Convolutional Neural Network for Cryptocurrency Market Pattern Identification

[1]Che-Wei Yu (游哲維), [1,*]Chiou-Shann Fuh (傅楸善)

[1]Graduate Institute of Networking and Multimedia,
National Taiwan University, Taipei, Taiwan,
[2]Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
stevenyu.supreme@gmail.com fuh@csie.ntu.edu.tw

ABSTRACT

The cryptocurrency market is getting increasingly popular during these years. However, as more and more retail investors attend the high volatility market when institutional investors make a rug pull, retail investors are being trapped. If we can find the critical patterns in the cryptocurrency market, then we can find what the market is happening to know whether the price is going to go up or down to make the right decision. We aim to experiment with unsupervised Convolutional Neural Networks (CNNs) to identify the stock market patterns. In the past years, experts have been diving into using labels to find out which patterns lead to the rising and fall of the stock markets and find many ways to label the rising and fall. However, in this paper, the critical distinction of our model is to use the unsupervised learning which extracts latent features without labels to find out the patterns of the stock market, which would be helpful for us to see whether the computer sees the patterns by itself instead of a human. Moreover, by those patterns, we can make an induction on which patterns lead to rising and which patterns lead to falling, which would benefit us, investors.

Keywords: *Financial Computing, Machine Learning, Deep Learning, Supervised Learning, Feature Extraction.*

1. INTRODUCTION

Price prediction of commodities has been a hot issue for both academic and investment fields. There have been many theories for the past 100 years related to the price prediction of commodities, for example, the Dow theory, Elliott wave theory, and Gann's theory. The Efficient Market Hypothesis (EMH)[1] is the most commonly learned in the department of Economics. EMH presumes that all market participants are rational, and they make optimal strategies in the market. However, the presumption of EMH is not practical in the real world, as the participants in the market can be retail investors like us, institutional investors, or some retail investors with a large number of assets that can influence the price. The Adaptive Market Hypothesis (AMH) is proposed by economists to fix the error of the efficient market hypothesis. The AMH considers the information asymmetry to be the critical reason for excessive return, which means the price of commodities is predictable. An ability to find out patterns in the commodities market will benefit both investors and financial researchers. Investors, as they are capable of identifying the patterns can know whether the trend is going up or down. Researchers can predict more accurately based on those patterns. In this essay, we do not make suggestions for investment. We just share the result of our classification models.

Cryptocurrency becomes increasingly popular during these years. In 2009, Satoshi Nakamoto invented Bitcoin providing a public distributed ledger in the blockchain and pursuing the task of decentralization. Bitcoin is a decentralized digital currency that can be transferred on the peer-to-peer Bitcoin network. The precision of the price of bitcoin is up to 10^{-8}. The word Bitcoin is defined in a white paper published on 31 October 2008. Those transaction records of Bitcoin are kept by a service called mining by computational power. Each block contains an SHA-256 cryptographic hash of the previous block, thus linking it to the previous block and giving the blockchain its name.

The production of Bitcoin will reduce to half every four years, where the production of the Bitcoin reduces, leading to the rise of the price of the Bitcoin.

Therefore, the price of Bitcoin has been rising from \$2 U.S. dollars to \$20,000 U.S. dollars during the time before 2017. However, in 2017, an economic bubble occurred, which cause many retail investors to be trapped at the highest point. Moreover, a trading platform Binance was hacked in 2017, which caused the loss of a large amount of money from the trading platform itself and many users of the trading platform. What is gratifying is that the system problem of Binance was fixed and in 2020, the Bitcoin price strikes up to \$20,000 U.S. dollars again and those retail investors can get away from the trap.

Besides Bitcoin, Ethereum is also a cryptocurrency that provides decentralized smart contracts. While Bitcoin is also called digital gold, Ethereum is called digital gas. Ethereum has more applications than Bitcoin due to the functionality of smart contracts. The functionality of smart contracts enables the development of Non-Fungible Tokens (NFT), which provide decentralized price bid action of artworks online. However, there exist some problems with Ethereum, such as the high price of the gas fees. Therefore, some alternative solutions for Ethereum are produced such as Cardano (ADA) and Polkadot (DOT).

