Pakistan Journal of Commerce and Social Sciences
2025, Vol. 19(3), 407-440
https://doi.org/10.64534/Commer.2025.511

Unsupervised Machine Learning Based Anomaly Detection in High Frequency Data: Evidence from Cryptocurrency Market

Muhammad Nouman Latif [1]*, Muhittin Kaplan[2] & Asad Ul Islam Khan[3]
[1,2,3] Department of Economics, School of Business, Ibn Haldun University, Istanbul, Turkiye
*Corresponding author: muhammad.latif@stu.ihu.edu.tr

Article History

Received: 20 June 2025 Revised: 15 Sept 2025 Accepted: 24 Sept 2025 Published: 30 Sept 2025

Abstract

The rapid integration of cryptocurrencies into the global financial ecosystem has introduced unprecedented challenges in market surveillance, risk management, and anomaly detection. While conventional statistical models such as ARIMA (Autoregressive Integrated Moving Average) and GARCH (Generalized Autoregressive Conditional Heteroscedasticity) have been widely used for anomaly detection, their reliance on assumptions of normality and stationarity often fails to capture the complexities of high-frequency, non-linear cryptocurrency trading. Furthermore, traditional risk metrics including down-to-up volatility, negative conditional skewness, and relative frequency may overlook short-term anomalies due to data aggregation limitations.

In order to address these issues, this paper proposes machine-learning model for detecting anomalies in cryptocurrency markets using Jupyter Notebook. We compare four advanced unsupervised machine learning models, i.e, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Isolation Forest (iForest), One-Class Support Vector Machine (OC-SVM), and Local Outlier Factor (LOF) for anomaly detection by using Monte Carlo simulations. The findings indicate that DBSCAN has the highest precision (79.7%) with the fewest false positives, making it ideal for supervisory monitoring. However, the high false positive rates of OC-SVM and Isolation Forest limit their use. By using data of six well-known cryptocurrencies at three different temporal resolutions (daily, hourly, and 15-minute) the performance of these four unsupervised learning techniques also examined and confirmed that the anomalies identified by DBSCAN are also consistent with the other three methods. Additionally, for robustness of results, we use UpSet Plots to incorporate the shared anomalies and found across the three unsupervised learning methods. Number of anomalies also depends on the volatility and time interval of cryptocurrencies, more volatile / high frequency more anomalies.

The study presents sound methodological approach for facilitating financial monitoring and mitigating risks in the cryptocurrencies market, and provides useful information for

market players, analysts and policymakers. These results emphasize the importance of choosing algorithms based on specific surveillance targets to promote greater stability in digital asset environments.

Keywords: Unsupervised machine learning models, anomaly detection, Monte Carlo simulations, Bitcoin, Dashcoin, Ethereum, Stellar, Tron, Litecoin.

1. Introduction

The rapid growth of cryptocurrencies as a basic component of the worldwide financial system has clearly complicated market behavior. Unlike traditional financial markets run under centralized management, cryptocurrency markets operate within a decentralized system. Although blockchain technology enhances security and transparency but it also increases volatility, making cryptocurrencies more prone to price manipulation, fraud and liquidity crises (Aysan et al., 2019). As a result, there is need of mechanism for recognizing anomalies and fostering more transparency and confidence is anomaly detection since the lack of governmental supervision in cryptocurrency markets raises investment risks (Corbet et al., 2019; Pérez-Cano & Jurado, 2025). Therefore, in recent years, professionals and pertinent stakeholders have stressed upon the requirement of improved analytical tools to both monitor and spot market irregularities (Başçı & Khan, 2023).

Many researchers applied standard statistical models such as Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and Autoregressive Integrated Moving Average (ARIMA) to detect anomalies. While effective in capturing volatility and projecting price movement, these models are limited by presumptions of normality, stationarity, and linearity, which often are out of step with the features of high-frequency cryptocurrency market (Kyriazis, 2021). Latif et al., (2025) uses risk measures to detect anomalies i.e, Relative Frequency (RF), Negative Conditional Skewness (NCS), and Down-to-Up Volatility (DUV). Given their consideration of skewness and non-linearity, these techniques are particularly suited for high-frequency data (Khan et al., 2024; Latif et al., 2025). They can oversight the anomalies since they depend on data aggregation from daily to monthly. Moreover, extremely high values are frequently categorized as anomalies by conventional risk assessments without differentiating between whether these variations are the result of manipulative trading or fraudulent activity and are normal components of market cycles. So, there is need of more flexible, data-driven methods that can recognize abnormalities instantly and consider the complexity of cryptocurrencies markets, as observed by Bhatia & Jain, (2025). The advanced Machine learning models can overcome these problems. These models have the capacity to detect unusual trends in the financial data with great accuracy (Kaleem et al., 2024).

