https://eventos.utfpr.edu.br//sicite/sicite2020

Aprendizado de Máquina para a Predição do Valor de Criptomoedas

Machine Learning Applied to Value Forecasting of Cryptocurrencies

João Vitor Breda
jvitor.breda@gmail.com
Universidade Tecnológica Federal do Paraná, Dois Vizinhos, Paraná, Brasil

Gustavo Henrique Paetzold
ghpaetzold@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil

RESUMO

O mercado de câmbio, ou *FOREX*, é o mercado que mais movimenta dinheiro diariamente, tendo empresas e profissões focadas em prever padrões e valores ao análisar os dados das moedas. Para esse estudo, foram utilizados modelos de regressão e classificação de Aprendizado de Máquina para desenvolver um algoritmo que faça previsão do preço de criptomoedas, mais especificamente do Bitcoin, considerada a mais importante dentre todas as criptomoedas. O objetivo deste trabalho é verificar a eficência dos algoritmos desenvolvidos para a predição do preço do Bitcoin e a geração de lucro quando aplicado em estratégias de compra e venda do mesmo. Para isso, os modelos foram desenvolvidos com camadas de LSTM, uma arquitetura de rede neural artificial capaz de operar sobre dados temporais a fim de identificar e armazenar padrões que ajudem na predição dos valores. Como os algoritmos desenvolvidos são simples e os dados do Bitcoin são muito voláteis e tem um vis de subida de preço desde sua criação, os resultados obtidos foram negativos, tendo ambos perdido dinheiro na simulação de compra e venda.

PALAVRAS-CHAVE: Aprendizado de Máquina. Mercado de Câmbio. Criptomoedas.

ABSTRACT

The foreign exchange market, or FOREX, is the market that moves the most money on a daily basis, with companies and professions focused on predicting patterns and values when analyzing currency data. For this study, Machine Learning's regression and classification models were used to develop an algorithm that predicts the price of cryptocurrencies, more specifically Bitcoin, considered the most important among all cryptocurrencies. The objective of this work is to verify the efficiency of the developed algorithms for Bitcoin price prediction and its profit generation when applied in buying and selling strategies. For this, the models were developed with layers of LSTM, an artificial neural network architecture capable of operating on temporal data in order to identify and store patterns that help in the prediction of values. As the algorithms developed are simple and Bitcoin data is very volatile and has seen a price hike since its inception, the results obtained were negative, both having lost money in the simulation of buying and selling.

KEYWORDS: Machine Learning. Foreign Exchange Market. Cryptocurrencies.

Recebido: 19 ago. 2020.
Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

INTRODUÇÃO

O mercado de câmbio, ou FOREX, é o maior mercado do mundo, tendo movimentado cerca de $6.6 trilhões de dólares diáriamente em abril de 2019 (Bank for International Settlements, 2019). Com tamanha importância monetária, vários estudos foram direcionados a fim de entender esse mercado e a possibilidade de prever seus valores.

De acordo com Murphy (1986, p. 2), a Análise Técnica é o estudo da ação de mercado, feita principalmente com o uso de gráficos e com o objetivo de prever as tendências de preços futuras. O termo "ação de mercado" é a definição das três principais fontes de informação disponíveis: preço, volume e juros em aberto.

Murphy (2019, p. 2) afirma ainda que a Análise Técnica é baseada em três premissas:

- Ação de Mercado Desconta Tudo: Qualquer informação que possa afetar o preço já foi levada em consideração. Isso significa que a análise do preço do mercado é tudo que é necessário, já que todas as informações estão refletidas no preço.

- Movimento de Preço em Tendências: O mercado tem tendências e por isso técnicas para identificação e categorização de gráficos são altamente utilizadas em sua análise.

- História se Repete: Muito do que se é análisado trata-se sobre a psicologia humana. Os padrões em gráficos que são identificados revelam a psicologia do mercado, tendo funcionado há anos, assume-se que irão continuar a funcionar no futuro.

