BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 72(4), 2024, Article number: e150117
DOI: 10.24425/bpasts.2024.150117

An analysis of the influence of famous people's posts on social networks on the cryptocurrency exchange rate

Sergii TELENYK[1], **Grzegorz NOWAKOWSKI**[1]*, **Olena GAVRILENKO**[2], **Mykhailo MIAHKYI**[2], and **Olena KHALUS**[2]

[1] Faculty of Electrical and Computer Engineering, Cracow University of Technology, Warszawska 24, 31-155 Cracow, Poland
[2] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Prosp. Peremohy 37, Kyiv, Ukraine

Abstract. In this work, the level of influence of the posts published by famous people on social networks on the formation of the cryptocurrency exchange rate is investigated. Celebrities who are familiar with the financial industry, especially with the cryptocurrency market, or are somehow connected to a certain cryptocurrency, such as Elon Musk with Dogecoin, are chosen as experts whose influence through social media posts on cryptocurrency rates is examined. This research is conducted based on statistical analysis. Real cryptocurrency exchange rate forecasts for the selected time period and predicted ones for the same period, obtained using three algorithms, are utilized as a dataset. This paper uses methods such as statistical hypotheses regarding the significance of Spearman's rank correlation coefficient and Pearson's correlation. It is confirmed that the posts by famous people on social networks significantly affect the exchange rates of cryptocurrencies.

Keywords: cryptocurrency exchange rate; forecasting algorithms; posts on social networks; methods of statistical analysis; information technology of intellectual analysis.

1. INTRODUCTION

Today, people can obtain the information they need through the Internet without leaving their homes. But there are problems with availability – the Internet is filled with a large quantity of non-informative data that has no value for the vast majority of users, or in other words, "garbage data". That is why every day it becomes increasingly difficult to find relevant information because huge amounts of data must be reviewed and analyzed.

This problem is inherent in all spheres of human activity, whether it is building one's own business or participating in society through governing bodies. In addition, the appearance of a large amount of non-informative data is, as always, the fault of humanity itself, because every entrepreneur seeks to preserve as much information as possible about the state of their business and at the same time strives to remain competitive, which is a vital factor for any enterprise.

In view of all this, methods for Internet content analysis must be constantly developed to facilitate finding the necessary data conveniently, and most importantly, quickly. The most expensive commodity today is information, and at the moment the main sources are social networks, through which people share their thoughts and plans. However, to use these posts for specific purposes, the data must be analyzed.

The collection and processing of data related to the financial sphere is of particular interest, especially the exchange rates of various currencies and cryptocurrencies. Cryptocurrency is gaining more popularity every day due to the relative ease of entry and the large amount of recommended information available about the process. Buying and selling cryptocurrency is an interesting process because certain conditions are met, an individual can increase their wealth several times, or even replace their main job with this activity. However, to truly make money from this process, it is necessary to conduct research regarding the chosen cryptocurrency, as well as its exchange rate and news concerning it.

The relevance of this research is due to the growing popularity of investing in cryptocurrency. Posts by famous people who have their own interest in this process have a significant influence on the formation of the prices of certain cryptocurrencies. When traders create forecasts regarding changes in the exchange rate of certain cryptocurrencies, they need an information system that can analyze the impact of such publications on changes in the exchange rates of cryptocurrencies and provide recommendations on market behaviour. This will increase the accuracy of the prediction created.

This work presents research that can be used by financial market participants to obtain high-quality forecasts of cryptocurrency rates, based on which they can make decisions about its purchase (or sale).

*e-mail: gnowakowski@pk.edu.pl

Manuscript submitted 2023-06-01, revised 2024-02-26, initially accepted for publication 2024-04-04, published in July 2024.

2. ANALYSIS OF LITERARY SOURCES AND FORMULATION OF THE PROBLEM

In general, the task of analyzing Internet posts is very important, because a well-analyzed publication can provide much more information than a simple question to the author such as: "What did you mean by this publication?". Detailed analysis of posts on social networks will allow for obtaining information about preferences and professional activity, as well as the users' circle of communication and their mutual influence.

In the paper [1], the process of the computer detection and categorization of opinions expressed in a piece of text to determine whether the writer's attitude towards a certain topic, product, etc., is positive, negative, or neutral is examined. Within the framework of the research presented in this paper, a detailed study of the analysis of moods and the cause-and-effect relationship of moods was carried out. In addition, using sentiment analysis, a generalized event can be identified based on mood and time. The results of the causality analysis can be used not only to determine the causes and effects, respectively, but also for their subsequent prediction of user sentiments. The main part of the publication is an overview of the combination of these two approaches, combined into a model that allows for determining the mood during future events, as well as creating a time forecast about the length of the interval between certain events. The average relative error was used to assess accuracy.

To view posts, you need to choose a place where a high number are available and stored in a single text format. A social network such as Twitter is useful for this. Special linguistic analysis and Twitter statistics are discussed in detail in [2]. This study aimed to identify criminal elements in the United States by modelling discussion topics and then incorporating them into a crime prediction model. A study was conducted on the impact of social media posts on future crimes.

In [3], a comprehensive reference for researchers and practitioners was considered, as well as coverage of all areas that contribute to the construction and analysis of social networks.

The paper [4] presents an integrated framework that offers the infrastructure needed to access, integrate, and analyze multilingual user-generated content from various social media sites.

