Expert Systems With Applications 200 (2022) 117017

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An automated cryptocurrency trading system based on the detection of unusual price movements with a Time-Series Clustering-Based approach

Faruk Ozer, C. Okan Sakar *

Department of Computer Engineering, Bahcesehir University, Istanbul 34353, Turkey

ARTICLE INFO

Keywords:
Price prediction
Dynamic time warping
Hierarchical clustering
Anomaly detection
Outlier detection
Machine learning

ABSTRACT

The cryptocurrency market, which has a rapidly growing market size, attracts the increasing attention of individual and institutional investors. While this highly volatile market offers great profit opportunities to investors, it also brings risks due to its sensitivity to speculative news and the unpredictable behaviour of major investors that can cause unsual price movements. In this paper, we argue that rapid and high price fluctuations or unusual patterns that occur in this way may negatively affect the functionality of technical signals that constitute a basis for feature extraction in a machine learning (ML)-based trading system and this may cause the generalization of the model to deteriorate. To address this problem, we propose an end-to-end ML-based trading system including a time series outlier detection module that detects the periods in which unusual price formations are observed. The training of the classification algorithms for the price direction prediction task was performed on the remaining data. We present the results related to the accuracy of the classification models as well as the simulation results obtained using the proposed system for real time trading on the historical data. The findings showed that the outlier detection step significantly increases return on investment for the machine learning-based trading strategies. Besides, the results showed that during the highly volatile periods the trading system becomes more profitable compared to the baseline model and buy&hold strategy.

1. Introduction

The concept of cryptocurrency emerged in 2009 with the birth of Bitcoin which can be regarded as the ancestor of cryptocurrencies (Nakamoto, 2008). Bitcoin is the world's first peer-to-peer electronic cash system that enables online payments to be transferred from one party to another without an intermediary financial institution. Transfer transactions are digitally encrypted with blockchain technology, an immutable, distributed database of transactions (Narayanan, Bonneau, Felten, Miller, & Goldfeder, 2016; Longo, Podda, & Saia, 2020; Niranjanamurthy, Nithya, & Jagannatha, 2019; Monrat, Schelén, & Andersson, 2019).

In the past 11 years, more than 10 thousand cryptocurrencies and tokens have been created with an increasing momentum, especially since 2017. Bitcoin price has risen from $1,000 at the beginning of 2017 to $64,000 by 2021, experiencing an exponential growth that has led to tremendous earning opportunities that none of the other financial asset class can provide. The cryptocurrency market has become a large financial and investment ecosystem with a market size of 1.5 trillion USD, a daily trading volume of 100 billion USD and nearly 400 exchanges around the world. This market, which is fast growing and developing, highly volatile, easy to access and open 24/7, continues to attract the attention of individual and institutional investors (Jalal, Alon, & Paltrinieri, 2021).

Traders invest in cryptocurrencies as an emerging alternative asset in different ways, through a variety of strategies: by including crypto assets in their portfolio, holding them in expectation of profits in the medium to long term, or by trading at high frequency. Particularly in high frequency trading, the traders seek to find the optimum entry and exit points in a financial time series with the intention of acquiring high returns with low risk. At this point, they benefit from technical methods and strategies as well as the machine learning techniques (Alonso-Monsalve, Suárez-Cetrulo, Cervantes, & Quintana, 2020).

Recently, machine learning (ML) and data mining techniques have been widely applied in the predicting of financial markets and bring better results than simple technical or fundamental analysis methods (Huang, Huan, Xu, Zheng, & Zou, 2019; Rundo, Trenta, di Stallo, & Battiato, 2019; Gan, Wang, & Yang, 2020; Paiva, Cardoso, Hanaoka, &

* Corresponding author.
E-mail address: okan.sakar@eng.bau.edu.tr (C. Okan Sakar).

https://doi.org/10.1016/j.eswa.2022.117017
Received 21 December 2021; Received in revised form 18 February 2022; Accepted 27 March 2022
Available online 2 April 2022

Duarte, 2019). However, the risks related with the cryptocurrency market makes it very sensitive to speculation, social media messages, and news feeds that cause distortions in price movements (Burnie & Yilmaz, 2019; Bizzi & Labban, 2019). In this study, considering that using the data containing such price movements for training may worsen the generalization ability of the model built to predict the price direction, we propose to integrate an outlier detection process into the ML-based trading system to remove the unusual price subseries from the training set.

In this context, in the first part of the study, we employed six classification methods utilizing supervised learning for prediction of next period price direction, up or down. For feature engineering, we created diverse set of features based on the commonly used financial technical indicators. To evaluate the generalization ability of the system, we included three cryptocurrencies, Bitcoin, Ethereum and Litecoin, which have a significant market cap in our experiments. Besides, we used two periods of data, 4-hour and 1-day frequency, to assess the success of the trading system in for different prediction horizons. We evaluate the accuracy and rate of interest (ROI) of the ML-based trading system without the outlier detection process.

In the second part of the study, we integrate the outlier detection process as an intermediate step of the trading system to assess its effect on accuracy and ROI. For this purpose, we first detect the periods that have price movements with unusual characteristic using the outlier detection approach and remove these parts from the dataset. The outlier detection process is a clustering-based approach that uses dynamic time warping technique to compute the similarities between the subseries of the price data. The results indicated that removing the subseries marked as outliers increases the ROI obtained with the trading system. We have also observed that the proposed trading system becomes more profitable than buy and hold strategy during the highly volatile periods.

The rest of this paper is organized as follows. Section 2 presents the related works. Section 3 presents the dataset description, the baseline strategies used for comparison, and the details of the proposed trading system. In Section 4, we give the details of the experiments and simulations performed to assess the performance of the trading system, the effect of the outlier detection process on the accuracy and ROI, and discussion. Finally, Section 5 presents the conclusions and future work.

2. Related works

In line with the short history of cryptocurrencies compared to other financial instruments, the development and use of machine learning-based price prediction and trading system in the cryptocurrency market is at an early stage. There are some research efforsts that attempt to idetify the factors affecting the price movements of the cryptocurrencies and use these factors to predict their future prices using machine learning methods. In these studies, the problem has been handled from different perspectives and machine learning techniques such as classification, regression and time series forecasting were applied on the price data. In this section, we provide an overwiew of the related studies.

Conrad, Custovic, and Ghysels (2018) stated that the effects of the US stock market (SP500 index) and the global stock market index (Nikkei 225 index) on Bitcoin price movements are not significant. Liu and Tsyvinski (2021) introduced research on the factors that affect the price movements of cryptocurrencies. The conclusion drawn in this study was that the macroeconomic factors, which generally determine the dynamics of currency, stock, and commodity markets, do not have a significant effect on the dynamics of the cryptocurrency market.

Phillips and Gorse (2017) built a trading strategy based on historical social network data and epidemic modelling. It has been shown that the trading strategy is superior to the buy-and-hold strategy. This study revealed that social media data can play an important role in predicting cryptocurrency price movements. Similarly, Valencia, Gómez-Espinosa, and Valdés-Aguirre (2019) found that it is possible to predict the cryptocurrency market using the price and social media (twitter) data of certain cryptocurrencies. In this study, it was stated that the prediction model based on artificial neural network (ANN) outperforms other machine learning algorithms for this task. Bouri, Lau, Lucey, and Roubaud (2019) examined the role of the trading volume in predicting the returns and volatility of several cryptocurrencies. In this study, it has been indicated that trading volume is useful in predicting the returns of the selected cryptocurrencies but limited in predicting their volatility. Misnik, Krutalevich, Prakapenka, Borovykh, and Vasiliev (2019) pointed out that increasing the data can improve the accuracy of the prediction.

Zhengyang, Xingzhou, Jinjin, and Jiaqing (2019) built two advanced machine learning regression models with a fully connected ANN and a Long-Short-Term-Memory (LSTM) network to predict the six major cryptocurrencies' prices. The results in this study showed that ANN in general outperforms LSTM. In addition, the prediction error decreases at a longer timescale (30 days). Besides, they found that the joint prediction of multiple cryptocurrency time series improves the prediction ability compared to training on each cryptocurrency individually. Sattarov et al. (2020) proposed a model by applying a deep learning-based reinforcement learning approach that can serve a trader in making trading decisions for the cryptocurrency market. According to the test results in this study, a trader investing in Bitcoin, Ethereum and Litecoin earnes a net income of 14.4%, 74% and 41% percent in one month, respectively. Koker and Koutmos (2020) introduced a model using reinforcement learning for active trading of five major cryptocurrencies and demonstrated how this model yields risk-adjusted returns comparing to the buy-and-hold strategy. Jeong and Kim (2019) also used deep Q-network algorithm that combines deep neural network and reinforcement learning. One of the main contributions of their study was to utilize transfer learning with the aim of improving the generalization ability of the neural network model.

As more relevant research to our study; Sun, Zhou, and Lin (2019) created features from some of formulas in Alpha101 (Kakushadze, 2016) and predicted the price movements of several cryptocurrencies using the random forest classifier. The results related to the performance of the random forest models showed that some factors defined in the context of Alpha101 play important roles in predicting the price movements. Besides, the models can make more accurate predictions over longer time intervals and the backtesting results performed in this study also supported this observation. Attanasio, Cagliero, Garza, and Baralis (2019) explored the use of the machine learning methods in comparison to the time series forecasting techniques to predict the next-day price of several cryptocurrencies. In this study, although the time series forecasting models generated more trade signals in average than the machine learning-based methods, the average return per trade achieved by the classification models was higher than those achieved by the time series forecasting methods. It was also stated that the classification models often produce more accurate signals but miss many profitable trades.

