CRYPTOCURRENCIES PRICE MOVEMENT PREDICTION USING MACHINE LEARNING

Shilpa Nair[1], Siddharth Nanda[2]
[1]U.G. Student, School of Engineering, Ajeenkya DY Patil University Pune, India -4121052
[2]Faculty, School of Engineering, Ajeenkya DY Patil University Pune, India -4121052

Journal	Samvakti Journal of Research in Information Technology https://www.sjrit.samvaktijournals.com ISSN : 2583-3979 (Online) Volume 4 Issue 2 Year of Volume 2023 Page No : 1 - 8
Discipline	Artificial intelligence and Machine Learning

Date Received : July 17, 2020 **Publication Date**: November 31, 2023
ID : 202007.51.277 **Paper Type** : Short research paper

ABSTRACT

Crypto currencies are receiving a lot of attention from media due to its importance. Several papers have similar prices but some others don't, but the time complexity of those forecasts is greater, so we use an artificial intelligence-linked algorithm called the least absolute shrinkage selection operator here in this paper to reduce the time complexity. The enormous section of research is crucial part focuses thoroughly on the nature of the cryptocurrency. In this observation, we introduce the usage of regular AI devices and accessible online networking information to predict the development of value of Bitcoin, Ethereum, Ripple and Litecoin's digital currency showcase developments. We are also using the neural networks, SVM and Random Forests while analyzing the components from Twitter.

Keywords: Crypto Currency, Machine Learning, Prediction, Sentiment Analysis, Price Movement.

INTRODUCTION

Cryptocurrency is basically a digital asset or resource which is notably produced to work as a medium of exchange where a person or an individual coin and compose partnership and control purchase records are and will be stored in very high secured computerized database with the help of strong cryptography to protect the transactions records. Given the very fact that there are a couple of researches and studies concerned with both the undertaking to estimate volatility in exchange costs also The

chances of financial status of some of these Cryptocurrencies adds very concerning details about particular markets, and projections consistent with which cryptocurrencies function in separate fields were also acknowledged Digital currency advertising appears to be behaving independently with the opposite capital markets. Asian markets economy definitely affects due to this. One of the exciting parts would be that about the mining of digital currency helps to choice in the customary markets, for example we can say gold. However, such qualities have drawn tons of resources, thus far there aren't many tests that have tried to form the cryptographic money advertise competitive exchange schemes. Another focus within the digital currency showcase is that the tremendous scale of obtainable open slant information, especially from interpersonal organisations. Maybe this information are often wont to deduce future act and may thus be wont to establish beneficial trading techniques, as has been seen in current attempts to seek out hidden air pockets within the digital money showcase using estimation research.

LITERATURE SURVEY

The market capitalizations for all the cryptocurrencies and digital assets have already topped over $1 trillion according to the data from Coindesk. Stock market expectation has always been viewed as tough and risky step that stood out.

Aggregate stock-market returns are not as expected using Huang Partial-Least-Squares Sentiment File as per Example, in (Bekiros, pp. 10-13), which sums up data from six proxies. (Mark, An-Sing, & Hazem, 2001) (Lahmiri & Bekiros, 2018) (Salim & Stelios, 2019) (Stelios & Dimitris, 2008) (Saad, Prokhorov, & Wunsch, 1996) (Ray, Yenshan, & Charles, 1998)

In [2] it’s observed that costs in the Bitcoin industry escalate and display riotous elements and level of uncertainty in returns significantly increased, rather than profits, during the big expenditure system era. The undertaking’s lack of predictability is also a contribution to such insecurity that affected the economy, including fiscal and political conditions. But in reality, anticipating the market stock contrasted with its worth can bring about higher profits [3]. Past few decades, Deep learning and AI techniques have been a major use to predict and forecast in stock market. The goal of creating this model is to demonstrate that how a prepared computer model will predict the cost of a digital asset on the off-risk if we get the perfect measure of knowledge and computational impact. It presents a map which contains the qualities anticipated. With significant measure of evidence being generated and reported on an ongoing basis, we eventually reached a time in which estimates can be reliable and can be made on the basis of real. These strategies rely on well-established and functional approaches (such as SVM’s) while apparently some others may not. Many observable procedures adopt the decision-making principle with a common probabilistic process that best represents knowledge inside a particular collection of models.

METHOD

Market Data

For creating this model and compiling it, one of the most crucial part was to understand the environment of trading and market knowledge. After this, the data from the top-performing cryptocurrency transactions and records was obtained using the free API cryptocompare.com, which helps us to view the data and gives permission to use it as a important detail for cryptocurrency trading and it also has every information about the transactions which in further will help to predict the movement. The entire data that would be reported will be obtained in feedback to public demand. The acquired information might be listed in a format of hour-to-hour or it can be even in day-by-day format. The given dataset will also include the initial cost, the most important expenditure, the least benefit, the cost of shutting and the amount of trade for each time stage.

