JOIV : Int. J. Inform. Visualization, 8(3-2): IT for Global Goals: Building a Sustainable Tomorrow - November 2024 1987-2001

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Prediction of ROI Achievements and Potential Maximum Profit on Spot Bitcoin Rupiah Trading Using K-means Clustering and Patterned Dataset Model

Rizky Parlika [a,b,*], R. Rizal Isnanto [a], Basuki Rahmat [b]

[a] *Doctoral Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia*
[b] *Department of Informatics, Faculty of Computer Science, University of Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia*
Corresponding author: [*]rizkyparlika@students.undip.ac.id

Abstract*—Since Satoshi Nakamoto first proposed the idea of bitcoin in 2009, the cryptocurrency and prediction methods for it have grown and changed exceptionally quickly. The Patterned Dataset Model was a valuable tool in earlier studies to explain how changes in the price of Bitcoin affect the movements of other cryptocurrencies in a digital trading market. Three different kinds of datasets are generated by this model: patterned datasets under full conditions, patterned datasets under dropping prices (Crash), and patterned datasets under rising prices (Moon). The K-means approach was then used to cluster these three datasets. Specifically, each dataset was split into two clusters, and the clustering score was determined by utilizing eight unique clustering metrics. Consequently, the best clustering score was found in the patterned dataset in the crash situation. Additionally, from 2022 to 2024, the raw data from this crash-condition-patterned dataset is used to determine the possibility of reaching maximum profit and return on investment (ROI) daily and monthly. According to the calculation results, the range computed over the course of a whole month (30 to 31 days) is significantly larger than the daily range (24 hours multiplied by one month), which represents the most significant profit and ROI attained before the emergence of the first diamond crash level. This research also covers the application of a deep learning model to forecast patterned datasets for crash scenarios that may occur many days in advance. The ConvLSTM2D Model performs better in predicting pattern dataset values for the subsequent crash scenario, according to the hyperparameter comparison between the Gated Recurrent Unit (GRU) Model and the 2D Convolutional Long Short-Term Memory Model.

***Keywords*— Patterned datasets; Bitcoin; Altcoin; cryptocurrency, k-means cluster, diamond crash, return of investment.**

Manuscript received 20 Aug. 2024; revised 9 Sep. 2024; accepted 28 Oct. 2024. Date of publication 30 Nov. 2024.

I. INTRODUCTION

With the expansion of cryptocurrency commodities trading, predictions about the movement of prices in cryptocurrencies are becoming more and more different. While it may seem quite challenging to turn a profit when trading in digital markets consistently, there are always chances because market activity is sometimes random, and patterns will always appear [1]. Time series forecasting is a fundamental technique in data science, and it is the most popular analytical tool used by businesses and organizations to plan strategies for the most likely future and make appropriate plans [2].

Since Satoshi Nakamoto [3] unveiled Bitcoin in 2009, there has been a growing interest in modeling and measuring the fundamental features of cryptocurrency marketplaces. Essential elements, including investor emotion and microstructure, as well as the impact of significant events and fundamental values, are among the items that drive fluctuations [4]. The trading behavior of retail traders willing to engage in speculative activity drives and predicts bitcoin returns, in addition to investor sentiment [5]. Bitcoin prices are influenced by sentiment, while alternative coin values are influenced by Bitcoin pricing.

According to research findings, price surges, liquidity, and FOMC pronouncements seem to impact the search for intraday return predictability patterns in the cryptocurrency market, particularly concerning Bitcoin [6]. As a digital asset that is exchanged on exchanges both domestically and internationally, bitcoin is incredibly liquid. Examining factors influencing liquidity indicates that the liquidity of Bitcoin is distinct from that of other asset classes, such as foreign exchange and stocks. A lot of factors unique to cryptocurrencies and blockchains affect the liquidity of the Bitcoin market [7].

A patterned dataset model is presented to predict future prices and market circumstances as a step to anticipate losses due to changes in the cryptocurrency digital trading market caused by movements in the price of bitcoin. Whenever there is a story about cryptocurrencies in the media on a global scale, opinions are formed that impact fluctuations in the price of bitcoin. Numerous cryptocurrencies will move differently in a single digital cryptocurrency trading market when the price of Bitcoin fluctuates. The price position of Bitcoin and Alternative Coins (Cryptocurrencies other than Bitcoin) in general, which are at daily lows, declining, increasing, and at daily highs, can always be described by the outcomes of descriptive analysis and visualization of structured datasets [8], [1], [9].

Before executing a trading transaction, more knowledge will be gathered about how patterned datasets are used to read the market and the overall positions of bitcoin prices. Due to the trials conducted for this study, the choice to begin purchasing bitcoin assets on the spot market can be supported when the diamond crash level arises, mainly if it happens multiple times on multiple distinct days. The purpose of utilizing patterned datasets is to achieve [8], [1], [9]. Institutional investors are more knowledgeable and self-assured than individual investors, according to research on the effect of investor attention on returns and volatility [10]. Therefore, to reduce the risk of loss and maximize the opportunity for profit, this research offers fresh perspectives on how to interpret the price position of Bitcoin and altcoins (or other cryptocurrencies) that Bitcoin impacts.

According to research on the association between scenarios and the relationship between significant shocks and Bitcoin values, "what-if" analysis is useful for estimating the risk associated with Bitcoin [11]. The various methods, techniques, and models that attempt to predict the future move in the price of the bitcoin commodity are competing with one another to provide the most precise and intuitive method. It is still uncommon to find techniques for developing new datasets that can read positions in the cryptocurrency trading market based on fluctuations in the price of bitcoin and determine when it is best to buy and maybe profit from selling.

Without first patterning the data, a number of prediction models, either alone or in combination, still frequently use raw cryptocurrency trading history data to predict the price of bitcoin and other cryptocurrencies. The raw data, which consists of the history of cryptocurrency trading transactions, will first be patterned before being put in the database, which is where the Patterned Dataset Model differs from other approaches. After that, this will be continuously mined in real time and combined into a dataset of time-sequential patterns. After undergoing descriptive and visual analysis, patterned datasets can forecast future cryptocurrency price movements, particularly within the same virtual trading market, even in the absence of additional prediction models. This is so that it can specifically display the current state of cryptocurrency values at any one time, thanks to the patterned dataset model. Concerning bitcoin pricing, the Patterned Dataset Model can provide a precise picture at any moment. Using equations (1) through (7), the Patterned Dataset Model first processes transaction data elements, such as the most recent price, daily highest price, daily lowest price, and transaction time. This creates a new dataset characterized by price increase conditions (referred to as the moon) or falling prices (referred to as a crash) [8], [1], [9].

As with other studies that employ machine learning [12] or deep learning [13] on raw transaction data to forecast and estimate cryptocurrency prices, the patterned dataset generated by the patterned dataset model may also be used as test data. Therefore, a hybrid prediction model will be created if statistical learning, machine learning, deep learning, technical analysis, sentiment analysis, or other techniques are used to apply prediction models to a patterned dataset.

Three primary types of datasets are generated by the Patterned Dataset model: Patterned Datasets in complete conditions, which include Patterned Datasets in Price Increase conditions (also known as Patterned Datasets in Moon conditions), and Patterned Datasets in Price Decrease conditions (also known as Patterned Datasets in **Crash** conditions) [8], [1]. Applying the K-means cluster method yields a candidate dataset with optimal clustering [14]. Subsequently, the ROI Potential Achievement value is computed using this chosen patterned dataset candidate as a historical dataset. Next, two deep learning models, the Gated Recurrent Unit (GRU) Model and the Long Short Term Memory Convolution 2d (LSTM-Conv2d), frequently used to forecast bitcoin and cryptocurrency trading, were chosen. Then, using nearly identical hyperparameter values, both are compared, and the metrics MAE [15], MSE [16], RMSE [15], and MAPE [15] are used to measure them. The outcome was a deep learning model that performed better in the future when anticipating patterned datasets in crash conditions.

II. Materials and Methods

A. Materials

1) Patterned Dataset Model:

The Patterned Dataset Model is a revolutionary technique for assessing and predicting the volatility of bitcoin prices. This model uses several processes to predict cryptocurrency values based on the particular effect that changes in the price of bitcoin have on other currencies in a market for trading digital currencies. The Indonesian Indodax Cryptocurrency Market provided the primary data, or raw data, initially utilized to build this model [17],[18]. This notion can be extended to all reputable digital cryptocurrency markets listed on Coinmarketcap through connection ties in the form of publicly accessible pages and API codes [19]. Users can also access sources of real-time trading transaction data with it. The basic formulas of the Patterned Dataset Model are as follows [9], [20], [8]:

$$\mathrm{R}=\mathrm{H}-\mathrm{L} \tag{1}$$

$$\mathrm{TR}=\mathrm{H}-\mathrm{C} \tag{2}$$

$$\mathrm{LR}=\mathrm{C}-\mathrm{L} \tag{3}$$

$$\mathrm{PTR} = \frac{\mathrm{TR}}{\mathrm{R}} \times 100\% \tag{4}$$

$$\mathrm{PLR} = \frac{\mathrm{LR}}{\mathrm{R}} \times 100\% \tag{5}$$

Using the following variables: H is the highest price, L is the lowest price, and C is the current price; TR stands for the top range, LR for the lower range, PTR for percent of the top range, and PLR for percent of the lower range. Eqs. 1 through 5 are used to pattern trade transaction data on each cryptocurrency asset at one time when the data is being stored in the MySQL database on the hosting service used [9], [20], [8]

In a cryptocurrency trading market, most cryptocurrency assets are either in a crash state or are seeing a price decrease (PTR value = 100 and PLR value = 0) [9], [20], [8]:

$$\text{Total\_Crash\_Condt} = \text{Count}\left(\sum_{i=1}^{n} \text{Crash\_Condt}(i)\right) \quad (6)$$

PLR value = 100 and PTR value = 0 indicate that most bitcoin assets in a cryptocurrency trading market are either in a moon state or are rising in value [9], [20], [8]:

$$\text{Total\_Moon\_Condt} = \text{Count}\left(\sum_{i=1}^{n} \text{Moon\_Condt}(i)\right) \quad (7)$$

Following that, two categories were created from the patterned dataset collected between May 2022 and December 2022: the moon and crash conditions.