In another study, Akyildirim, Goncu, and Sensoy (2021) analyzed the predictability of the most liquid twelve cryptocurrencies at the daily and minute level frequencies using 12 classification algorithms. In this study, the historical prices with technical indicators of the cryptocurrencies were used as input, and the prediction accuracies found on average at the daily or minute level frequencies of the classifiers reached up to 55%-65%. Borges and Neves (2020) resampled the market data of the cryptocurrencies analyzed according to a closing price threshold, then calculated a number of technical indicators and fed them as input to four machine learning algorithms and combined the predictions with an ensemble approach using the resampled and original data. In this study, the unweighted average yielded the best overall results, namely accuracies up to 59.26% for the resampled data. Carta, Corriga, Ferreira, Recupero, and Saia (2019) also combined an ensemble learning approach with an optimized feature selection step to build a trading system that can be applied automatically in different market conditions. They applied the proposed system on different real-world futures markets and showed the effectiveness of the proposed ensemble learning-

based approach. Brzeszczyński and Ibrahim (2019) aimed at analyzing the effect of foreign information signals in domestic trades. They used Relative Strength Index (RSI) to quantify the domestic momentum. Anghel (2021) compared the predictive ability of the machine learning and technical analysis trading rules in the cryptocurrency market and found that statistically significant positive excess returns are rarely achieved with these approaches.

In these studies, we see that ML has a significant potential to be used for trading in crtptocurrency market. However, high volatility and speculative market conditions in this emerging market cause noisy price movements which may affect the generalization ability of the ML models built to predict the market prices or price direction. To fill this gap, in this study we investigate the effect of detecting the periods with unusual price formations and removing the related parts from the training data to improve the model's prediction capability.

3. Materials and methods

In this section, first we give an overview of the datasets used in our experiments. Then, we briefly summarize the widely used trading strategies and a baseline strategy with which the proposed trading system is compared in terms of ROI. Finally, we explain the steps of the proposed trading system.

3.1. Dataset description

In this study, four-hour (4 h) and daily (1D) frequency historical market data of Bitcoin (BTC), Ethereum (ETH) and Litecoin (LTC) cryptocurrencies, which have significant market caps in the market, were used. The data belongs to the period between July 1, 2017 and April 30, 2021 as shown in Fig. 1. The market data of the selected three cryptocurrencies were retrieved through a script written using the API (Application Programming Interface) provided by a cryptocurrency exchange.

The data retrieved from the exchange consists of 5 columns in US dollars: Open, High, Low and Close Price (OHLC), Volume and Milliseconds. There are 8,394 number of data points in the 4 h datasets and 1,399 number of data points in the 1D datasets. Table 1 shows some important statistics of the retrieved data. Here, the return column shows the return of the B&H strategy between July 1, 2017 and April 30, 2021. The market capitalization (market cap) column represents the total value of all the coins that have been mined for the related cryptocurrency as of the date of dataset acquisition (April 30, 2021) in this study. This value is calculated by multiplying the market price of a single coin and the number of coins in circulation on April 30, 2021.

3.2. Baseline strategy

The performance of the proposed ML based trading strategy we designed in this study was compared with two baseline strategies. The performance metric of these strategies is the Return on Investment (ROI) value which was computed based on backtesting (simulation) performed by simulating under real trading conditions as detailed in Section 3.3.6. We calculate the ROI at the end of a simulation period as.

$$ROI = \frac{FinalValueofInvestment - InitalValueofInvestment}{InitalValueofInvestment} \cdot 100\% \quad (1)$$

The two strategies to be a baseline, Buy and Hold (B&H) and Simple Technical Strategy (STS), are briefly explained below.

i. Buy and Hold (B&H) Strategy: Buy and hold is a passive investment strategy in which securities that are generally considered to have an increase potential as a result of a fundamental analysis are bought and kept for medium to long term. Investors who use this strategy have no concern for short-term price movements and technical indicators. It is usually preferred by investors in cryptocurrency markets. Considering that cryptocurrencies, in general, have an increasing price trend in long term as also seen in Fig. 1, it is important to design a trading strategy that is more profitable than B&H strategy.
ii. Simple Technical Strategy (STS): A simple custom strategy was created using technical analysis methods for this study. In this strategy, the RSI indicator, one of the most widely used technical indicator, was used to generate buy and sell signals. Using the generated signals, we simulated four different three-month periods that we determined for the test. For each test period, the parameter set of the RSI indicator (any of 9, 14 or 26 as an RSI period; any of 70, 75 or 80 as an overbought/sell point; any of 30, 25 or 20 as an oversold/buy point) was tuned by simulating on the previous 12 months. After finding the optimal values for overbought/oversold parameters, we apply those values on the next three months. This approach is applied for each of the four test periods. The flowchart of the STS-based trading strategy is shown in Fig. 2.

In Table 2, the ROI results of the two strategies obtained for the test periods using the 4 h and 1d frequency data are summarized. Overall, a bull market in which the prices show an increasing trend has occurred in

Table 1
Summary statistics of the cryptocurrencies (01 Jul 2017–30 Apr 2021).

Cryptocurrency	Abbrev.	High Price in USD	Low Price in USD	Last Price in USD	Return	Market Cap.*
Bitcoin	BTC	63,537	1,925	53,565	2,183 %	49.49%
Ethereum	ETH	2,758	85	2,758	982 %	14.27%
Litecoin	LTC	355	23	255	590 %	0.82%

***Source:** https://coinmarketcap.com/

Fig. 1. Price performance of the cryptocurrencies (01 Jul 2017–30 Apr 2021).

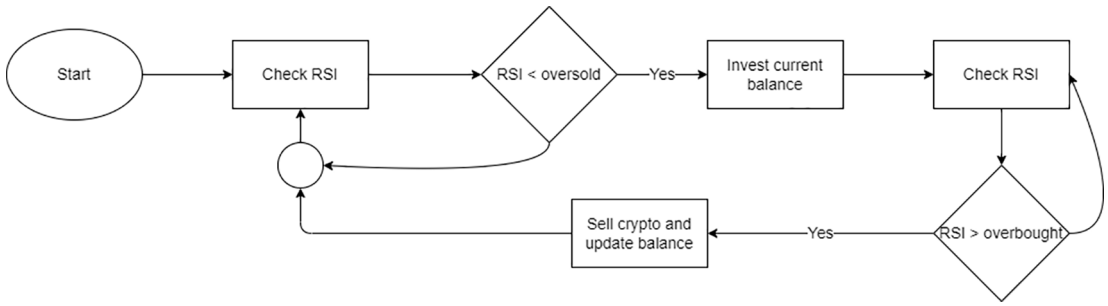

Fig. 2. Overview of the trading simulation performed with STS strategy.

Table 2
Results of return simulations of the two baseline strategies: STS and B&H.

Crypto.	Period	STS Parameter Tuning	Test Period	B&H ROI	STS ROI	
					4 h	1D
Bitcoin	1	2019/05/01–2020/04/30	2020/05/01–2020/07/31	31%	15%	13%
	2	2019/08/01–2020/07/31	2020/08/01–2020/10/31	21%	18%	0%
	3	2019/11/01–2020/10/31	2020/11/01–2021/01/31	142%	62%	0%
	4	2020/02/01–2021/01/31	2021/02/01–2021/04/30	71%	59%	38%
			Avg.	**66%**	**39%**	**13%**
Ethereum	1	2019/05/01–2020/04/30	2020/05/01–2020/07/31	66%	11%	44%
	2	2019/08/01–2020/07/31	2020/08/01–2020/10/31	10%	36%	0%
	3	2019/11/01–2020/10/31	2020/11/01–2021/01/31	240%	188%	53%
	4	2020/02/01–2021/01/31	2021/02/01–2021/04/30	110%	132%	110%
			Avg.	**107%**	**92%**	**52%**
Litecoin	1	2019/05/01–2020/04/30	2020/05/01–2020/07/31	24%	5%	4%
	2	2019/08/01–2020/07/31	2020/08/01–2020/10/31	−6%	2%	0%
	3	2019/11/01–2020/10/31	2020/11/01–2021/01/31	136%	73%	13%
	4	2020/02/01–2021/01/31	2021/02/01–2021/04/30	108%	108%	108%
			Avg.	**66%**	**47%**	**31%**

the past 12 months. In this uptrend market, the B&H strategy yielded better returns than a simple technical strategy (STS). However, as seen Table 2, in the second test period in which prices are volatile but flat or downtrend, the STS strategy yielded comparable or better performance than the B&H for trading at 4-hour frequency. Also, for the STS approach, it should be noted that trading at the 4-hour frequency is more profitable than daily trading.

