Bitcoin Price Modeling and Forecast

•••

Springboard Data Science Track - Capstone Project 2

Julio Moreno
February 28, 2021

Project objective:

Build a model that can determine what fundamental variables have the most impact on the bitcoin price, and that can forecast the bitcoin price within a 90%+ accuracy.

Why should we care about the Bitcoin Price?

Rank	Name	Market Cap
1	Gold GOLD	$11.008 T
2	Apple AAPL	$2.036 T
3	Saudi Aramco 2222.SR	$2.025 T
4	Microsoft MSFT	$1.753 T
5	Amazon AMZN	$1.557 T
6	Silver SILVER	$1.459 T
7	Alphabet (Google) GOOG	$1.368 T
8	Bitcoin BTC	$871.60 B
9	Tencent TCEHY	$858.88 B
10	Facebook FB	$733.61 B

https://companiesmarketcap.com/assets-by-market-cap

Bitcoin's current market cap is similar to the most valuable companies and assets on earth.

In its 12 years of existence, it has climbed in from $0 to above $50,000 per coin.

Bitcoin's value has increased along the growth in its usage (users, transactions, etc) and its intrinsic monetary properties (decreasing supply of new bitcoins, decentralization, censorship-resistant, etc).

Who cares?

Companies, Investors

Companies, HNWI, investment funds, etc, have added bitcoin to their balance sheets or have a financial interest in bitcoin.

General Public

Around 100 million people around the world have put part of their savings into bitcoin.

Data source and Analysis

We use demand and supply metrics to model the price

Usage/demand features

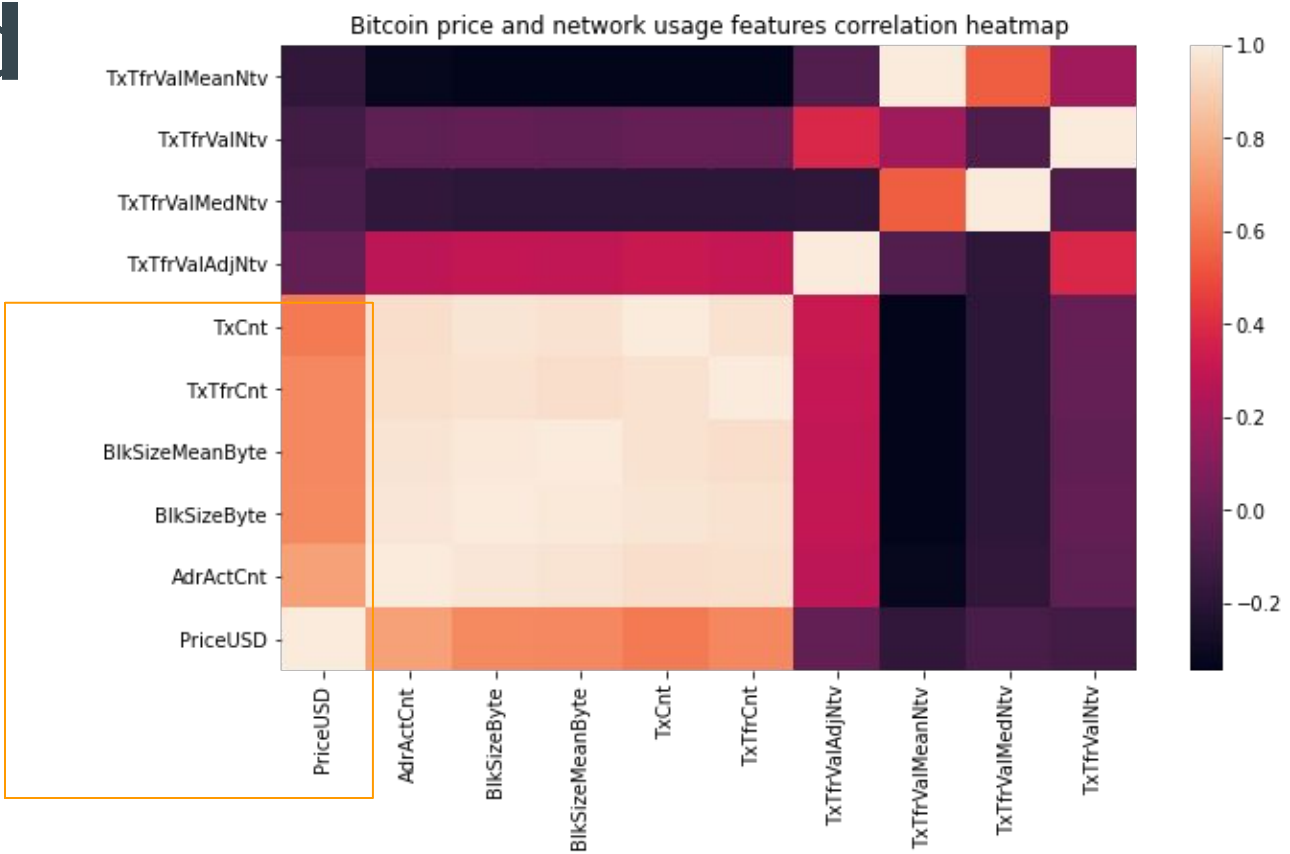
- Addresses (proxy for number of users).
- Block size
- Transactions (number, value, transfers).

