

BITCOIN OPENING VALUES ANALISYS

Filipe Barreto de Oliveira

Luciana Conceição Dias Campos

filipe.bsc@gmail.com

luciana.campos@ice.ufjf.br

Universidade Federal de Juiz de Fora

Juiz de fora, Minas Gerais, Brasil

Abstract. A crypto-currency is a decentralized exchange means that uses blockchain technology, which is a type of distributed log book operated on a point-to-point network of thousands of computers, where everyone has an equal copy of the entire transaction history, preventing a central entity from promoting changes to the registry or to the software unilaterally, and from cryptography to ensure the validity of transactions and the creation of new currency units. Bitcoin, the first decentralized digital currency in history, was created in 2009. The idea with Bitcoin was to replicate the properties of physical money in a digital environment, the digitization of money in the fullest sense of the term. Since then, many other crypto-coins have been created, and the challenge has been raised to discover which crypto-coins are most promising to invest. Investing in the financial market involves buying and selling financial assets, and every investor expects to find standards in the market that allow him to buy low value's assets and sell high value's assets. The objective of this work is to propose a deep learning method to learn the behavior of increasing or decreasing the opening values of each day of the Bitcoin crypto-currency in an interval of up to 20 days. We used Bitcoin purchase values from April 2013 to February 2018. Different forecasting strategies were studied. The experimental results showed relevance in the price trend of Bitcoin.

Keywords: Bitcoin, crypto-currency, Deep Learning, Long short-term memory networks

1 Introdução

No ano de 2008, o artigo escrito por Nakamoto [1] apresentava ao mundo o primeiro conceito de moeda totalmente virtual e regulada somente pelo usuário que a detém, a Bitcoin. A Bitcoin é a realização de décadas de estudos em criptografia e sistemas distribuídos e inclui 4 inovações tecnológicas que a transformaram hoje em um gigantesco mercado de mais de 180 bilhões de dólares, segundo o site CoinMarketCap [2]. A Bitcoin compreende as tecnologias de rede ponto-a-ponto; uso de registro público virtual de transações, a *blockchain*; emissão de moeda de maneira determinística e descentralizada, a mineração; sistema de verificação de transações descentralizada.

Unidades da moeda chamadas bitcoins são usadas para guardar ou transferir valor entre os participantes da rede Bitcoin, que por sua vez se comunicam entre si primariamente pela Internet. A participação na rede Bitcoin é simples e fácil, se dá através de softwares livres de carteiras virtuais da criptomoeda, que provam a posse de bitcoins pelo usuário, por meio de uma chave criptográfica que deve ser resguardada por este.

Os programas clientes da Bitcoin podem ser de três modalidades, a primeira é a *full client* ou *full node* e possui o histórico completo de transações da Bitcoin, gerencia a carteira virtual do usuário e é capaz de iniciar transações diretamente na rede Bitcoin; a segunda é a *lightweight client*, que gerencia a carteira virtual do usuário, porém acessa serviços de terceiros para acessar a rede da Bitcoin; e a última é a *web client*, que é acessada através de navegadores e armazenam a carteira virtual do usuário em terceiros.

Contudo, a principal inovação advinda do trabalho de Nakamoto [1] é a completa descentralização do sistema de dinheiro virtual, totalmente independente de um órgão central para emitir novas unidades da moeda e validar as transações. A *blockchain* é responsável por registrar cada movimentação de bitcoins já realizadas a partir de transações válidas. A validação destas transações é feita por um algoritmo, chamado de “*proof-of-work*”, que realiza uma espécie de votação entre os usuários a cada 10 minutos para escolher quais transações serão adicionadas à *blockchain*. Na prática, o algoritmo resolve o problema de gasto duplo, onde uma mesma unidade de moeda é gasta duas vezes, e ao validar o conjunto de transações, bonifica um dos usuários com novas bitcoins. O algoritmo também é capaz de se autoajustar para manter o tempo de 10 minutos para cada nova validação e emissão de moedas.

Conforme mais usuários participavam da “eleição” e utilizavam energia elétrica para manter seus computadores executando o “*proof-of-work*”, e sendo recompensados por isso, mais difícil acabava sendo “sequestrar” a rede e adicionar transações fraudulentas. Como já havia sido informado pelo artigo de Nakamoto [1], a emissão de novas unidades, que começou com 50 bitcoins, é reduzida pela metade a cada 4 anos, e chegará a zero unidade em 2140, o que dá à bitcoin um comportamento deflacionário em seu valor.

