

Cryptocurrency	Price in USD	Price in BRL
BTC Bitcoin 	\$2,218.6 -5.00% (\$110.00) -10.00% (\$200.00)	1.00 +0.00% (\$0.00) +0.00% (\$0.00)
ETH Ethereum	\$197.04 -2.00% (\$4.00) -17.70% (\$40.00)	0.0000 +0.00% (\$0.00) +0.00% (\$0.00)
LTC Litecoin 	\$42.42 -8.12% (\$2.00) -7.00% (\$3.00)	0.0010 +0.00% (\$0.00) +0.00% (\$0.00)
ETC Ethereum Classic 	\$16.85 -4.00% (\$0.70) +4.00% (\$0.70)	0.0070 +0.00% (\$0.00) +0.00% (\$0.00)

explained_forecast()



Bruno Rodrigues



Pedro Gengo



Rafael Riciardi

Objetivo

- Entender o impacto de uma criptomoeda em outra;
- Entender o impacto dos índices de bolsa mundiais no valor das criptomoedas;
- Realizar previsões dos preços com diferentes modelos de machine learning;
- Explicar os resultados encontrados;
- Avaliar a performance dos modelos em séries temporais com alta volatilidade.

★ Tecnologias utilizadas



Metodologia

Para entender o impacto que as criptomoedas tinham entre si e também o impacto dos índices de bolsas no valor delas, realizamos duas abordagens:

- Análise de correlação das séries;
- Análise dos valores [shapley](#) para os modelos testados, a fim de entender como cada variável impacta a predição.

Para avaliação dos modelos, utilizamos uma métrica clássica para modelos de regressão, o RMSE (Root Mean Squared Error), o qual nos dá uma visão, em escala real da série, do quanto estamos errando. Sua fórmula é dada por:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$



SHAP

Modelos utilizados

Nossa abordagem foi baseada no conceito da [navalha de Ockham](#), ou seja, testamos modelos extremamente simples, de baixa complexidade computacional, até modelos mais complexos (*Deep Learning*), a fim de realizar uma comparação entre performance vs complexidade.

Os modelos utilizados para nossa análise foram:

- Modelo de média móvel;
- Regressão Linear;
- Regressão Linear penalizada;
- SVR (Support Vector Machines para Regressão);
- KNN (K-Nearest Neighbors);
- MLP (Multi Layer Perceptron);
- RNN utilizando blocos de LSTM.



Tivemos como inspiração para tal análise o paper de Spyros Makridakis, *Statistical and Machine Learning forecasting methods: Concerns and ways forward*.

Base de dados

Para realizar nossas análises utilizamos duas bases de dados diferentes (ambas públicas e disponíveis no Kaggle).

1. [Base](#) histórica dos valores minuto a minuto de pares de criptomoedas. Tal base estava no formato *parquet* e apresentava os seguintes dados:

open_time	open	high	low	close	volume	quote_asset_volume	number_of_trades
2017-08-17 04:00:00	301.1300	301.1300	301.1300	301.1300	0.42643	128.4108	2

2. [Base](#) histórica diária dos índices dos grandes mercados mundiais. Arquivos em *csv* com os seguintes dados:

date	close	open	high	low	volume	var
2018-05-30	77239.75	76779.38	78168.54	75524.43	6,23M	0,63%