Besides Bitcoin and Ethereum, Ripple Coins (XRP) were invented to reach the goal of the fast transactions between two countries far away such as Taiwan and the United States. Bitcoin and Ethereum and those coins derived from these two are collectively called cryptocurrencies, which are currencies related to cryptography. The price action of cryptocurrencies is drastic due to the decentralization. The price of a decentralized market is not supervised and is just determined by peer-to-peer transactions. Therefore, the price is subject to media no matter whether the information is large or small. Even if there is no information, large price movements will happen. It can give those with a large number of assets to have a large space for manipulation of the prices of cryptocurrencies.

In recent years, not only Layer 1 solutions but there are also Layer 2 solutions have been provided such as Avalanche (AVAX). However, as more and more cryptocurrencies are being traded on the trading platforms, the market participants have changed and the patterns of the candlesticks of cryptocurrency prices have changed. Moreover, more and more derivatives are pushed into the crypto market such as futures and options.

Machine learning is a very useful tool for dealing with many problems such as image classification and language processing divided into three parts: supervised machine learning, semi-supervised machine learning, and unsupervised machine learning. The difference between these three is whether we have labels for each datum, some data, or no data. Traditional machine learning ways include Supported Vector Machines (SVMs), Random Forests (RFs), and ensemble learning.

Although machine learning has been a useful tool for a long time, deep learning gains popularity in recent years. Traditional deep learning methods include Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Convolutional Neural Networks are used in the task of image classification, while Recurrent Neural Networks are used in the task of language modeling. Image processing has been very successful in recent years. GoogleNet, AlexNet, ResNet, and so on, have been very successful in image classification. The language modeling evolves from RNNs to Long-Short Term Memory (LSTM) to Bidirectional Encoder Representations from Transformers (BERT).

Compared with Convolutional Neural Network, Recurrent Neural Network would be more ideal for the problem of stock price prediction, because the input is the stock price data. However, the biggest problem of Recurrent Neural Network is the problem of gradient vanishing and the debugging process is troublesome for the reason that the structure is complicated especially in LSTM[2] whose unit is composed of a cell, an input gate, an output gate, and a forget gate. Moreover, it is hard to explain the information from the hidden state which would harden the process of debugging. Furthermore, the stock price market is periodic, which means that it is meaningful to slice the market data into different time frames of the same size. Therefore, in this paper, we apply a supervised convolutional neural network.

In this paper, we apply a supervised convolutional neural network to find out the critical patterns. The cryptocurrency prices will be cut into windows of fixed time elapse and will be identified out of patterns. Considering different periodicities, we should use ensemble learning with many

Convolutional Neural Networks feeding different inputs with different time frames to see different data.

2. LITERATURE REVIEW

Before deep learning becomes the main method, machine learning has been a good way of financial data analysis. SVM was a good way to make predictions because it can find some important factors from Principal Component Analysis (PCA). Moreover, SVM [3] can do the work of classification to determine whether the stock price would rise and fall in the short term. Besides SVM, there are other methods to predict the stock trends, such as random forest, ensemble learning, or artificial neural network. Moreover, [4] a genetic algorithm with the neural network was also used to make predictions, which achieves satisfactory results.

However, in recent years, deep learning becomes the most common method in dealing with problems, especially image classification and natural language processing. CNNs are usually used for image problems, while RNNs are usually used for time series problems. ResNet110 has reached 97% accuracy in image classification of dogs and cats. BERT is the common model for language tasks. Using deep learning methods has shown its superiority when compared with machine learning methods in both image classification and natural language processing tasks. The efficiency of the former is better than the latter. Moreover, the feature extraction of a convolutional neural network plus explainable AI can help us see which parts of the figure are used.

3. METHOD

Our proposed frame would be a linear neural network and a supervised convolutional neural network with many layers. As input, those cryptocurrency prices will be cut into different time frames of the same size. The time frame cannot be too large or too small. If too large, it is hard to predict due to national policy, and if too small, it is hard to predict due to too much random noise. We label each input data by the rising or the falling ratio of the next day and categorize them into three types: rise, fall, and horizontal price movement.

