Modelos de previsão do preço do Bitcoin baseados em Machine Learning

Aluno: **Luiz Carlos**

Professora: Karla Figueiredo

Universidade do Estado do Rio de Janeiro – UERJ

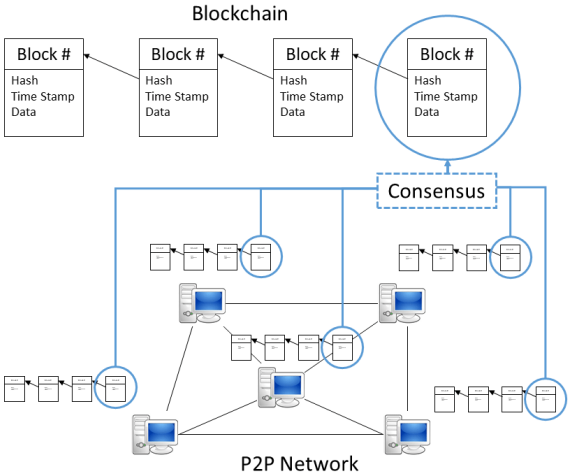
Trabalho da disciplina de Inteligência Artificial

Rio de Janeiro, 8 de Agosto de 2022

Agenda

Introdução

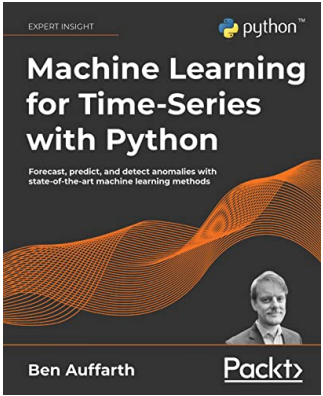
O que é a blockchain do Bitcoin?



O Preço dos bitcoins

- O que é a cotação dos bitcoins?
- É possível prever o seu preço?

- Séries temporais;
- Algoritmos de *Machine Learning*;
- Algoritmos de *Machine Learning* aplicados a séries temporais.



Motivação

Motivação

- O Bitcoin como um ativo financeiro
  - Cotação diária em $ USD
  - Variáveis intrínsecas à blockchain do Bitcoin
    - difficulty
    - hash_rate
    - n_transactions
    - block_size
  - Variáveis extrínsecas
    - câmbio
    - cotação da bolsa
    - notícias
    - pesquisas no Google e na Wikipedia
    - publicações no Twitter
- Aplicação de modelos de previsão de preço de ativos.

Objetivo

Objetivos

- Objetivo geral:
  - Geração de modelos de *Machine Learning* capazes de prever com um nível de confiança o preço do Bitcoin.
- Objetivos específicos:
  - Esclarecer fundamentos a respeito de modelos de *Machine Learning* aplicados à previsão de séries temporais;
  - Comparar os resultados obtidos pelo modelo treinado com os resultados obtidos por outros métodos vistos em literatura especializada.

Revisão Bibliográfica

Predicting the price of Bitcoin using Machine Learning

Sean McNally
x15021581
MSc Reseach Project in Data Analytics

9th September 2016

Abstract

This research is concerned with predicting the price of Bitcoin using machine learning. The goal is to ascertain with what accuracy can the direction of Bitcoin price in USD can be predicted. The price data is sourced from the Bitcoin Price Index . The task is achieved with varying degrees of success through the implementation of a Bayesian optimised recurrent neural network (RNN) and Long Short Term Memory (LSTM) network. The LSTM achieves the highest classification accuracy of 52% and a RMSE of 8%. The popular ARIMA model for time series forecasting is implemented as a comparison to the deep learning models. As expected, the non-linear deep learning methods outperform the ARIMA forecast which performs poorly. Wavelets are explored as part of the time series narrative but not implemented for prediction purposes. Finally, both deep learning models are benchmarked on both a GPU and a CPU with the training time on the GPU outperforming the CPU implementation by 67.7%.