kaggle

Resultados - Correlações

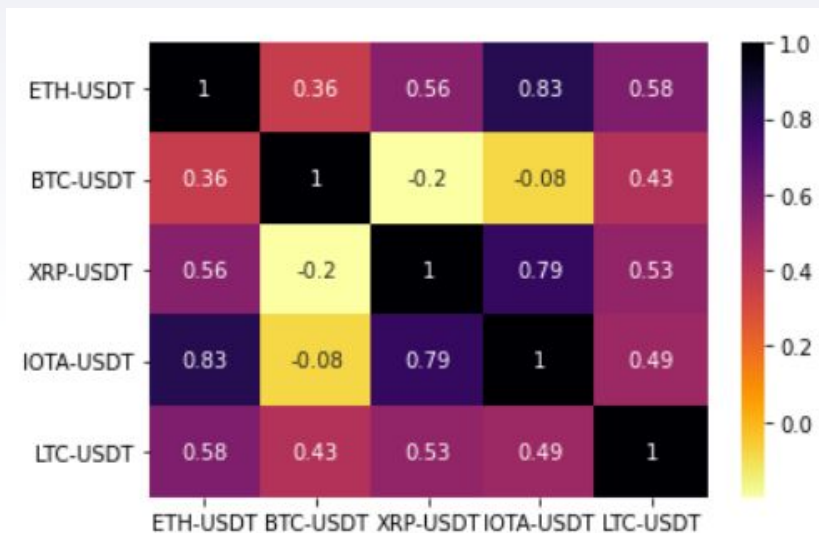


Fig.1 - Correlação de criptomoedas com preço em dólar

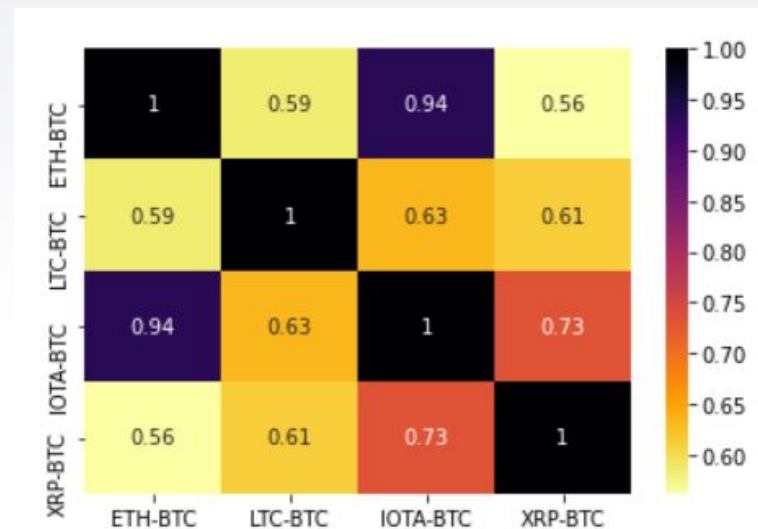


Fig.2 - Correlação de criptomoedas com preço em BTC

Existem dois pontos importante a se notar com base nos mapas de calor das correlações:

- As correlações variam dependendo do valor base de preço que se utiliza;
- No geral, a correlação das criptomoedas aumenta quando olhamos seu valor em preço de BTC.