2) Mean Absolute Error:

One way to calculate the average absolute error is to use the mean absolute error measure. This metric is a risk measure that matches the absolute error loss's predicted value [15].

$$MAE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i| \quad (8)$$

The mean absolute error (MAE) measures how much the actual and anticipated values differ. This provides a sense of the degree to which the positive or negative error is aligned with the real number.

3) Mean Squared Error

An assessment metric called mean squared error calculates the average of the squared discrepancies between the actual value and the expected value. larger numbers denote larger mistake rates. It is the average of the deviations between expected and actual values [16].

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2 \quad (9)$$

The mean square error (MSE) measures how much the actual and anticipated values differ. This provides a sense of the deviation between the expected and actual values. One drawback of MSE is that its unit of measurement, which is the square of the original target variable's unit of measurement, might be challenging to understand intuitively. Because of this, the square root of MSE (RMSE) is sometimes employed as a substitute evaluation metric.

4) Root Mean Squared Error:

Because the result is the square root of the mean squared error, or MSE, RMSE offers an error measurement proportionate to the variable's target unit [15].

$$RMSE = \sqrt{MSE} \quad (10)$$

The RMSE calculates the degree to which the model's predictions and data values agree.

5) Mean Absolute Percentage Error (MAPE:)

A metric for evaluation used in regression problems is mean absolute percentage deviation (MAPD), sometimes called mean absolute percentage error (MAPE). The target variable's global scaling does not affect this metric, which is sensitive to relative error [15].

$$MAPE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{|y_i - \hat{y}_i|}{max(\epsilon, |y_i|)} \quad (11)$$

The relative error of the prediction, or the prediction error expressed as a percentage of the correct value, is measured by MAPE. When assessing a forecasting model's quality, MAPE is frequently utilized, and the outcomes are frequently given as a percentage.

6) Mutual Information Score:

Information shared by two clusters. Mutual information is a metric for comparing two labels' similarity on the same data set. The Mutual Information between clustering **U** and **V** is stated as follows, where **|U_i|** is the number of samples in cluster **U_i** and **|V_j|** is the number of samples in cluster **V_j**.

$$MI(U, V) = \sum_{i=1}^{|U|} \sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} \log \frac{N|U_i \cap V_j|}{|U_i||V_j|} \quad (12)$$

A permutation of the class or cluster label values does not affect the score value, indicating that this metric is independent of the absolute values of the labels. Furthermore, this metric is symmetric, meaning that substituting V (i.e., label_pred) for U (i.e., label_true) will yield the same score. When the true ground truth is unknown, this can be helpful in assessing the agreement between two independent label assignment algorithms on the same dataset [21].

7) Adjusted Mutual Information Score:

The Mutual Information (MI) score is adjusted to account for chance, and this is known as Adjusted Mutual Information (AMI). Regardless of whether there is indeed more information shared, it explains why the MI is typically higher for two clusterings with a larger number of clusters. When two partitions are identical or perfectly matched, the AMI returns a result of 1. When independent labels, or random partitions, have an AMI average of around 0, then both can be negative. The AMI for two clustering, **U** and **V**, is provided as follows [21]:

$$\text{AMI}(U, V) = \frac{[\text{MI}(U,V) - \text{E}(\text{MI}(U,V))]}{[\text{avg}(\text{H}(U),\text{H}(V)) - \text{E}(\text{MI}(U,V))]} \quad (13)$$

vms is an abbreviation of v measure score [22].

8) Normalized Mutual Information Score:

Mutual Information that is Normalized between two clusters. To scale the results between 0 (no mutual information) and 1 (perfect correlation), the Mutual Information (MI) score is normalized to create Normalized Mutual Information (NMI). The average_method defines a generalized mean of **H(labels_true)** and **H(labels_pred)** that is used to standardize mutual information in this function. Chance adjustment is not applied to this measure. Consequently, adjusted_mutual_info_score might be the better choice. A permutation of the class or cluster label values has no effect on the score value, indicating that this metric is independent of the absolute values of the labels. Additionally, this metric is symmetric, meaning that substituting **label_true** with **label_pred** will yield the same score. Measuring the agreement between two separate label assignment algorithms on the same dataset can be helpful [21].

9) Rand index Score

By considering all sample pairings and counting pairs assigned in the same or different clusters in the anticipated and true clustering, the Rand Index calculates a similarity measure between two clusterings [23]. The unaltered RI score is:

$$\mathrm{RI} = \frac{(\text{number of agreeing pairs})}{(\text{number of pairs})} \qquad (14)$$

10) Adjusted Rand Index Score:

The Rand Index calculates a similarity measure between two clustering by considering all sample pairings and counting pairs assigned in the same or different clusters in the anticipated and accurate clustering [23]. Thus, with random labeling, the modified Rand index is guaranteed to be near 0.0 regardless of the number of clusters and samples. When the clustering is identical (up to a permutation), it is precisely 1.0.

$$\mathrm{ARI} = \frac{(\mathrm{RI} - \text{Expected\_RI})}{(\max(\mathrm{RI}) - \text{Expected\_RI})} \qquad (15)$$

11) Fowlkes Mallows Score:

Calculate how similar two groupings of points are to each other. The definition of the Fowlkes-Mallows index (FMI) is the geometric mean of the precision and recall:

$$\mathrm{FMI} = \frac{\mathrm{TP}}{\mathrm{sqrt}((\mathrm{TP} + \mathrm{FP}) * (\mathrm{TP} + \mathrm{FN}))} \qquad (16)$$

Where FP stands for False Positive (i.e., the number of pairs of points that belong in the same clusters in labels_true and not in labels_pred), FN for False Negative (i.e., the number of pairs of points that belong in the same clusters in labels_pred and not in labels_True), and TP for True Positive (i.e., the number of pairs of points that belong in the same clusters in both labels_true and labels_pred). The possible scores are 0 to 1. When two clusters are highly similar, they are said to be well-matched [24].

12) Homogeneity Score

Measure of homogeneity for a cluster labeling concerning a ground truth. When all of a clustering result's clusters contain only data points from one class, the result fulfills homogeneity. A permutation of the class or cluster label values does not affect the score value, indicating that this metric is independent of the absolute values of the labels. This measure is not symmetric; if label true is switched to label pred, the completeness score will generally change [22].

13) V-measure Score

Cluster labeling using V-measure with a ground truth. When the 'arithmetic' option for averaging is selected, this value is the same as the Normalized Mutual Info value. The harmonic mean between completeness and homogeneity is the V-measure [22].

$$\mathrm{V} = \frac{(1 + beta) * homogeneity * completeness}{(\text{beta} * \text{homogeneity} + \text{completeness})} \qquad (17)$$

14) K-means Cluster:

Data can be grouped according to variables or attributes using the K-means clustering method, a cluster analysis approach. Finding the number of clusters, choosing a random point to serve as the initial cluster center, figuring out how far off each data point is from the cluster center, putting the data into the closest cluster, and then recalculating the cluster center are the stages involved. Until the cluster center remains unchanged, the procedure is repeated. Maximizing differences between clusters and similarities within clusters is the aim. This approach uses the cluster center's distance from the data to group the data [14]. The K-means Cluster algorithm is as follows.

K-Means (data, K):

1. As cluster centers, randomly establish the starting positions.
2. Carry going till convergence:
   a. Determine the distance to every cluster center for every data point:
      1) *for each data_point in data:*
      2) *for each cluster_center in cluster_centers:*
      3) *calculate the distance between data_point and cluster_center*
      4) *store the distance into an array*
   b. Find the nearest cluster for every data point:*for each data_point in data:*
      1) *find the cluster index with the smallest distance*
      2) *assign data_point to appropriate cluster*
   c. Take the average of the data within the cluster to recalculate the cluster center.
      1) *for each cluster_center in cluster_centers:*
      2) *calculate the average of all data included in the cluster*
      3) *set a new value for cluster_center*
   d. End the iteration if there is no change in the cluster center.
3. Give back the resulting cluster set.

Depending on the requirements of the data analysis, the formula for determining the distance between the data (data_point) and the cluster center (cluster_center) can employ Euclidean distance or alternative distance metrics. The Eucledian distance formula can be used to implement the distance computation process, which is not explicitly specified in this pseudocode, as follows:

$$d = \sqrt{[\,(x_1 - x_2)^2 + (y_1 - y_2)^2]} \qquad (18)$$

with, $(\boldsymbol{x_1}, \boldsymbol{y_1})$ are the coordinates of the first point, (x_2, y_2) is the coordinate of the second point, and d is the distance between the points (x_1, y_1) with (x_2, y_2).

15) Model Deep Learning in Predicting Cryptocurrency Price

Further refinement of the research findings can also involve investigating methods to improve accuracy in reducing the danger of loss and boosting earnings. Additionally, some researchers integrate deep learning [13] and machine learning [12] to construct combination models of many technical indicators that improve profits and decrease the risk of losses. One sort of deep learning model that is frequently used to estimate changes in the price of bitcoin

[25], cryptocurrency [26],[27],[15], and investor sentiment [28] is the short-term memory model. This model uses economic data, such as movement data on the price of gold, as features to estimate the value of Bitcoin as a risk management tool [25]. This method can overcome overfitting in a model's output and yield estimates that are better and more accurate [30]. Another example is the use of a combination model to predict cryptocurrency prices by normalizing the root mean square error (RSME) value to boost accuracy [29]. This model combines a two-way LSTM model (Bi-LSTM) with a gate recurrent unit (GRU) [30]. Long Short-Term Memory (LSTM), a deep learning model, has been extensively utilized for cryptocurrency (Bitcoin and Altcoin) prediction.

B. Methods

1) First Section of The Research Design: This study consists of two sections. An explanation of the top clustering outcomes from the patterned dataset is provided in Segment 1. The dataset that results from the Patterned Dataset Model's implementation is called the Patterned Dataset. Based on the response to Bitcoin price changes, the Patterned Dataset Model can read the price position of most cryptocurrencies at any given time, regardless of whether there is a general price increase or decline. Three patterned datasets are generated by this approach and classified into three conditions: complete, price increase (Moon), and price decline (Crash). There are two sections to this study. The pattern dataset is clustered in three conditions: complete, price increase (Moon), and price reduction (Crash). The Patterned Dataset test data in the price reduction condition (Crash) is used in the second section to illustrate how the two deep learning techniques, LSTM Conv2d and Gated Recurrent Unit (GRU), compare their hyperparameters. The test dataset was preferred in the price reduction condition since, according to the K-means clustering results, it produced the best clustering results when eight unique clustering metrics were calculated. Figure 1's flowchart displays the first research design.