Social Data

Social Media related gathered details regarding the data as Direct Messages. The following integration measures are selected when implementing tweets:

1. During the timeframe, the review was undertaken: in any situation, previous tweets are not mentioned if they affect present action. These data will be used for further study.

2. Using the label, for example ‘Litecoin’ or ticker impression ‘ltc’ for one of the broken commercial norms in one of the fields of information or labels: this provides a expressing degree of certainty that the tweet is in some case related to one of the cryptocurrencies to which the connection is established.

3. Is not duplicated: that means that re-tweet have been allowed because this may signify a sentimental pattern, copied tweets have not been taken into consideration as bot accounts actually display this management. Although the venture of each other affecting cryptocurrencies, we also learned that tweets from another cryptocurrency would be less important than tweets when it comes to breaking down the cryptocurrency. To make Twitter's details one of a kind to boost cryptocurrency, tweets were not to a certain extent seen as what cryptocurrency they had with when they were getting.

FRAMEWORK

A model is created to assemble each collected data from the different yet valid sources of information to create a single indicator of information that combines both business and social statistics. For this, given the market or can be called as the business index and social statistics, an element vector V for a specific timeframe t is defined as:

$$[\boldsymbol{V(t)} = (\boldsymbol{ne, nor, ng, ps, pl, cls, hg, lw, opn, vlmto})]$$

Equation 1 : Feature Vectors Equation

Where,

ne = $\frac{\sum_{i}^{n}=1^{t}ne}{n}$

ng = $\frac{\sum_{i}^{n}=1^{t}ng}{n}$

nor = $\frac{\sum_{i}^{n}=1^{t}nor}{n}$

ps = $\frac{\sum_{i}^{n}=1^{t}ps}{n}$

pl = $\sqrt{V_{ps}V_{ng}}$

cls is the expenses that would be spent while closing transaction in the time span
hg is the capital cost also known as maximum cost in the time span
lw is the base cost also known as minimum cost in the time span
opn is the price of security in the beginning of trading day in the time span
vlmto is the volume of financial assets for the time span

Support Vector Machines

SVM also known as Support Vector Machines is a supervised Machine Learning Algorithm. Its used for two group classification problems. Classifications and recurrence problems also can be solved with the help of SVM's. SVM will alter the data that is available to a high-dimensional object zone using a bit of effort. The choice capacity is:

$$\boldsymbol{k = sgm(\sum_{i=1}^{n} k_i \partial_i K(x_i, x) + \rho)}$$

Equation 2 Decision Function SVM

Where,

k is to characterize the label for classification (1 or −1),
m is the number of training vectors used in Support Vector Machines,
∂ is a Lagrange multiplier,
$\boldsymbol{K(x_i, x)}$ is the Kernel function,
ρ is for the maximum margin decision boundary intercept.

Testing of Model

In Machine Learning, data is usually divided in two parts. Some of the data from the collection would be used for training the model so that it can get acquainted and the other data would be used for testing the model. Now, for time the information index also known as data index; included a business schedule and Twitter information. For making, preparing and training, the data and information collection was separated through a 70–30 split where 70 per cent of the expertise is used to prepare and 30 per cent is used for study.

RESULTS

The precision, precision, recall, and f1 scores are used to assess each model's robustness as stated below:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{t_p} + \boldsymbol{t_n}}{\boldsymbol{t_p} + \boldsymbol{t_n} + \boldsymbol{f_p} + \boldsymbol{f_n}}$$

Equation 3 :Accuracy Equation

$$\boldsymbol{Precision} = \frac{\boldsymbol{t_p}}{\boldsymbol{t_p} + \boldsymbol{f_p}}$$

Equation 4 : Precision

$$\boldsymbol{Recall} = \frac{\boldsymbol{t_p}}{\boldsymbol{t_p} + \boldsymbol{f_n}}$$

Equation 5 :Recall

$$f_1 = 2\frac{\boldsymbol{precision} * \boldsymbol{recall}}{\boldsymbol{precision} + \boldsymbol{recall}}$$

Equation 6 :f1 scores equation

Where,

$\boldsymbol{t_p}$= Number of true positive values

$\boldsymbol{t_n}$= Number of true negative values

$\boldsymbol{f_p}$ = Number of false positive values

$\boldsymbol{f_n}$ = Number of false negative values.

(Valencia, 2019)

Precision calculates the expanse of all cadent correctly, faultlessness of the expanse of all attempts are necessary. All the redeemed occurrences in the test are exam mined in percentage of suitable graded evaluations in the overall data and consistent data. F1 can be described as the basic diadem of accuracy and study. Consistency still is the most important metric, as it indicates how good we were correct about our predictions and can work out what kind of market demand a plan will generate.