Supply features

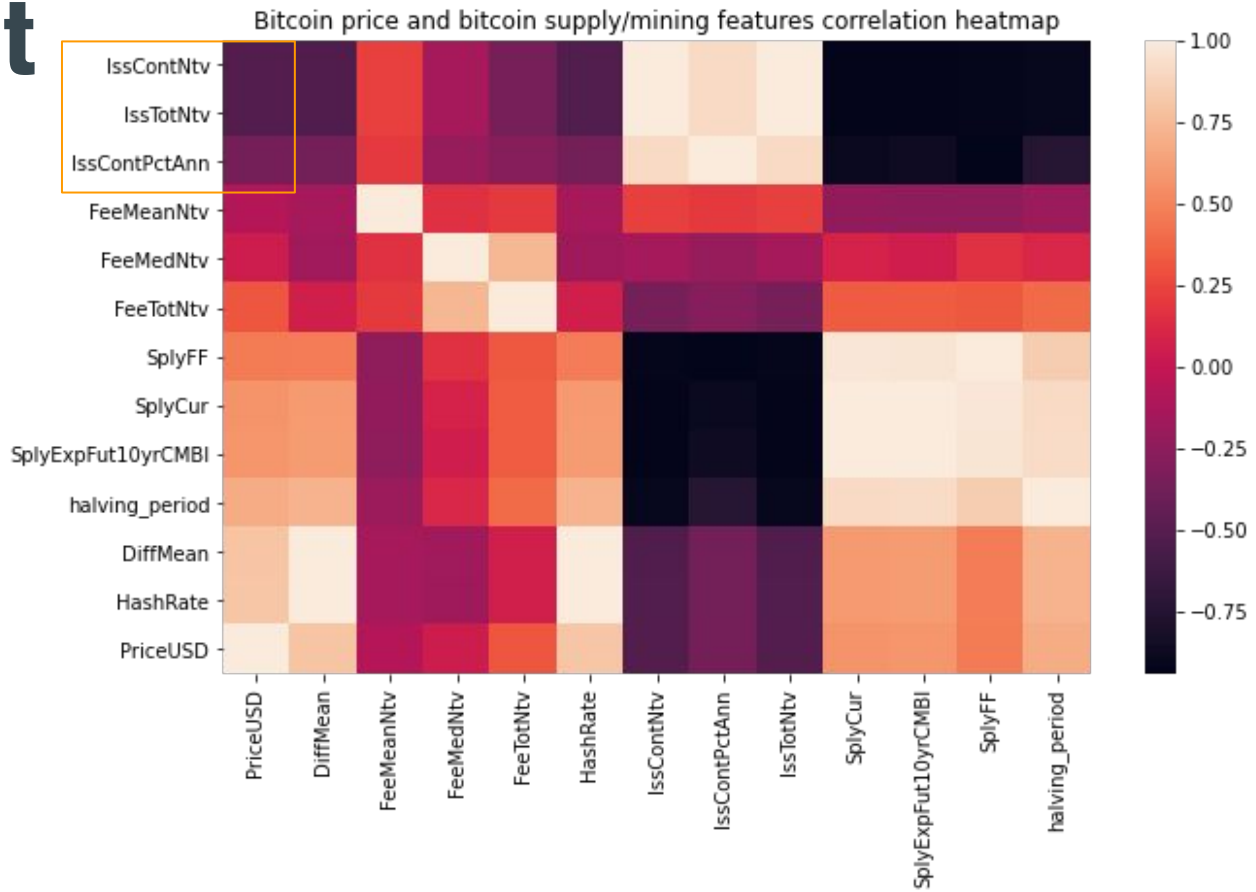
- Hashrate (bitcoin miners activity).
- Difficulty of mining.
- Features related to issuance or the growth in new bitcoins created.