A invenção da Bitcoin foi um marco na história e inspirou vários novos estudos em diversas áreas, como nos campos de computação distribuídas e nas áreas de economia. Sua alta taxa de valorização e volatilidade inspiraram este trabalho e levantaram o desafio: será possível descrever o comportamento futuro do valor da bitcoin a partir de seus valores passados?

O presente trabalho tem como proposta o uso de métodos de aprendizagem profunda para prever a tendência de alta ou queda nos valores de abertura¹ de cada dia da criptomoeda Bitcoin. O modelo escolhido baseado em aprendizagem profunda é o *Long short-term memory*, ou LSTM Hochreiter and Schmidhuber [3]. Segundo Brownlee [4], redes neurais recorrentes e profundas, como a LSTM, são capazes de lidar com dependências sequenciais mesmo em um alto volume de dados de treinamento como é o caso das séries diárias da Bitcoin.

1.1 Trabalhos relacionados

Grande parte dos estudos acerca de previsões de valores estão associados à bolsa de valores, como em Mithani et al. [5], onde é proposto um modelo modificado de rede neural de retropropagação para

¹Valor da primeira transação realizada

predizer valor futuro e o máximo de ganho para o titular do ativo, e complementa comparando seu método com outros métodos utilizados para a mesma classe de problemas, como métodos de regressão e as máquinas de vetor suporte (SVM).

Nelson [6] utiliza redes LSTM para atuar no problema de previsão da tendência de ações, utilizando tanto dados quantitativos (série temporal do valor da ação), quanto dados qualitativos (indicadores de análise técnica), obtendo resultados promissores em termos de retorno sobre o investimento.

Estudos recentes discutem a viabilidade de modelos de predição em relação à séries temporais de valores de criptomoedas. Shah and Zhang [7] propõem um método de regressão bayesiana, com o intuito de entender se a série histórica da Bitcoin guarda alguma informação relevante, capaz de ajudar a prever flutuações em seu preço; e Stenqvist and Lönnö [8] analisa o impacto que *tweets*² têm na tendência do preço da mais conhecida criptomoeda.

2 Modelo LSTM para análise dos valores de abertura da Bitcoin

2.1 Redes Neurais Recorrentes LSTM

A aprendizagem profunda (*deep learning*) representa a área de aprendizado de máquina capaz de gerar modelos de abstrações de alto nível de dados usando grafos com várias camadas de processamento. Segundo LeCun et al. [9], tais métodos melhoraram o estado-da-arte em relação a reconhecimento de fala, reconhecimento de objetos visuais, detecção de objetos e até em descoberta de drogas. Redes convolucionais profundas trouxeram avanços no processamento de imagens, vídeo, fala e áudio, enquanto as redes recorrentes esclareceram dados sequenciais, como texto e fala.

A rede *long short-term memory* (LSTM) é um modelo recorrente e profundo de rede neural, segundo Hochreiter and Schmidhuber [3] “capaz de aprender e interligar intervalos de tempo de mais de 1000 períodos, sem que haja perda na capacidade de memória de curto prazo. Isso se tornou possível através de um algoritmo eficiente baseado em gradiente para um fluxo de erro constante”. Sua motivação foi em resolver o “problema de desaparecimento de gradiente” Hochreiter [10] que redes recorrentes possuem ao lidar com grande volume de dados, e acontece devido ao gradiente desaparecer ao ser retropropagado ao longo do tempo.

Em resumo, mantendo o fluxo de erro constante através de filtros no interior do neurônio foi possível resolver o problema do desaparecimento de gradiente, ao permitir o ajuste de pesos e truncar a sequência quando a informação é desnecessária. Este efeito pode ser interpretado como um esquecimento. Na Figura 1 é mostrado a organização interna de um neurônio da rede LSTM.