The paper [5] demonstrates that Twitter messages (tweets) can be reliably classified based on flu-related keywords; the spread of flu can be predicted with high accuracy; and there is a way to monitor the spread of flu in selected cities in real time. We propose an approach to efficiently mine and extract data from Twitter streams, reliably classify tweets based on their sentiment, and visualize the data using an interactive real-time map.

The paper [6] shows what topics are being voiced by individuals and groups about the pandemic. It is determined whether there are any noticeable thematic trends, and if so, how these topics change over time and in response to important events. Using an improved sequential latent distribution model, the twelve most popular topics present in the Twitter dataset collected from April 3 to 13, 2020 in the United States are identified and their growth and changes are discussed.

It is also impossible not to highlight Internet blogs, in which many people express their own opinions and visions of certain problems, etc. Therefore, in [7], a study was conducted on the identification of hate groups. The proposed approach is semi-automatic and consists of four modules, namely: blog spider, information retrieval, network analysis, and visualization. This investigation was conducted on the blogging site Xanga. The results of the analysis revealed some interesting demographic and topological characteristics in hate groups and identified at least two large communities in addition to the smaller ones. The proposed approach is also appropriate for the examination of hate groups and other related blog communities.

In terms of business and the financial market, the process of analyzing large amounts of data and understanding the needs of the majority of people is very important, as it directly affects the income of the company and individuals. In [8], a study of the dominant factors that lead to currency crises was conducted. Within the framework of the research presented in this publication, the nature and characteristics of currency crises were identified and the forecasting of possible currency crises at an early stage was conducted. This can save managers some time in improving crisis management policies and corrective actions.

The work of [9] investigated the dynamics of linear and non-linear serial dependencies in financial time series within the framework of a moving window. In particular, the focus was on identifying episodes of statistically significant two- and three-tribal correlation in the returns of several leading exchange rates, which may offer some potential for their predictability. A moving-window approach was used to capture correlation dynamics for different window lengths and to analyze the distribution of periods with statistically significant correlations. It was found that for sufficiently large window lengths, these distributions correspond well to the power law. Predictability itself was measured by the hit rate, i.e. the level of agreement between the actual return features and their predictions obtained using a simple correlation-based predictor.

It should be noted that in all the cited works, the research is general and the results of forecasting currency rates, in particular cryptocurrencies, were not provided. Accordingly, the factors affecting them were not investigated.

In [10], a study of the main macroeconomic indicators of the influence on the US dollar exchange rate in Ukraine was carried out, considering the purchase/sale of cash and non-cash currency, the balance of these, inflation in the current year, and nominal and real gross domestic product, as well as purchases/sales by bank clients, transactions between banks, gross and net international reserves, unemployment rates, and accounting (interest) rates, in addition to the balance of foreign exchange interventions, and the volume of nominal value transactions. Using the method of main components, the main economic components of the formation of the exchange rate were determined. With the help of an autoregressive integrated moving average (ARIMA), exponential smoothing, and singular spectrum analysis statistical models, the values of the selected influencing factors were predicted. The values of currency rates were predicted using regression models built by fast tree, fast forest, fast tree Tweedie, and generalized additive model algorithms, and the obtained values were studied for accuracy. This work did not forecast the exchange rates of cryptocurrencies in particular and did not investigate the influence of factors such as posts on social networks.

The paper [11] analyzed methods, areas of application, and approaches to the analysis of publications and forecasting events based on the collected data, as well as the concept of the influence of publications on changes in the cryptocurrency exchange rate. The justification of the topicality of the topic was presented and the possibilities of appropriate application of the results of the work were described. The main stages of working with event forecasting data were defined, namely: the pre-processing of data, their further analysis, and forecasting. This work did not investigate the level of influence of posts by famous people on social networks on the cryptocurrency exchange rate. Within the framework of the research presented in the previously mentioned papers [10, 11], information systems were created to implement the above-described tasks of intellectual data analysis.

The above discussion indicates that the influence of certain factors on the exchange rate of cryptocurrencies, especially that of posts by famous people on social networks, has not been sufficiently studied to date and requires further study.

3. THE PURPOSE AND OBJECTIVES OF THIS STUDY

The purpose of this study is to examine the level of influence of posts by famous people on social networks on the cryptocurrency exchange rate. This will make it possible to increase the reliability of the forecast of the exchange rate of cryptocurrencies. To achieve this purpose, the following aims were set:

- To form statistical samples based on real exchange rates of the selected cryptocurrency, as well as their forecasts based on ARIMA, exponential smooth algorithms, and an algorithm for taking into account posts on social networks (ATAPSN).
- To calculate forecasting errors for all considered models and choose the most accurately forecasted cryptocurrency rates.
- To establish the level of dependence between the predicted and real exchange rates of cryptocurrency.
- To determine the statistical significance of the forecasted cryptocurrency exchange rates.

4. RESEARCH MATERIALS AND METHODS

The object of this study is the level of influence of posts on social networks on the exchange rate of cryptocurrencies. The necessary information includes forecasts for a selected period, taking into account the influence of posts by famous people on social networks (ATAPSN) [11], using the ARIMA [12] and exponential smoothing algorithms [13], as well as real cryptocurrency rates for the same period.

Time-limited impact: The first and most important argument is that publications and posts have the greatest impact on decision-making during the period of greatest interest. In the first hours, days, or weeks after publication, information can be very relevant and important for decision-making.

