
CRYPTO CURRENCIES - MARKET ANALYSIS

Predicting crypto currencies' prices with Machine Learning algorithms

Author: Jane Nikolova

Date: August, 2023



[This Photo](#) by Unknown author is licensed under [CC BY-SA](#).

CONTENT



I. Objectives of the analysis

II. Initial Analysis

- Extreme values – Bitcoin and Bitcoin Cash
- Treatment of extreme values and random sampling – Approach to Analysis
- Data quality

III. Predictions & Validation:

- Linear Regression, Support vector machines, Stochastic Gradient Descent, Gradient Boosting, Neural Networks

IV. Conclusion

- Recommendation for the best performing predictive model for modeling of crypto currencies High daily prices.
 - Comment of efficiency of algorithms
-

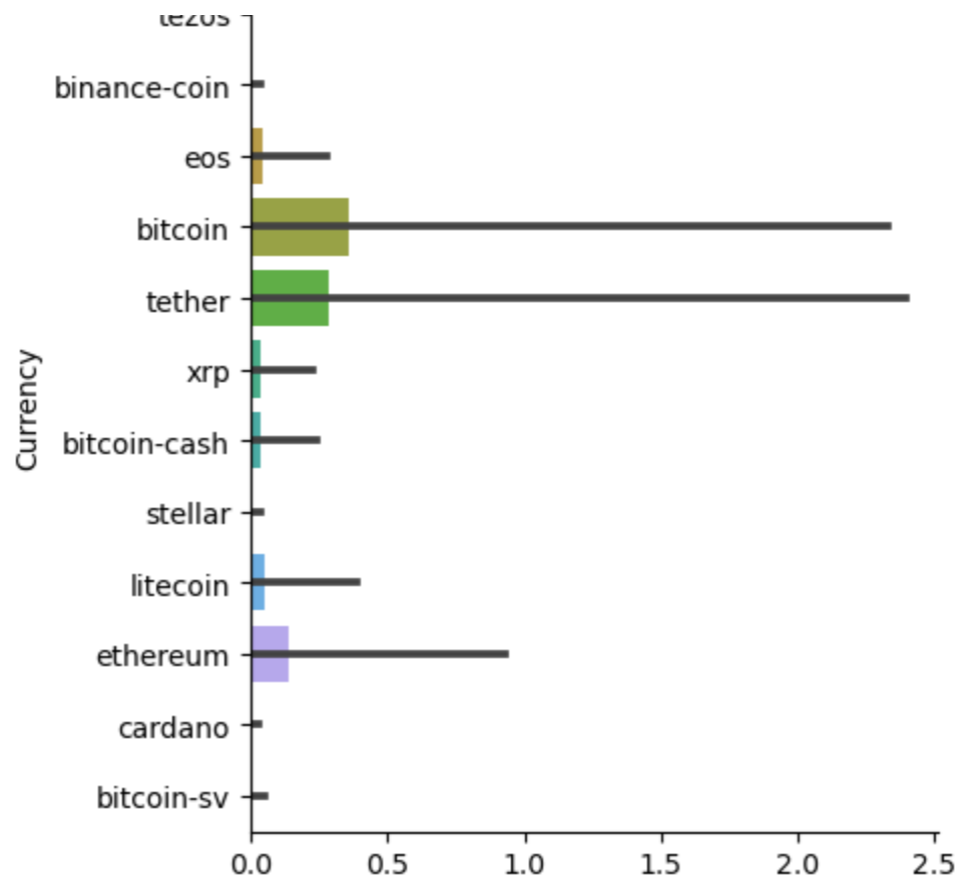
OBJECTIVES

- The main goal of this analysis is to predict highest daily price for each crypto currency from the list.
- To achieve this, several ML (Machine learning) algorithms are explored and a best performing algorithm is recommended based on its accuracy rate.