Os elementos do neurônio LSTM da Figura 1 são descritos em:

- Filtro f_t : responsável por decidir qual informação, advinda do estado do neurônio anterior h_{t-1} , é pertinente para a rede ou deve ser ignorada;
- Filtro i_t : possui um “portão de entrada” onde são calculados novos valores, que são adicionados ao estado do neurônio caso sejam relevantes;
- Filtro C_t : calcula um novo estado baseado no estado do neurônio anterior C_{t-1} , que é usado para calcular a saída do neurônio nesse passo de tempo, e também repassado para o próximo neurônio;
- Filtro o_t : faz uso de todas as informações passadas pelos filtros anteriores para gerar a saída, composta por este e vários passos de tempo anteriores.

2.2 Base de dados

A base de dados utilizada neste trabalho foi extraída do site Kaggle [11], sendo escolhidos somente os valores diários de transação de abertura³ da bitcoin no período compreendido entre abril de 2013 até fevereiro de 2018.

Para a análise dos dados, utilizou-se os valores de transação da Bitcoin como entrada do nosso mod-

²Mensagens compartilhadas por usuários da rede *Twitter*

³Valor da primeira transação em dólares no período

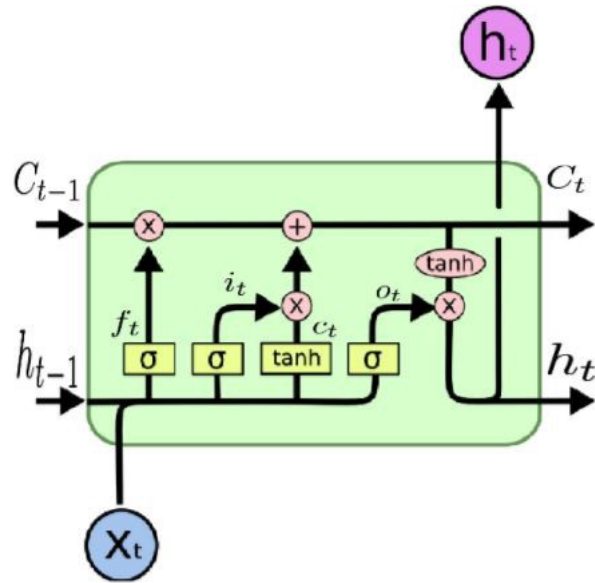


Figure 1. Esquema de um neurônio da rede LSTM

elo LSTM, que irá prever o próximo valor de transação na sequência. Por isso, é realizada a técnica de janelamento dos dados dessa série temporal, que consiste em unir os dados de vários dias anteriores para a previsão de um dia depois. Após a escolha da quantidade de dias da janela, é necessário normalizar os dados para facilitar tanto a convergência do treinamento quanto a generalização da previsão do método.

O esquema do janelamento de dados é realizado conforme apresenta a Tabela 1

Dados de Treinamento					Target de Treinamento
1	z_{t-m}	...	z_{t-1}	z_t	z_{t+1}
2	z_{t-m+1}	...	z_t	z_{t+1}	z_{t+2}
...
n	z_{t-m+n}	...	z_{t+n-1}	z_{t+n}	z_{t+n+1}

Table 1. Exemplo de disposição dos dados no janelamento

onde em cada iteração n , é utilizado a observação z na instância de tempo t e seus m dados anteriores, para prever o valor z no instante posterior $t + 1$. O valor de m , que representa o tamanho da janela, foi determinado a partir de testes empíricos. Como pode ser visto na tabela, os dados de treinamento são separados do *target*, indicando que o aprendizado do modelo é supervisionado.

Devido a capacidade da rede LSTM de guardar, e “esquecer”, memória de curto prazo ao realizar treinamento com grandes entradas, ela se ajusta bem ao problema de previsão de valores, que faz parte da análise dos dados que estamos propondo nesse trabalho.

Em estudos preliminares, utilizou-se a função de ativação ReLU, de Nair and Hinton [12], para o modelo LSTM. No entanto, experimentos realizados mostraram que uma parcela dos resultados retornaram valores zerados para a saída. Tal comportamento é devido a ReLU possuir a desvantagem de produzir no neurônio valores zeros caso a soma ponderada se torne negativa, fazendo com que a fase de treinamento pare de surtir efeito na rede. Então Maas et al. [13] propuseram a inserção de uma pequena inclinação α para a parte negativa do domínio da ReLU. Esse α é responsável por determinar a região positiva, denominada vazamento, da derivada da região negativa, o que originou a função de ativação *Leaky ReLU*, escolhida para o modelo.