Patterns of influence: Some studies show that the impact of publications and posts on decision-making decreases rapidly over time. People may react more quickly to news and information that has appeared recently, and their reactions may be more intense.

Competition for attention: In an information society, competition for consumer attention is intense. Over time, new events and information can displace old ones, and the impact of publications can be significantly reduced.

Importance of freshness of information: Some topics or industries may require fresh information to make accurate decisions. For example, in the financial sector, information can become outdated quickly, and making decisions based on outdated information can be more risky.

Of course, the length of the period for a study depends on the specific context and objectives of the study. For example, when providing certain services, it is necessary to take into account the opinions of users from social networks based on their posts.

The application of a mathematical apparatus using the rank correlation coefficient allows for using this information to determine the relationship between the real and predicted cryptocurrency rates. Therefore, the following information is required to form the dataset: the real exchange rates of the chosen cryptocurrency for a certain period, as well as those predicted using various algorithms.

For ATAPSN, well-known individuals who are knowledgeable in the field of finance in general and cryptocurrencies in particular, or whose activities are somehow related to a certain cryptocurrency, were chosen as experts. As part of the conducted research, the Dogecoin cryptocurrency was selected to form the dataset and Elon Musk's posts on the Twitter social network were taken into account.

A fragment of the dataset is given in Table 1.

Table 1
A fragment of the input dataset

Hours	Real rates	Algorithm ATAPSN	Algorithm ARIMA	Exponential smoothing
1	467	466.19	497.8707174	502.2597
2	475	473.44	502.1439974	463.9021
3	516	513.78	506.4172773	490.0742
4	533	534.78	510.6905573	515.4008
5	508	508.31	514.9638372	523.1350
6	510	508.60	519.2371172	528.5218
7	525	525.67	523.5103971	518.3018
8	512	510.85	527.7836771	515.8382
9	514	512.55	532.0569570	529.6971
10	514	515.43	536.3302370	521.8438

Table 1 shows forecasts of Dogecoin rates, which were obtained using three algorithms (ATAPSN, ARIMA, exponential smoothing), for 10 hours, as well as its real rates for the specified period. The data were taken from the site of the crypto exchange Binance [14].

The dataset assembled in this way formed the input information for this study. As part of the conducted research, the following steps were necessary:

- Obtain forecasts based on ATAPSN and classic ARIMA and exponential smoothing algorithms and then form statistical samples based on them.
- Calculate forecasting errors for all considered models and choose the most accurately forecasted cryptocurrency rates.
- Establish the level of dependence between the predicted and real exchange rates of cryptocurrency using Spearman and Pearson correlation coefficients.
- Determine the statistical significance of the predicted cryptocurrency rate using the Student's t-test.

The use of the specified methods of statistical analysis guarantees obtaining reliable results when forecasting cryptocurrency rates and researching how they are influenced by the posts of famous people on social networks.

Software was developed to perform statistical analysis and obtain results based on the specified methods. It consisted of two main parts: the client (site) and the server. The client part was implemented using the Angular framework and was designed to display the resulting data, as well as to configure the search part.

The server part was implemented using the Python language and consisted of the following main blocks: the search block of publications – with the main goal to collect data about the publications of a certain person for a certain time; the cryptocurrency search block – chiefly intended to collect data on cryptocurrency exchange rates for the specified period; the algorithmic block – aimed at analyzing and forecasting future cryptocurrency rates; the communication unit – with the main purpose of data transfer between the client and server part and other blocks.

5. STUDY OF THE LEVEL OF INFLUENCE OF POSTS BY FAMOUS PEOPLE ON SOCIAL NETWORKS ON THE CRYPTOCURRENCY RATE

5.1. Formation of statistical samples based on real rates of the selected cryptocurrency, as well as their forecasts based on the ARIMA, exponential smoothing, and ATAPSN algorithms

Formulation of the problem: Based on the input dataset, form the following statistical samples X, Y_i $(i = \overline{1,3})$ of volume n each (n is the number of experiments (forecasts) made during the selected time [15]. The value of n is set depending on the desired accuracy of forecasting: hourly, daily, etc.):

X is a set of real cryptocurrency rates.

Y_1 is a set of forecasted cryptocurrency rates obtained using ATAPSN.

Y_2 is a set of forecasted cryptocurrency rates obtained using the ARIMA algorithm.

Y_3 is a set of predicted cryptocurrency rates obtained using the exponential smoothing algorithm.

Justification. Algorithm for forecasting the cryptocurrency exchange rate taking into account posts on social networks (ATAPSN) [11]. The idea of the algorithm is to calculate the coefficient of significance of the expert's post c_j, which is calculated according to the formula:

$$c_j = k_j \cdot ch_j\,, \tag{1}$$

where ch_j is the assessment of the tonality of the expert's post:

$$ch_j = \begin{cases} 1, & \text{if the post is positive,} \\ 0, & \text{if the post is neutral,} \\ -1, & \text{if the post is negative,} \end{cases} \tag{2}$$

k_j is the correctness of the forecast made at the previous moment in time:

$$k_j = |y_j - x_j|, \tag{3}$$

where y is the predicted value of the cryptocurrency exchange rate obtained using time series; x_j is the actual value of the cryptocurrency exchange rate; j is the moment in time.