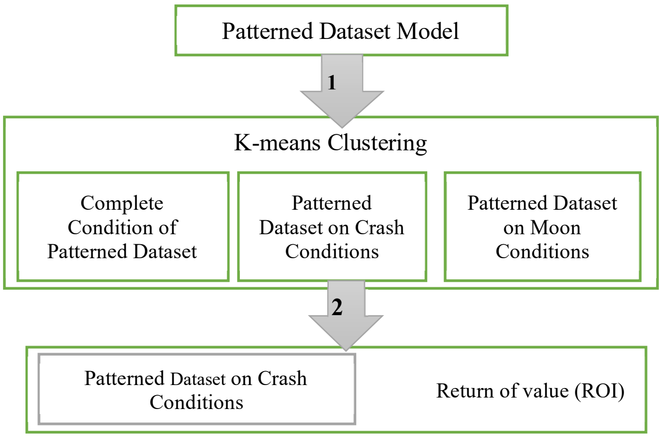


Fig. 1 The flow of research design 1

K-means cluster is used for this process. Next, eight unique clustering metrics adjusted mutual info score [21], adjusted rand score [23], Fowlkes mallows score [24], homogeneity score [22], mutual info score [21], normalized mutual info score [21], rand score [23], and v measure score [22] are used to visualize and calculate the clustering results for each patterned dataset that has been clustered using K-means. Consequently, every metric employed demonstrates that the patterned dataset in crash conditions yields the optimal grouping outcomes.

2) Second section of the research design: In the second section, two deep learning models that are frequently used to forecast the price of Bitcoin and other currencies are compared based on the outcomes of their hyperparameter tests. The Gated Recurrent Unit (GRU) Model and the LSTM Conv2d Model are the two models. When evaluated with a patterned dataset under crash conditions, the LSTM-Conv2d model yields lower error levels as a result as shown in Fig. 2.

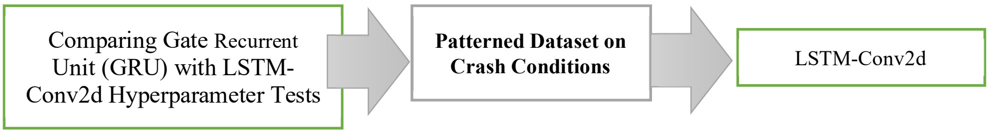


Fig. 2 The flow of research design 2

3) Dataset utilized for study: A comprehensive example of a patterned dataset is displayed in Figure 3, with price increases (shown in green) and price decreases (shown in red). A patterned dataset visualization, spanning from May 2022 to the end of December 2022, is presented as an example dataset.

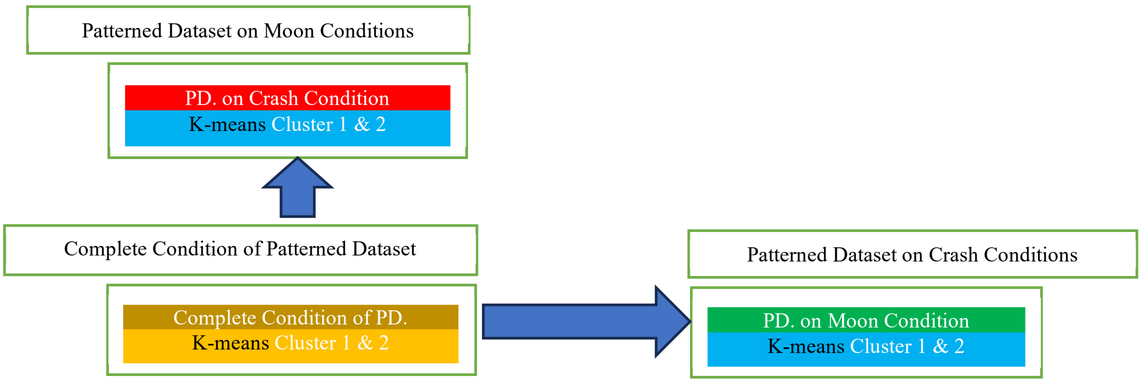


Fig. 3 K-means clustering on datasets with moon patterns, crashes, and completeness

III. Results and Discussion

A. K-means Evaluating Metrics

Crash and moon conditions are always present in the stored data. In general, the price of BTC IDR and other IDR cryptocurrencies tends to decrease when the total signal value rises during a crash; conversely, the price of BTC IDR and other IDR altcoins tends to rise when the total signal value increases during a moon condition. Next, one of the three patterned datasets is selected as the most suitable for further testing aimed at determining the highest potential for profit and daily as well as monthly ROI. The three datasets were analyzed using the K-means clustering method to determine which had the best grouping, as shown in Figure 5. Figure 6 shows the K-means clustering procedure applied to the Crash patterned datasets. At the top, information markers are arranged from left to right as follows: BTC IDR Price, Patterned Dataset, Use of the Min-Max Scaler function to combine the BTC IDR dataset with the Patterned Dataset, Results of clustering using K-means for two clusters: detailed graphics showing the outcomes of applying K-means to the first cluster and the second cluster.

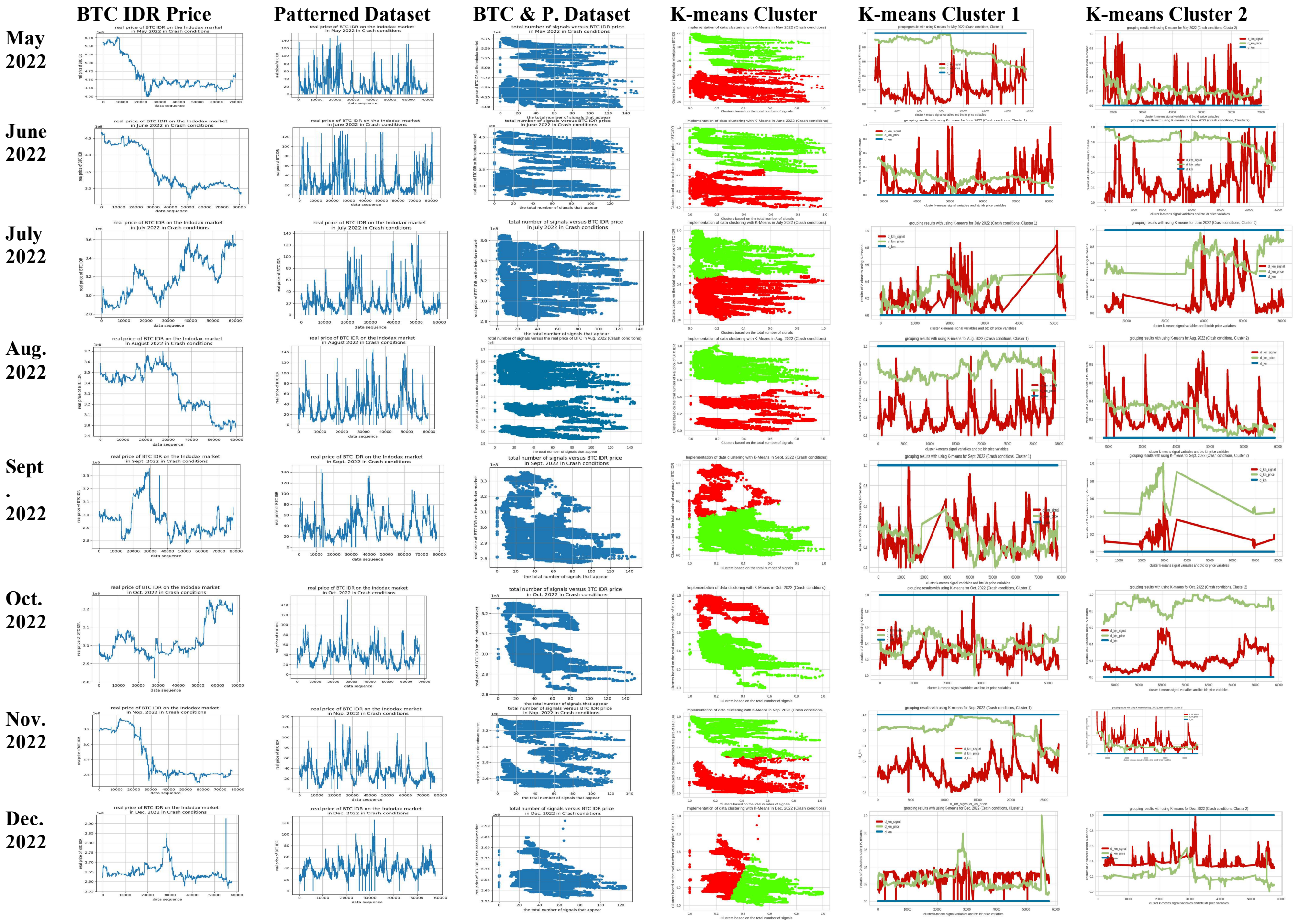


Fig. 4 K-means grouping of fundamentally structured datasets under crash scenarios

TABLE I

Clustering Metrics Result from K-means Calculations on Patterned Datasets on All Conditions in 2022

Month	Total data	Clustering metrics								
		amis	ars	cs	fms	hs	mis	nmis	rs	vms
05	124,210	0.1569	0.0022	0.5933	0.0133	0.2649	2.3692	0.3663	0.9742	0.3663
06	139,986	0.1754	0.0034	0.6167	0.0198	0.2725	2.4648	0.3780	0.9729	0.3780
07	107,165	0.1594	0.0047	0.6135	0.0208	0.2820	2.5127	0.3864	0.9779	0.3864
08	104,502	0.1771	0.0057	0.5669	0.0223	0.2773	2.3512	0.3724	0.9795	0.3724
09	121,712	0.2017	0.0085	0.5862	0.0286	0.2943	2.5096	0.3918	0.9826	0.3918
10	112,138	0.1967	0.0051	0.5391	0.0185	0.2719	2.2428	0.3615	0.9813	0.3615
11	120,039	0.2107	0.0071	0.5247	0.0212	0.2727	2.186	0.3589	0.9807	0.3589
12	101,573	0.1805	0.0060	0.4305	0.0190	0.2298	1.7198	0.2996	0.9763	0.2996

