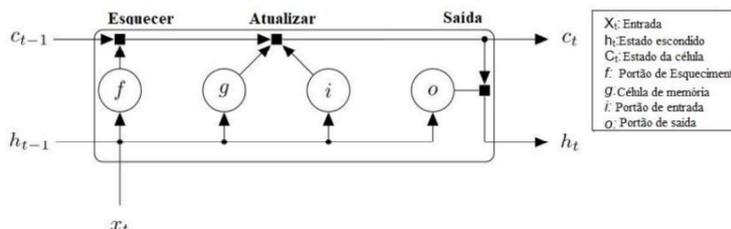


Figura 2. Bloco LSTM

Fonte:

https://www.mathworks.com/content/mathworks/www/en/discovery/lstm/jcr:content/mainParsys/image_28882819.adapt.full.medium.jpg/1548156994542.jpg

Cada portão possui pesos e um bias associado, sendo estas grandezas responsáveis por controlar cada um dos portões. A alteração dos pesos e bias do portão de entrada implica em quanto um novo valor pode fluir para dentro da célula. A alteração destes fatores para o portão de esquecimento controla até que ponto um determinado valor permanecerá no bloco. Por sua vez, a alteração destes fatores para o portão de saída controlará até que ponto o valor armazenado na célula poderá ser usado para calcular o sinal de ativação da saída do bloco LSTM.

2.3. bi-LSTM

De acordo com Sharfuddin, Tihami e Islam (2018), nas redes neurais LSTM unidirecionais, as informações fluem apenas de trás para frente, já nas LSTM bidirecionais (bi-LSTM) a informação flui de trás para frente, bem como de frente para trás. Esta bidirecionalidade dos dados faz com que este tipo de rede possa compreender melhor o contexto.

Na implementação deste tipo de rede, são utilizadas duas camadas de LSTM, conforme pode ser visualizado na Figura 3. Uma camada é responsável pelos estados passados, enquanto a outra camada é responsável pelos estados futuros.

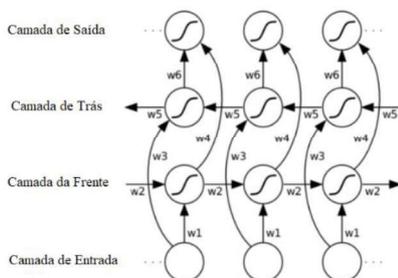


Figura 3. Modelo da rede bi-LSTM

Fonte: SHARFUDDIN, TIHAMI e ISLAM (2018, p. 2)

3. Metodologia

Para treinamento e teste das redes bi-LSTM e MLP, utilizou-se como base a rotina do Matlab desenvolvida por Nejatian (2019). Os dados para treinamento e teste da rede consistiu do preço diário do Ether em dólares, extraído por meio de um arquivo com extensão csv fornecido por Etherscan (2019).

De toda a série temporal coletada no arquivo csv, foi selecionado o período de tempo de 30/07/2015 até 10/12/2016 para realização de treinamento e teste, totalizando assim 500 amostras da série. O conjunto de dados foi dividido em 80% para treinamento e 20% para teste. Após a realização de testes de execução das redes definiu-se a utilização de um n_p de 30. Todos os dados foram normalizados com base no desvio padrão.

Na Figura 4, tem-se um esquema da rede MLP utilizada, a mesma foi treinada por meio do algoritmo *Bayesian Regularization Backpropagation*. Foi adotada a quantidade de 100 épocas, como a condição de parada para o treinamento. Este valor foi definido com base na observação da curva de treinamento da rede, bem como com a realização de comparação de desempenho de previsões para valores diferentes de épocas de treinamento.

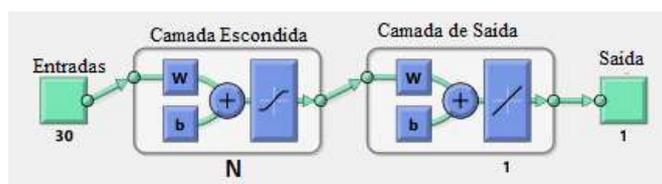


Figura 4. Rede MLP Utilizada

Fonte: Elaborado Pelos Autores

Pela Figura 4, verifica-se que a rede possui 30 entradas, referentes aos últimos trinta valores da série, sendo que a saída corresponde a estimativa do próximo valor da série. Na camada escondida utilizou-se a função de ativação Tangente Sigmóide Hiperbólica. Por sua vez, na camada de saída utilizou-se a função de ativação linear.