After determining the coefficient c_j from (1)–(3) a forecast of the change in the cryptocurrency exchange rate will be created based on the available data on the expert's posts in the selected social network for the specified time:

$$y'_{j+1} = y_{j+1} + c_j\,. \tag{4}$$

Algorithm for forecasting the cryptocurrency exchange rate using the ARIMA algorithm [12]. The ARIMA(p, d, q) model for a non-standard time series looks like this:

$$\Delta^d y_j = c + \sum_{k=1}^{p} a_k \Delta^d y_{j-k} + \sum_{k=1}^{q} b_l \varepsilon_{j-l} + \varepsilon_j\,, \tag{5}$$

where ε_j is a stationary time series; c, a_k, b_l are model parameters; Δ^d is a time series difference operator of order d.

Algorithm for forecasting the cryptocurrency exchange rate using the exponential smoothing algorithm [13]. The exponential smoothing model is as follows:

$$y'_j = \begin{cases} y_1 : & j = 1, \\ y'_{j-1} + \alpha\left(y_j - y'_{j-1}\right) : & j > 1, \end{cases} \tag{6}$$

where y'_j is smoothed series; y_j is primary series; α is the smoothing coefficient.

Statistical samples X, Y_i $(i = \overline{1,3})$ of volume n each are formed based on forecasts obtained n times during the considered time using formulas (4)–(6).

The representation of cryptocurrency rates, both real and forecasted, in the form of a statistical sample, facilitates the application of statistical analysis methods to calculate forecasting errors, determine the level of dependence between forecasted and real rates of the selected cryptocurrency, and establish the statistical significance of forecasted cryptocurrency rates.

The choice of forecasting algorithms is determined by the need to compare forecasts using ATAPSN [11], the focus of this study, with classical forecasts made utilizing time series [16] including ARIMA and exponential smoothing [12, 13].

It should also be noted that the quality of the obtained samples depends on the accuracy of forecasting. Therefore, before conducting this research, to obtain correct forecasts of cryptocurrency exchange rates, it is necessary to conduct pre-processing of the data (reject abnormal forecasts, establish a sufficient volume of samples, normalize data, etc.).

Result. Based on the dataset presented in Table 1, the following statistical samples in volume were obtained $n = 10$:

$$X = (467.0, 475.0, 516.0, 533.0, 508.0, 510.0, 525.0, 512.0, 514.0, 514.0),$$
$$Y_1 = (466.19, 473.44, 513.78, 534.78, 508.31, 508.6, 525.67, 510.85, 512.55, 515.43),$$
$$Y_2 = (497.8707174289546, 502.14399737658755, 506.4172773242205, 510.69055727185344, 514.9638372194864, 519.2371171671194, 523.5103971147524, 527.7836770623854, 532.0569570100184, 536.3302369576514),$$
$$Y_3 = (502.259726, 463.902197, 490.07416, 515.400778, 523.134984, 528.521778, 518.301774, 515.838192, 529.697092, 521.84384).$$

5.2. Calculating the forecasting error for all considered models and choosing the most accurately forecasted cryptocurrency rates

Formulation of the problem. Let the pairs of statistical samples (X, Y_i), $i = \overline{1, 3}$, be given (see Section 5.1).

For each such pair of samples, it is necessary to calculate the deviation of the elements of the sample X from the corresponding elements of the sample Y_i.

Justification. The deviation of the elements of the sample X from the corresponding elements of the sample Y_i is calculated as a relative error according to the formula:

$$R_{ij} = \frac{|x_j - y_{ij}|}{x_j} 100\%, \tag{7}$$

where x_j are elements of the sample X; y_{ij} are elements of the sample Y_i; $i = \overline{1, 3}$, $j = \overline{1, n}$; n is the volume of samples X and Y_i [17].

Then, the average relative error of the sample is calculated by the formula:

$$R_i = \frac{\sum_{j=1}^{n} R_{ij}}{n}. \tag{8}$$

The proposed approach makes it possible to assess the accuracy of cryptocurrency forecasts to assess the quality of samples Y_1, Y_2, and Y_3. It also allows for choosing from the proposed algorithms the one that gives the highest accuracy, to further use the forecasts obtained with the help of this particular algorithm.

Results. Values for relative errors and their average values obtained using formulas (7) and (8) for the samples presented in Section 5.1 are provided in Table 2.

Table 2
Values of forecasting errors

X	Y_1	Y_2	Y_3
467	466.19	497.8707174	502.2597
475	473.44	502.1439974	463.9021
516	513.78	506.4172773	490.0742
533	534.78	510.6905573	515.4008
508	508.31	514.9638372	523.1350
510	508.60	519.2371172	528.5218
525	525.67	523.5103971	518.3018
512	510.85	527.7836771	515.8382
514	512.55	532.0569570	529.6971
514	515.43	536.3302370	521.8438
R_i	0.23406855	3.27736587	3.1429464

In Table 2, it is easy to see that the sample Y_1 has the smallest average relative error (0.23%), followed by the sample Y_3 (3.14%), and Y_2 has the largest error (3.28%).

5.3. Setting the level of dependence between the predicted and real cryptocurrency rates

Formulation of the problem. Let the pairs of statistical samples (X, Y_i), $i = \overline{1, 3}$, be given (see Section 5.1).

It is necessary to establish the level of dependence between pairs of samples (X, Y_i), taking into account the sequence of elements in the sample.