TABLE II

CLUSTERING METRICS RESULT FROM K-MEANS CALCULATIONS ON PATTERNED DATASETS ON MOON CONDITIONS IN 2022

Month	Total data	Clustering metrics								
		amis	ars	cs	fms	hs	mis	nmis	rs	vms
05	54,081	0.1741	0.0032	0.7366	0.0217	0.2946	2.6219	0.4209	0.9605	0.4209
06	58,797	0.1882	0.0040	0.7608	0.0285	0.2968	2.6736	0.4270	0.9583	0.4270
07	46,808	0.1742	0.0088	0.7647	0.0375	0.3411	3.0416	0.4717	0.9763	0.4717
08	44,957	0.1928	0.0087	0.7049	0.0377	0.3138	2.6627	0.4343	0.9717	0.4343
09	43,321	0.1795	0.0079	0.7116	0.0317	0.3328	2.8492	0.4535	0.9783	0.4535
10	44,593	0.1998	0.0059	0.6694	0.0237	0.3201	2.6426	0.4331	0.9764	0.4331
11	43,286	0.2092	0.0072	0.6551	0.0250	0.3151	2.5482	0.4255	0.9768	0.4255
12	43,837	0.2095	0.0076	0.5577	0.0266	0.2681	2.0210	0.3621	0.9685	0.3621

TABLE III

CLUSTERING METRICS RESULT FROM K-MEANS CALCULATIONS ON PATTERNED DATASETS ON CRASH CONDITIONS IN 2022

Month	Total data	Clustering metrics								
		amis	ars	cs	fms	hs	mis	nmis	rs	vms
05	70,129	0.2191	0.0058	0.7448	0.0264	0.3511	3.1270	0.4772	0.9801	0.4772
06	81,189	0.2414	0.0094	0.7638	0.0429	0.3576	3.2189	0.4872	0.9793	0.4872
07	60,357	0.2215	0.0100	0.7608	0.0394	0.3584	3.1691	0.4872	0.9790	0.4872
08	59,545	0.2466	0.0131	0.7125	0.0427	0.3633	3.0590	0.4812	0.9826	0.4812
09	78391	0.2766	0.0191	0.7112	0.0542	0.3667	3.0907	0.4839	0.9839	0.4839
10	67,545	0.2741	0.0122	0.6808	0.0360	0.3528	2.8832	0.4648	0.9830	0.4648
11	76,753	0.3303	0.0159	0.6516	0.0384	0.3497	2.7614	0.4551	0.9819	0.4551
12	57,736	0.2434	0.0131	0.5550	0.0338	0.2987	2.2094	0.3884	0.9761	0.3884

Each patterned dataset in this study is split into just two clusters, and each dataset's clustering is computed using eight distinct metrics. These 8 metrics include: Mutual Information Score, Adjusted Mutual Information Score, Normalized Mutual Information Score, Rand index Score, Adjusted Rand Index Score, Fowlkes Mallows Score, Homogeneity Score, Completeness Score, and V-measure Score. The visualization results of each patterned dataset, both before and after K-means clustering, are shown below.

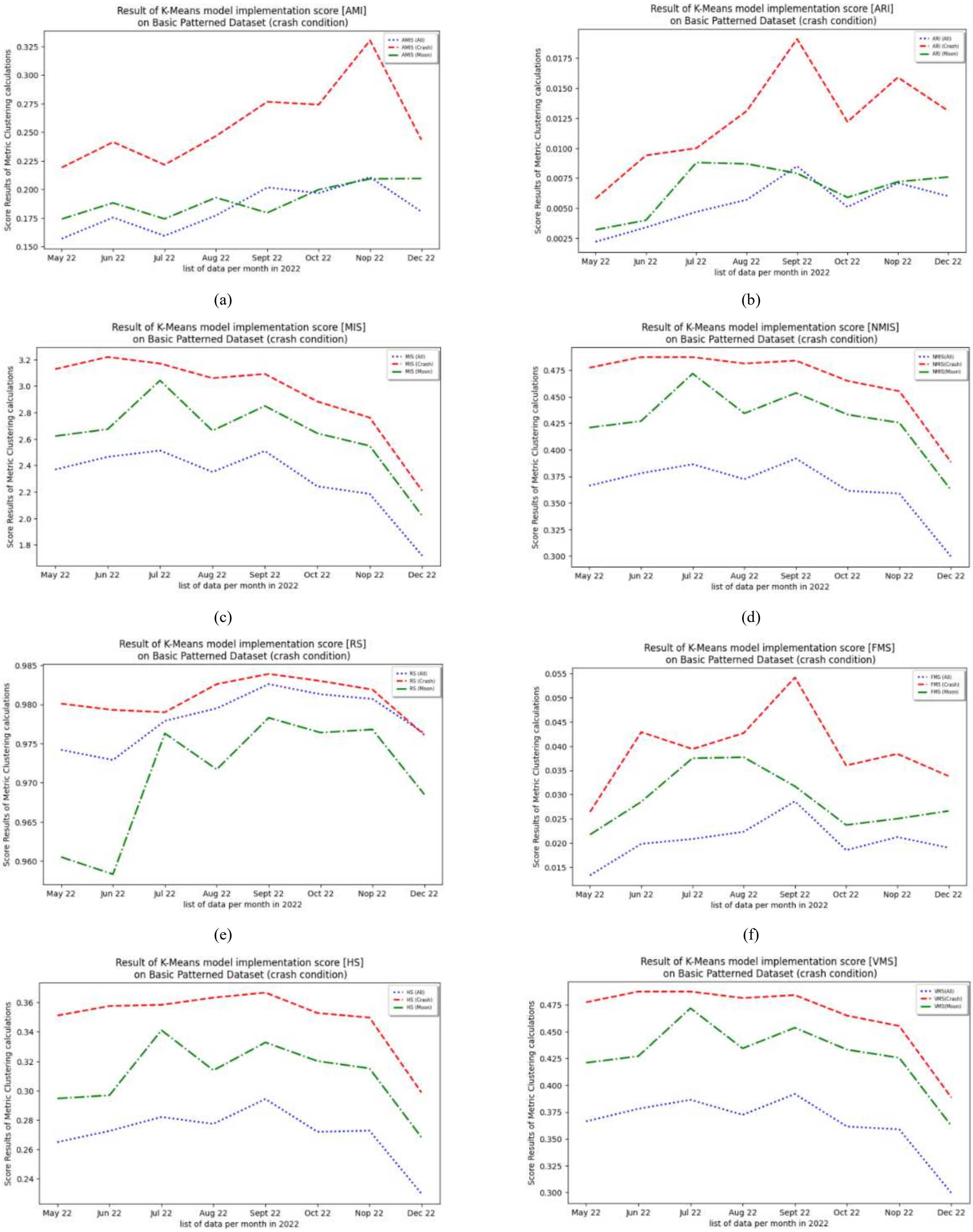
Table 1 displays the clustering metrics measured by K-means Cluster in Patterned Dataset Clustering in Complete Conditions (Price Decrease and Increase Conditions). Table 2 displays the outcomes of calculating clustering metrics for pattern dataset clustering under crash conditions (price decline) using K-means clustering. Table 3 displays the clustering parameters measured by Kmeans Cluster in Patterned Dataset Clustering under Moon Conditions (Price Increase).A visualization of the clustering metric measurement data was made and is shown in Figure 7 per metric to make Tables 1, 2, and 3 easier to read. The description of the variable In 2022, the month signifies the month code. The terms amis, ars, fms, hs, and nmis refer to different abbreviations for different types of scores: amis for adjusted mutual info score [21], ars for adjusted rand score [23], fms for Fowlkes mallows score [24], hs for homogeneity score [22], mis for mutual info score [21], nmis for normalized mutual info score [21], rs for rand score [23], and vms for v measure score [22]. In the clustered pattern dataset, every condition has a unique color given to it. Basic Patterned Datasets are colored blue when they are in the Complete Condition, green when they are in the Rising (Moon) Condition, and red when they are in the Decreasing Condition (Crash). The pattern dataset in a crash state receives the greatest score after computations using the eight clustering metrics are performed on the dataset from May to December 2022. As a consequence, it is determined that this dataset has the best clustering outcomes when utilizing the K-Means Cluster method.

B. *Possibility of Obtaining Daily and Monthly Goals of Profit and Maximum ROI*

Utilizing the Google Collaboration Editor and Python programming, a program was developed to show a historical time series of the prices of Bitcoin and Rupiah combined with a dataset based on crash scenarios. In particular, notes were taken on the day that the Diamond Crash level was made available for play for the first time that month. This study employed a dataset based on crash situations that occurred between May 2022 and April 2024.

When we get to the Diamond Crash level, it frequently appears multiple times in a single day. The first-ever diamond crash data on a single day was utilized as the data source for the whole days' worth of diamond crashes in the research's Return of Value (ROI) and Profit Calculation Procedure. Hence, even though multiple Diamond Crash levels may be released on the same day, the first one to do so was employed in this investigation. When prices are declining, Diamond Crash is the highest-level value in the Patterned Dataset (Crash). This DC level indicates that a price decrease to the daily lowest point in a 24-hour period is occurring simultaneously for most Alternative Coins (Cryptocurrency other than Bitcoin) on the same digital cryptocurrency trading platform.