Data Source: COINMETRICS

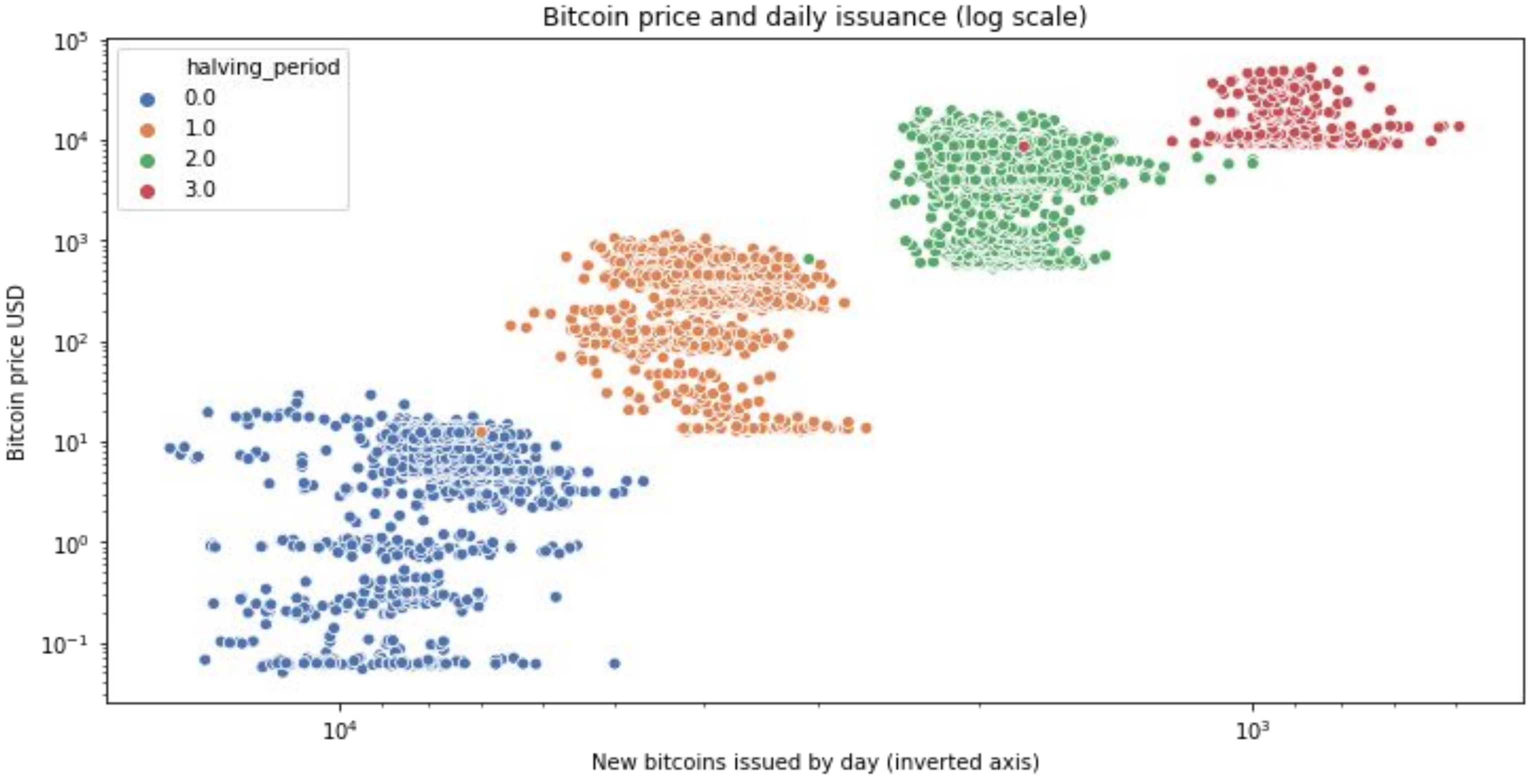
Addresses, block size and transactions have the highest positive correlation with bitcoin price.