3.3. Proposed trading system

3.3.1. Classification for price direction prediction

In simplest terms, a trader, after deciding on the security to invest in, tries to predict the price direction to buy/sell at the right time. In this study, for this purpose we developed a machine learning based trading strategy. The first main step of this strategy was to predict the price direction. Therefore, we formulated the prediction part of the problem as a classification task that predicts the price direction. Then, in the second step the predicted price direction was used as a part of a strategy to give buy/hold/sell decisions.

We used six machine learning algorithms (classifiers) based on a supervised learning approach for the prediction of the next period price direction. These classifiers are Logistic Regression and the following nonlinear methods: k-Nearest Neighbor (KNN), Support Vector Machines (SVM) (Cortes & Vapnik, 1995) with radial basis function (RBF) kernel, Random Forest (RF) (Ho, 1998), Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016), and CatBoost (Prokhorenkova, Gusev, Vorobev, Dorogush, & Gulin, 2017). The hyperparameter of the classifiers, except for the Logistic Regression, were tuned at all training stages using cross validation and grid search technique. In Table 3, the range of the hyper-parameter values for each of the classifiers are presented.

3.3.2. Data labeling

As the first part of the proposed trading strategy was formulated as a classification task, it is required to design a labeling strategy to obtain the labels that will guide the price direction prediction task. Considering that there is not an explicit discrete value representing the future price direction, the data labeling approach that extracts the price direction from the fluctuating future price series is one of the challenging parts of

Table 3
Range of the hyper-parameter values of the ML algorithms used to predict the future price direction.

Classifier	Hyperparameter Set	Total Combination
KNN	k = {4, 6, 8, 12, 16}	5
Support Vector Machine	C = {2, 8, 16, 32, 64} sigma = 0.03	5
Random Forest	ntree = {500, 1500} mtry = {2, 3, 4}	6
XGBoost	max_depth = 3 nrounds = {500, 1500, 2500} eta = {0.01, 0.05} colsample_bytree = 0.75 subsample = 0.75	6
CatBoost	depth = 3 iterations = {500, 1500, 2500} learning_rate = {0.01, 0.05} border_count = 128 rsm = 0.75 l2_leaf_reg = {0.01, 0.1}	12

ML-based trading systems.

The basic approach within the scope of this problem is that if the price of the cryptocurrency has increased in the next period, it is labeled positive, otherwise negative. Various methods have been used in different studies. Mostly, a binary classification task is performed by labeling the samples with a future price change exceeding a predetermined threshold as positive, otherwise negative. In addition, it is also common to perform a multi-class classification by including another class label representing the future periods in which a significant price change does not occur. In general, the threshold value is determined according to the risk/reward ratios and market volatility by also taking the commission rates into account.

In this study, we also determined a threshold value and constitue a three-class dataset representing positive, neutral, and negative values for buy, hold, and sell strategies, respectively. However, unlike the common approach, in this study the samples with neutral class label whose price changes fall between the threshold values determined for positive and negative labels were excluded from the training sets. The resulting dataset consisted of the data points in which a significant price change has occurred. Then, this dataset was split into train and test set periods by considering the temporal structure of the data. Both train and test sets include two classes, and hence the accuracies of the classifiers were computed and reported for the binary class labels. With this approach, our motivation was to train the models with the trading periods in which a meaningful price change occurred. We have determined various thresholds and tested the corresponding models for the 4 h and 1D datasets. The optimal values were determined as one percent (1%) and one and a half percent (1.5%) for 4 h and 1d datasets, respectively. The resulting labeling strategy used in this study is shown below:

$$r_p = \ln(Close_{p+1}/Close_p)$$

$$class_p = \begin{cases} 1, r_p > threshold \\ -1, r_p < -threshold \\ excluded, otherwise \end{cases} \quad (2)$$

where $Close_p$ represents the closing price of a cryptocurrency for period p, r_p represents the price change, and $class_p$ represents the label for the corresponding period. We should note that the whole dataset was used during trading simulations and the ROIs of the models were computed including all data points as a real-world trading scenario. The holding decision during trading simulations was given based on the probability estimates as detailed in Section 3.3.5. In Fig. 3, the data labeling process is shown for four exemplary data points on Bitcoin price time series (4 h frequency data). The blue dashed lines indicate ±1% threshold bands.

3.3.3. Feature engineering

Features of the ML model built to predict the future price direction were created based on the technical indicators that investors frequently refer to in technical analysis. Technical indicators take one or more parameters (usually time period n) and produce different results depending on different parameter sets given. For an accurate prediction model, in addition to choosing the right technical indicators, it is crucial to choose suitable parameter values for each of them. In our study, as a part of the feature engineering task, the parameters were determined by trials during training. Finally, 27 features were obtained from the selected technical indicators and their different parameters.

The technical indicators produce signals of similar nature, such as overbought, oversold, and trend change, and thus are often correlated with each other. Therefore, the features with very high correlation (>95%) were excluded from the datasets. Table 4 summarizes the technical indicators used, their respective parameters, and the number of obtained features.

3.3.4. Outlier detection with Time-Series clustering

An outlier is a data point that deviates too much from the other data points as if produced by a different mechanism (Wang, Bah, & Hammad, 2019). In other words, outliers are observations that are far outside the general trend in the dataset. Outliers in a dataset have a significant effect on the generalization ability of the prediction model especially in noisy environments. The fundamental factors which can be considered as the hidden factors that are not included in the feature set are important noise sources in highly volatile cryptocurrency market which is extremely sensitive to speculation, social media messages, and news feeds from public or private institutions. Therefore, we argue that rapid

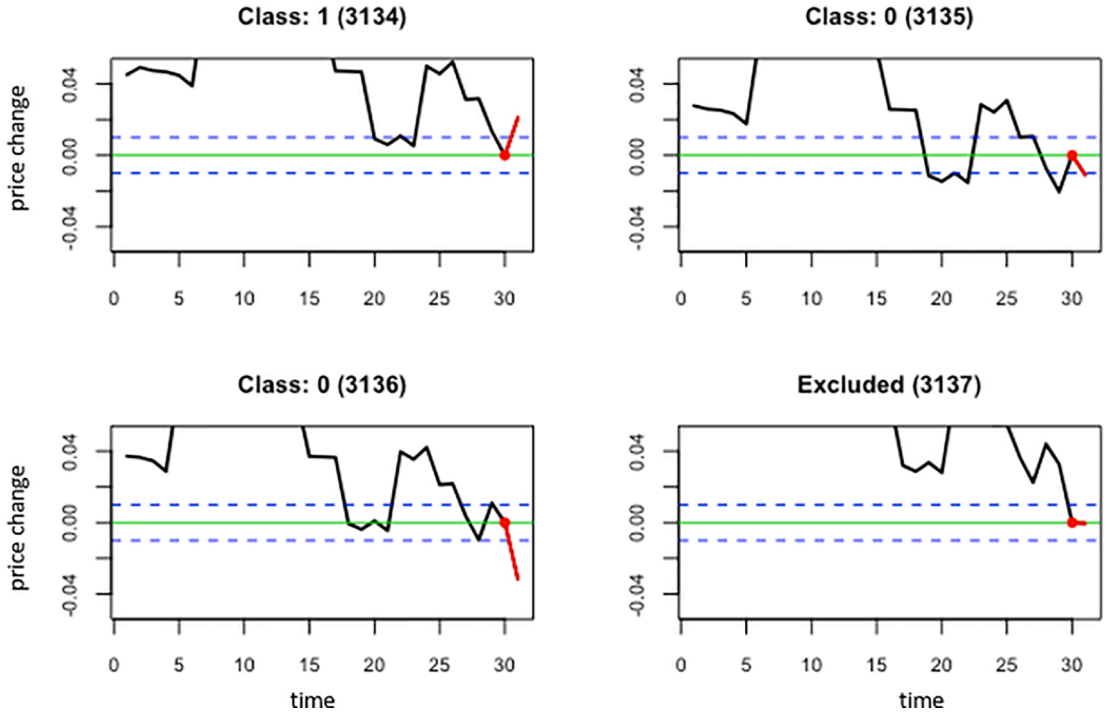

Fig. 3. Data labeling process applied on Bitcoin price time series.

Table 4
The feature set generated from technical indicators.

Technical indicator	Abbreviation	Parameter (4 h/1D)	Number of features
Rate of Change	ROC	{1, 3, 42/7} periods	3
Exponential Moving Average	EMA cross	{84/14, 168/28,} periods	2
Relative Strength Index	RSI	{6, 14, 26} periods	3
	RSI previous	Lag n = 2 RSI	3
Moving Average Convergence Divergence	MACD histogram	Fast 12, Slow 26, Signal 9 periods	1
	MACD hist. previous	Lag n = 2 MACD histogram	1
Commodity Channel Index	CCI	{10, 20} periods	2
Bollinger Bands	BB %b	{10, 20} periods	2
	BB band width	{10, 20} periods	2
Stochastic Oscillator	SO fast %D	Fast %K {6, 14}, Fast % D and Slow %D 3 periods	2
	SO histogram	Fast %K {6, 14}, Fast % D and Slow %D 3 periods	2
Price Volatility	Volatility	20 periods	1
Volume	Net volume	42/7 periods	1
	Volume change	{42/7, 84/14} periods	2

and high price fluctuations or unusual patterns that occur in this way will negatively affect the functionality of technical signals and reduce the success of the model.