Adicionalmente ao método de janelamento proposto, também utilizou-se a estratégia de imputar

os valores de compra do dólar do mesmo período dos valores de transação da Bitcoin nos dados de entrada do modelo para prever o valor esperado da Bitcoin, a fim de analisar se o valor do dólar ajuda a determinar o comportamento da criptomoeda.

Para medir o desempenho do modelo LSTM utiliza-se o Erro Médio Percentual Absoluto (MAPE) que é definido na equação 1:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - F_t}{F_t} \right| \quad (1)$$

onde Y_t representa o valor real a ser alcançado no instante de tempo t e F_t a previsão de Y_t , e n é a quantidade de dados.

De acordo com De Myttenaere et al. [14], o MAPE se mostra mais relevante em cenários com variações relativas de valor, como na finança, onde uma mudança percentual é mais perceptível do que mudanças de valor absoluto. Ademais, De Myttenaere et al. [14] cita que o MAPE se mostra bem condicionado a problemas de previsão, principalmente em situações onde há quantidade satisfatória de dados no problema.

3 Estudo de Caso

Os experimentos foram realizados em uma máquina com processador Intel Core i7 de 3ª Geração de 2,1 GHz (4 núcleos e 8 threads) e 8 GB de memória RAM DDR3 1600 MHz.

Conforme o tamanho da janela é aumentado, foram retirados, proporcionalmente, a quantidade de dias na saída. Isso resulta em menos valores encontrados na janela de 7 dias do que na janela de 1 dia, onde aquele resultou em uma saída de 15 dias e este em uma saída de 21 dias.

Para cada janelamento experimentado, foi escolhida de forma empírica o uso de 300 épocas e realizada 30 iterações do experimento. Ao final das iterações, é guardado o modelo que apresentou o melhor resultado e registrado seu MAPE.

A Tabela 2 mostra os valores do pior e melhor MAPE resultantes de uma iteração do algoritmo, e por fim o MAPE médio para todas as iterações, sendo estes para cada janelamento.

Tamanho da janela	Melhor MAPE (%)	Pior MAPE (%)	MAPE total médio (%)
1 dia	1.17	3.34	2.10
2 dias	0.55	3.73	2.39
3 dias	0.73	4.07	2.39
5 dias	0.88	4.21	2.11
7 dias	0.94	5.20	2.24

Table 2. Tabela com valores de MAPE de cada janelamento

O janelamento de 1 dia obteve, no seu melhor resultado um MAPE com valor de 1.17%, o mais alto valor de MAPE entre todos os melhores resultados dos outros janelamentos apresentados na Tabela 2. Porém, seu MAPE médio, somado de todas as iterações, obteve o valor de 2.10%, o menor entre todos os MAPEs médios. A Figura 2 apresenta uma comparação da série predita com a série histórica utilizando o modelo que forneceu o pior MAPE (gráfico à direita) e a topologia que forneceu o melhor MAPE (gráfico à esquerda) com o janelamento de 1 dia.

O janelamento de 2 dias obteve como melhor resultado MAPE de 0.55%. Observe na Tabela 2 que é o melhor resultado em relação aos outros janelamentos realizados. No entanto seu MAPE médio medido,

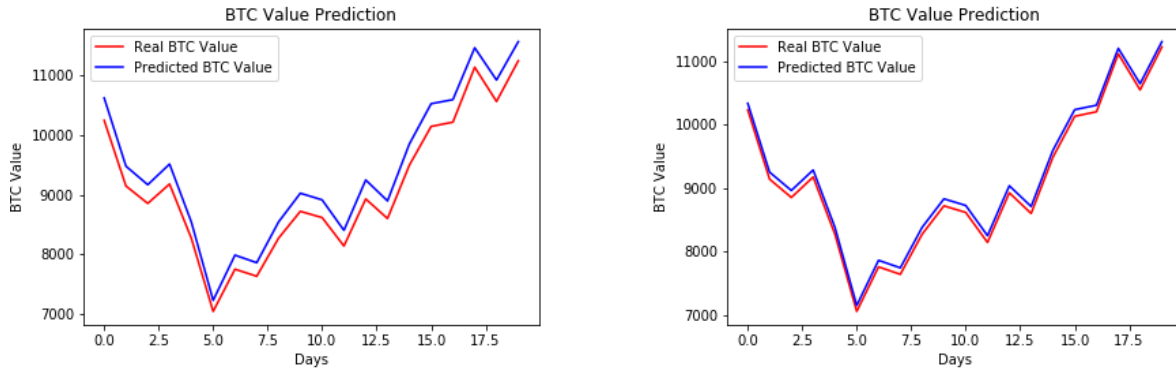


Figure 2. Gráficos com o pior e melhor resultados do janelamento de 1 dia, respectivamente