O algoritmo *Bayesian Regularization Backpropagation*, funciona de maneira similar ao *Levenberg-Marquardt*, sendo que a retropropagação é utilizada para calcular o jacobiano da função de desempenho, com relação aos pesos e bias da rede. Por padrão, este algoritmo de treinamento desabilita a validação.

Foram testados os valores de (10, 20, 30, 40, 50, 60 e 70) neurônios na camada escondida de modo a determinar a melhor configuração para a rede. Para tanto, em cada configuração foi efetuado 5 coletas da raiz quadrada do erro quadrático médio (RMSE) do conjunto de teste, e a execução com menor RMSE médio foi selecionada para comparação com a bi-LSTM.

Para a bi-LSTM após a realização de algumas execuções e análise do gráfico de evolução do erro no treinamento, fixou-se a quantidade máxima de épocas de treinamento em 250, com 3 iterações por época e a quantidade de neurônios das duas camadas LSTM foram variadas com relação aos mesmos valores descritos para a MLP, sendo assim obtido a configuração com melhor RMSE médio. Foram utilizadas as

funções de ativação tangente hiperbólica para o estado da célula e a sigmoide para os portões.

Posteriormente foram selecionadas as melhores execuções para as duas redes, foram plotados gráficos de ajustes e de erro, bem como foram obtidos o RMSE e a média do erro percentual absoluto (MAPE) para os melhores resultados de teste para cada uma das redes. Estas informações foram comparadas para estabelecer a rede que obteve o melhor desempenho preditivo.

4. Resultados e Discussões

Com relação aos testes de RMSE médio para cada quantidade de neurônios variada, verificou-se que a MLP teve o melhor RMSE médio de US\$ 0,42, para a configuração com 50 neurônios na camada oculta. Por sua vez, o melhor RMSE médio para a bi-LSTM foi de US\$ 0,33, para a configuração com 10 neurônios em cada camada LSTM.

Analisando-se agora os treinamentos de melhor resultados, tem-se nas Figuras 5 e 6, os gráficos da evolução do erro no processo de treinamento das redes MLP e bi-LSTM respectivamente.