Resultados - Correlações

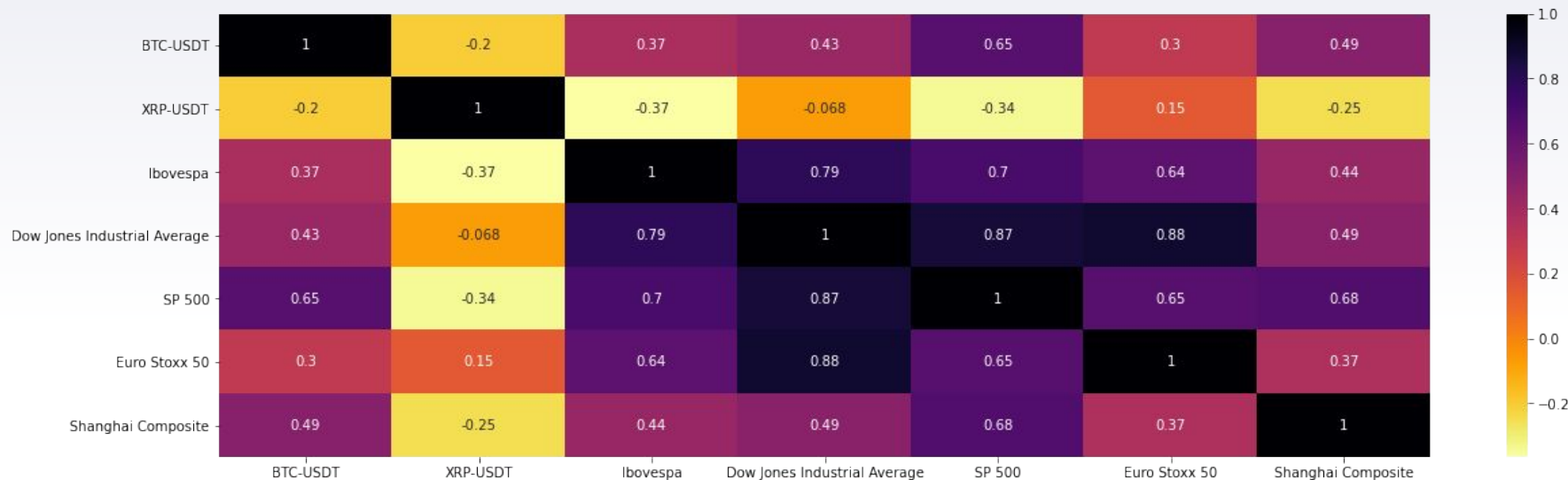


Fig.3 - Correlação de criptomoedas e índices das bolsas

- Vemos que o Bitcoin tem baixas correlações com os índices das bolsas, o que pode ser compreendido devido ao fato de não possuir lastro como os índices.

Resultados - Forecast

Para realização do forecast decidimos selecionar apenas algumas criptomoedas para análise de impacto, e escolhemos duas para realizar o forecast. Além disso tratamos todas as bases de forma diária. Mostraremos apenas os resultados mais relevantes, devido ao limite de slides.

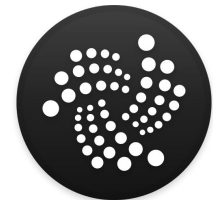
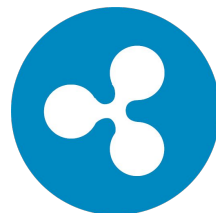
As moedas escolhidas para predição foram:

- Bitcoin
- Ripple

E utilizamos as seguintes moedas para agregar mais variáveis e entender o impacto de uma na outra:

- Litecoin
- IOTA
- Ethereum

Além disso, como já comentado, também introduzimos algumas variáveis dos índices de grandes mercados.



Resultados - Forecast (individual)

Quando realizamos a predição do valor futuro usando apenas variáveis atrasadas da moeda em estudo (**XRP**), encontramos uma relação que já esperávamos: **a variável de um dia atrás é a primeira mais importante, explicando praticamente toda predição.**

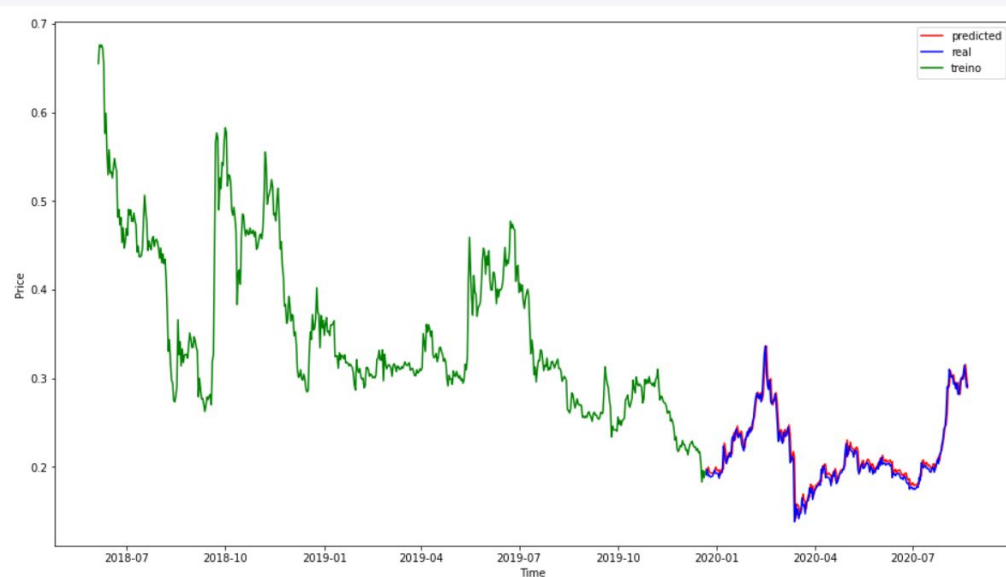


Fig.4 - Visualização da base utilizada para treino, da base utilizada para teste e dos resultados preditos por meio da regressão linear. RMSE = 0.0098, representando um erro de cerca de 4.5%