To handle this problem, we identified the outlier periods in the raw data (historical price series) of the cryptocurrencies, removed them from the dataset, and then trained the models on the remaining data. For this purpose, we derived N subseries of length m from the historical price series of the cryptocurrency. Each of the subseries slides to the right side by a certain length and overlaps the next subseries. The subseries length m and slide length s were tuned during model training. The technique used to generate subseries is illustrated in Fig. 4.

The derived subseries were first scaled by z-score and then divided into k clusters using clustering algorithms. For clustering, we applied k-Means clustering and hierarchical clustering to identify the subseries with similar price movements. Commonly in clustering algorithms, the similarity scores between observations are calculated using Euclidean distance metric. However, the Euclidean distance metric does not fit to the nature of the temporal signals since it only considers one-to-one distances between the aligned points of the series. This method is sensitive to distortions in the time axis. Therefore, in our study the dynamic time warping (DTW) technique was used in the calculation of distances (similarity scores). With this method, the sequences that are similar but locally out of phase can be associated by taking the non-linear alignments into account (Benkabou, Benabdeslem, & Canitia, 2018).

The DTW algorithm can be applied as follows. Suppose that $X(x_1, x_2, \ldots, x_n)$ and $Y(y_1, y_2, \ldots, y_n)$ represent two series of length n and m, respectively. Let a matrix D of size $n \times m$ keeps the pairwise correspondence relationship between X and Y. Here, each element of matrix D, D_{ij}, represents the distance between data points x_i and y_j. Based on these definitions, the following time warping path can be used to represent the pointwise alignment between series X and Y.

$$W(w_1, w_2, \ldots, w_n), max(m, n) \leqslant K < m + n - -1 \tag{3}$$

where $w_k = (i, j)$ is the alignment between x_i and y_j. Then, the optimal path W_o can be defined as the path which corresponds to the minimum value of the following warping cost:

$$DTW(X, Y) = \min\left\{ \sum_{k=1}^{K} w_k \right. \tag{4}$$

After clustering the subseries based on the use of DTW distances, the mutual DTW scores of the central series representing the cluster center and each intra-cluster series were calculated for each cluster. Series that are a far from the central series were determined as outliers and excluded from the training datasets. This outlier detection procedure is shown in Algorithm 1. The hyper-parameters shown in Table 5 with their range of values were optimized during model training and the set of hyper-parameter values with the best prediction and ROI results were selected.

Figures 5 and 6 show the clustering results of 319 subseries derived from Bitcoin's 4 h frequency historical prices from July 2017 to January 2021 obtained with hierarchical and k-means algortihms, respectively. The length of each subseries (window size) is 48 and the slide length is 24. As seen, each cluster consists of a set subseries having a specific trend characteristic such as uptrend, downtrend, flat, and their variants. However, as seen in Fig. 6, some subseries with similar characteristics were scattered to different clusters by k-means. On the other hand, Fig. 5 shows that as the number of clusters increases in hierarchical clustering, clusters containing subseries with unusual characteristics emerged. Based on this finding, the subseries grouped in Cluster 7, 11 and 12 with hierarchical clustering were considered directly as outliers.

In Fig. 7 and Fig. 8, the centroid of each cluster and the subseries marked as outliers in each cluster are shown for hierarchical and k-means clustering, respectively. The number of subseries detected as outliers with hierarchical and k-means algorithms are 52 and 74, respectively. As seen in Fig. 7, all subseries grouped in Cluster 11 and Cluster 12 have unusual patterns and were considered as outliers.

3.3.5. ML-Based trading strategy

The proposed trading system uses the outputs of the ML-based prediction module to perform trading decisions. The decision-module operates at a frequency of four hours, or at the end of the day, according to the price frequency of the dataset on which the prediction model was trained. The probability estimates generated by the price direction prediction module were converted into "buy", "sell" and "hold" signals. The trading strategy constituted to generate trading signals is shown in Eq. (5).

$$signal = \begin{cases} 1, probabilty \geq 0.50 \\ -1, probabilty < 0.45 \\ 0, otherwise \end{cases} \tag{5}$$

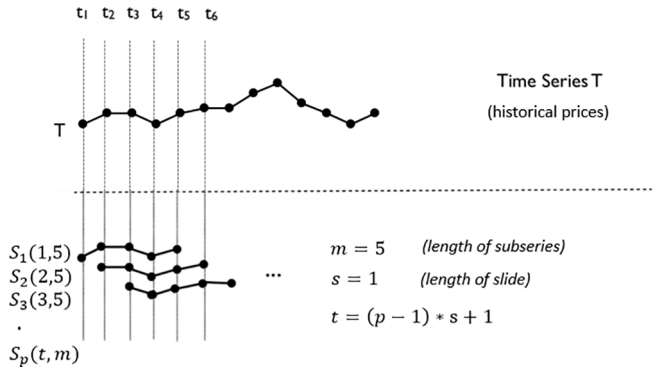

Fig. 4. Subseries generation technique used for the outlier period detection task.

Table 5
Outlier detection hyper-parameters to be tuned.

Parameter Name	Parameters (4 h / 1D)
Method	{Hierarchical, K-Means}
k Cluster	{6, 9, 12}
Slide Length (s)	{18/3, 24/4, 30/5} periods
Window Length (m)	2*s
Percentile (p)	{75, 80, 85, 90, 95}

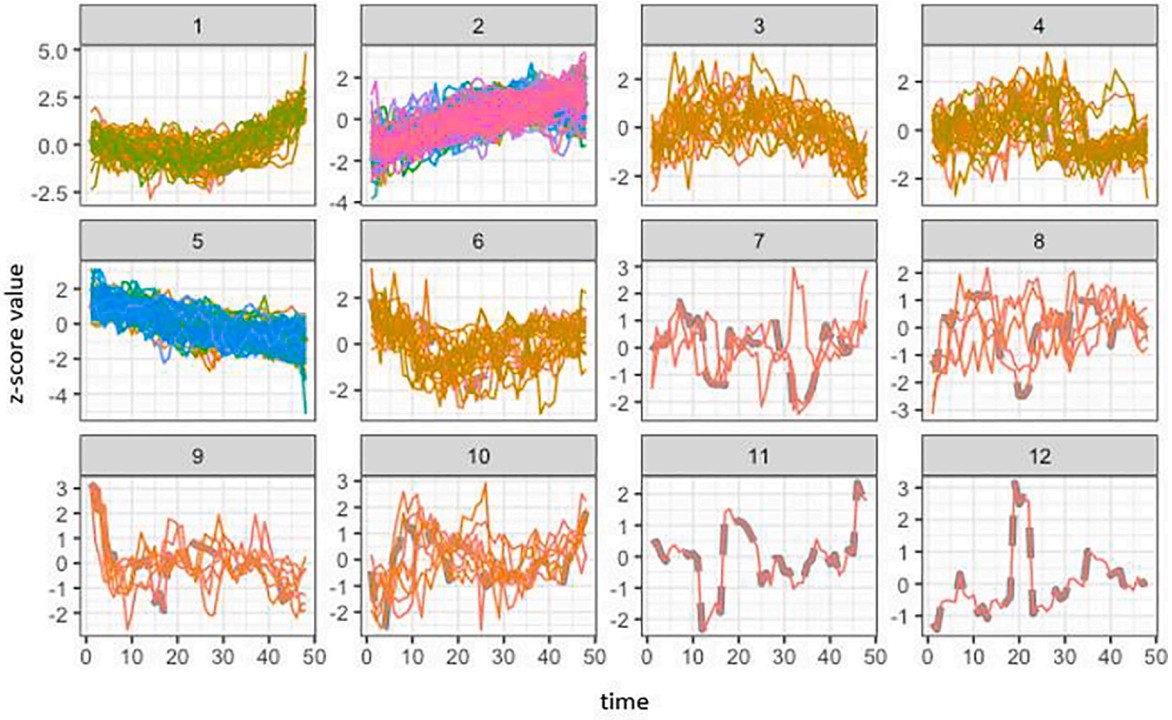

Fig. 5. Subseries divided into 12 clusters using hierarchical clustering and DTW.