Figure 3: Long Short Term Memory[6]

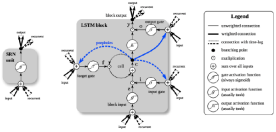
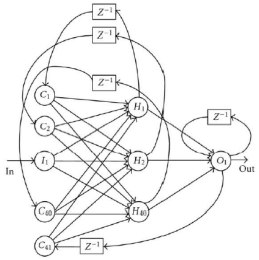


Figure 2: Recurrent Neural Network



An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information

HUISU JANG AND JAEWOOK LEE
Department of Industrial Engineering, Seoul National University, Seoul 151742, South Korea
Corresponding author: Jaewook Lee (jaewook@snu.ac.kr)

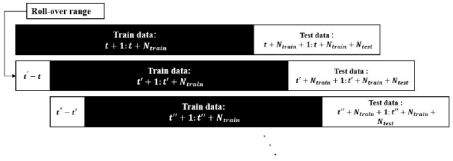
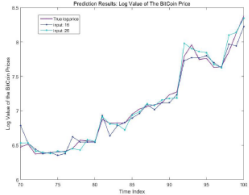
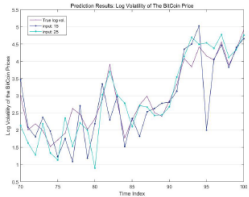


FIGURE 3. the formation of the Blockchain.



(a)



(b)

FIGURE 7. Prediction results of (a) the log value of the Bitcoin price and (b) the log volatility of the Bitcoin price.

Journal of *Risk and Financial Management*

MDPI

Article

Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning

Takuya Shintate and Lukáš Pichl *

Graduate School of Arts and Sciences, International Christian University, Osawa 3-10-2, Mitaka, Tokyo 181-8585, Japan; g196729a@icu.ac.jp

* Correspondence: lukas@icu.ac.jp; Tel.: +81-422-33-3286

Received: 25 December 2018; Accepted: 17 January 2019; Published: 21 January 2019

check for updates

Table 2. Model evaluation scores on accuracy, recall, precision, and F1 measure. Bitcoin price in U.S. dollar (BTCUSD) from the OkCoin market was used as the dataset. We used the last 120k data points for the evaluation, which were not used in training and validation.

	Accuracy	Recall	Precision	F1 Score
MLP	0.5559	0.4945	0.4978	0.4786
LSTM	0.5759	0.5464	**0.5717**	0.5034
RSM (*ours*)	**0.6264**	**0.5538**	0.5488	**0.5367**

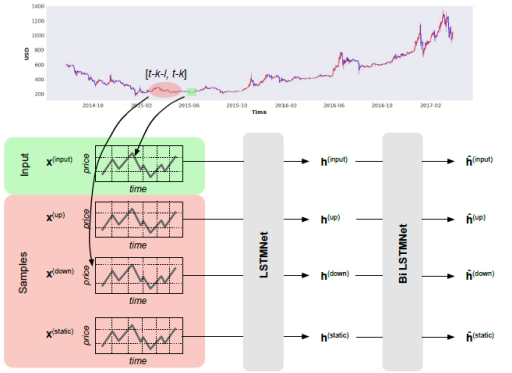


Figure 2. Visualization of the pipeline. Given inputs $\mathbf{x}_i^{(input)}$, sequences $\mathbf{x}_i^{(up)}$, $\mathbf{x}_i^{(down)}$, $\mathbf{x}_i^{(static)}$ are randomly sampled in the red window. Then, the input and samples are independently encoded with LSTMNet (Equation (3)) and bi-directional LSTMNet (Equation (4)). Refer to the text for more details.

autores	métodos utilizados
Shah and Zhang	bayesian regression
McNally et al.	ARIMA, recurrent neural networks (RNNs), long short-term memory (LSTM)
Jang and Lee	bayesian neural network (BNN), linear regression, support vector regression (SVR)
Jang et al.	rolling window LSTM model
Shintate and Pichl	deep learning-based random sampling model