In the future, there will be an opportunity to conduct more thorough research to decide the impact of using the most recent DC, average DC, highest value DC, lowest value DC, most frequently occurring DC, and other variations when multiple DC levels are released on the same day. Nevertheless, as it is not always possible to predict whether the next DC would occur after the first one appears in real-time, only the Diamond Crash data that was released initially was recorded and utilized as a reference for computations in this study.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5 Cluster Metric Measurement Visualization Findings: nmis, (d) amis, (e) rs, (f) fms, (g) hs, (h) vms, (b) ars, (c) mis

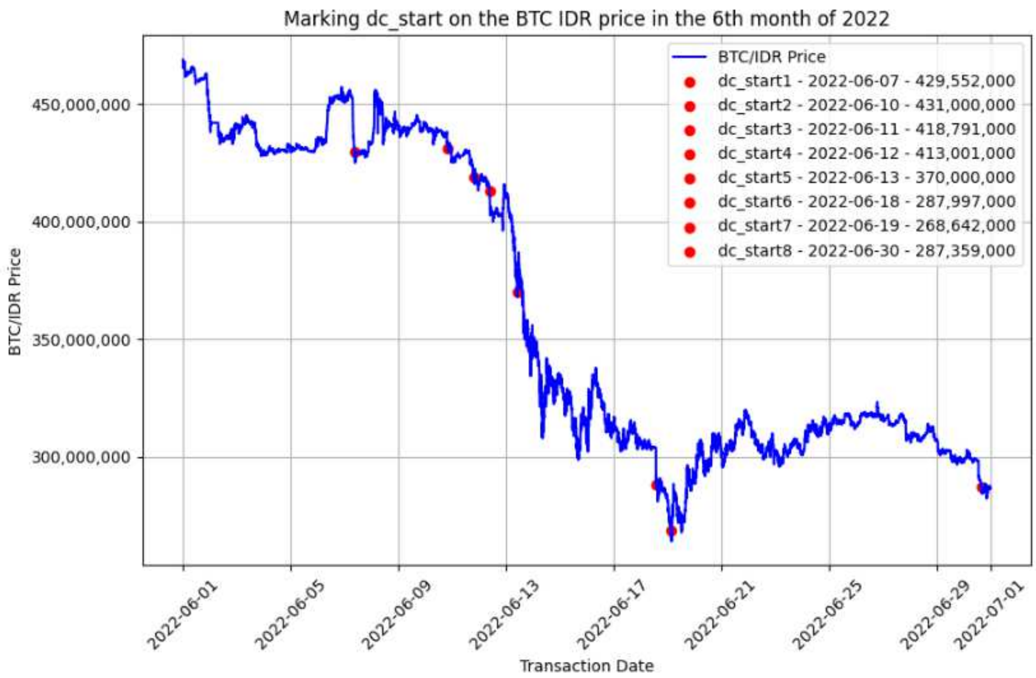


Fig. 6 Marking of the Diamond Crash Level on the BTC IDR chart in June 2022

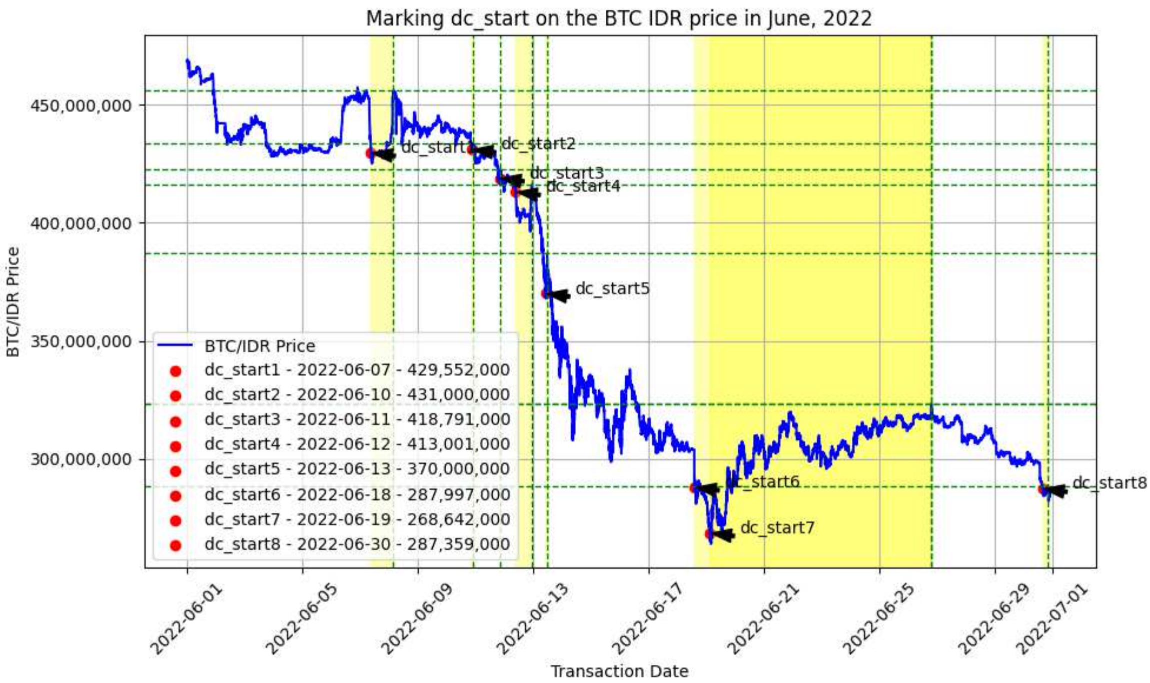


Fig. 7 Diamond Crash Level Marking and Maximum Price area in June 2022

The Diamond Crash level, for instance, is marked and displayed in Figure 8 on the BTC IDR price line in June 2022. It is evident that 8 days during the course of that one-month period encountered the Diamond Crash condition. The sample data displayed is data for June 2022. This month, the Diamond Crash level was recorded on 8 different days, namely the 7th, 10th, 11th, 12th, 13th, 18th, 19th, and 30th. Observers can also see how much the price of Bitcoin (BTC) pair Rupiah (IDR) was when the Diamond Crash level came out.

Additionally, Figure 9 illustrates how the program will record the greatest price of Bitcoin in IDR that is attained following the appearance of the diamond crash level. An acronym beginning with dc_start will be used to designate each diamond crash point. The next section, indicated in yellow, is the window of opportunity to turn a profit.

Following the marking of the diamond crash point, the next maximum price of BTC IDR will be displayed by the dotted green line. The precise moment the highest possible price of Bitcoin was attained is indicated by the intersection of the dotted green lines.

TABLE IV
SUMMARY OF COMPUTATION OUTCOMES FOR M ROI (%), M MAX PRICE, M PROFIT TIME, AND M MAX PROFIT

dc_start	Date	Start Price	M Profit Time	M Max Price	M Max Profit	M ROI (%)
1	07/06/2022	429.552.000	0 **days** 18:34:59	456.001.000	26.449.000	6,16
2	10/06/2022	431.000.000	0 **days** 01:27:00	433.500.000	2.500.000	0,58
3	11/06/2022	418.791.000	0 **days** 00:32:25	422.416.000	3.625.000	0,87
4	12/06/2022	413.001.000	0 **days** 13:07:00	415.999.000	2.998.000	0,73
5	13/06/2022	370.000.000	0 **days** 01:22:10	386.946.000	16.946.000	4,58
6	18/06/2022	287.997.000	8 **days** 04:49:41	323.375.000	35.378.000	12,28
7	19/06/2022	268.642.000	7 **days** 16:03:59	323.375.000	54.733.000	20,37
8	30/06/2022	287.359.000	0 **days** 03:44:30	288.409.000	1.050.000	0,37
				Total	**143.679.000**	**45,93**

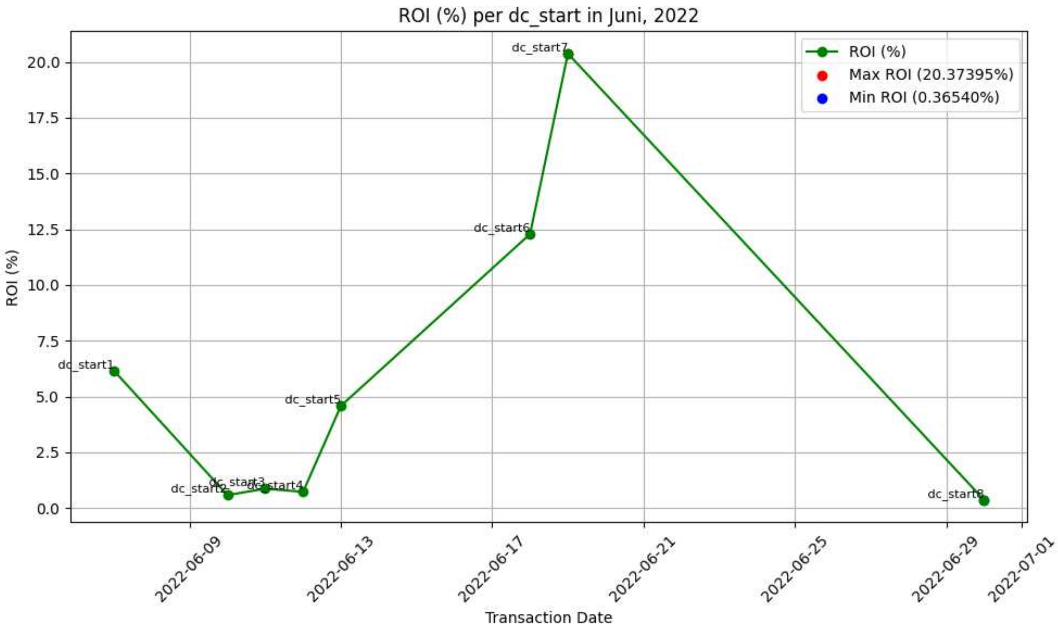


Fig. 8 Max and Min ROI Potential in June 2022

Next, a profit-finding computation is performed, which involves deducting the initial price of the diamond crash when it initially appeared from the greatest BTC IDR price that was reached thereafter. The outcome is the highest possible profit on the BTC-IDR trading pair. The possible maximum profit of BTC IDR can then be divided by the original price of BTC IDR at the time of the first diamond fall, and then multiplied by 100% to determine the potential maximum return on investment (ROI). After then, a single table contains a summary of all the computation results. Table 4 shows the following information in chronological order: the time profit was made after the diamond point appeared to crash in the same month range (M Profit Time), the price of BTC IDR at its maximum price following the appearance of the diamond crash in the same month range (M Max Price), the maximum potential profit made in the same month (M Max Profit), and the maximum potential return of value achieved in percent units also in the same month (M ROI).