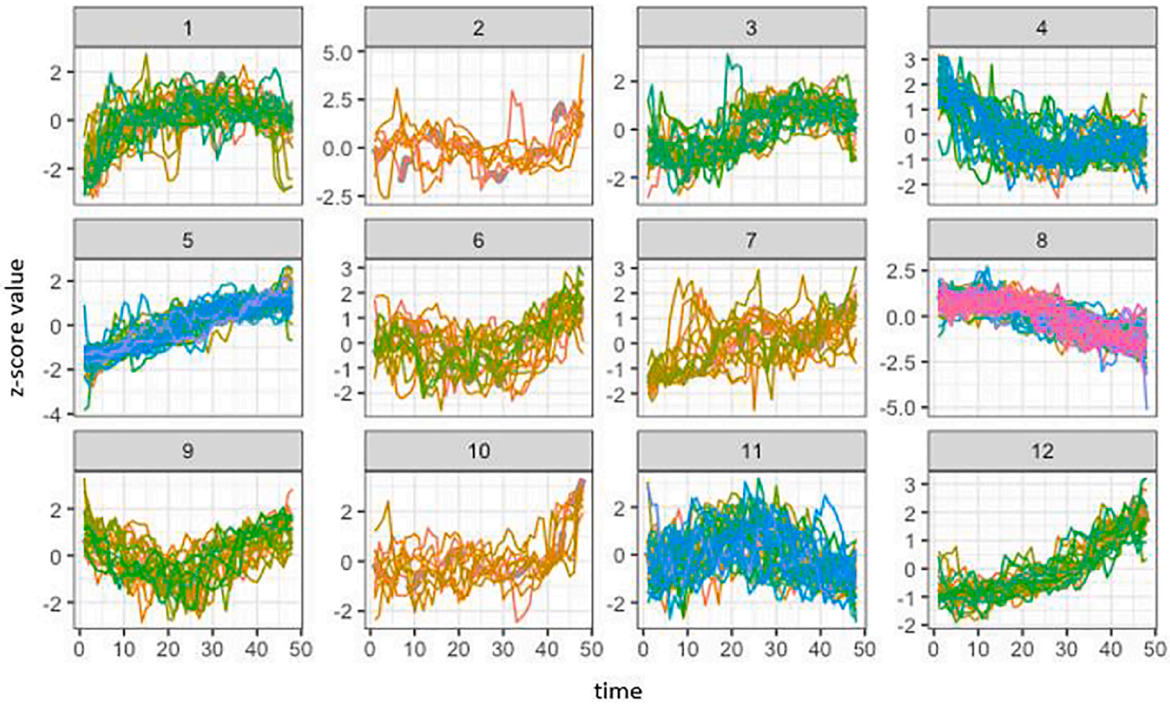

Fig. 6. Subseries divided into 12 clusters using k-Means and DTW.

The signals "1", "-1" and "0" represent "buy", "sell", and "hold" signals, respectively. The threshold values for the probability estimates given in Eq. (5) were determined using grid search technique as a part of the hyper-parameter optimization procedure. For this purpose, various combinations for the buy and sell threshold values were determined. Since these hyper-parameters are common to all classifiers, the probability estimate pair with the highest average ROI was selected to be used in the final trading framework. The overview of the proposed end-to-end trading system is shown in Fig. 9. When a buy signal is generated, the signal variable is set to 1 and a position is opened if there is no open position that has been entered but not closed with an opposing trade yet. In Fig. 9, the position variable indicates the amount invested in the currently open position. The value of zero for the position variable indicates that there is no open position at that time. An open position is closed when the system generates a sell signal. As seen in Fig. 9, the system has at most one open position at a time.

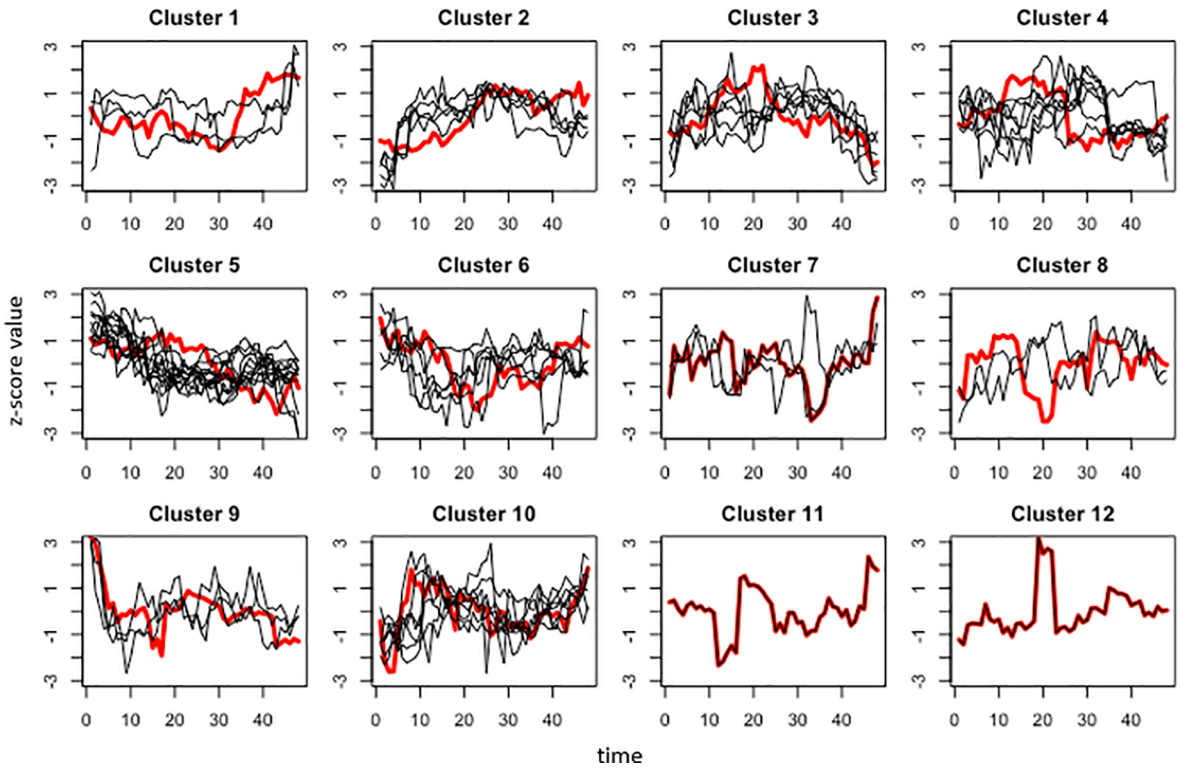

Fig. 7. Outliers detected in each of the clusters using hierarchical clustering.

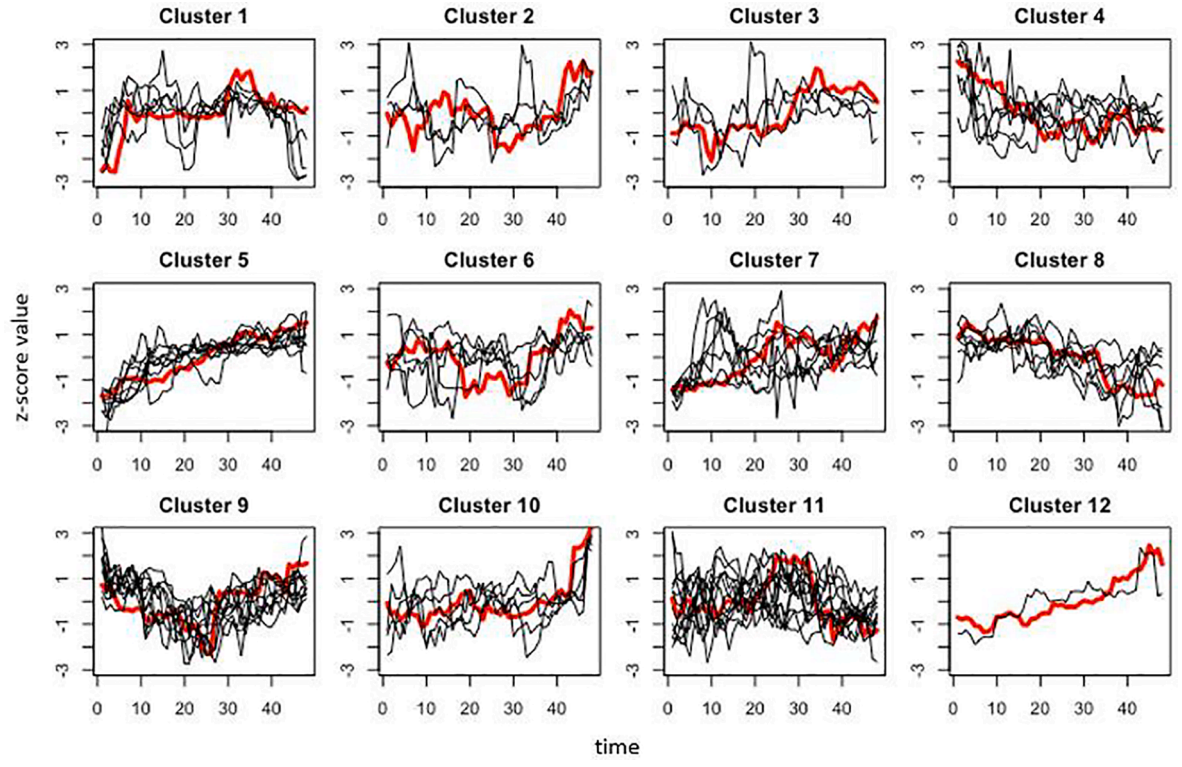

Fig. 8. Outliers detected in each of the clusters using k-means clustering.