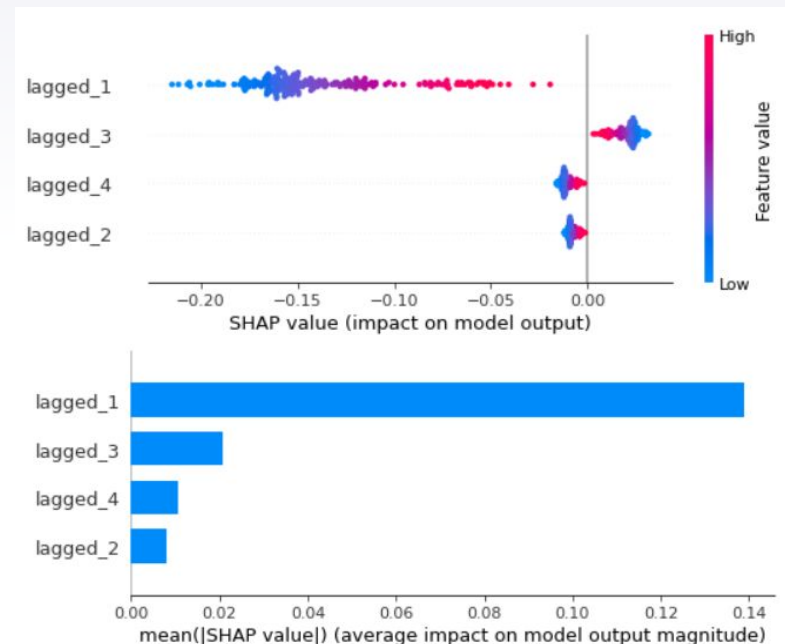


Fig.5 - Impacto das variáveis. Vemos que o dia anterior é que o apresenta maior impacto na predição. Enquanto a variável de 2 dias atrás é a que apresenta a menor importância.

Resultados - Forecast (individual)

Quando analisamos o **Bitcoin** e suas variáveis atrasadas, encontramos o mesmo tipo de resultado relatado no caso da XRP, **a variável de 1 dia atrás era a mais representativa para a predição do modelo**. Contudo, vemos uma diferença, onde a ordem de dias atrasados representa exatamente a ordem de importância.



Fig.6 - Visualização da base utilizada para treino, da base utilizada para teste e dos resultados preditos por meio da regressão linear. RMSE = 340.69, representando um erro de cerca de 3,86%

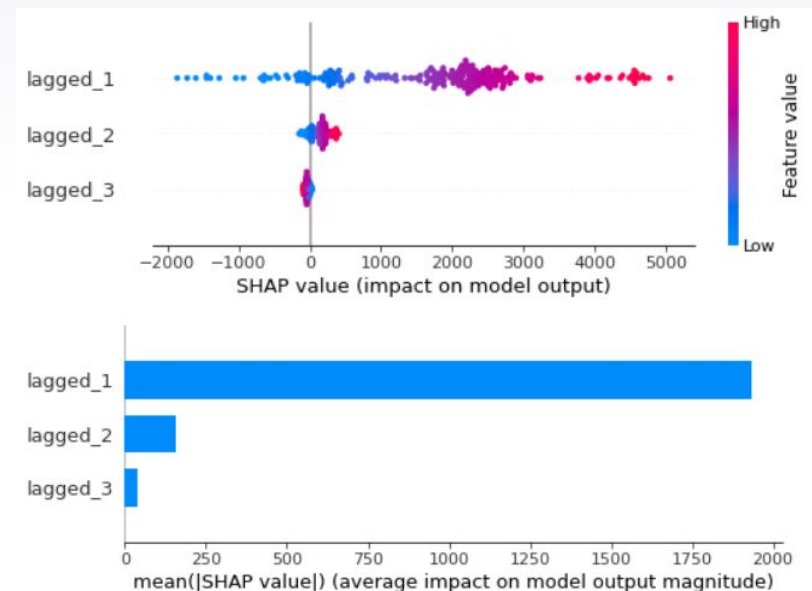


Fig.7 - Impacto das variáveis. Vemos que o dia anterior é que o apresenta maior impacto na predição, seguidos pelos dias anteriores em ordem de atraso.

Resultados - Forecast (pares)

Nessa etapa realizamos a predição de uma criptomoeda utilizando variáveis atrasadas de si mesma e de outra criptomoeda. Depois, invertemos a predição para analisar o impacto de uma sobre a outra. Começaremos mostrando a análise do **Bitcoin** frente às demais moedas.