Antes dominado por métodos lineares estátisticos, desde meados de 1990, os modelos de Aprendizado de Máquina vem se estabelecendo como uma forte opção para a predição de valores em séries temporais (Bontempi, Taieb, Borgne, 2013).

Já as criptomoedas surgiram após a criação do Bitcoin em 2008 pelo pseudônimo "Satoshi Nakamoto" e tem seu funcionamento similar a mesma. Conforme Schär e Berentsen (2018), o Bitcoin é uma moeda sem representação fisíca que existe apenas em um arquivo de dados que contém o histórico de todas as transações já feitas, sua blockchain. A blockchain consiste em uma sequência de blocos de dados onde cada bloco é criado com base em seus antecessores, adicionando os dados de novas transações.

Levando em consideração essas informações, este trabalho tem como objetivo a aplicação e avaliação de técnicas de Aprendizado de Máquina na predição dos valores de criptomoedas que, por sua vez, foram escolhidas pela sua importância que vem aumentando desde sua criação em 2009 com o Bitcoin e sua alta volatilidade em decorrência do menor volume de transações, o que torna a identificação de padrões mais complexa.

MATERIAIS E MÉTODOS

Para o desenvolvimento deste trabalho, a linguagem de programação Python foi utilizada, bem como suas bibliotecas: Keras, que inclui métodos de aprendizado

de máquina, e Pandas, Numpy e Matplotlib para análise e manipulação de dados. Para a análise de preços, o Bitcoin foi escolhido por ser a criptomoeda mais importante e a que possui a maior quantidade de valores históricos.

O trabalho então foi definido e executado na seguinte sequência de passos:

COLETA E LIMPEZA DE DADOS

A coleta de dados foi feita do site CoinMarketCap e incluí os dados do Bitcoin de 28 de abril de 2013 até 31 de outubro de 2019, resultando em dados de 2378 dias. Os preços utilizados são em dólares Americanos. A Tabela 1 demonstra a estrutura dos dados coletados.

Tabela 1 – Demonstração dos dados dos 3 primeiros dias

Data	Abertura	High	Low	Close	Volume	Marketcap
28/04/2013	135.30	135.98	132.10	134.21	-	1488566728
29/04/2013	134.44	147.49	134.00	144.54	-	1603768865
30/04/2013	144.00	146.93	134.05	139.00	-	1542813125

Fonte: Autores (2020).

Os dados então foram separados em uma base de treinamento e outra de testes, para utilização futura no algoritmo. Os dados de treinamento incluem os dados de 31 de abril de 2013 até 31 de julho de 2019. Os dados de teste incluem os dados de 1 agosto de 2019 até 31 de outubro de 2019.

Também foi definida a utilização apenas do valor de abertura inicialmente, mantendo o algoritmo simples e podendo utilizar todos os dados originais sem que seja necessário fazer o tratamento das linhas que possuem campos sem valores.

DEFINIÇÃO DO ALGORITMO

Definimos a utilização de dois tipos de algoritmos, um regressor e um classificador, pois são as duas principais categorias de modelos de Aprendizado de Máquina.

MODELO DE REGRESSÃO LINEAR

No algoritmo de regressão, os dados foram normalizados entre 0 e 1 utilizando-se da função MinMaxScaler do Keras, como demonstrado na Tabela 2.

Tabela 2 – Demonstração dos dados dos 3 primeiros dias após tratamento

Abertura
0.003442
0.00339769
0.00389029

Fonte: Autores (2020).

Para que o algoritmo possa prever o preço de um dia, foi definida a utilização dos 90 dias anteriores ao mesmo. O valor de 90 dias foi definido com base em tentativa e erro. Com isso, o total de dados de treinamento diminui de 2286 para 2196 na base de treinamento, pois os 90 primeiros não possuem 90 dados

anteriores que o algoritmo possa utilizar para prever seu valor atual. A quantidade de dados de teste que devem ser previstos manteve-se em 92 dias, porém, os dados dos 90 últimos dias da base de treinamento foram utilizados para que se possa fazer a predição dos 90 primeiros da base de testes.