3.3.6. *Backtesting*

Backtesting is the simulation of the ML-based trading system given in Fig. 9 performed to calculate ROI during the test period. The simulations were performed on the four three-month test periods between May 2020 and April 2021 and ROI were calculated for each period. During the simulations, the trading commissions for both buy and sell actions were assumed to be one-thousandth (0.1%) and deducted from the calculated returns. The trading simulations performed with both baseline strategies and proposed trading framework were started with 1000 USD at the beginning of each test period. The position is opened with 100% of the

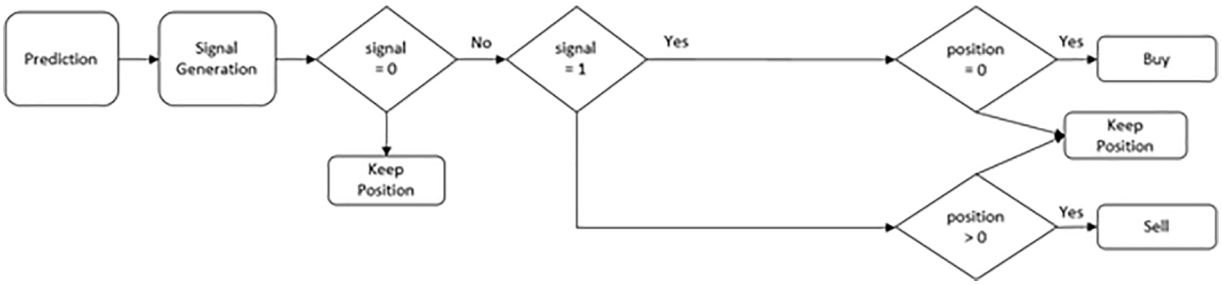

Fig. 9. Overview of the proposed end-to-end trading system.

remaining balance when a buy signal is generated, and an open position is completely closed with the sell signal.

4. Experiments

In this section, we first give the experimental setup used throughout the experiments. We give the details of the cross-validation and hyper-parameter optimization procedure. Then, we present the comparative results. We also present the trading simulation results performed with an initial equity of 1000 USD. Our aim is three folds: (1) We give comparative results of the models built with six different classifiers with the aim of assessing how well each classifier can model the price direction prediction problem in this market, (2) We compare the proposed end-to-end trading system with two baseline strategies, B&H and STS, in terms of the average ROIs for the related periods, and (3) We aim to show the effect of outlier detection step on accuracy and ROI of the trading system.

4.1. Experimental setup

In our experiments, the classification algortihms were trained with two different period data, 4-hour (4 h), and 1-day. In this way, we assess the performance of the trading system for different trading strategies (daily or intraday). The models were tested to predict the price direction of each of the three cryptocurrencies namely Bitcoin, Ethereum and Litecoin. Thus, we aim to analyze the generalization ability of the trading system on different crtprocurrencies.

We divided the cryptocurrency price data into four training and test periods sequentially according to the timestamp information. In this way, we tested the performance of the trading systems on the last 12 months data consisting of four different periods and gave an assessment of how our technique performs in different market conditions occur during the year. In Table 6, we give the details of the training and test data we used for each period.

In all training phases, we applied cross validation on time slices and grid search techniques for hyperparameter tuning. We used the logarithmic loss (Log-loss) performance metric for hyper-parameter optimization. Log-loss is indicative of how close the prediction probability is to the corresponding actual value. The Log-loss calculation for binary classification is.

$$LogLoss = \frac{1}{N}\sum_{i=1}^{N}[y_i \ln(p_i) + (1-y_i)\ln(1-p_i)] \quad (6)$$

where N is the total number of samples, y_i and p_i denote the actual value (1 or 0) and probability estimate for sample i, respectively.

As we mentioned in section 3.3.5, our ML-based trading system takes the probability estimations as input generated by the classification model that was built to predict the price direction and generates trading buy, sell, and hold signals accordingly within a certain rule. Therefore, the success of the system highly depends on the accuracy of the prediction module. However, the classifiers do not produce a numerical prediction for the future prices but only a probability estimate related to the price direction that may not be highly correlated with the price increase/decrease. Considering that the main goal is to choose the model that provides the maximum return on the test set for real-time trading in the market conditions, in all test phases, ROI and accuracy were used together for the evaluation of different models.

4.2. Results

4.2.1. Results without outlier detection

Table 7 shows the summary of the test results for the four test periods obtained with B&H and STS strategies as the baseline trading systems and the proposed trading system without outlier detection step. We show the prediction results obtained with the data retrieved at 4 h and 1d frequency. As seen in Table 7, trading at the 4 h frequency is more profitable than daily trading and in overall the classifiers performed more consistently on this comparably shorter-term trading strategy. We also observed that the prediction accuracy at the 4 h frequency is higher for Bitcoin than the other cryptocurrencies.

The results in Table 7 indicated that logistic regression yielded the highest average accuracy in the price movement direction prediction task. The highest accuracies for Bitcoin and Litecoin were obtained with logistic regression-based classification model for both 4 h and 1d data. It is also seen that for all cryptocurrencies, the logistic regression model was ranked in top-2 for 4 h data in terms of max drawdown and max gain values. However, for 4 h and 1d data of Bitcoin, Random Forest yielded the highest accuracy. On the other hand, the ROI results obtained with the trading simulations performed on the historical data showed that SVM-based trading strategy gave the most profitable results for 1d trading and second highest for 4 h trading. Considering that, as seen in Eq. 12, the trading strategy is based on the use of probability estimates of the classifiers, this finding indicates that the probability estimations of SVM are more accurate in general than the other classifiers. We should also note that the ROIs of the best models for both 4 h and 1d data of all

Table 6
Training and test periods.

Period	Training Period	Training Length	Test Period	Test Length	Price Change (BTC)	Market Characteristic of the Test Periods		
						1st month	2nd month	3rd month
1	2017/07/01–2020/04/30	34 months	2020/05/01–2020/07/31	3 months	31%	Bearish	Bearish	Bullish
2	2017/07/01–2020/07/31	37 months	2020/08/01–2020/10/31	3 months	21%	Bearish	Bearish	Bullish
3	2017/07/01–2020/10/31	40 months	2020/11/01–2021/01/31	3 months	142%	Bullish	Bullish	Bearish
4	2017/07/01–2021/01/31	43 months	2021/02/01–2021/04/30	3 months	71%	Bullish	Bullish	Bearish

Table 7
Performance of the baseline strategies and the proposed trading system with different classifiers for different crptocurrencies and data frequency.

Method	Bitcoin				Ethereum				Litecoin				Average			
	Avg. Accuracy %		Avg. ROI %		Avg. Accuracy %		Avg. ROI %		Avg. Accuracy %		Avg. ROI %		Avg. Accuracy %		Avg. ROI %	
	4 h	1D	4 h	1D	4 h	1D	4 h	1D	4 h	1D	4 h	1D	4 h	1D	4 h	1D
Logistic Regression	56.5	53.1	78	65	54.3	49.7	114	77	54.6	54.0	98	69	55.1	52.2	97	70
KNN	52.2	48.7	28	45	51.4	44.7	61	39	49.6	49.7	37	113	51.1	47.7	42	66
SVM	56.4	48.1	69	73	52.3	50.9	112	106	53.6	53.2	104	94	54.1	50.7	95	91
Random Forest	54.0	47.5	56	51	56.2	53.1	133	84	50.8	48.7	79	47	53.7	49.8	89	61
XGBoost	54.2	43.5	57	40	54.7	51.7	96	81	50.4	48.5	60	58	53.1	47.9	71	59
CatBoost	54.8	49.7	53	57	55.1	50.3	123	83	51.2	49.9	44	67	53.7	50.0	73	69
Average	**54.7**	**48.4**	**57**	**55**	**54.0**	**50.1**	**106**	**78**	**51.7**	**50.7**	**70**	**75**	**53.5**	**49.7**	**78**	**69**
Max	**56.5**	**53.1**	**78**	**73**	**56.2**	**53.1**	**133**	**106**	**54.6**	**54.0**	**104**	**113**	**55.1**	**52.2**	**97**	**91**
B&H Strategy			66	66			107	107			66	66			79	79
STS Strategy			39	13			92	52			47	31			59	32

Method	**Bitcoin**				**Ethereum**				**Litecoin**				**Average**			
	Max Draw-Down %		Max Gain %		Max Draw-Down %		Max Gain %		Max Draw-Down %		Max Gain %		Max Draw-Down %		Max Gain %	
	4 h	**1D**	**4 h**	**1D**	**4 h**	**1D**	**4 h**	**1D**	**4 h**	**1D**	**4 h**	**1D**	**4 h**	**1D**	**4 h**	**1D**
Logistic Regression	18	28	132	145	40	38	263	167	23	27	191	114	27	31	195	142
KNN	8	9	67	126	13	−3	110	120	−19	28	118	354	1	11	98	200
SVM	19	11	119	159	37	12	265	238	1	27	286	145	19	17	223	181
Random Forest	13	4	123	145	33	16	249	225	−11	14	166	126	12	11	179	165
XGBoost	7	14	119	76	11	12	182	226	−19	8	127	174	0	11	143	159
CatBoost	2	24	88	117	17	21	273	252	–22	20	96	180	−1	22	152	183
Average	**11**	**15**	**108**	**128**	**25**	**16**	**224**	**205**	**−8**	**21**	**164**	**182**	**10**	**17**	**165**	**172**
Max	**19**	**28**	**132**	**159**	**40**	**38**	**273**	**252**	**23**	**28**	**286**	**354**	**27**	**31**	**223**	**200**
B&H Strategy	21	21	142	142	10	10	240	240	−6	−6	136	136	8	8	173	173
STS Strategy	15	0	62	38	11	0	188	110	2	0	108	108	9	0	119	85

crptocurrencies are higher than the ROIs obtained with B&H and STS strategies.