It is evident from the above table that there were eight separate days in June 2022 when there was a diamond crash. The fifth diamond crash yielded the fastest profit, which was realized in 1 hour, 22 minutes, and 10 seconds. On the other hand, the sixth diamond crash yielded the longest profit, which took 8 days, 4 hours, 49 minutes, and 41 seconds to realize. The seventh diamond crash produced the highest possible maximum profit and the lowest potential maximum ROI, whereas the eighth diamond crash produced the highest potential maximum profit and the lowest potential maximum ROI. A diagram showing the maximum ROI value, minimum ROI value, and possible ROI achievement (%) in June 2022 is presented to aid in understanding.

Next, the highest profit and return on investment realized in the 24 hours after the day the diamond crash level initially occurs are computed. Short-term trading within the same day is covered by this trading pattern. Refer to Figure 10 for further information. In the years 2022, June 7, 10, 11, 12, 13, 18, 19, and 30, the visualization is displayed. Every day at 1:00 pm Western Indonesian time, trade lasts until 23:59:59 seconds the next day.

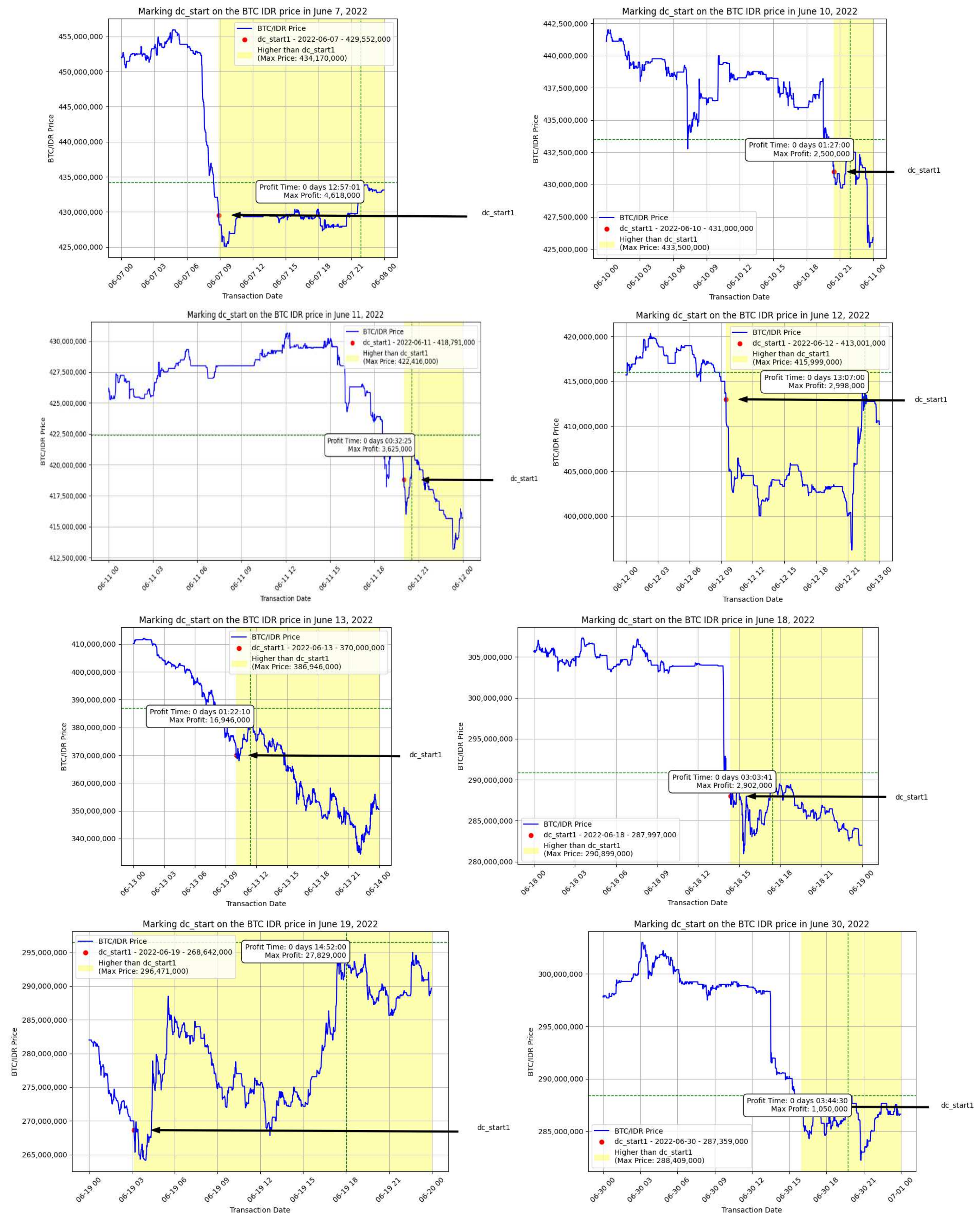


Fig. 9 Diamond Crash Level Marking per day, Maximum Price, and Maximum Profit Potential in June 2022

Figure 11 provides detailed information, including a blue-marked monthly history of the price of BTC IDR. A black arrow points to the red-marked spot on that day where the diamond crash first appeared. The region shown in yellow represents the price of Bitcoin IDR, which is higher now than it was during the diamond meltdown.

TABLE V
A RECAP OF THE POSITIVE DAY-WIDE MAXIMUM RETURN AND PROFIT FOR JUNE 2022

dc_start	Date	Start Price	D Profit Time	D Max Price	D Max Profit	D ROI (%)
1	07/06/2022	429.552.000	12:57:01	434.170.000	4.618.000	1,08
2	10/06/2022	431.000.000	01:27:00	433.500.000	2.500.000	0,58
3	11/06/2022	418.791.000	00:32:25	422.416.000	3.625.000	0,87
4	12/06/2022	413.001.000	13:07:00	415.999.000	2.998.000	0,73
5	13/06/2022	370.000.000	01:22:10	386.946.000	16.946.000	4,58
6	18/06/2022	287.997.000	03:03:41	290.899.000	2.902.000	1,01
7	19/06/2022	268.642.000	14:52:00	296.471.000	27.829.000	10,36
8	30/06/2022	287.359.000	03:44:30	288.409.000	1.050.000	0,37
				Total	**62.468.000**	**19,56**

The greatest price of BTC IDR that day, the duration of the profit in hours, minutes, and seconds, and the amount of profit that resulted from subtracting the price at dc_start1 are all displayed at the junction of the dotted green line. Figure 11 shows the specifics of each day in June 2022 that the Diamond Crash initially emerges as follows:

a. The first diamond crash occurred on June 7, 2022, and was identified by dc_start1. With a maximum profit potential of 4,618,000 Rupiah, it will take 12 hours, 57 minutes, and 1 second to reach that potential.
b. There is a maximum profit potential of 2,500,000 Rupiah on June 10, 2022, and it will take 1 hour and 27 minutes to reach that potential.
c. On June 11, 2022, the time to achieve maximum profit potential is 32 minutes and 25 seconds, with a maximum profit potential of 3,625,000 Rupiah.
d. On June 12, 2022, the time to achieve maximum profit potential is 13 hours and 7 minutes, with a maximum profit potential of 2,998,000 Rupiah.
e. The time to reach the maximum profit potential of 16,946,000 Rupiah on June 13, 2022, is one hour, twenty-two minutes, and ten seconds.
f. On June 18, 2022, the time to achieve maximum profit potential is 3 hours, 3 minutes, and 41 seconds, with a maximum profit potential of 2,902,000 Rupiah.
g. 14 hours and 52 minutes on June 19, 2022, will be needed to reach the maximum earning potential of 27,829,000 Rupiah.
h. On June 30, 2022, the time to achieve maximum profit potential is 3 hours, 44 minutes, and 30 seconds, with a maximum profit potential of 1,050,000 Rupiah.

Table 6 and Figure 12 show the possibility for reaching maximum profit between the accumulation per day (D) and range per month (M) in June 2022, as well as the potential for achieving maximum return on investment (ROI). It is evident that there is room for maximal performance. In the same month, the monthly return on investment (ROI) exceeds the daily accumulation (D).

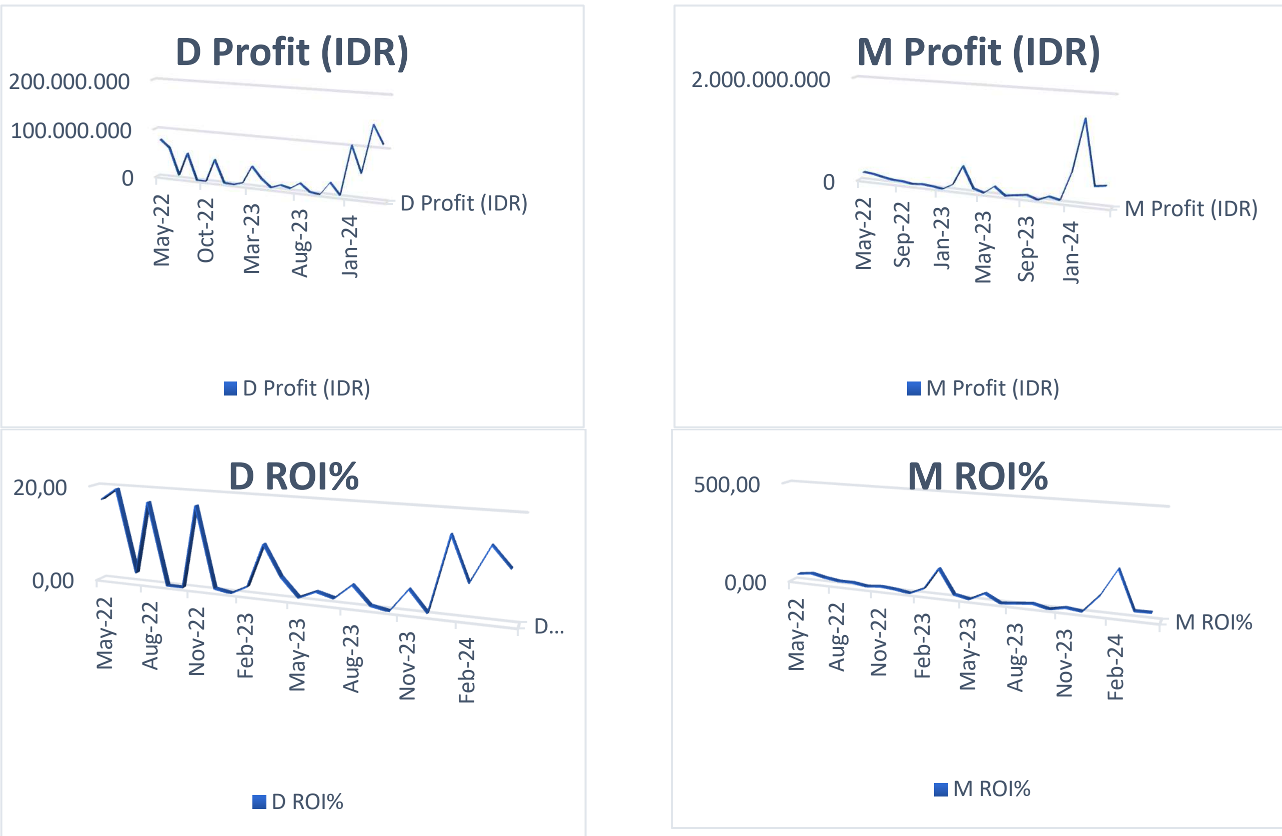


Fig. 10 Comparing the Potential Maximum ROI & Profit Achievables for the Month (M) & Day (D) Ranges 2022-2024