3.1. Data Preparation

We prepare those data from ftx.com, a website providing automatic trading, and let us crawl data from it. The trading pair we crawled is ETH-USDT (Ethereum - United States Dollars Tether). The price data include high, low, open, close, and trading volume, while the trading volume is normalized after the log. The price of the historical datasets of ETH-USDT covered from 2019-7-20 to 2022-6-1 every trading day from the trading platform API, which is 5 years. We split the training data to be 2019-07-20 to 2020-12-01 and the test data to be 2021-01-01 to 2022-06-01. The size of training and testing data is about 1:1. Moreover, we crawled the trading pair of BTC-USDT (Bitcoin - United States Dollars Tether) for four years for the reason that the price of bitcoin is highly correlated to the price of Ethereum.

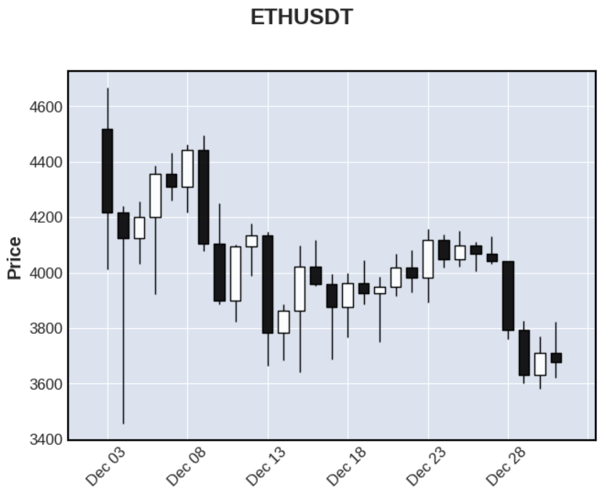

Figure 1 The daily price chart of ETHUSDT from Dec. 1 to 31, 2021.

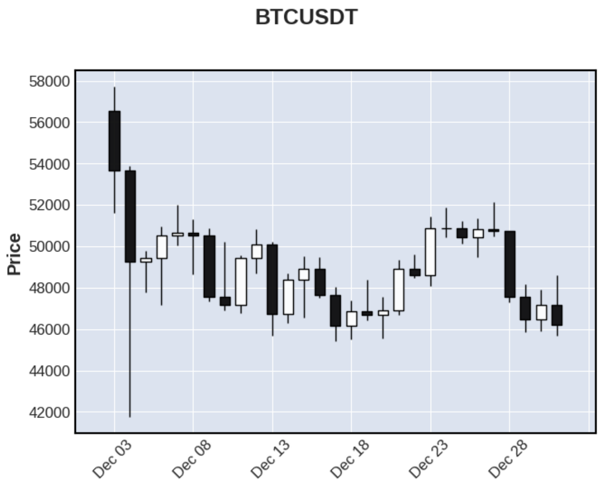

Figure 2 The daily price chart of BTCUSDT from Dec. 1 to 31, 2021.

	startTime	open	high	low	close	volume
0	2021-01-01T00:00:00+00:00	737.14	750.80	713.28	730.43	3.772505e+08
1	2021-01-02T00:00:00+00:00	730.43	790.31	716.42	775.34	7.701209e+08
2	2021-01-03T00:00:00+00:00	775.09	1022.65	770.54	982.91	2.054540e+09
3	2021-01-04T00:00:00+00:00	982.91	1176.14	886.61	1045.45	3.175210e+09
4	2021-01-05T00:00:00+00:00	1045.45	1141.33	975.11	1105.28	2.212057e+09
...	...	...	...	...	...	...
528	2022-05-28T00:00:00+00:00	1725.80	1810.00	1720.30	1790.60	1.552242e+09
529	2022-05-29T00:00:00+00:00	1790.60	1826.80	1755.30	1811.50	1.044130e+09
530	2022-05-30T00:00:00+00:00	1811.50	2013.70	1801.30	1998.70	2.713999e+09
531	2022-05-31T00:00:00+00:00	1998.70	2015.70	1922.00	1940.70	2.052387e+09
532	2022-06-01T00:00:00+00:00	1940.70	1972.00	1760.70	1815.40	2.905796e+09

533 rows × 6 columns

Figure 3 The chart of ETHUSDT from 2021-01-01 to 2022-06-01.