- tezos
- binance-coin
- eos
- bitcoin
- tether
- xrp
- bitcoin-cash
- stellar
- litecoin
- ethereum
- cardano
- bitcoin-sv



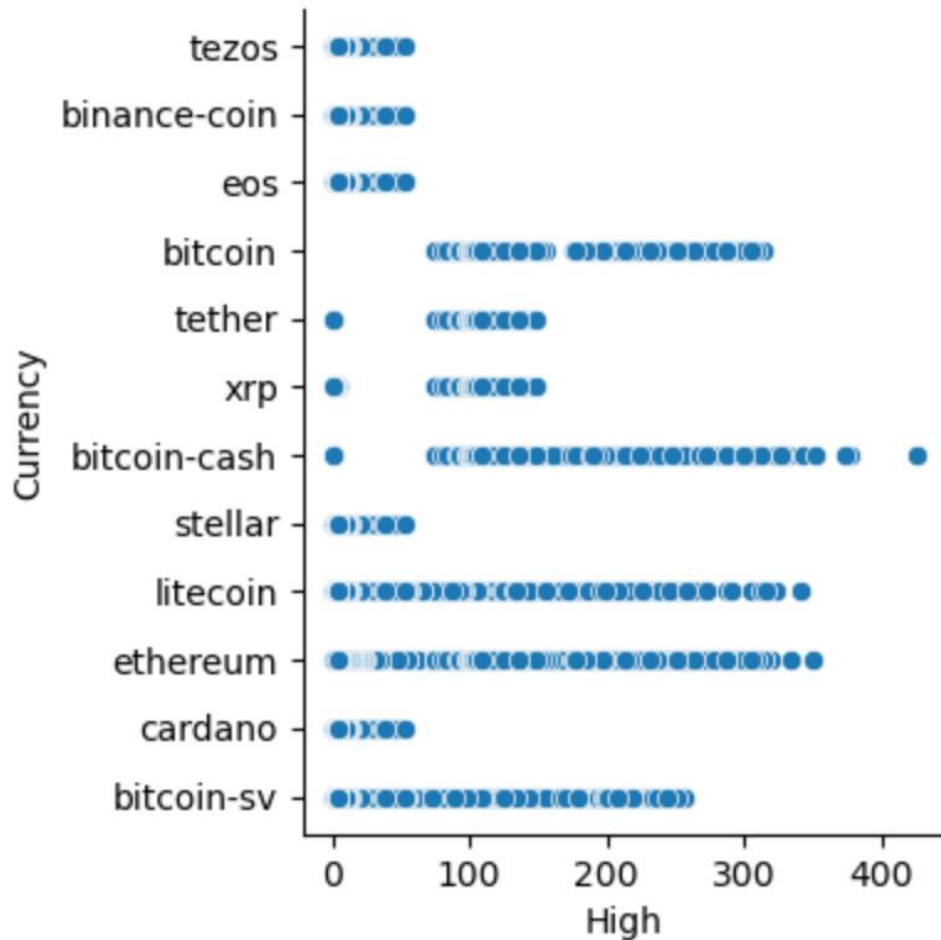
This Photo by Unknown author is licensed under [CC BY-SA](#).



INITIAL ANALYSIS

Extreme values:

- The dataset has 28,944 records from April 2013 to December 2019. Bitcoin is an outlier in the data. It makes sense to analyze this crypto currency on its own.
- However, for the sake of looking at the entire crypto market, all currencies are included in the analysis and outliers are cleaned if Z-Score is higher than 3.
- This changes the predicted market values of Bitcoin but also protects from risky or speculative investment in the same, considering the entire crypto market trends and current value.



INITIAL ANALYSIS

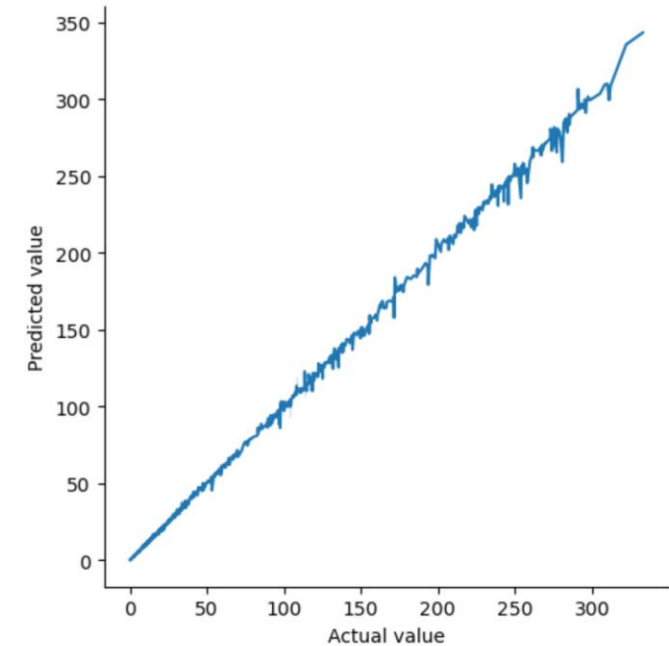
Treatment of extreme values and random sampling:

- After cleaning the data from outliers, the number of observations becomes 25,324. For some algorithms and visualizations, this means that a random sample must be drawn to complete all calculations.
- As mentioned, after pre-processing, the data is 'zoomed' to all currencies except for Bitcoin and Bitcoin Cash which have extreme price values and are treated as outliers.
- Hence, the focus of this analysis is intentionally shifted towards the analysis of the rest 10 currencies and Bitcoin and Bitcoin Cash are recommended for individual analysis (outside of the scope of this work).

PREDICTIONS – LINEAR REGRESSION

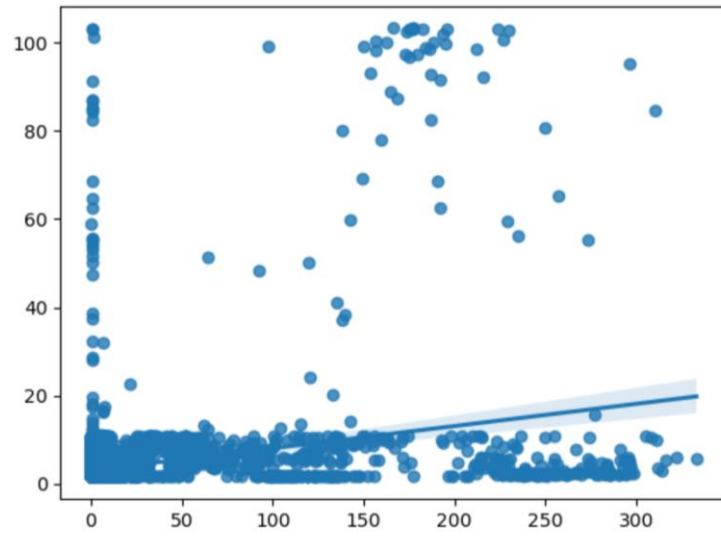
Steps of the advanced analysis:

- Encode the categorical variables
 - Standardize/Scale the numerical variables to make them comparable.
 - Split the data into model training and model testing (validation) sets using random sampling and 20% size for the testing (validation) set.
 - Generate a LR model with the training sets
 - Fit the model
 - Estimate the model accuracy with the testing dataset
-

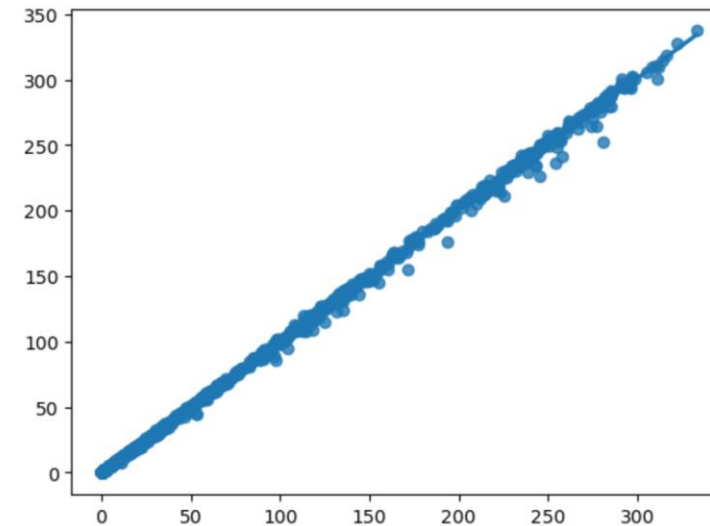


- Percent variance explained: 99.93%
- Mean Absolute Error: 0.33 (low)
- Mean Square Error: 1.39 (acceptable)
- Root Mean Square Error: 1.18 (acceptable)