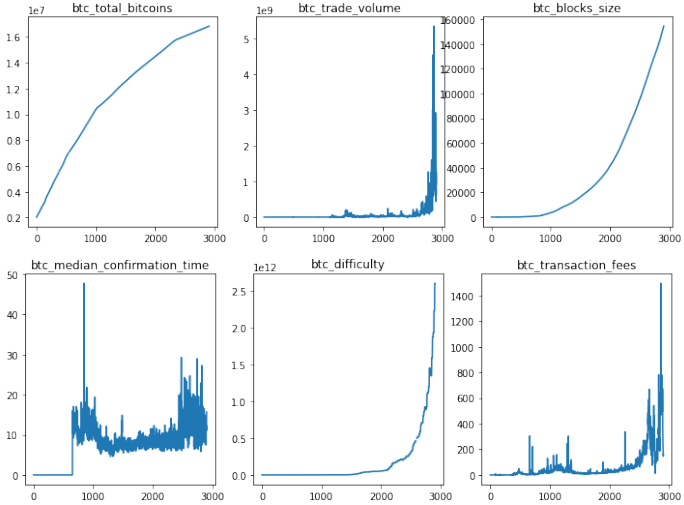
Análise dos Dados

Base de dados - Campos

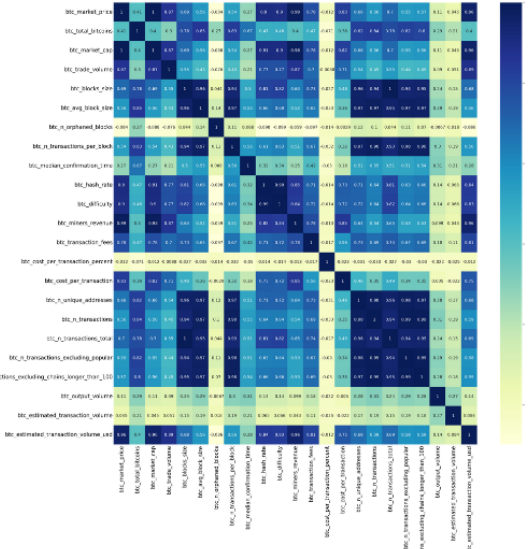
```
dados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2920 entries, 0 to 2919
Data columns (total 24 columns):
 #   Column                                                Non-Null Count  Dtype  
---  ------                                                --------------  -----  
 0   Date                                                  2920 non-null   object 
 1   btc_market_price                                      2920 non-null   float64
 2   btc_total_bitcoins                                    2920 non-null   float64
 3   btc_market_cap                                        2920 non-null   float64
 4   btc_trade_volume                                      2899 non-null   float64
 5   btc_blocks_size                                       2920 non-null   float64
 6   btc_avg_block_size                                    2920 non-null   float64
 7   btc_n_orphaned_blocks                                 2920 non-null   float64
 8   btc_n_transactions_per_block                          2920 non-null   float64
 9   btc_median_confirmation_time                          2920 non-null   float64
 10  btc_hash_rate                                         2920 non-null   float64
 11  btc_difficulty                                        2920 non-null   float64
 12  btc_miners_revenue                                    2920 non-null   float64
 13  btc_transaction_fees                                  2920 non-null   float64
 14  btc_cost_per_transaction_percent                      2920 non-null   float64
 15  btc_cost_per_transaction                              2920 non-null   float64
 16  btc_n_unique_addresses                                2920 non-null   float64
 17  btc_n_transactions                                    2920 non-null   float64
 18  btc_n_transactions_total                              2920 non-null   float64
 19  btc_n_transactions_excluding_popular                  2920 non-null   float64
 20  btc_n_transactions_excluding_chains_longer_than_100   2920 non-null   float64
 21  btc_output_volume                                     2920 non-null   float64
 22  btc_estimated_transaction_volume                      2920 non-null   float64
 23  btc_estimated_transaction_volume_usd                  2920 non-null   float64
dtypes: float64(23), object(1)
memory usage: 547.6+ KB
```