a partir da soma de todas a iterações, foi de 2.39%, ficando acima ou igual aos valores de MAPE médio encontrados nos outros janelamentos. A Figura 3 apresenta uma comparação da série predita com a série histórica utilizando o modelo que forneceu o pior MAPE (gráfico à direita) e a topologia que forneceu o melhor MAPE (gráfico à esquerda) com o janelamento de 2 dias.

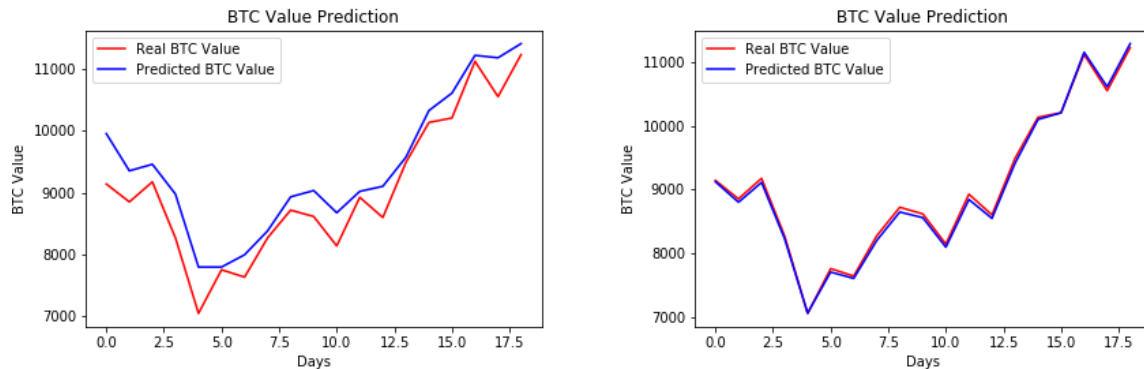


Figure 3. Gráficos com o pior e melhor resultados do janelamento de 2 dias, respectivamente

O janelamento de 3 dias obteve um MAPE médio de suas iterações semelhante ao do janelamento de 2 dias, de 2.39%, porém com valores, tanto do melhor MAPE quanto do pior, acima dos encontrados no janelamento de 2 dias. A Figura 4 apresenta uma comparação da série predita com a série histórica utilizando o modelo que forneceu o pior MAPE (gráfico à direita) e a topologia que forneceu o melhor MAPE (gráfico à esquerda) com o janelamento de 3 dias.

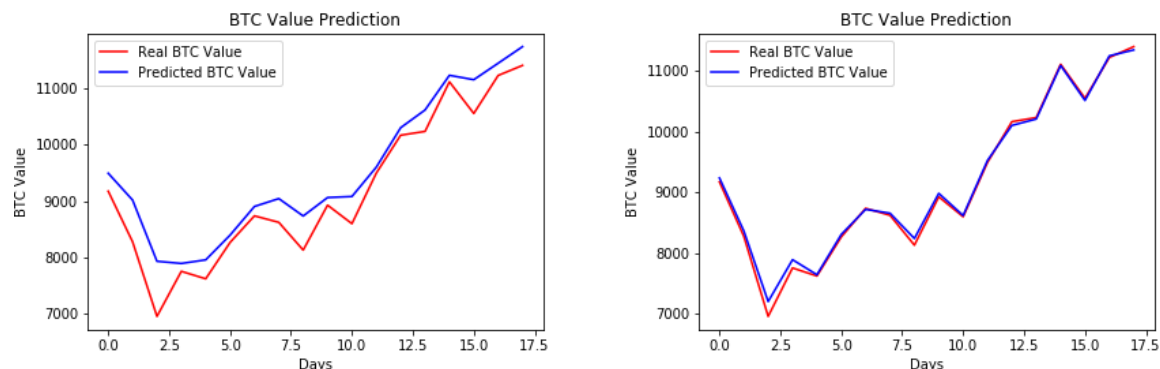


Figure 4. Gráficos com o pior e melhor resultados do janelamento de 3 dias, respectivamente