Justification. In general, the Pearson correlation coefficient is used to establish the level of dependence between two samples [18]:

$$r_i = \frac{\sum_{j=1}^{n} (x_i - \bar{x}) (y_{ij} - \bar{y_j})}{\sum_{j=1}^{n} (x_i - \bar{x})^2 \sum_{j=1}^{n} (y_{ij} - \bar{y_j})^2}, \tag{9}$$

where x_j are elements of sample X; y_{ij} are elements of the sample Y_i; $\bar{x}$, $\bar{y_j}$ are their sample averages; $i = \overline{1, 3}$; $j = \overline{1, n}$; n is the volume of samples X and Y_i.

Since the sequence of items in the samples is important for this research (real and forecasted cryptocurrency rates are considered and compared at the appropriate time), using the rank correlation coefficient, in particular the Spearman coefficient, is recommended [19].

Spearman's rank correlation coefficient is calculated using the formula:

$$\rho_i = 1 - \frac{6 \sum_j d_j^2 + T_X + T_{Y_i}}{n(n^2 - 1)}, \tag{10}$$

where $d_j = x_j - y_{ij}$ is the rank difference for the j-th element of the sample Y_i; $i = \overline{1, 3}$; $j = \overline{1, n}$; n is the volume of samples X and Y_i.

If there are connected ranks in the samples, i.e. the ranks of the elements of the sample are repeated:

$$T_X = \frac{N_X^3 - N_X}{12} \quad \text{and} \quad T_{Y_i} = \frac{N_{Y_i}^3 - N_{Y_i}}{12}, \tag{11}$$

(N_X and N_{Y_i}, respectively, is the number of repeated ranks in the samples X and Y_i, $i = \overline{1,3}$). For a sample that does not have repeated elements, the corresponding coefficient from formula (11) is zero.

The value of both correlation coefficients lies in the interval $[-1, 1]$.

It should be noted that Spearman's rank correlation coefficient (10) is less accurate compared to Pearson's correlation coefficient (9) since its calculation does not take into account the quantitative values of sample elements, but only their order. This indicates the need to consider the Pearson correlation coefficient in a more in-depth determination of the level of dependence of samples X and Y_i, $i = \overline{1,3}$. It is implemented using Table 3.

Table 3
Chaddock's scale of identifying the strength of the relationship by the value of the paired correlation coefficient (rank correlation)

The value of the correlation coefficient (by module)	0.1–0.3	0.3–0.5	0.5–0.7	0.7–0.9	0.9–0.99
Characteristics of the bond strength	Weak	Moderate	Notable	Strong	Very strong

Note that if the correlation coefficient (rank correlation) is equal to 1, there is a very strong (up to functional) directly proportional relationship between random variables. If it is equal to –1, then there is a very strong (to functional) inversely proportional relationship between random variables.

The proposed approach makes it possible to assess how close a correlation exists between the real and forecasted exchange rates of the chosen cryptocurrency for the specified period of time.

Result. For the samples presented in Section 5.1, the results are shown in Table 4.

In this table, the correlation coefficients are calculated according to formulas (9)–(11), and the strength of the connection with the sample X is established using Table 3.

The data given in Table 2 emphasizes that the sample Y_1 has the strongest correlation with the sample X among all considered samples (very strong relationship). The sample Y_2 has an average level of connection with the sample X, while the sample Y_3 has the worst indicators.

5.4. Calculating the forecasting error for all considered models and choosing the most accurately forecasted cryptocurrency rates

Formulation of the problem. Let (X, Y_i), $i = \overline{1,3}$, be a given pair of samples. We know the coefficient of rank correlation between them (Spearman's rank correlation coefficient ρ_i or Pearson correlation coefficient r_i, the values of which were obtained in Section 5.3).

Table 4
Correlation analysis between pairs of samples (X, Y_i), $i = \overline{1,3}$

The sample	Spearman's rank correlation coefficient	Pearson's correlation coefficient	The strength of bond with the sample X
Y_1	0.97828283	0.99829258	A very strong connection both in terms of the similarity of the elements of the samples and in terms of taking into account their order of follow-up
Y_2	0.39040404	0.581244474	A noticeable connection from the point of view of the similarity of sample elements, and moderate from the point of view of taking into account their sequence
Y_3	0.09949495	0.58275805	A noticeable connection in terms of the similarity of the elements of the samples, but almost absent in terms of taking into account their order of follow-up

It is necessary to test the hypothesis about the significance of the corresponding rank correlation coefficient at the level of significance α.

Justification. To solve the formulated problem, the following rules were used, which are part of the Student's t-test [20, 21].

Rule 1. To test the hypothesis about the significance of the Spearman rank correlation coefficient at the significance level α, it is necessary to calculate the observed value

$$t_i = t_{\text{cr}}(\alpha, k)\sqrt{\frac{1-\rho_i^2}{n-2}}, \tag{12}$$

where the critical value $t_{\text{cr}}(\alpha, k)$ is taken from the Student distribution table according to the level of significance α and the degree of freedom $k = n-2$; ρ_i is the Spearman's rank correlation coefficient; $i = \overline{1,3}$, n is the volume of samples X and Y_i.

If $|\rho_i| \le t_i$, then the hypothesis is accepted. If not, it is rejected, that is, there is a significant correlation between the samples.