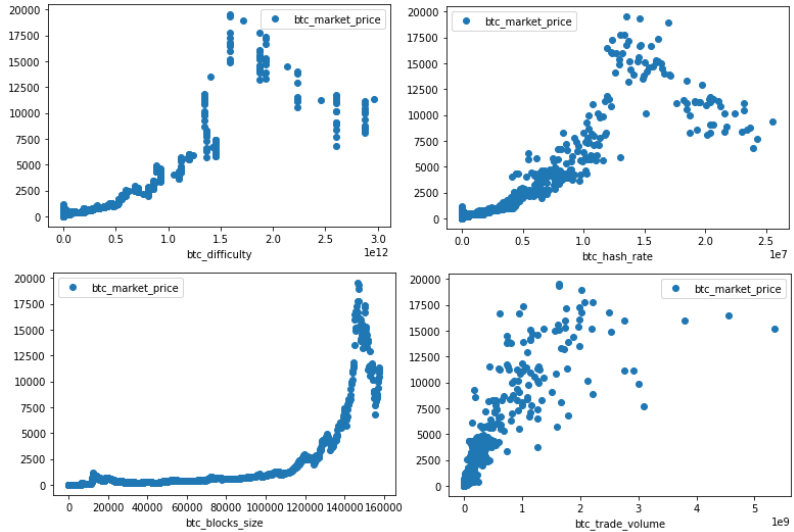
Base de dados - Gráficos das principais variáveis



Matriz de correlação

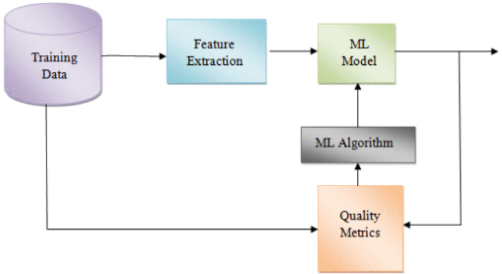


Dispersão dos dados em função do preço de mercado do BTC



Metodologia

1. Geração da base de dados;
2. seleção dos atributos a serem utilizados para treinar o modelo;
3. verificação e tratamento dos dados faltantes (*missing values*);
4. normalização dos dados;
5. treinamento do modelo;
6. análise dos resultados obtidos.



Pré-processamento: função feature_importance x correlograma

btc_n_orphaned_blocks	0.000006
btc_median_confirmation_time	0.000337
btc_estimated_transaction_volume	0.000369
btc_cost_per_transaction_percent	0.000517
btc_output_volume	0.000704
btc_n_transactions_excluding_chains_longer_than_100	0.000913
btc_n_transactions	0.001720
btc_n_transactions_excluding_popular	0.002391
btc_n_transactions_per_block	0.005801
btc_transaction_fees	0.006594
btc_n_unique_addresses	0.014158
btc_avg_block_size	0.037979
btc_total_bitcoins	0.052907
btc_trade_volume	0.073569
btc_cost_per_transaction	0.089108
btc_blocks_size	0.139657
btc_n_transactions_total	0.159972
btc_hash_rate	0.201639
btc_difficulty	0.211659

Modelos testados

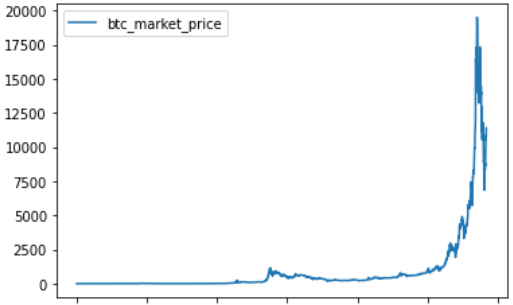
- Os seguintes modelos de regressão foram testados (com base nos modelos estudados na disciplina):
  - Regressão Linear
  - Regressão Polinomial
  - Support Vector Machine
  - Regressão por Árvores de Decisão
  - K-Nearest Neighbors (KNN)
- Estatísticas teste consideradas: R^2 e Erro Quadrático Médio (EQM)