O janelamento de 5 dias obteve como o melhor MAPE e o pior MAPE resultados acima de outros

janelamentos, como pode ser visto na Tabela 2. Contudo, observe que o seu MAPE médio, soma de todas as suas iterações, foi de 2.11%, valor bem próximo do melhor resultado, encontrado no janelamento de 1 dia. A Figura 5 apresenta uma comparação da série predita com a série histórica utilizando o modelo que forneceu o pior MAPE (gráfico à direita) e a topologia que forneceu o melhor MAPE (gráfico à esquerda) com o janelamento de 5 dias.

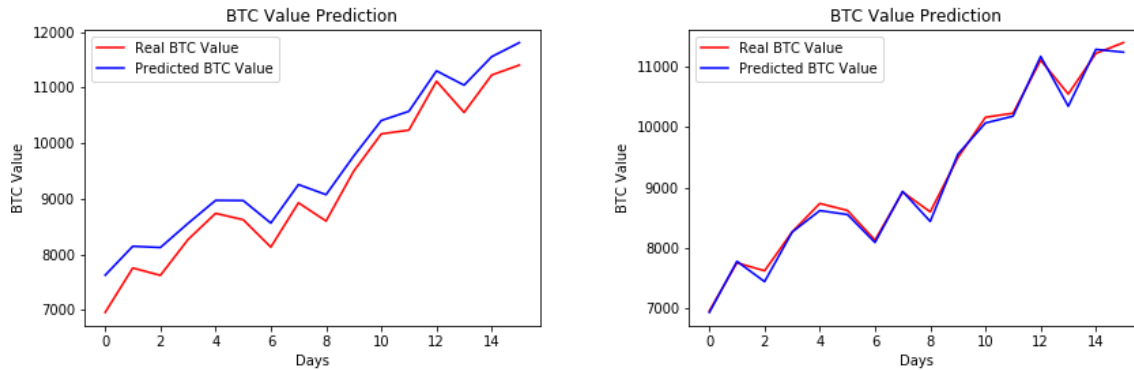


Figure 5. Gráficos com o pior e melhor resultados do janelamento de 5 dias, respectivamente

O janelamento de 7 dias obteve o pior resultado geral de pior caso, com um MAPE de 5.20%, e no melhor resultado só não foi maior que o janelamento de 1 dia. Entretanto, seu valor de MAPE médio, somado de todas as iterações foi de 2.24%, valor menor do que os janelamentos de 2 e 3 dias. A Figura 6 apresenta uma comparação da série predita com a série histórica utilizando o modelo que forneceu o pior MAPE (gráfico à direita) e a topologia que forneceu o melhor MAPE (gráfico à esquerda) com o janelamento de 5 dias.

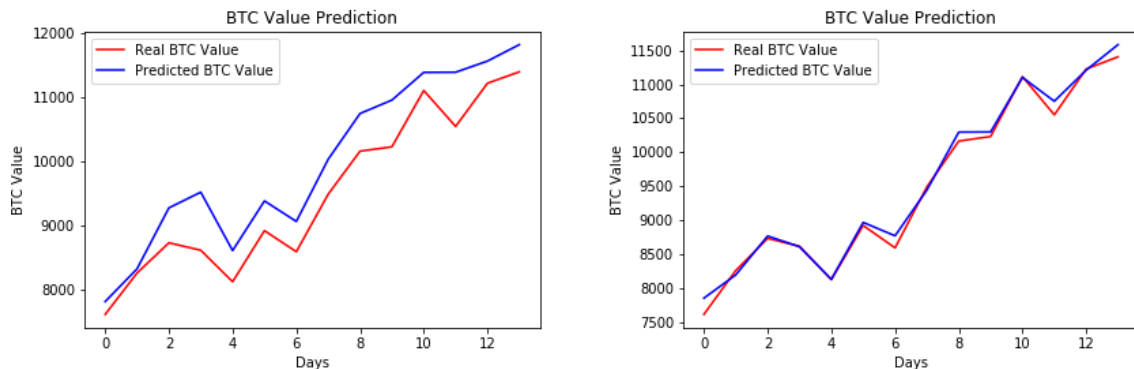


Figure 6. Gráficos com o pior e melhor resultados do janelamento de 7 dias, respectivamente

Ao observar a tabela 2, pode ser notado que os valores do MAPE dos piores resultados encontrados para cada janelamento crescem conforme o aumento do número de dias. Já nos melhores resultados, o comportamento do janelamento de 1 dia é que possui o pior melhor resultado entre todos os janelamentos explorados. Os demais apresentam uma piora quando se aumenta o número de dias, conforme acontece no pior caso. Tal comportamento confirma que o número de dias da janela influencia no aprendizado do modelo. É possível notar que apenas um dia não é suficiente para a rede absorver o conteúdo dos dados.