Our task is to classify instead of regressing because, from previous research, regression does not give a satisfactory result and is worse than binary classification. Labeling would be a hard problem for financial data. In our methods, we classify the time-series data into three classes: rise, fall, and horizontal price movement. The rise, fall, and horizontal price movement is a reference indicator for whether it is suitable for day trading the next day. For rise and fall, it is suitable. However, for horizontal price movement, it is not suitable, since it is hard to manipulate and is subject to losing money, because the futures market maker is absorbing coins. Here, we use the way of epsilon and when the price movement ratio is larger than an epsilon then we build the position long, and when the price movement ratio is smaller than a minus epsilon then we build the position short. When the price movement is between epsilon and minus epsilon then we do not do anything. Moreover, we can build leverage in our position.

Moreover, the input data should be normalized to percentage relative to the open price, because as we want to identify the patterns, the actual number of prices is not important but the pattern of the price. The number of prices is very high, especially in recent 1-2 years, ETH is about US$2,000 and BTC is about US$50,000. If we do not normalize, then the weight would converge very slowly because the weight should be very small, and the learning rate should be adjusted, plus the *xmin* bias. The normalization formula: $r=(price-open)/open$, where *open* indicates the open price, and *price* indicates the opening, closing, high, and low prices of each *k*-line.

The time frames should be cut into some important periods, such as days, especially weeks, and months because weeks and months reflect human behavior and human usually reference the candles of weeks and months to find out the future trend. In our research, we let the timeframe be 128 for convenience to input into the fully connected neural network and Convolutional Neural Network. From BraveNewCoin Liquid Index for Bitcoin (BLX), we can find some kind of periodicity from 2011-1-1 to now. The BLX after log scale is shown in Figure 3, and the price shows some periodicity. This machine learning model works because of market periodicity.

Figure 3 BLX price in log scale.

3.2. Model Structure

We first use a fully connected neural network as the model structure to find out the feature extraction of the stock patterns. Then, we use a Convolutional Neural Network as the model structure to find out the feature extraction of the stock patterns. Here, we always let the time length be 128 days. Moreover, we try ensemble learning because ensemble learning is a group of classifiers to become a big classifier, and we think those sub-classifiers votes to have better results.

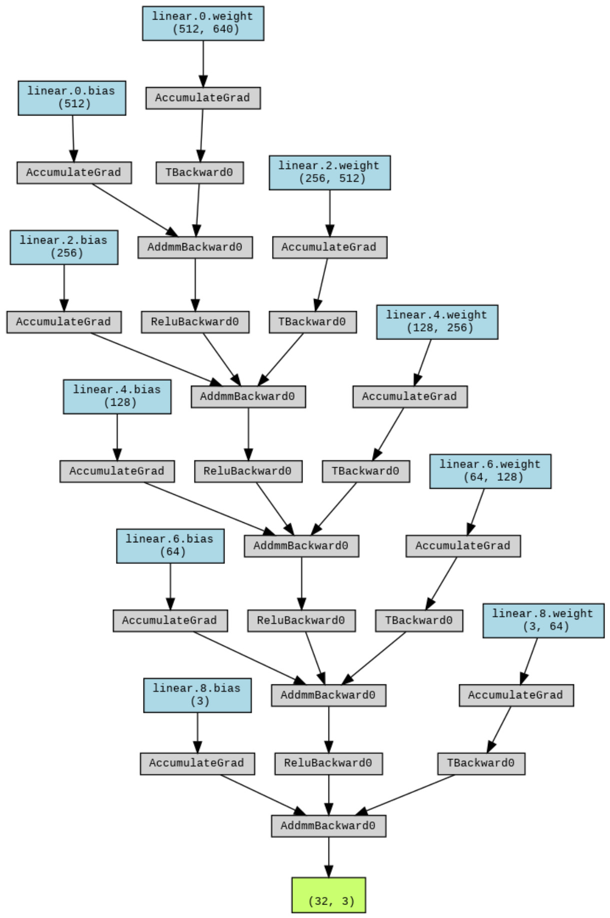

Figure 4 Linear model architecture.