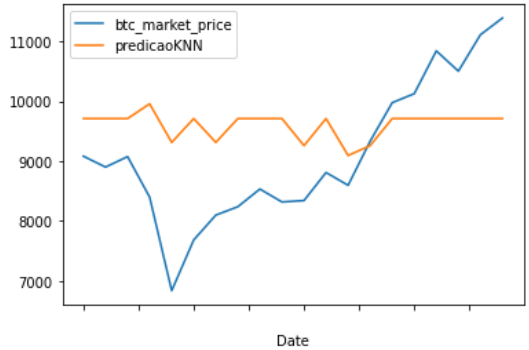
Resultados

Matriz de avaliação dos modelos testados

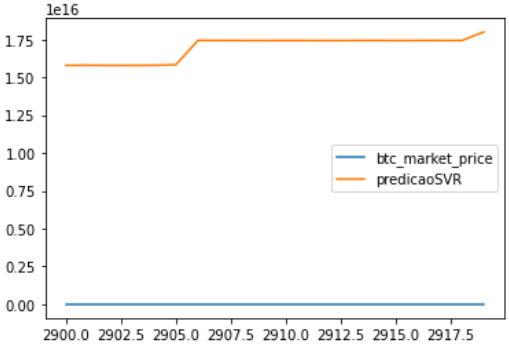
	R-quadrado	EQM
Reglin	0.889775	636028.043702
knnN	0.983700	94053.339735
svrLnr	0.865279	777376.902766
arvoreDecisao	0.978034	126746.621656
ply	0.969387	176642.690674

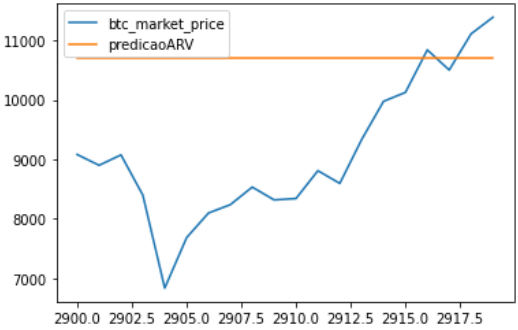
Predição do modelo com maior acurácia: K-Nearest Neighbors (KNN)





A título de comparação...





Conclusão

Conclusão

- O desenvolvimento do trabalho confirmou a tese de que, apesar da grande volatilidade no preço dos bitcoins, é possível estimar o seu preço no curto(íssimo) prazo com variáveis intrínsecas à própria blockchain do Bitcoin

- Em relação aos modelos preditivos, a regressão estatística do tipo K-Nearest Neighbors (KNN) forneceu os melhores resultados, tanto em termos de precisão quanto de previsão

- Os resultados apresentados neste trabalho contribuem para o desenvolvimento de estratégias de compra e venda de bitcoins, podendo ser estendido, se consideradas as restrições, para outras criptomoedas

Referências

Referências I

Sean McNally, Jason Roche, and Simon Caton.
Predicting the price of bitcoin using machine learning.
In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pages 339–343, 2018.

Devavrat Shah and Kang Zhang.
Bayesian regression and bitcoin, 2014.

Jang Huisu.
Predicting bitcoin prices by using rolling window lstm model.
2018.

Takuya Shintate and Lukáš Pichl.
Trend prediction classification for high frequency bitcoin time series with deep learning.
Journal of Risk and Financial Management, 12(1), 2019.

Referências II

Huisu Jang and Jaewook Lee.
An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information.
IEEE Access, 6:5427–5437, 2018.

Luiz Carlos da Silva Leão.
Jupyter notebook para o artigo trabalho final da disciplina de ia 2022-1 do ccomp/uerj.
https://github.com/lcsl88/IA-CCOMP_20221/blob/main/projeto_artigo_IA.ipynb.

Sylvain Saurel.
Bitcoin is the best-performing asset in 2019 and the best is still coming for 2020.
https://medium.com/swlh/bitcoin-is-the-best-performing-asset-in-2019-and-the-best-is-still-coming

Modelos de previsão do preço do Bitcoin baseados em Machine Learning

Aluno: **Luiz Carlos**

Professora: Karla Figueiredo

Universidade do Estado do Rio de Janeiro – UERJ