Features about bitcoin issuance have a high negative correlation with price.



Price seems to cluster as the number of bitcoins being issued (IssContNtv) decreases over time.



The dataset had to be transformed in various ways to make the features stationary, necessary for time series modeling

To make the series stationary we performed a series of transformations to the dataset:

- Calculating first differences.
- Scaling the data (using sklearn StandardScaler) and calculating first differences.
- Calculating 4-year % change.

Under all transformations, price and most features series of interest resulted stationary. Stationarity was tested using Statsmodels' Augmented Dickie-Fuller test.

Feature engineering: 2 new features were created

Number of active addresses squared (AdrActCnt2)

Test if the price of bitcoin could be modeled per Metcalfe law, which states the value of a network is directly proportional to the square of its users.

Stock-to-Flow ratio (S2F)

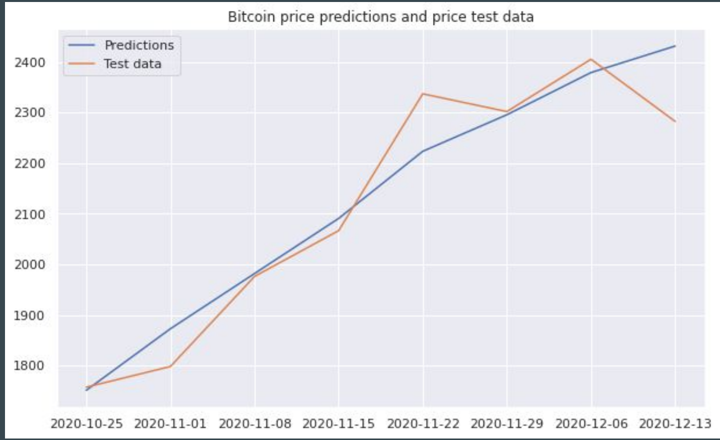
S2F is the number of bitcoins in existence divided by the annualized issuance of new bitcoins (SplyCur / IssCntNtv).

This feature captures bitcoin's scarcity, measuring how big the current stock is compared to the growth in the supply of bitcoin.

Modeling and model selection

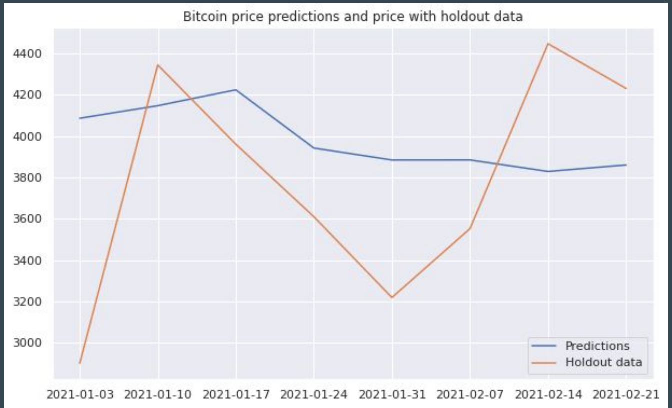
The best model according to the test set scores only included the square of addresses (AdrActCnt2) and the bitcoin price (PriceUSD) as the explanatory variables.

Model	Vector Autoregressions (VAR)
Features	AdrActCnt2 and PriceUSD
Data units and periodicity	4-year % change, weekly data.
Test set scores	r2: 0.91 Explained variance: 0.91 MAE: 50.7 MSE: 5,262 RMSE: 72.5



...but scores worsened significantly when predicting with the holdout set.

Model	Vector Autoregressions (VAR)
Features	AdrActCnt2 and PriceUSD
Data units and periodicity	4-year % change, weekly data.
Holdout set scores	r2: -0.24 Explained variance: -0.10 MAE: 495 MSE: 336,857 RMSE: 580.3



Conclusions

Most important feature

- Under our model, the number of Addresses (proxy for users) squared is the feature under which the model performed best.

Model performance

- Our model performed well in the test set, but significantly decreased its performance in the holdout set.

FUTURE ANALYSIS

More data granularity	• Future work could be done by having access to more granular data of the network.
Use other type of features.	• As bitcoin has become a major financial asset, price movements could now be more influenced by features outside the network.
Test other types of models.	• These models may include ARMA, ARIMA, Non-linear models, etc. Other option is also to try classification models for time series.