Rule 2. To test the hypothesis of the significance of the Pearson correlation coefficient at the significance level α, it is necessary to calculate the observed value:

$$t_i = \sqrt{\frac{r_i^2(n-2)}{1-r_i^2}}, \tag{13}$$

where n is the volume of samples X and Y_i; r_i is Pearson correlation coefficient; $i = \overline{1,3}$.

Next, it is necessary to compare it with the tabular critical value of this criterion $t_{\text{cr}}(\alpha, k)$ (which is taken from the Student distribution table [20, 21] taking into account the given level of significance α, which is a sufficient level to obtain reliable results, and the number of degrees of freedom $k = n - 2$).

If $t_i \le t_{cr}(\alpha, k)$, then the hypothesis is accepted. If not, it is rejected, that is, there is a significant correlation between the samples.

By a significant correlation between the predicted rates of cryptocurrency, i.e. samples Y_i are well consistent with real rates at similar moments of time, i.e. with the sample X. In particular, if we consider a pair of samples (X, Y_1), within the framework of the proposed approach, it is possible to track how much the forecast of the cryptocurrency exchange rate, taking into account the posts by famous people on social networks, correlates with the real exchange rates for a similar period. This makes it possible to assess the level of influence of these posts on cryptocurrency rates, which is the main goal of this study.

Result. For the samples presented in Section 5.1, taking into account the rank correlation coefficients defined in Section 5.3, the results are shown in Tables 5 and 6.

Table 5

Results of testing the hypothesis about the significance of the Spearman rank correlation coefficient between pairs of samples (X, Y_i) at $\alpha = 0.05$ and $t_{cr}(0.05, 8) = 2.31$

The sample	Spearman's rank correlation coefficient ρ_i	Value of t_i	Conclusions about accepting or rejecting the hypothesis
Y_1	0.97828283	0.16	Significant correlation
Y_2	0.39040404	0.75	Non-significant correlation
Y_3	0.09949495	0.81	Non-significant correlation

Table 6

Results of testing the hypothesis about the significance of the Pearson rank correlation coefficient between pairs of samples (X, Y_i) at $\alpha = 0.05$ and $t_{cr}(0.05, 8) = 2.31$

The sample	Pearson correlation coefficient r_i	Value of t_i	Conclusions about accepting or rejecting the hypothesis
Y_1	0.99829258	48.34	Significant correlation
Y_2	0.581244474	2.02	Non-significant correlation
Y_3	0.58275805	2.03	Non-significant correlation

In these tables, correlation coefficients are calculated according to formulas (9)–(11), and critical values according to formulas (12) and (13), respectively.

The calculation results shown in Tables 5 and 6 confirm that there is a significant correlation in the pair of samples (X, Y_1), while the correlation in the pairs of samples (X, Y_2) and (X, Y_3) is insignificant.

6. THE INTRACTABLE PROBLEM OF MINIMIZING THE TOTAL TARDINESS OF PARALLEL MACHINE COMPLETION TIMES REGARDING THE COMMON DUE DATE WITH MACHINE RELEASE TIMES

The use of statistical information, the formation of a dataset (Table 1), and the calculation of informative indicators based on the methods of statistical analysis allow for solving the formulated research tasks.

It is worth emphasizing that the fact that both samples were obtained from real and predicted cryptocurrency rates over the same period and at the same moments is crucial. In addition, the order in which they were obtained should be taken into account. This makes it possible to justify the legitimacy of the transition from time series to statistical samples and, thus, to correctly calculate the average relative forecasting error.

The accuracy of the obtained results depends on the size and quality of samples X, Y_1, Y_2, and Y_3, presented in Section 5.1. In this regard, it is recommended to carry out thorough data processing before starting the research.

The classic ARIMA and exponential smoothing algorithms, which have proven themselves in solving forecasting problems, were used to create samples Y_2 and Y_3. This study showed that these algorithms do not make it possible to track the influence of any specific factor, unlike the ATAPSN algorithm, based on which the sample Y_1 was created. The advantage of the proposed method of presenting the dataset (Table 1) in the form of statistical samples is the possibility of expanding the area of use of the results of this study far beyond the boundaries of finance. An example of the application of the proposed approach to the analysis of public services is given in [22]. This approach was used to study the relationship between some pillars of European regional competitiveness depending on the quality of regional institutions [23].

The proposed approach can also be applied in various studies related to the design, implementation, management, and development of services and in general supporting the life cycle of services in high-tech industries [24, 27], and to determine which appliances can be considered for the Demand Response programme [25].

Checking the accuracy of the received forecasts, and therefore the quality of samples Y_1, Y_2, and Y_3, was carried out by calculating the average value of the relative error (see Section 5.2). The results of the analysis of the received forecasting errors showed that forecasts based on the sample Y_1 are the most accurate (error 0.23%) (see Table 2). That is, the algorithm, which takes into account posts by famous people on social networks, makes it possible to obtain better results in terms of accuracy.

Considering the fact that the input dataset (see Table 1) was formed from predicted and real cryptocurrency rates for a certain period when studying the correlation between pairs of samples (X, Y_i), $i = \overline{1, 3}$, it is appropriate to use rank correlation coefficients. The Pearson correlation coefficient was also utilized to clarify the obtained results.