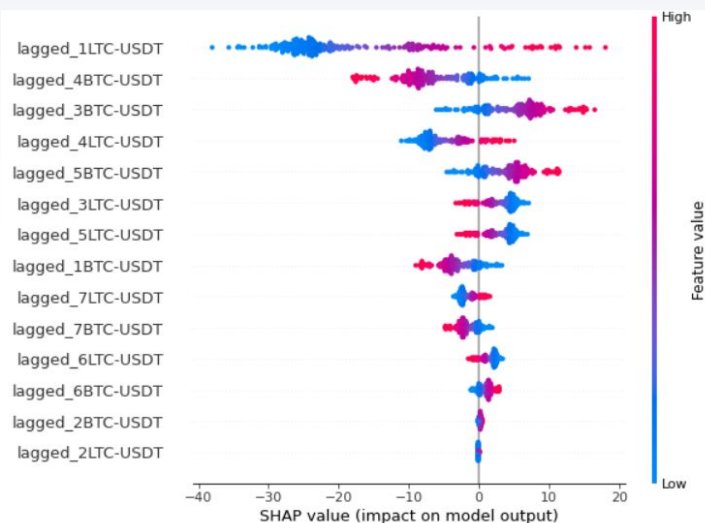


Fig.8 - Predição do Litecoin usando variáveis atrasadas do Bitcoin. Podemos verificar que o Bitcoin tem uma grande importância para essa predição. Verificamos que valores positivos do Bitcoin de 4 dias atrás impactam mais negativamente o valor do Litecoin atual do que valores positivos. Já para o caso do valor de 3 dias atrás do Bitcoin, vemos que valores positivos impactam positivamente o preço do Litecoin. Ou seja, possivelmente existe uma relação entre o preço atual do Litecoin com as variações dos preços de dias atrás do Bitcoin.

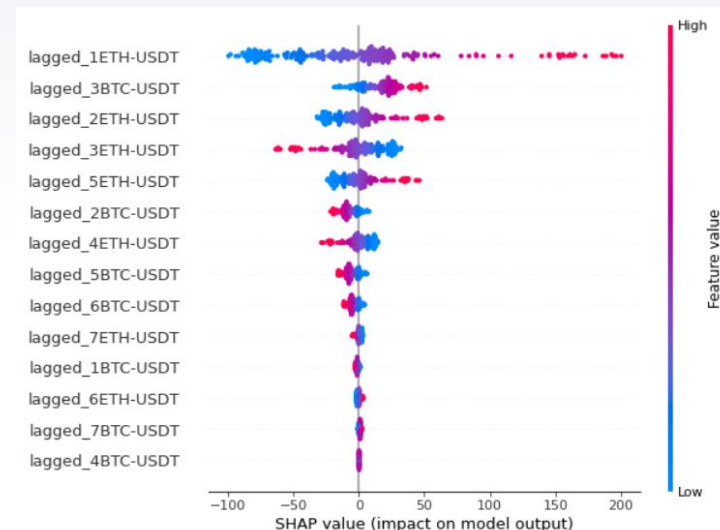


Fig.9 - Predição da Ethereum usando variáveis atrasadas do Bitcoin. Podemos verificar que os valores atrasados do Bitcoin tem bastante relação com os valores atuais da Ethereum. Mais precisamente, vemos que valores de 3 dias atrás impactam o modelo de maneira semelhante, mas em escala menor, que o valor de 1 atrás da Ethereum.

Resultados - Forecast (pares)

Já para o

Fig.4 - Visualização da base utilizada para treino, da base utilizada para teste e dos resultados preditos por meio da regressão linear. $RMSE = 0.0098$, representando um erro de cerca de 4.5%

Fig.5 - Impacto das variáveis. Vemos que o dia anterior é que o apresenta maior impacto na predição. Enquanto a variável de 2 dias atrás é a que apresenta a menor importância.

Resultados - Comparação de modelos

Já para o

Webapp

- Protótipo inicial de ferramenta para auxiliar na análise e forecast de criptomoedas;
- Desenvolvido utilizando streamlit;
- Pode ser expandido para análises mais complexas e também podem ser adicionados novos modelos.