Layer (type)	Output Shape	Param #
Conv1d-1	[-1, 64, 128]	1,024
BatchNorm1d-2	[-1, 64, 128]	128
ReLU-3	[-1, 64, 128]	0
MaxPool1d-4	[-1, 64, 64]	0
Conv1d-5	[-1, 128, 64]	24,704
BatchNorm1d-6	[-1, 128, 64]	256
ReLU-7	[-1, 128, 64]	0
MaxPool1d-8	[-1, 128, 32]	0
Conv1d-9	[-1, 256, 32]	98,560
BatchNorm1d-10	[-1, 256, 32]	512
ReLU-11	[-1, 256, 32]	0
MaxPool1d-12	[-1, 256, 16]	0
Conv1d-13	[-1, 512, 16]	393,728
BatchNorm1d-14	[-1, 512, 16]	1,024
ReLU-15	[-1, 512, 16]	0
MaxPool1d-16	[-1, 512, 8]	0
Conv1d-17	[-1, 512, 8]	786,944
BatchNorm1d-18	[-1, 512, 8]	1,024
ReLU-19	[-1, 512, 8]	0
MaxPool1d-20	[-1, 512, 4]	0
Linear-21	[-1, 1024]	2,098,176
ReLU-22	[-1, 1024]	0
Linear-23	[-1, 512]	524,800
ReLU-24	[-1, 512]	0
Linear-25	[-1, 3]	1,539

==

Total params: 3,932,419
Trainable params: 3,932,419
Non-trainable params: 0

Input size (MB): 0.00
Forward/backward pass size (MB): 1.01
Params size (MB): 15.00
Estimated Total Size (MB): 16.01

Figure 5 Convolutional neural network architecture.

In our research, the input data would be *5 * time length*, for the 5 being high, low, open, close, and volume. Therefore, the input data would be a 1D image of length *time length* with *5* channels: high, low, open, close, and volume. First, we input the data into the stock prediction model. Because the stock price is nonlinear, we should simulate the nonlinear function by activation function. However, gradient vanishing problems can be brought up by activation functions such as sigmoid because when the gradient is multiplied by the value less than 1 then the norm of the gradient will become smaller and smaller and finally lead to vanishing. Hence, we use ReLU. ReLU is a good function for solving gradient vanishing problems. Second, the feature maps will be the result of convolution and will be the input of the pooling layers. Last, the output of the last layer will be input to the fully connected layer, and the loss will be Shannon's cross entropy loss: softmax function.

4. RESULTS AND ANALYSIS

A. Analysis of Proposed Labeling Method

In most prior works, rise and fall are the most common methods in identifying the trend of the stock markets. However, in this work, we also define a trend called horizontal price movement when the large investors are absorbing chips and let the retail investors be afraid and get down the market. This trend is very important for the trend analysis because for a single side trend, rise and fall, it is easy to do but for the horizontal price movement, it is hard to manipulate and is subject to severe loss if the position is too large. In our research, we want to give a result that gains more investment credibility and is more useful for investors.

However, still missing: we do not consider the case where it is in a horizontal price movement stage and the difference between high and low is large while the difference between open and close is small. This would affect the result and the explanation because when the difference between high and low is large then the retail investors are subject to lose money due to the horizontal price movement with drastic volatility.

B. Parameter Settings of Experiments

We have tried some neural networks such as linear network and ResNet and put those neural networks into ensemble learning classifiers. We use PyTorch implementation of ensemble learning boosting classifiers. Moreover, we have adjusted the learning rate after each step of training, because as the weight point is close to the minima in the loss curve, the original learning rate may be too large and we have to lessen the learning rate to let the model become more accurate in training. We adjust the learning rate from 1e-3 to 1e-9 and lessen the learning rate by multiplying 0.1. for each step. We set the number of estimators to be 10 which is just right for the 10 estimators to vote for the result.