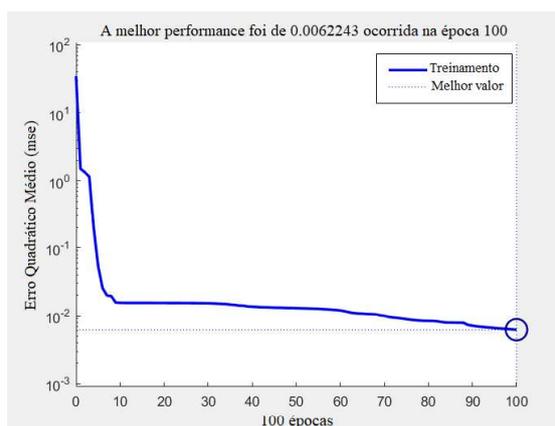


Figura 5. Evolução do MSE no treinamento da MLP

Fonte: Elaborado Pelos Autores

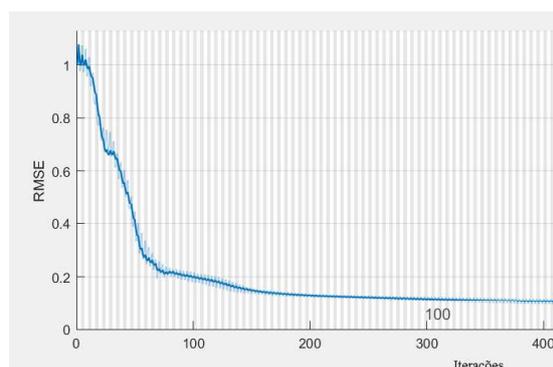


Figura 6. Evolução do RMSE no treinamento da bi-LSTM

Fonte: Elaborado Pelos Autores

Com relação a Figura 5, percebeu-se que a rede convergiu em 100 iterações para o valor de MSE 0,0062243. Considerando a Figura 6, percebeu-se que a rede convergiu em um RMSE de aproximadamente 0,1 após 400 iterações. A Figura 7 apresenta o conjunto de dados de treinamento da rede, correspondente ao período de 30/07/2015 até 01/09/2016 (400 amostras). Para o conjunto de teste, adotou-se o período de 02/09/2016 até 10/12/2016, sendo que as primeiras 30 amostras foram utilizadas para realizar a primeira predição, desta forma as comparações serão realizadas com 70 amostras.

As Figuras 8 e 9, apresentam os resultados da aplicação do conjunto de teste nas redes MLP e bi-LSTM, respectivamente, que apresentaram o melhor resultado. Comparando-se elas, conclui-se que a rede bi-LSTM apresentou melhores resultados de previsão, se comparado com a MLP, para o intervalo de tempo selecionado, uma vez que a bi-LSTM apresentou um MAPE 0,32% menor e um RMSE US\$ 0,03 menor.

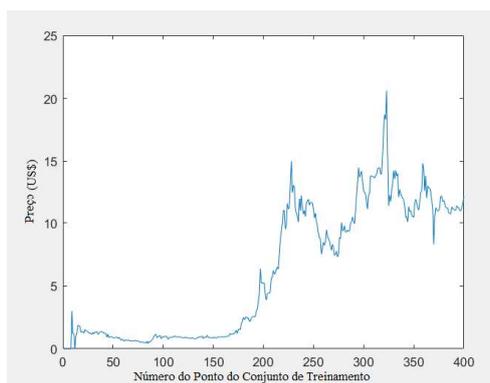


Figura 7. Dados do conjunto de treinamento

Fonte: Elaborado Pelos Autores

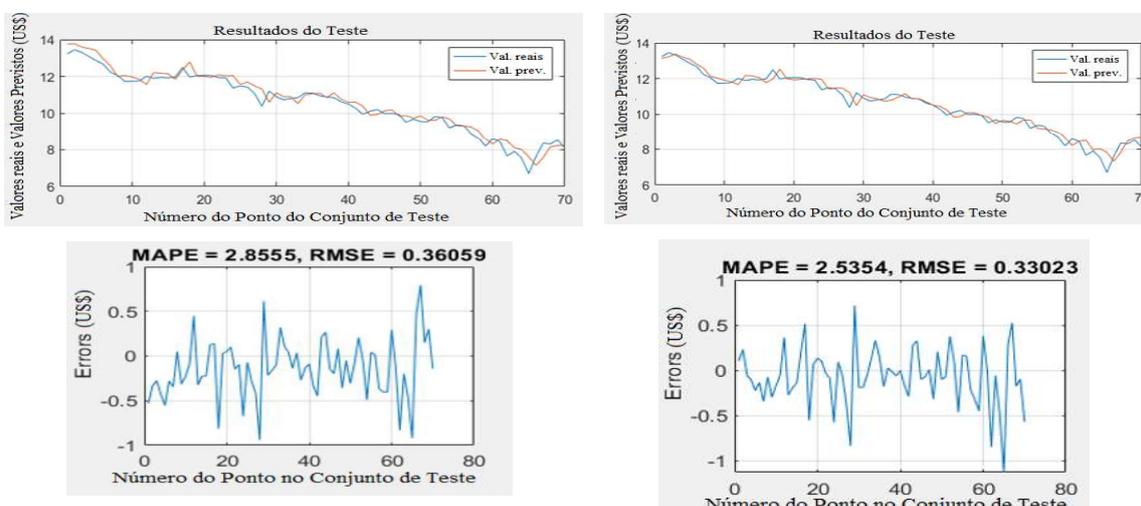


Figura 8. Resultados da MLP para o conjunto de teste

Fonte: Elaborado Pelos Autores

Figura 9. Resultados da bi-LSTM para o conjunto de teste

Fonte: Elaborado Pelos Autores

É importante observar que a diferença de desempenho das redes foi baixa, desta forma, pode-se afirmar que ambas cumpriram seu papel de maneira adequada. Comparando-se os melhores RMSE médios obtidos nas configurações da MLP e bi-LSTM analisadas, verifica-se que a bi-LSTM apresentou uma melhoria de desempenho mais evidente, com um RMSE US\$ 0,09 menor.