The trading simulations on Bitcoin data at 4-h frequency using Logistic Regression over four test periods (see Table 6 for the test periods) is shown in Fig. 10. The y-axis values in the left side represents the return over time. The black and blue lines represent the returns over time of the proposed ML-based and B&H strategies, respectively. The return of the B&H strategy is the same as the price change of the cryptocurrency, and the second vertical Y-axis in the right side shown in blue color represents the price of the cryptocurrency. For the ML strategy, the vertical red and green bands show the selling and buying signals, respectively.

As seen in Fig. 10, our ML-based strategy provided better returns

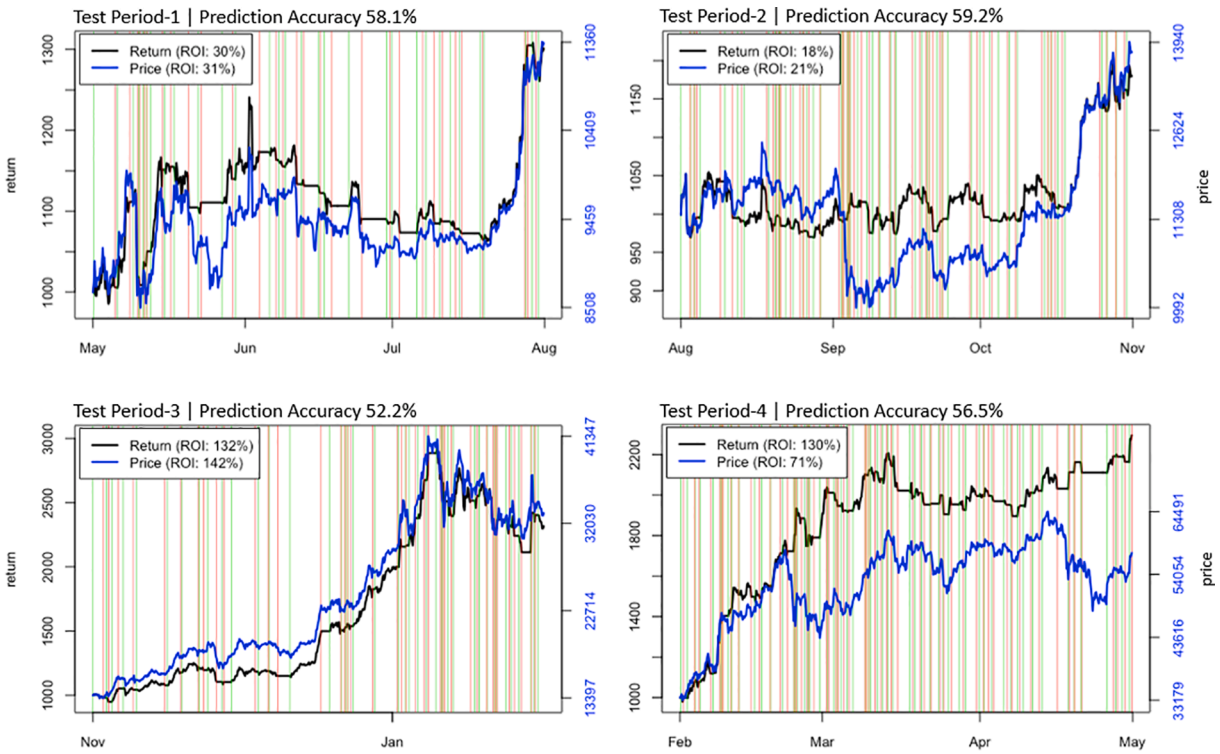

Fig. 10. Simulation of Bitcoin trading at 4-h data using Logistic Regression without outlier detection process.

than B&H strategy in trendless and volatile (zigzag) periods. We should also note that as our algorithm performs many buy and sell transactions, the commission expense creates a disadvantage compared to B&H strategy. In the four periods simulated in Fig. 10, the number of transactions was 61, 99, 81, and 118, respectively, and 13% commission was paid in average with a 0.1% commission rate for both buy and sell transactions.

4.2.2. Results with outlier detection

The features fed to the classification model that predicts the price movement direction are based on the technical indicators. However, as mentioned in Section 3.3.4, the cryptocurrency market is extremely sensitive to speculation, social media messages, and news feeds from public or private institutions which may cause price movements that cannot be modeled by the technical indicators. To address this problem, we first detected the subseries that have an unusual characteristic with the outlier detection approach mentioned in Section 3.3.4 and removed these subseries from the dataset. The training process was applied using the remaining data.

The results presented in Section 4.2.1 showed that the proposed ML-based trading strategy provides higher ROI on 4-h data. Therefore, in this section, we give the results obtained on 4-h data in a comparative manner with the results obtained without the outlier detection step. Table 8 shows the summary average of the test results obtained on the test periods shown in Table 6. As the results in Table 8 indicates, ROIs significantly increase with the outlier detection process for Bitcoin and Litecoin data while it is comparable for Ethereum. We should also note that the outlier detection step did not improve the prediction accuracy of the price movement direction task. Logistic Regression was superior to the other classifiers in both prediction accuracy and ROI. In parallel to the results presented in Table 7, SVM also performed well giving the second highest ROI after the outlier detection process.

The trading simulation performed in Section 4.2.1 was repeated using the logistic regression-based model with the outlier detection process. As seen in Fig. 11, the outlier detection process increased the ROIs of the proposed trading strategy for the first, third, and fourth test periods while did not make a positive or negative impact on the second test period. The highest increase in ROI was observed in the fourth period from 130% to 157%. The highest prediction accuracy of 63.3% was obtained on the second test period. However, the results demonstrated that the higher prediction accuracy does not always yield higher ROI showing that the characteristic of the price movements has a significant impact on the ROI.

4.2.3. Discussion

The classifiers we used to predict the next-period price direction produced different results in different test periods. In general, the results indicated that the logistic regression-based prediction models perform better than more complex non-linear models. This finding is mostly because the feature set that was extracted from the price data contains non-linear features and the capacity of the model based on the linear combination of these non-linear features is succifient to explain the relationship between the features and the price direction.

The trading simulations indicated that the ROI of the proposed trading framework are, on average, superior to the buy-and-hold and the baseline trading strategy built based on the use of a common technical indicator. The analysis of the results on different test periods also demonstrated that the trading system becomes more profitable than B&H strategy during the highly volatile periods while it performs comparably to the B&H strategy when there is an increasing or decreasing trend in the market.

In our study, unlike the common approach, the neutral samples whose price changes fall between the positive and negative values of the threshold were excluded from the training sets and the classification models were trained using only the samples that belong to positive and negative classes. This decision is based on the results obtained in our initial experiments. Our findings showed that using the neutral samples during training decrease the ROI of the trading framework. The main reason for this situation is that although the closing price used for labeling is not significantly different from the previous period for neutral samples, there are upward or downward movements in some of these samples within the relevant period. These samples, which can be seen as mislabeled samples, reduce the generalization ability of the classification model.

5. Conclusion and future work

The aim of this study was to develop an end-to-end ML-based trading system for cryptocurrency market and investigate the effect of using an outlier detection procedure on the prediction and trading performance. For this purpose, in the first part of the study, we developed and tested our machine learning-based strategy without the outlier detection as a baseline model. In this context, we employed six classification methods to predict the next period price direction. To assess the generalization ability of the system, we applied it on three cryptocurrencies with significant market caps. Besides, we used 4-h and 1-d frequency data of these cryptocurrencies to evaluate the performance of the trading system with different data frequency. The simulation results showed that the proposed ML-based strategy gives more profitable results in average than the baseline strategies. Also, it was observed that trading on the 4-h frequency provides higher ROI than 1-d.

In the second part of the study, we proposed to integrate an outlier detection step into the trading system to improve the ML-based strategy. For this purpose, we applied a clustering-based outlier detection approach on the time series data using the DTW technique to calculate the similarity scores between the price subseries. The results indicated that the ROI obtained with the trading system increases by removing the outlier subseries from the training set.

The proposed trading framework was trained to generate a buy signal only when an increase of more than 1% is predicted. Especially in periods with an upward trend in prices, this approach can result in profitable trades much higher than 1%. Thus, high ROIs can be achieved

Table 8
Comparative summary results after the outlier detection process (BO: before the outlier detection, AO: after the outlier detection).