TABLE VI
MAXIMUM PROFIT AND ROI ACHIEVEMENTS CALCULATION RESULTS BY DAY AND MONTH

Y – M	M Profit (IDR) × 10^3	M ROI (%)	D Profit (IDR) × 10^3	D ROI (%)	T DC
Apr-24	398.878	37,60	108.810	10,29	12
Mar-24	368.931	36,04	145.088	14,22	4
Feb-24	1.568.147	219,74	50.397	6,97	6
Jan-24	596.959	90,04	102.563	15,61	12
Dec-23	42.747	6,66	3.153	0,49	1
Nov-23	88.258	15,71	26.266	4,67	3
Oct-23	0	0,00	0	0,00	0
Sep-23	70.038	17,59	2.647	0,66	3
Aug-23	40.075	9,89	18.078	4,36	3
Jul-23	7.476	1,70	5.819	1,32	2
Jun-23	163.763	41,03	10.134	2,24	3
May-23	17.456	4,24	2.836	0,69	2
Apr-23	77.939	18,70	18.692	4,44	4
Mar-23	489.830	139,05	41.873	10,73	7
Feb-23	109.255	32,69	6.064	1,76	3
Jan-23	0	0,00	0	0,00	0
Dec-22	30.475	11,63	1.575	0,60	1
Nov-22	48.299	17,61	47.570	17,32	3
Oct-22	31.557	10,75	107	0,04	1
Sep-22	61.511	21,40	177	0,06	2
Aug-22	70.257	20,88	54.779	17,33	6
Jul-22	101.813	30,98	7.165	2,13	3
Jun-22	143.679	45,93	62.468	19,56	8
May-22	164.812	36,45	78.337	17,21	7

M Profit (IDR) × 10^3 is a Chance of Realizing the Highest Profit Every Month. M Profit happens when, on any given day in a given month, Diamond Crash makes a first appearance. This results in the purchase of one bitcoin, for example, which is then sold for a price greater than the bitcoin's purchase price in that same month. The greatest price that is more than the bitcoin purchase price for the remaining portion of the same month is subtracted to determine profit. If D profit is only allowed to last for a day, then M profit can be made that day or another day later in the same month. **D Profit (IDR) × 10^3** is a Chance of Obtaining the Highest Profit Every Day. D Profit happens when on a specific day in a specific month, the Diamond Crash level first emerges and is followed by the acquisition of, say, one bitcoin. Additionally, the maximum daily profit potential is realized if, within the next 24 hours, the price of bitcoin reaches its peak and surpasses the purchase price. Later on in the same month, this D profit will be added together and compared to the technical computation of M profit. The year and month that the dataset crashes is explained by **Y - M.** In percentage terms, M ROI % indicates the possibility of making maximum profits every month. In percentage terms, D ROI % illustrates the possibility of achieving daily maximum earnings. What is recorded is the initial Diamond Crash level that occurs on a single day inside the same month. The monthly total of diamond crash levels is displayed by TDC.

C. Hyperparameter Comparison on GRU and the LSTM-Conv2d

The patterned dataset in the crash condition was chosen in the preceding section as the dataset with the best grouping when compared to the moon condition and the complete condition. The Patterned Dataset in Crash Conditions will be utilized as test data in this section to forecast the upcoming price. The number of error values that are produced using the Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) is then displayed in the test results. Prior to generating predictions, two deep learning models LSTM-Conv2d and GRU are used to compare the hyperparameter test settings in order to determine which prediction model performs best. A basic pattern dataset in crash settings from November 2022 was selected as example data, and the LSTMConv2d model construction system will be used to test it.

Table 7 then displays the findings of contrasting two deep learning models employing MSE, RMSE, and MAE metric values, which will be utilized to forecast pattern dataset values under crash scenarios. The end findings demonstrate that in terms of predicting pattern dataset values in the subsequent crash circumstance, the LSTM-conv2d model performs better.

D. Discussion

The study of how technology affects economic phenomena can be done by combining computer science with other interdisciplinary subjects. It also seeks to mitigate the drawbacks of cryptocurrency use[31]. According to certain study, cryptocurrency specialists play a significant role in the price discovery process in the Bitcoin market and increase market efficiency when they make non-positive predictions [32]. Naturally, this is consistent with the findings of this study's K-means clustering, which reveal that the Patterned Dataset in Crash Conditions exhibits the best clustering outcomes.

While certain individual bitcoin returns can be predicted with the help of informed trading, overall market projections remain largely inaccurate [33]. Because of this, the purpose of the patterned dataset model is to clearly illustrate when the decrease in Bitcoin prices corresponds with the negative emotion that causes the rise in the crash line. The next course of action is to wait for the crash line to reach its highest point, which is the diamond crash [9], [1], [8], and then take appropriate action, such as gradually increasing your position in the long term or buying on the spot trading market, in preparation of the impending price collapse. in fact took place. This implies that you should wait to trade until the Crash Condition Patterned Dataset reaches its maximum growth, which will cause the bulk of cryptocurrencies to collapse.

TABLE VII
LSTM-CONV2D MODEL HYPERPARAMETER METRIC TEST RESULTS COMPARING WITH GRU MODEL

Deep Learning Model on Patterned Crash Dataset, November 2022		Resample (Minutes)	Epoch	Batch Size	Activation	MSE	RMSE	MAE	Execution time
LSTM-Conv2d	Validation Split 0.2	2	10	128	RELU	0.0001	0.0075	0.0041	3 min
GRU	Learning Rate 0.4					1.2767	1.1299	0.7551	1 min

A significant factor in the decisions made by cryptocurrency traders is public opinion. Bullish emotion was linked to higher returns, larger trade sizes, and more deposits and withdrawals, according to data from two million transactions [34]. The lower the crash values that emerge while employing a Patterned Dataset Model in Crash Conditions, the more bullish the market is shown to be [9], [1], [8]. Research on the application of heuristic reasoning for topological data analysis trends in identifying important transition early warning signals in financial time series on samples of positive and negative bubbles in historical Bitcoin values is consistent with this study [35].

The incidence of extreme falls is consistent with research on the origins of price surges in the cryptocurrency market and how to identify investor attention that increases the likelihood of jumps in predicting excess returns in the future [36]. A patterned dataset under crash conditions indicates the possibility of subsequent price reversals in the bitcoin/rupiah pair. Before completing a transaction, this moment can serve as a marker and a further decision [9], [1], [8].

E. Future development

Future research on the ultimate outcomes of the Patterned Dataset development in this study will focus, among other things, on its impact on particular cryptocurrency communities or groups, hence reducing the overall impact of Bitcoin on particular altcoins [37]. Moreover, testing the basic and hybrid models from this Patterned dataset is very open to be conducted on various reputable Cryptocurrency Digital Markets, which are indexed on coinmarketcap [19], [19], and openly share data access via API connections [18], [8], as a bridge for access to transaction data in real time [38]. This is true even though the original source of raw data for the Patterned dataset model was the Indodax Cryptocurrency Digital Market [9].

Ultimately, the goal of this research is to help bitcoin traders become more adept at understanding how to turn a profit as a method for adjusting to uncertainty [39]. Models from patterned datasets may be utilized as test data in the future to implement techniques and models like sentiment analysis, technical analysis, machine learning, statistical learning, and deep learning. These techniques and models are associated with patterned datasets in both conditions of price increase and decrease (crash). More information about research on Patterned Dataset Models can be accessed on the ppdsi github page [40].

V. Conclusion

In comparison to Patterned Dataset Clustering in a Price Increase (Moon) Condition as well as in a Complete Condition, the Experimental Results of Clustering a Basic Patterned Dataset with a Price Decrease (Crash) Condition using K-means Cluster yielded the greatest results. The ROI per month (M) is always significantly larger than the Potential for Achieving Maximum Profit and Return of Value (ROI) per day (D) accrued in the same month, according to the results of the Potential for Achieving Maximum Profit and Return of Value (ROI) calculation.

After determining the highest profit and ROI, the next step is to evaluate which model, LSTM-Conv2d or GRU, better predicts pattern datasets under crash conditions. The LSTM-Conv2d model, with a validation split of 0.2, outperforms the GRU model with a learning rate of 0.4 across MSE, RMSE, and MAE, despite the GRU model's slightly faster execution time.

In the future, LSTM-Conv2d could improve accuracy in predicting patterned datasets during time-sequential crashes. Further research should compare predictions using various models and techniques, such as statistical learning, machine learning, deep learning, technical analysis, and sentiment analysis, to enhance pattern dataset testing in crash scenarios.

References

[1] R. Parlika, Mustafid, and B. Rahmat, "Detect Areas of Upward and Downward Fluctuations in Bitcoin Prices Using Patterned Datasets," in *2023 IEEE 9th Information Technology International Seminar (ITIS)*, 2023, pp. 1–6. doi: 10.1109/ITIS59651.2023.10420129.