According to the results obtained in Section 5.3, it can be stated that the largest correlation exists between the samples from the pair (X, Y_1), which indicates that the predictions of the

Dogecoin cryptocurrency exchange rates, taking into account the posts by Elon Musk on social networks, are most consistent with the real values of these rates. At the same time, it is worth noting that a very strong correlation between samples from the pair (X, Y_1) was confirmed using both correlation coefficients, which guarantees the correctness and accuracy of the obtained results.

Obtaining high-quality forecasts by the ATAPSN algorithm, which are used to form the sample Y_1, depends on the correctness of the selected experts, and for this purpose, it is necessary to choose the social network that best suits the specifics of the study [26].

In addition, the data given in Table 4 confirm that for pairs of samples (X, Y_2) and (X, Y_3) Spearman's rank correlation coefficient is smaller than Pearson's correlation coefficient. This is because the elements in samples Y_2 and Y_3 do not differ too much from the elements of the sample X, which is confirmed by the value of 3% relative error (see Table 2) for both samples. But the orders of placement of elements for pairs of samples (X, Y_2) and (X, Y_3) have moderate and almost no correlations, respectively, because samples Y_2 and Y_3 were formed using algorithms that are based on the use of time series that are not able to track the influence of a specific factor on the dependent variable, in contrast to ATAPSN. Therefore, such forecasts achieved a sufficient level of accuracy, but at the same time, they did not take into account the trends of constant growth or decline of cryptocurrency rates.

The assessment of the correlation between the real and predicted cryptocurrency rates can also be carried out using the Kendall rank correlation coefficient [19], which could be the subject of further research.

To check the statistical significance of the Pearson correlation coefficient (Spearman's rank correlation) (see Section 5.4), classical stationary criteria were used at the significance level $\alpha = 0.05$ (see Tables 5 and 6). As a result of testing these hypotheses, it was established that the existing correlation is significant in the pair of samples (X, Y_1), and this fact was confirmed by two statistical criteria at once. This means that a factor such as a post made by famous people on social networks significantly affects the formation of the cryptocurrency rate at the time when the post was made. It should also be noted that for sample pairs (X, Y_2) and (X, Y_3), the correlation is insignificant.

The advantage of the performed research is the relative simplicity in the formation of the input dataset and the correctness of the proposed methods of statistical analysis. In addition, this approach can be used when checking the level of influence of any other factor on the formation of the rate of the selected cryptocurrency.

Disadvantages include the need to prepare a high-quality dataset since an insufficient number of forecasts, the presence of anomalous data (atypical forecasts), or inaccurate forecasts in the input dataset can significantly affect the accuracy of the results. It should also be noted that the accuracy of the forecast can be negatively affected by an incorrectly selected time interval for which the forecast is made since it is not known in advance how long the expert's post will affect the cryptocurrency rate. This indicates the need for the constant monitoring of both cryptocurrency rates and experts' posts on social networks.

An alternative way to determine the level of influence of one or another factor on the cryptocurrency exchange rate is the method of principal components. But to use this method, it is necessary to have a clear list of all factors that in one way or another affect the cryptocurrency rate. That is, the process of forming the input dataset would be much more complicated. In addition, this method would be more difficult from the point of view of software implementation.

7. CONCLUSIONS

1. As part of this research, it has been established that posts by famous people are an influential factor in the formation of the cryptocurrency exchange rate.
2. For this study, a model was created that best suited the goals and objectives of the research. The authors did not consider such well-known models as Granger Causality and LSTM, as they contradict the main goal of the study, which is to identify the impact of celebrity posts on cryptocurrency rates.
3. Since the current research investigates the influence of only one factor, namely, the posts of celebrities on social media on the cryptocurrency rate, there is no need to use multivariate models that would complicate the forecasting process.
4. The obtained results have been used to increase the accuracy of forecasting cryptocurrency rates.
5. The reliability of the obtained results is guaranteed by the quality of the input dataset and the correctness of the statistical analysis methods used.
6. In the course of the research, software has been developed that, in combination with other software developed by the authors of the study, can form an information system for forecasting cryptocurrency rates and studying the impact of various factors on their formation.
7. In further research, the authors plan to study the impact of other factors on cryptocurrency rates and investigate the impact of posts by several experts on a particular cryptocurrency, as well as the impact of posts by the same expert (group of experts) on the rates of different cryptocurrencies. In addition, it is planned to consider a wider range of cryptocurrencies and influencers to increase the generalizability and relevance of the study, conduct sentiment analysis to determine the tone and sentiment of social media posts and their correlation with market movements and conduct comparative case studies involving different social media platforms or financial markets to provide a broader context for the findings.

ACKNOWLEDGEMENTS

Funding: This research was funded by the Faculty of Electrical and Computer Engineering, Cracow University of Technology, and the Ministry of Science and Higher Education, Republic of Poland (grant no. E-1/2024).

REFERENCES

[1] P. Preethi, V. Uma, and A. Kumar, "Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction," *Procedia Comput. Sci.*, vol. 48, pp. 84–89, 2015, doi: 10.1016/j.procs.2015.04.154.

[2] M.S. Gerber, "Predicting crime using Twitter and kernel density estimation," *Decis. Support Syst.*, vol. 61, pp. 115–125, 2014, doi: 10.1016/j.dss.2014.02.003.

[3] R. Alhajj and J. Rokne, *Encyclopedia of Social Network Analysis and Mining*, Springer, 2020, p. 2200, doi: 10.1007/978-1-4614-7163-9.