4 Conclusão

O artigo proposto teve como premissa propor um modelo de previsão capaz de se aproximar do comportamento real da tendência do preço da criptomoeda Bitcoin, utilizando como base redes LSTM, e é sustentado pela ideia de que um algoritmo de rede recorrente robusto como a LSTM poderia fornecer

melhores previsões, justamente por ter a capacidade de guardar informação em um espaço de tempo de curto prazo.

Como forma de validar os resultados foram realizados testes empíricos utilizando dados extraídos diretamente do registro de transações da Bitcoin, através do site Kaggle [11]. Pode ser observado que o modelo proposto teve uma alta eficácia basicamente em todos os janelamentos testados, porém este resultado pode ser indício de uma pseudo predição dos valores, onde o modelo só se mostra capaz de memorizar comportamentos anteriores da bitcoin, ao invés de generalizar o conhecimento adquirido através do treinamento.

4.1 Trabalhos Futuros

O campo de estudos de previsões é uma área com crescente número experimentações de métodos de aprendizado de máquina, num primeiro momento com a bolsas de valores, que oferece uma infinidade de possibilidades de pesquisa. Desta maneira, esse artigo pode ser estendido de diversas formas de maneira análoga.

Um dos principais pontos de melhoria está na qualidade das previsões, pois problemas de aprendizado de máquina sempre estão sujeitas à qualidade da entrada. Com isso, uma das possibilidades é melhorar os atributos utilizados, como por exemplo, extrair o retorno como informação de dois dias com valores de transação, assim reduzindo o ruído da base de dados e estudar melhores métodos de normalização.

A inclusão de atributos com natureza mais qualitativa, como por exemplo indicadores de sentimento de notícias relacionadas às criptomoedas, podem ajudar a refinar os dados de treinamento, melhorando o aprendizado da rede por consequência.

Agradecimentos

Agradecimentos à Universidade Federal de Juiz de Fora (UFJF), por ter proporcionado ensino de qualidade através de seus professores altamente qualificados, e especialmente à professora que apoiou desde o início esta pesquisa e percebeu o potencial deste artigo, Professora Doutora Luciana Conceição Dias Campos.

References

- [1] Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system.
- [2] CoinMarketCap, 2009. Bitcoin. <https://coinmarketcap.com/currencies/bitcoin/>. Access 2019-07-23.
- [3] Hochreiter, S. & Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, vol. 9, n. 8, pp. 1735–1780.
- [4] Brownlee, J., 2016. Time series prediction with lstm recurrent neural networks in python with keras.
- [5] Mithani, F., Machchhar, S., & Jasdanwala, F., 2016. A modified bpn approach for stock market prediction. In *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, pp. 1–4. IEEE.
- [6] Nelson, D. M. Q., 2017. Uso de redes neurais recorrentes para previsão de séries temporais financeiras.
- [7] Shah, D. & Zhang, K., 2014. Bayesian regression and bitcoin. In *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pp. 409–414. IEEE.
- [8] Stenqvist, E. & Lönnö, J., 2017. Predicting bitcoin price fluctuation with twitter sentiment analysis.
- [9] LeCun, Y., Bengio, Y., & Hinton, G., 2015. Deep learning. *nature*, vol. 521, n. 7553, pp. 436.

- [10] Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, n. 02, pp. 107–116.
- [11] Kaggle, 2018. Bitcoin,ethereum,litecoin exchange price. Accessed: 2018-11-12.
- [12] Nair, V. & Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.
- [13] Maas, A. L., Hannun, A. Y., & Ng, A. Y., 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, pp. 3.
- [14] De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F., 2016. Mean absolute percentage error for regression models. *Neurocomputing*, vol. 192, pp. 38–48.