A construção do modelo então foi feita com os seguintes parâmetros:

- Camada de entrada LSTM com dropout de 0.3 e 100 células
- 3 camadas ocultas LSTM com dropout de 0.3 e 50 células
- Camada de saída densa com 1 célula e função de ativação linear
- Função de optimização *RMSprop*, função de custo *Mean squared error*
- 50 épocas e 32 de *batch size*

O *Mean Absolute Error* foi a métrica de avaliação utilizada para auxiliar na avaliação do treinamento do modelo. A LSTM foi utilizada pois é um tipo de rede neural artificial que retém memória com base em suas conexões de feedback, importante para identificar e gravar padrões nos preços das moedas.

O modelo sequencial, sua camada de saída do tipo densa, a função de ativação e as funções de optimização, custo e métricas são comumente utilizadas em modelos de regressão linear. Os demais parâmetros foram definidos com base em tentativa e erro.

MODELO DE CLASSIFICAÇÃO

No modelo de classificação utilizamos o mesmo método de prever um preço com base nos 90 dias anteriores, porém, é necessária a mudança dos preços para valores binários. Mudamos um preço para 1 se o preço é maior no próximo dia, ou 0 se o preço é menor no próximo dia. Com isso o último dado não pode ser utilizado, pois não sabemos se o preço sobe ou desce no próximo dia, resultando em 2195 dados de treinamento.

A construção do modelo então foi feita com os seguintes parâmetros:

- Camada de entrada LSTM com dropout de 0.2 e 100 células
- 2 camadas ocultas LSTM com dropout de 0.2 e 50 células
- Camada de saída densa com 1 célula e função de ativação softmax
- Função de optimização *adam*, função de custo *binary crossentropy*
- 50 épocas e 32 de *batch size*

A Acurácia foi a métrica de avaliação utilizada para auxiliar na avaliação do treinamento do modelo. Utilizamos a LSTM pelo mesmo motivo do modelo de regressão. O modelo sequencial, sua camada de saída do tipo densa, a função de ativação e as funções de optimização, custo e métricas são comumente utilizadas em modelos de classificação. Os demais parâmetros foram definidos com base em tentativa e erro.

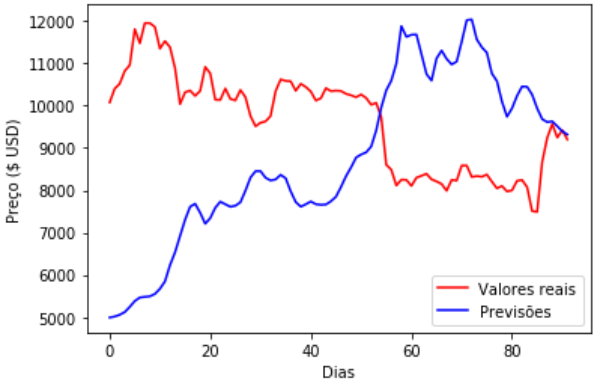
RESULTADOS E DISCUSSÃO

Após o treinamento, as bases de teste foram aplicadas em seus respectivos modelos, obtendo os seguintes resultados:

RESULTADO DO MODELO DE REGRESSÃO

O modelo de regressão teve como resultado da métrica de avaliação *Mean Absolute Error* o valor de 8725.13, o que demonstra sua incapacidade em predizer o preço do Bitcoin. Se análisarmos o gráfico da Figura 1, podemos ver claramente a grande diferença entre o preço previsto diariamente e o preço real.

Figura 1 – Comparação do preço real do Bitcoin com a previsão do algoritmo



Fonte: Autores (2020).