Method	Bitcoin				Ethereum				Litecoin				Average			
	Avg. Accuracy %		Avg. ROI %		Avg. Accuracy %		Avg. ROI %		Avg. Accuracy %		Avg. ROI %		Avg. Accuracy %		Avg. ROI %	
	BO	AO	BO	AO	BO	AO	BO	AO	BO	AO	BO	AO	BO	AO	BO	AO
Logistic Regression	56.5	58.7	78	88	54.3	55.5	114	123	54.6	53.4	98	120	55.1	55.9	97	110
KNN	52.2	50.2	28	34	51.4	54.7	61	74	49.6	51.1	37	56	51.1	52.0	42	55
SVM	56.4	54.8	69	70	52.3	52.1	112	112	53.6	52.7	104	102	54.1	53.2	95	95
Random Forest	54.0	53.6	56	60	56.2	53.9	133	110	50.8	52.0	79	94	53.7	53.1	89	88
XGBoost	54.2	55.1	57	67	54.7	54.5	96	105	50.4	51.2	60	56	53.1	53.6	71	76
CatBoost	54.8	53.3	53	58	55.1	53.0	123	119	51.2	51.3	44	66	53.7	52.5	73	81
Average	**54.7**	**54.3**	**57**	**63**	**54.0**	**53.9**	**106**	**107**	**51.7**	**52.0**	**70**	**82**	**53.5**	**53.4**	**78**	**84**

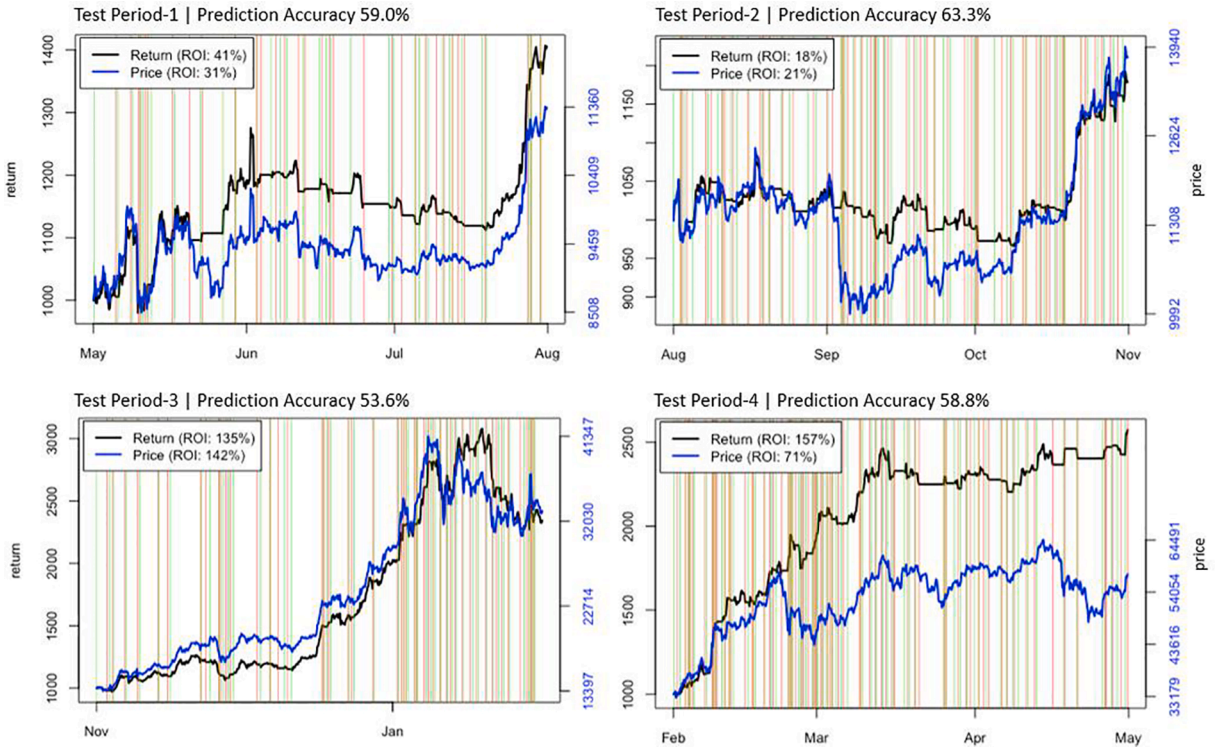

Fig. 11. Simulation of Bitcoin trading at 4-h data using Logistic Regression with outlier detection process.

by the model with ~60% accuracy, as an average of 60 out of 100 decisions made by the model result in a profitable trade with a return of 1% or more.

As a future research direction, considering that the cryptocurrency market is very sensitive to speculative messages on social media, a sentiment analysis module can be integrated into the system to improve the prediction ability and the outlier detection module. Besides, the trading system can be executed on various cryptocurrencies simultaneously and the probability estimations can be used to develop a portfolio management system that optimizes the amounts of balance that will be invested to each of the cryptocurreny.

CRediT authorship contribution statement

Faruk Ozer: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **C. Okan Sakar:** Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research, 297*(1), 3–36.

Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications, 149*, Article 113250.

Anghel, D. G. (2021). A reality check on trading rule performance in the cryptocurrency market: Machine learning vs. technical analysis. *Finance Research Letters, 39*, Article 101655.

Attanasio, G., Cagliero, L., Garza, P., & Baralis, E. (2019). Quantitative cryptocurrency trading: Exploring the use of machine learning techniques. In *In Proceedings of the 5th Workshop on Data Science for Macro-modeling with Financial and Economic Datasets* (pp. 1–6).

Benkabou, S. E., Benabdeslem, K., & Canitia, B. (2018). Unsupervised outlier detection for time series by entropy and dynamic time warping. *Knowledge and Information Systems, 54*(2), 463–486.

Bizzi, L., & Labban, A. (2019). The double-edged impact of social media on online trading: Opportunities, threats, and recommendations for organizations. *Business Horizons, 62*(4), 509–519.

Borges, T. A., & Neves, R. F. (2020). Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods. *Applied Soft Computing, 90*, Article 106187.

Bouri, E., Lau, C. K. M., Lucey, B., & Roubaud, D. (2019). Trading volume and the predictability of return and volatility in the cryptocurrency market. *Finance Research Letters, 29*, 340–346.

Brzeszczyński, J., & Ibrahim, B. M. (2019). A stock market trading system based on foreign and domestic information. *Expert Systems with Applications, 118*, 381–399.

Burnie, A., & Yilmaz, E. (2019). Social media and bitcoin metrics: Which words matter. *Royal Society open science, 6*(10), Article 191068.

Carta, S., Corriga, A., Ferreira, A., Recupero, D. R., & Saia, R. (2019). A holistic auto-configurable ensemble machine learning strategy for financial trading. *Computation, 7*(4), 67.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management, 11*(2), 23.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Gan, L., Wang, H., & Yang, Z. (2020). Machine learning solutions to challenges in finance: An application to the pricing of financial products. *Technological Forecasting and Social Change, 153*, Article 119928.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832–844.

Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: A survey. *Enterprise Information Systems, 13*(1), 132–144.

Jalal, R. N. U. D., Alon, I., & Paltrinieri, A. (2021). A bibliometric review of cryptocurrencies as a financial asset. *Technology Analysis & Strategic Management*, 1–16.

Jeong, G., & Kim, H. Y. (2019). Improving financial trading decisions using deep Q-learning: Predicting the number of shares, action strategies, and transfer learning. *Expert Systems with Applications, 117*, 125–138.

Kakushadze, Z. (2016). 101 formulaic alphas. *Wilmott, 2016*(84), 72–81.

Koker, T. E., & Koutmos, D. (2020). Cryptocurrency trading using machine learning. *Journal of Risk and Financial Management, 13*(8), 178.

Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies, 34*(6), 2689–2727.
Longo, R., Podda, A. S., & Saia, R. (2020). Analysis of a consensus protocol for extending consistent subchains on the bitcoin blockchain. *Computation, 8*(3), 67.
Misnik, A., Krutalevich, S., Prakapenka, S., Borovykh, P., & Vasiliev, M. (2019). In *September). Impact Analysis of Additional Input Parameters on Neural Network Cryptocurrency Price Prediction* (pp. 163–167). IEEE.
Monrat, A. A., Schelén, O., & Andersson, K. (2019). A survey of blockchain from the perspectives of applications, challenges, and opportunities. *IEEE Access, 7*, 117134–117151.
Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review, 21260*.
Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). *Bitcoin and cryptocurrency technologies: A comprehensive introduction*. Princeton University Press.
Niranjanamurthy, M., Nithya, B. N., & Jagannatha, S. J. C. C. (2019). Analysis of Blockchain technology: Pros, cons and SWOT. *Cluster Computing, 22*(6), 14743–14757.
Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications, 115*, 635–655.
Phillips, R. C., & Gorse, D. (2017). Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In 2017 IEEE symposium series on computational intelligence (SSCI) (pp. 1-7). IEEE.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2017). CatBoost: unbiased boosting with categorical features. arXiv preprint arXiv: 1706.09516.
Rundo, F., Trenta, F., di Stallo, A. L., & Battiato, S. (2019). Machine learning for quantitative finance applications: A survey. *Applied Sciences, 9*(24), 5574.
Sattarov, O., Muminov, A., Lee, C. W., Kang, H. K., Oh, R., Ahn, J., … Jeon, H. S. (2020). Recommending cryptocurrency trading points with deep reinforcement learning approach. *Applied Sciences, 10*(4), 1506.
Sun, J., Zhou, Y., & Lin, J. (2019). Using machine learning for cryptocurrency trading. In 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS) (pp. 647-652). IEEE.
Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy, 21*(6), 589.
Wang, H., Bah, M. J., & Hammad, M. (2019). Progress in outlier detection techniques: A survey. *IEEE Access, 7*, 107964–108000.
Zhengyang, W., Xingzhou, L., Jinjin, R., & Jiaqing, K. (2019). Prediction of cryptocurrency price dynamics with multiple machine learning techniques. In *In Proceedings of the 2019 4th International Conference on Machine Learning Technologies* (pp. 15–19).