C. Experimental Results

We have tried different parameters and produced such results. These are the results of our experiments from different parameters. Below is the chart for different model architectures and their corresponding training and validation accuracies. Indeed, the experimental result is higher than our expectation.

For the linear models, we can find that the prediction accuracy is very high. It can tell us that the linear model is very effective, because in linear models, all of the parameters are passed to the next neuron. For the ensemble of linear models, the training accuracy is not given by the log, but the validation accuracy 0.93536 is near the validation accuracy of a single linear model.

For convolutional models, we can find that the prediction accuracy is also very high. It can tell us that the convolutional model is also very effective, because in convolutional models, the filter is extracting useful information from model so accuracy can be also high enough.

We have compared the general results and the results after ensemble learning. We find that the results after ensemble learning are similar to the results after a single model because the training results of a single model is good enough. Here, we do not do the investment profitability analysis because this classifier is for the assessment of day trading of the next day not for the portfolio optimization.

	Accuracy
Linear Model	Train/valid = 0.96263/0.93514
Ensemble of Linear Models	Train/valid = –/0.94074
Convolutional Neural Network	Train/valid = 0.96154/0.94105
Ensemble of Convolutional Neural Networks	Train/valid = –/0.94074

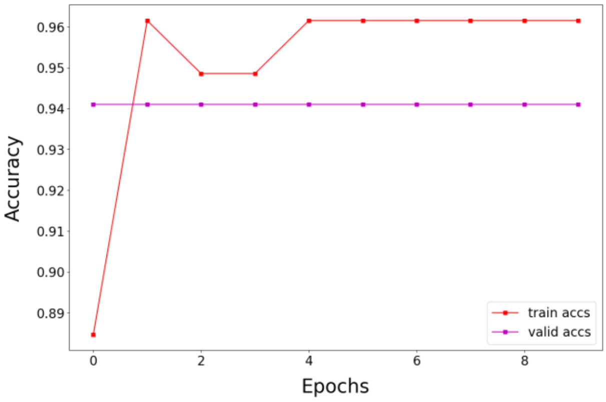

Figure 6 Accuracy vs. epochs for linear models.

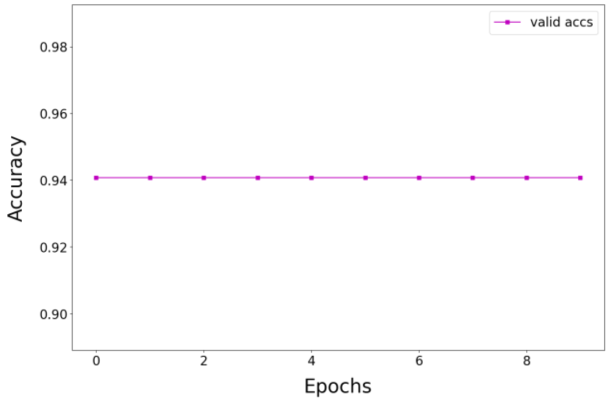

Figure 7 Accuracy vs. epochs for the ensemble of linear models.

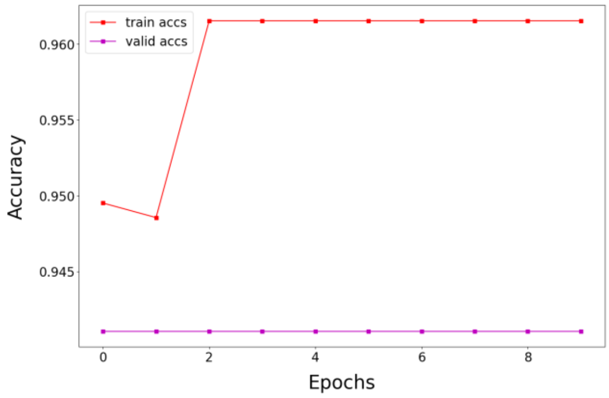

Figure 8 Accuracy vs. epochs for Convolutional Neural Network.