RESULTADO DO MODELO DE CLASSIFICAÇÃO

O modelo de classificação atingiu uma acurácia de 43.96%, porém, se analisarmos os resultados na Tabela 3, podemos ver que a razão desse índice baixo de desempenho é pelo fato do modelo estar prevendo em todos os dias que o preço irá subir, possívelmente pelo fato do preço histórico do Bitcoin ter subido mais do que descido desde sua criação.

Tabela 3 – Comparação do número de ocorrências de mudança do preço real com a previsão do algoritmo

Tipo de valor	Mudança do preço no próx. dia	
	Aumento	Redução
Valor real	40	51
Previsão	91	0

Fonte: Autores (2020).

SIMULAÇÃO DE COMPRA E VENDA

Utilizamos os modelos para simular a compra e venda de Bitcoin no período de teste, a fim de analisar o quanto o algoritmo lucraria, ou perderia, nesses 90 dias.

A primeira simulação foi feita com a estratégia mais comum em que o algoritmo compraria Bitcoin quando ele prevesse que o preço aumentaria no dia seguinte. Com isso, o modelo regressor perderia US$1825.34 e o modelo classificador perderia US$654.98

A segunda simulação foi feita com a estratégia de *shorting*, em que se pega valores emprestado e vende-os, a fim de comprar os mesmos valores novamente por um preço menor no futuro para devolve-los a quem se pegou emprestado. Ou seja, o algoritmo venderia o Bitcoin "emprestado" quando prevesse que o preço desceria no dia seguinte, significando que o preço atual está alto, e pagaria o Bitcoin quando verificasse que o preço aumentaria no próximo dia, significando que o preço atual está baixo. Com isso, o modelo regressor perderia US$950.35 e o modelo classificador US$0.0, uma vez que, como discutido na subseção de resultados do modelo de classificação, o mesmo prediz uma subida de preço do Bitcoin para todos os dias do conjunto de testes.

CONCLUSÃO

Os modelos de Aprendizado de Máquina possuem grande potencial na previsão de valores em séries temporais, porém, a grande volatilidade no preço de criptomoedas requer modelos mais robustos, utilizando-se de estratégias de inteligência artificial mais sofisticadas. Modelos simples de regressão e classificação não são o suficiente para se prever com sucesso o preço de uma criptomoeda como o Bitcoin, ambos tendo falhado em obter qualquer lucro em uma simulação de compra e venda realística.

Como trabalhos futuros, pode ser feita uma experimentação com o tamanho das redes neurais artificiais, utilização de técnicas como oversampling e undersampling, além de inserir no modelo padrões de gráficos já identificados e classificados que são utilizados com sucesso no mercado de câmbio, ações e outros.

AGRADECIMENTOS

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq – Brasil, da Fundação Araucária FA – Paraná/Brasil e da Universidade Tecnológica Federal do Paraná/Brasil.

REFERÊNCIAS

Bank for International Settlements. **Foreign Exchange Turnover in April 2019**. Disponível em: https://www.bis.org/statistics/rpfx19_fx.htm. Acesso em: 1 set.

BERENTSEN, A.; SCHÄR, F. **A Short Introduction to the World of Cryptocurrencies**. Disponível em: https://doi.org/10.20955/r.2018.1-16. Acesso em: 1 set. 2020.

BONTEMPI, G.; TAIEB, S. B.; BORGNE, Y. L. **Machine Learning Strategies for Time Series Forecasting**. Disponível em: https://www.researchgate.net/publication/236941795_Machine_Learning_Strategies_for_Time_Series_Forecasting. Acesso em: 1 set. 2020.

CoinMarketCap. **Historical data for Bitcoin**. Disponível em: https://coinmarketcap.com/currencies/bitcoin/historical-data/. Acesso em: 1 set. 2020.

MURPHY, J. J. **Technical Analysis Of The Financial Markets**: A Comprehensive Guide To Trading Methods And Applications. Nova York: New York Institute of Finance, 1999. p. 1-5.