These are the results of applying SVM using the information we obtained from the Twitter and the Cryptocurrency market data, which we have used to forecast the daily price movement of bitcoin

Model	Accuracy(95% CI)	Faultlessness	Recall	F_1 Score
SVM for Twitter Data	0.50(±0.03)	0.29	0.50	0.37
SVM for Market Data	0.55(±0.03)	0.53	0.56	0.47
SVM Twitter and Market	0.55(±0.03)	0.31	0.56	0.40
Random	0.50(±0.28)	0.49	0.50	0.50
Majority	0.55(±0.0)	0.31	0.56	0.40

Table 1 : Results of Bitcoin

These are the results of applying SVM using the information we obtained from the Twitter and the Cryptocurrency market data, which we have used to forecast the daily price movement of Ethereum.

Model	Accuracy(95% CI)	Faultlessness	Recall	F_1 Score
SVM for Twitter Data	0.39(±0.03)	0.13	0.37	0.20
SVM for Market Data	0.39(±0.03)	0.13	0.37	0.20
SVM Twitter and Market Data	0.39(±0.03)	0.13	0.37	0.20
Random	0.50(±0.28)	0.52	0.48	0.47
Majority	0.61(±0.0)	0.35	0.59	0.44

Table 2 : Results for Ethereum

These are the results of applying SVM using the information we obtained from the Twitter and the Cryptocurrency market data, which we have used to forecast the daily price movement of Ripple

Model	Accuracy(95% CI)	Faultlessness	Recall	F_1 Score
SVM for Twitter Data	0.56(±0.04)	0.62	0.58	0.51
SVM for Market Data	0.51(±0.04)	0.55	0.52	0.43
SVM Twitter and Market	0.57(±0.04)	0.25	0.52	0.34
Random	0.51(±0.28)	0.57	0.52	0.47
Majority	0.53(±0.0)	0.23	0.52	0.34

Table 3 : Results for Ripple

These are the results of applying SVM using the information we obtained from the Twitter and the Cryptocurrency market data, which we have used to forecast the daily price movement of Litecoin

Model	Accuracy(95% CI)	Faultlessness	Recall	F_1 Score
SVM for Twitter Data	0.52(±0.04)	0.50	0.52	0.42
SVM for Market Data	0.52(±0.04)	0.25	0.52	0.31
SVM Twitter and Market	0.66(±0.04)	0.80	0.66	0.61
Random	0.50(±0.28)	0.50	0.51	0.51
Majority	0.50(±0.0)	0.25	0.54	0.34

Table 4 : Results for Litecoin

CONCLUSION

In this paper, we determined that machine learning and a well created model can be very profitable and beneficial to predict and forecast the value of different kind of cryptocurrencies and bases on the industries. We also determined about the various prediction models based on different type of cryptocurrencies like Bitcoin, Ethereum, Ripple's, Support Vector Machines with the help of the data we collected from the twitter as re tweets and using sentimental analysis. There's a lot of research to be completed in there. Redundancy cuts and material removals from bots or ads can increase consistency. The utilization of content for certain public organizations, such as Reddit or Twitter, is feasible and would definitely be beneficial, as seen previously. There are even different kinds of networks that would be willing, as it means, to pick and alter the underlying "state of mind" of the market.

REFERENCES

[1] Bekiros, G. K. (n.d.). A non-Linear approach for predicting stock returns and volatility with the use of investor sentiment indices.

[2] Kazuhiro, K., Tsutomu, I., Yoshimi, F., & Yukihiro, N. (1998). Stock Price Prediction Using Prior Knowledge and Neural Networks. *doi.org* .

[3] Lahmiri, S., & Bekiros, S. (2018). Randomness and Multi-Fractality in Bitcoin Market. *Chaos Solitons Fractals* .

[4] Mark, T., An-Sing, C., & Hazem, D. (2001). Application of Neural Netwroks to an Emerging Financial Market. *Computers & Operations Research* .

[5] Ray, T., Yenshan, H., & Charles, C. (1998). Forecasting S&P 500 Stock Index Futures With a Hybrid AI System. *doi.org* .

[6] Saad, E., Prokhorov, D., & Wunsch, D. (1996). Advanced Neural Netwrok Training Methods For Low False Alarm Stock Trend Prediction. *International Conference on Neural Networks.* Washington, DC, USA: IEEE.

[7] Salim, L., & Stelios, B. (2019). Cryptocurrency Forecasting with Deep Learning Chaotic Neural Networks. *Choas, Solitons & Fractals* , 35-40.

[8] Stelios, B., & Dimitris, A. G. (2008). Direction-of-change forecasting using a volatility-based recurrent neural network. *Journal of forecasting* .

[9] Tsibouris, G., & Zeidenberg, M. Back Propagation as a Test of the Efficient Markets Hypothesis. *Twenty Fifth Hawaii International Conference On System Scienes.* Kauai, HI, USA.

[10] Valencia, F. (2019). Price Prediction Movement using Machine Learning. *MDPI* , 12.

END