McNally, Roche e Caton (2018), ao aplicarem a rede LSTM para prever o preço do Bitcoin, concluíram que o ganho de desempenho preditivo se comparado ao da MLP para o caso analisado por eles, não apresentou melhoria significativa. Desta forma, em consonância ao trabalho citado, este trabalho leva a uma conclusão semelhante, de que o emprego do bi-LSTM não apresentou melhoria de desempenho preditivo significativo se comparado a MLP para o caso do preço do Ether.

5. Conclusões

Por meio deste trabalho foi possível constatar que a rede neural bi-LSTM, conforme era de se esperar, apresentou um bom desempenho na predição da série temporal do Ether, embora o ganho de desempenho com relação a MLP não tenha sido significativo.

Com relação aos trabalhos futuros, sugere-se a análise de desempenho das mesmas redes em outros períodos de tempo da mesma série temporal. Pode-se também efetuar testes com outros tipos de redes neurais em busca de melhores resultados, como por exemplo a MLP recorrente e a rede RBF, bem como aplicar algoritmos inteligentes para otimização dos parâmetros da rede.

É possível também aplicar algumas transformações na série temporal, como por exemplo, o logaritmo do preço, em busca de melhores resultados de predição. Em adição, pode-se empregar também outras medidas de desempenho para avaliar a qualidade de predição de cada rede, como por exemplo a acurácia e precisão, considerando neste caso o problema de classificação com relação ao aumento ou diminuição do preço a cada passo de tempo.

6. Referências

- ALMEIDA, N. L. and CUNHA E. L. (2018) “O Comportamento da Bitcoin no Mercado Financeiro”. [S.l.]: UEA.
- BITCOIN, M. (2018). “O que é e como funciona o blockchain do Bitcoin”. Recuperado em 16 março, 2019, de <https://blog.mercadobitcoin.com.br/o-que-%C3%A9-blockchain-a087538e9550>.
- COINMARKETCAP, S. (2019) “Top 100 Criptomoedas por Capitalização de Mercado”. Recuperado em 13 março, 2019, de <https://coinmarketcap.com/pt-br/>.
- ETHEREUM. (2019). “What is Ether ?”. Recuperado em 12 março, 2019, de <https://www.ethereum.org/ether>.
- ETHERSCAN. (2019). “Ether Price History (USD)”. Recuperado em 12 março, 2019, de <https://etherscan.io/chart/etherprice>.
- MALLQUI, D. C. A. and FERNANDES, R. A. S. (2017). “Previsões da Taxa de Câmbio de Bitcoin por meio de Técnicas de Aprendizado de Máquina”. São José do Rio Preto: DINCON.
- MATHWORKS. (2019). “Long Short-Term Memory (LSTM)”. Recuperado em 12 março, 2019, de <https://www.mathworks.com/discovery/lstm.html>.
- MCNALLY, S., ROCHE, J. and CATON, S. (2018). “Predicting the price of Bitcoin using Machine Learning”. In *26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, p. 339-343.
- NEJATIAN, A. (2019). “Time Series Prediction”. Recuperado em 12 março, 2019, de <https://www.mathworks.com/matlabcentral/fileexchange/69506-time-series-prediction>.
- SHARFUDDIN, A. A., TIHAMI, M. N. and ISLAM, M. S. (2018). “A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification”. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*.
- SILVA, I. N., SPATTI, D. H. and FLAUZINO, R. A. “Redes neurais artificiais para engenharia e ciências aplicadas”. São Paulo: Artliber.
- SOARES, W. and TEIXEIRA, J. V. S. (2018). “Previsão do Índice da Bolsa de Valores do Estado de São Paulo utilizando Redes Neurais Artificiais”. Lavras: UFLA.