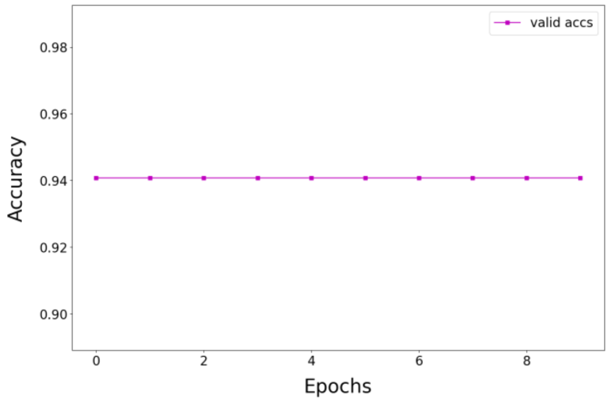

Figure 9 Accuracy vs. epochs for the ensemble of Convolutional Neural Networks.

CONCLUSION

In this paper, we use the linear and the CNN models and the ensemble of those models which is the combination of traditional and recent models, for the ensemble be traditional and deep learning models be recent, to make stock predictions based on the historical price data. We used the labeling "rise", "fall", and "horizontal price movement" to categorize the trends into three to help investors to make decisions. We experimentally compare different models on the prediction accuracy. However, what is a pity is that we did not use ExplainableAI to check whether the model is reliable enough. Moreover, there are some future directions for future study. Here, we assume that the future price is highly related to the historical prices. We can not only analyze the historical prices but some data from GlassNode which is not money-free.

There are some further directions for future study:

Labeling methods:

(1) Slope detection labeling method [6] is the labeling method by defining F_d to be the average price of the next K days after the day d, while B_d to be the average closing price in the past K days.

Modeling:

(1) We can dive into some behavior of institutional investors or someone with a large amount of money.
(2) We can reference some immature and mature markets and compare the Bitcoin market with those markets. Some immature markets such as stocks in developing countries and some mature markets such as stocks in developed countries.
(3) We can make the number of models larger or the time become longer in datasets to make the prediction better when our resources are much larger.
(4) Using Reinforcement Learning (RL) may also be a good idea, for the stock market can be viewed as an infinite poker game with four different numbers every time interval. Therefore, we can view it as the game problem in infinite space.
(5) Meta-Learning is also a good idea to learn many attributes of the learning algorithm. The meta-training process is implemented in two phases. The first phase updates the parameters of each task learner. The second phase updates the parameters of the meta-learner [5].
(6) Time-series forecasting of Bitcoin prices using high-dimensional features and feeding them into LSTM is a good approach [6].

REFERENCES

[1] E. F. Fama, “Random Walks in Stock Market Prices,” *Financial Analysis Journal*, Vol. 51, pp. 75–80, 1995.

[2] K. A. Althelaya, E. M. El-Alfy, and S. Mohammed, “Evaluation of Bidirectional LSTM for Short-and Long-Term Stock Market Prediction,” *Proceedings of International Conference on Information and Communication Systems*, Irbid, Jordan, pp. 151- 156, 2018.

[3] K. Kim, “Financial Time Series Forecasting Using Support Vector Machines,” *Neurocomputing*, Vol. 55, pp. 307–319, 2003.

[4] G. Armano, M. Marchesi, and A. Murru. “A Hybrid Genetic-Neural Architecture for Stock Indexes Forecasting,” *Information Sciences*, Vol. 170, No. 1, pp. 3-33, 2005.

[5] S. H. Chang, C. W. Hsu, H. Y. Li, W. S. Zeng, and J. M. Ho, “Short-Term Stock Price-Trend Prediction Using Meta-Learning,” https://arxiv.org/abs/2105.13599, 2022.

[6] M. Mudassir, S. Bennbaia, D. Unal, M. Hammoudeh, “Time-Series Forecasting of Bitcoin Prices Using High-Dimensional Features: A Machine Learning Approach,” *Neural Computing and Applications*, https://doi.org/10.1007/s00521-020-05129-6, 2020.