[4] Y. Dang *et al.*, "An integrated framework for analyzing multilingual content in Web 2.0 social media," *Decis. Support Syst.*, vol. 61, pp. 126–135, 2014, doi: 10.1016/j.dss.2014.02.004.

[5] K. Byrd, A. Mansurov, and O. Baysal, "Mining Twitter data for influenza detection and surveillance," *SEHS '16: Proceedings of the International Workshop on Software Engineering in Healthcare Systems*, 2016, pp. 43–49, doi: 10.1145/2897683.2897693.

[6] G. ChengHe, "Dynamic topic modeling of twitter data during the COVID-19 pandemic," *PLoS ONE,* vol. 17, no. 5, p. 22, 2022, doi: 10.7910/DVN/YXIAEK.

[7] M. Chau and J. Xu, "Mining communities and their relationships in blogs: A study of online hate groups," *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 1, pp. 57–70, 2007.

[8] D. Karahoca, A. Karahoca, and Ö. Yavuz, "An early warning system approach for the identification of currency crises with data mining techniques," *Neural Comput. Appl.*, vol. 23, pp. 2471–2479, 2013, doi: 10.1007/s00521-012-1206-9.

[9] M. Žukovič, "Dynamics of episodic transient correlations in currency exchange rate returns and their predictability," *Open Phys.*, vol. 10, pp. 615–624, 2012, doi: 10.2478/s11534-011-0120-6.

[10] O. Gavrilenko, K. Novakivska, and O. Shumeiko, "Selection of the most influential economic factors for forecasting the US dollar exchange rate," *NTU Bull.*, vol. 54, pp. 26–35, 2022, doi: 10.33744/2308-6645-2022-4-54-026-035.

[11] O. Gavrylenko, M. Miahkyi, and Y. Zhurakovskyi, "The task of analyzing publications to build a forecast for changes in cryptocurrency rates," *Adapt. Syst. Autom. Control,* vol. 2 no. 41, pp. 90–99, 2022, doi: 10.20535/1560-8956.41.2022.271349.

[12] ARIMA Model – Complete Guide to Time Series Forecasting in Python – Resources [Online] Available: https://www.machinelearningplus.com/%20time-series/arima-model-time-series-forecasting-python/. (Accessed 05.04.2024).

[13] R.J. Hyndman, A.B. Koehler, J.K. Ord, and R. D. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, 2008, p. 362, doi: 10.1007/978-3-540-71918-2.

[14] Binance cryptocurrency exchange – Resources [Online] Available: https://www.binance.com/ (Accessed 05.04.2024).

[15] M. Kartashov, *Probability, processes, statistics*. Kyiv University, 2007, p. 504.

[16] O. Valenzuela, F. Rojas, L.J. Herrera, H. Pomares, and I. Rojas, *Theory and Applications of Time Series Analysis and Forecasting*. Springer, 2023, p. 333, doi: 10.1007/978-3-031-14197-3.

[17] A.D. Helfrick and W.D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*. Prentice Hall of India, 2008, p. 460.

[18] H. Abdi, V. Guillemot, A. Eslami, and D. Beaton, *Canonical Correlation Analysis*. Springer, 2017, pp. 1–16, doi: 10.1007/978-1-4614-7163-9_110191-1.

[19] Rank correlation – Resources [Online] Available: https://moodle.znu.edu.ua/ (Accessed 05.04.2024).

[20] W.J. Ewens and K. Brumberg, *Introductory Statistics for Data Analysis*. Springer, 2023, p. 273, doi: 10.1007/978-3-031-28189-1.

[21] B. Blaine, *Introductory Applied Statistics*. Springer, 2023, p. 190, doi: 10.1007/978-3-031-27741-2.

[22] O. Gavrilenko *et al.*, "The principle for forming a portfolio of public services based on the analysis of statistical information," *East.-Eur. J. Enterprise Technol.*, vol. 3, no. 3(117), pp. 57–64, 2022, doi: 10.15587/1729-4061.2022.260136.

[23] J. Ascorbebeitia, E. Ferreira, and S. Orbe, *Testing conditional multivariate rank correlations: the effect of institutional quality on factors influencing competitiveness*. Springer, 2022, pp. 931–949, doi: 10.1007/s11749-022-00806-1.

[24] V. Chymshyr, S. Telenyk, Î. Rolik, and E. Zharikov, "The platform for supporting the life cycle of services in the information systems of information and communication service providers," *Adapt. Syst. Autom. Control*, vol. 1 no. 42, pp. 205–226, 2023, doi: 10.20535/1560-8956.42.2023.279172.

[25] P. Kapler, "The application of the fuzzy rule-based Bayesian algorithm to determine which residential appliances can be considered for the demand response program," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 71, no. 4, p. e146106, 2023, doi: 10.24425/bpasts.2023.146106.

[26] O. Gavrilenko and M. Myagkyi, "Forecasting the cryptocurrency exchange rate based on the ranking of expert opinions," *Innov. Technol. Sci. Sol. Ind.*, no. 1, no. 27, pp. 18–25, 2024, doi: 10.30837/ITSSI.2024.27.018.

[27] G. Nowakowski, S. Telenyk, and Y. Vovk, "Chatbots Lifecycle Support Platform," *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Germany, 2023, pp. 308–319, doi: 10.1109/IDAACS58523.2023.10348794.