[2] D. O. Oyewola, E. G. Dada, and J. N. Ndunagu, "A novel hybrid walk-forward ensemble optimization for time series cryptocurrency prediction," *Heliyon*, vol. 8, no. 11, p. e11862, 2022, doi:10.1016/j.heliyon.2022.e11862.

[3] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," p. 9, [Online]. Available: https://bitcoin.org/bitcoin.pdf

[4] B. Yang, Y. Sun, and S. Wang, "A novel two-stage approach for cryptocurrency analysis," *Int. Rev. Financ. Anal.*, vol. 72, no. October 2019, p. 101567, 2020, doi: 10.1016/j.irfa.2020.101567.

[5] K. Dunbar and J. Owusu-amoako, "Journal of Behavioral and Experimental Finance Predictability of crypto returns : The impact of trading behavior ☆," *J. Behav. Exp. Financ.*, vol. 39, p. 100812, 2023, doi: 10.1016/j.jbef.2023.100812.

[6] Z. Wen, E. Bouri, Y. Xu, and Y. Zhao, "Intraday return predictability in the cryptocurrency markets: Momentum, reversal, or both," *North Am. J. Econ. Financ.*, vol. 62, no. June, p. 101733, 2022, doi:10.1016/j.najef.2022.101733.

[7] A. Brauneis, R. Mestel, R. Riordan, and E. Theissen, "Bitcoin unchained: Determinants of cryptocurrency exchange liquidity," *J. Empir. Financ.*, vol. 69, no. August, pp. 106–122, 2022, doi:10.1016/j.jempfin.2022.08.004.

[8] R. Parlika, Mustafid, and B. Rahmat, "Use of Patterned Datasets (Minimum and Maximum) to predict Bitcoin and Ethereum price movements," Technium: Romanian Journal of Applied Sciences and Technology, vol. 16, pp. 137–142, Oct. 2023, doi: 10.47577/technium.v16i.9972.

[9] R. Parlika, Mustafid, and B. Rahmat, "Minimum, Maximum, and Average Implementation of Patterned Datasets in Mapping Cryptocurrency Fluctuation Patterns," *Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 378–386, 2024, doi: 10.62527/joiv.8.1.1543.

[10] M. Ozdamar, A. Sensoy, and L. Akdeniz, "Retail vs institutional investor attention in the cryptocurrency market," *J. Int. Financ. Mark. Institutions Money*, vol. 81, no. February, p. 101674, 2022, doi:10.1016/j.intfin.2022.101674.

[11] C. Cai, W. Li, H. Han, and M. Liu, "Risk scenario-based value estimation of Bitcoin," *Procedia Comput. Sci.*, vol. 199, pp. 1198–1204, 2021, doi: 10.1016/j.procs.2022.01.152.

[12] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," *Phys. A Stat. Mech. its Appl.*, vol. 551, p. 124569, 2020, doi:10.1016/j.physa.2020.124569.

[13] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions," *J. Inf. Secur. Appl.*, vol. 55, no. May, p. 102583, 2020, doi: 10.1016/j.jisa.2020.102583.

[14] L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica," *Tutor. Quant. Methods Psychol.*, vol. 9, no. 1, pp. 15–24, 2013, doi:10.20982/tqmp.09.1.p015.

[15] I. Nasirtafreshi, "Forecasting cryptocurrency prices using Recurrent Neural Network and Long Short-term Memory," *Data Knowl. Eng.*, vol. 139, no. March, p. 102009, 2022, doi:10.1016/j.datak.2022.102009.

[16] A. A. Ewees, M. A. Elaziz, Z. Alameer, H. Ye, and Z. Jianhua, "Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price

volatility," *Resour. Policy*, vol. 65, no. November 2019, p. 101555, 2020, doi: 10.1016/j.resourpol.2019.101555.

[17] R. Parlika, Mustafid, and B. Rahmat, "The next cryptocurrency price movement prediction application uses patterned datasets," *AIP Conf. Proc.*, vol. 3116, no. 1, p. 30003, May 2024, doi: 10.1063/5.0210203.

[18] R. Parlika and P. W. Atmaja, "Rizubot Version 1.0 algorithm: How to read the price movements of Crypto Currency Using the API to find a good purchase price," vol. 1, no. Icst, pp. 1045–1049, 2018, doi:10.2991/icst-18.2018.211.

[19] R. Parlika and A. Pratama, "Use of the Web API as a basis for obtaining the latest data on bitcoin prices at 30 exchange places," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 12035, 2021, doi:10.1088/1757-899X/1125/1/012035.

[20] R. Parlika, Mustafid, and B. Rahmat, "Detect Areas of Upward and Downward Fluctuations in Bitcoin Prices Using Patterned Datasets," 2023 IEEE 9th Information Technology International Seminar (ITIS), pp. 1–6, Oct. 2023, doi: 10.1109/itis59651.2023.10420129.

[21] N. X. Vinh, "Information Theoretic Measures for Clusterings Comparison : Variants , Properties , Normalization and Correction for Chance," vol. 11, pp. 2837–2854, 2010.

[22] A. Rosenberg and J. Hirschberg, "V-Measure : A conditional entropy-based external cluster evaluation measure," no. June, pp. 410–420, 2007.

[23] C. Yeh and M. Yang, "Evaluation measures for cluster ensembles based on a fuzzy generalized Rand index," *Appl. Soft Comput. J.*, vol. 57, pp. 225–234, 2017, doi: 10.1016/j.asoc.2017.03.030.

[24] D. Chicco and G. Jurman, "A statistical comparison between Matthews correlation coefficient (MCC), prevalence threshold , and Fowlkes – Mallows index," *J. Biomed. Inform.*, vol. 144, no. June, p. 104426, 2023, doi: 10.1016/j.jbi.2023.104426.

[25] E. Koo and G. Kim, "Centralized decomposition approach in LSTM for Bitcoin price prediction," *Expert Syst. Appl.*, vol. 237, no. PA, p. 121401, 2024, doi: 10.1016/j.eswa.2023.121401.

[26] P. J. R and V. S. Das, "Cryptocurrency Price Prediction Using Long-Short Term Memory Model," *Int. J. Res. Sci. Innov.* |, vol. V, no. Vii, pp. 70–72, 2018, [Online]. Available: www.rsisinternational.org

[27] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas, "An Advanced CNN-LSTM Model for Cryptocurrency Forecasting," 2021. doi: 10.3390/electronics10030287.

[28] C. Wang, D. Shen, and Y. Li, "Aggregate Investor Attention and Bitcoin Return: The Long Short-term Memory Networks Perspective," *Financ. Res. Lett.*, vol. 49, no. May, p. 103143, 2022, doi:10.1016/j.frl.2022.103143.

[29] S. H. Othman, R. Z. Radzi, D. Stiawan, and T. Sutikno, "Hybrid gated recurrent unit bidirectional-long short-term memory model to improve cryptocurrency prediction accuracy," vol. 12, no. 1, pp. 251–261, 2023, doi: 10.11591/ijai.v12.i1.pp251-261.

[30] M. J. Hamayel, "A Novel Cryptocurrency Price Prediction Model Using GRU , LSTM and bi-LSTM Machine Learning Algorithms," pp. 477–496, 2021.

[31] Y. Yue, X. Li, D. Zhang, and S. Wang, "How cryptocurrency affects economy? A network analysis using bibliometric methods," *Int. Rev. Financ. Anal.*, vol. 77, no. 71988101, p. 101869, 2021, doi: 10.1016/j.irfa.2021.101869.

[32] D. F. Gerritsen, R. A. C. Lugtigheid, and T. Walther, "Can Bitcoin Investors Profit from Predictions by Crypto Experts?," *Financ. Res. Lett.*, vol. 46, May 2022, doi: 10.1016/j.frl.2021.102266.

[33] Y. Wang, C. Wang, A. Sensoy, S. Yao, and F. Cheng, "Can investors' informed trading predict cryptocurrency returns? Evidence from machine learning," *Res. Int. Bus. Financ.*, vol. 62, no. August 2021, p. 101683, 2022, doi: 10.1016/j.ribaf.2022.101683.

[34] J. Bowden and R. Gemayel, "Sentiment and trading decisions in an ambiguous environment: A study on cryptocurrency traders," *J. Int. Financ. Mark. Institutions Money*, vol. 80, no. January, p. 101622, 2022, doi: 10.1016/j.intfin.2022.101622.

[35] S. W. Akingbade, M. Gidea, M. Manzi, and V. Nateghi, "Communications in Nonlinear Science and Numerical Simulation Why topological data analysis detects financial bubbles ?," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 128, no. October 2023, p. 107665, 2024, doi: 10.1016/j.cnsns.2023.107665.

[36] Z. Zhang and R. Zhao, "International Review of Financial Analysis Good volatility , bad volatility , and the cross section of cryptocurrency," *Int. Rev. Financ. Anal.*, vol. 89, no. June, p. 102712, 2023, doi: 10.1016/j.irfa.2023.102712.

[37] H. Chaudhari and M. Crane, "Cross-correlation dynamics and community structures of cryptocurrencies," *J. Comput. Sci.*, vol. 44, p. 101130, 2020, doi: 10.1016/j.jocs.2020.101130.

[38] R. Parlika and P. W. Atmaja, "Realtime monitoring of Bitcoin prices on several Cryptocurrency markets using Web API, Telegram Bot, MySQL Database, and PHP-Cronjob," *Proceeding - 6th Inf. Technol. Int. Semin. ITIS 2020*, pp. 253–257, 2020, doi:10.1109/ITIS50118.2020.9321109.

[39] R. Gemayel and A. Preda, "International Review of Financial Analysis Performance and learning in an ambiguous environment : A study of cryptocurrency traders ☆," *Int. Rev. Financ. Anal.*, vol. 77, no. July, p. 101847, 2021, doi: 10.1016/j.irfa.2021.101847.

[40] R. Parlika, "Crypto-Patterned-Dataset https://bit.ly/dsi5." [Online]. Available: https://github.com/pddsi5/Crypto-Patterned-Dataset/blob/main/README.md