PREDICTIONS – SUPPORT VECTOR MACHINES AND STOCHASTIC GRADIENT DESCENT

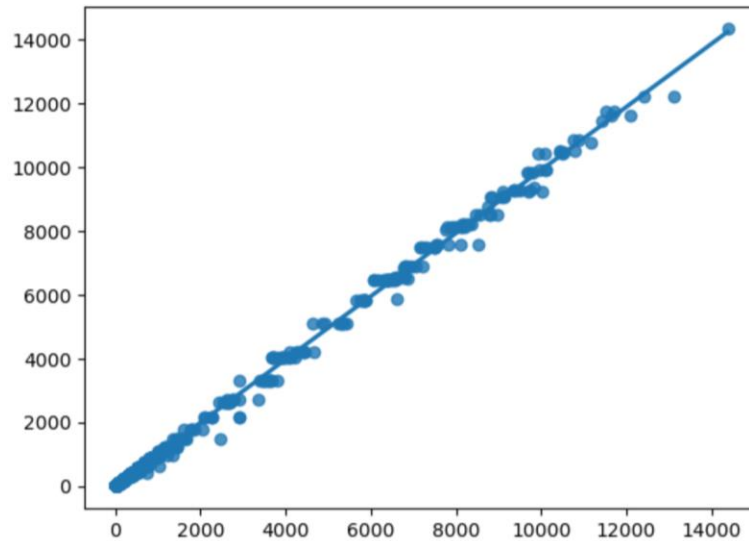


- MAE: 21.60
- MSE: 3275.40
- RMSE: 57.23

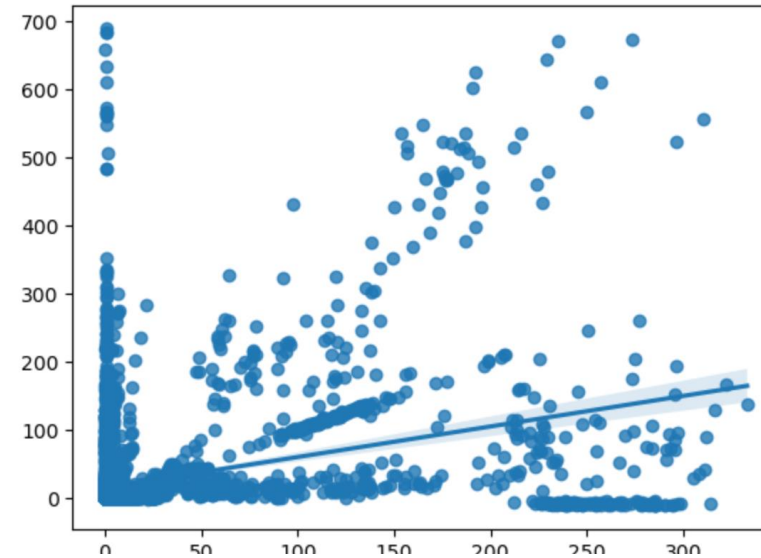


- MAE: 0.39
- MSE: 1.7
- RMSE: 1.29

PREDICTIONS – GRADIENT BOOSTING AND NEURAL NETWORKS



- MAE: 0.91
- MSE: 4.97
- RMSE: 2.23



- MAE: 28.31
- MSE: 5656.53
- RMSE: 75.21

SUMMARY OF RESULTS

- The best performing model is Linear Regression
- One may choose and more conservative approach to avoid OVERFITTING and use Stochastic Gradient Descent or Gradient Boosting instead.

Algorithm	MAE	MSE	RMSE
Linear Regression	0.33	1.39	1.18
Support Vector Machines	21.60	3275.40	57.23
Stochastic Gradient Descent	0.39	1.7	1.29
Gradient Boosting	0.91	4.97	2.23
Neural Networks	28.21	5656.53	75.21



CONCLUSION

- Each dataset has its specifics and not all algorithms perform well on a target dataset.
- The crypto currencies high daily price can be modeled using supervised learning algorithms such as linear regression and gradient descent and boosting.
- A separate analysis is required to find the best algorithm that can model Bitcoin pricing.
- Data enrichment, such as the addition of S&P500 data or other market data, can be applied to this analysis to enhance results.

ABOUT THE AUTHOR

- Jane Nikolova
- *LinkedIn:*
www.linkedin.com/in/nikolova2023
- *GitHub:* <https://github.com/joanna-2023>