Machine Learning Models classified into two categories i. supervised machine learning models and ii. Unsupervised machine learning models. Supervised Machine learning models requires labeled data in order to train the models. Whereas, unsupervised machine learning models does not rely on pre labeled datasets. Furthermore, it discovers deviations free of prior knowledge of anomalies by learning the inherent structure in the data

automatically (Cholevas et al., 2024; Eskin et al., 2002). There are many unsupervised machine learning models but four are mostly used i.e., Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Isolation Forest, One-Class Support Vector Machine (OC-SVM), and Local Outlier Factor (LOF). They are different in nature as LOF essentially discovers density-based anomalies by measuring the departure of value from its neighbors (Breunig et al., 2000a). On the other hand, the kernel-based OC-SVM approach consider both regular and anomalous (Schölkopf et al., 2001). Isolation Forest uses the boundary-based approach that deviates from density (Liu et al., 2024). Fourth method is DBSCAN, which is effective for detecting anomalies by using localized density fluctuation. They are using different parameters to detect anomalies in the data. Many studies done to detect anomalies in financial data by using different unsupervised machine learning models (Agyemang, 2024; Cholevas et al., 2024) but they did not compare the above-mentioned unsupervised machine learning models by using synthetic data. For comparison, we use confusion matrix and four predominant evaluation criteria for machine learning models, i.e., Accuracy, Precision, Recall and F-1 score (Agyemang, 2024). For robustness of results, we use Monte Carlo simulations and repeat the process 500 times to get comprehensive results. As the real-time detection of anomalies helps to reduce financial risk, combat fraud, and stabilize markets (Urquhart & Yarovaya, 2024).

Many studies used financial data for anomaly detection by implying unsupervised machine learning models (Poutré et al., 2024). But there is no evidence in literature that these methods on real dataset over three different time intervals of six leading cryptocurrencies i.e., Litecoin (LTC), Tron (TRX), Ethereum (ETH), Dashcoin (DSH), Bitcoin (BTC), and Stellar (XLM) has been utilized. These six cryptocurrencies hold a particular value with regard to stability, acceptance, and market capitalization (Golnari et al., 2024). The first and most well-known cryptocurrency, Bitcoin, maintains a consistent market position with an estimated value of $1.5 trillion in 2025. Not far behind with corresponding values of $190 billion and $23.3 billion are Ethereum (ETH) and Tron (TRX). Although Ethereum and Tron (TRX) function similarly, the former exhibits higher price volatility. The remaining three cryptocurrencies, with lower market capitalization also display noteworthy volatility due to their trading activity and market conditions (Naz et al., 2023). Furthermore, UpSet Plots are used to have a clear visual comparison of the overlap and unique findings (Blum & Gelfman, 2023). By systematically comparing these machine-learning techniques, our study aims to identify the most effective anomaly detection approach to provide valuable insights for investors, analysts and regulatory bodies in cryptocurrency market.

2. Literature Review

In order to effectively study anomaly detection in the cryptocurrency market, this section presents an analysis of the relevant theoretical frameworks and empirical reviews that support this research.

2.1 Theoretical Framework

The theoretical understanding for anomaly detection in cryptocurrency markets, particularly for Bitcoin, Dashcoin, Ethereum, Litecoin, Tron and Stellar is based on numerous significant frameworks that serve as the basis for this study. These include the Efficient Market Hypothesis, the density-based anomaly detection theory, statistical learning theory and deep learning representation theory.

The Efficient Market Hypothesis (EMH) traditionally assumes that financial markets are efficient and they usually include all available information, making it difficult to consistently achieve returns exceeding the market average (Fama, 1970). However, cryptocurrency markets have often refuted this assumption as they exhibit significant deviations from EMH principles due to their dynamic and evolving nature (Fan et al., 2022). Cryptocurrency has repeatedly displayed varying degrees of market efficiency in recent years. According to Kyriazis (2021b), crypto-based currencies frequently show abnormalities in returns that create detectable anomalies. These anomalies are predominantly noticeable during periods of high volatility, creating opportunities for anomaly detection algorithms to identify irregular patterns that may indicate market manipulation or fraudulent activities.

The density-based anomaly detection theory, on the other hand, is mostly applicable in analyzing cryptocurrency data with high frequencies. This approach identifies anomalies by examining the local density deviation of a data point with respect to its neighbors (Breunig et al., 2000b). In cryptocurrency markets where trading patterns often form clusters with varying densities, this method allows for the identification of outliers that significantly deviate from usual trading behaviors (Zhu et al., 2017). Therefore, this anomaly detection approach is mainly effective for cryptocurrencies with diverse market capitalization and trading volumes. For example, the higher liquidity nature of Bitcoin generates denser trading clusters compared to Stellar, necessitating adaptive density thresholds when applying algorithms like DBSCAN (Yahia, Mouhssine, El Alaoui, et al., 2024).

Another helpful theoretical tool in anomaly-centered studies is the statistical-learning theory. This theory offers the mathematical basis for machine learning models and allows the development of algorithms that can identify abnormal patterns in price movements and trading volumes of cryptocurrencies (Schölkopf et al., 2001). In the case of Bitcoin, Dashcoin, Litecoin, Tron, Ethereum and Stellar, statistical learning method enables the designing of decision boundaries that separate normal from abnormal trading behaviors. Since deep learning methods can independently create hierarchical representations from raw data (Wang et al., 2023), they are perfect for spotting particular traits of the Bitcoin. It is observed that using clustering-based methods helped us to identify changes in the Bitcoin. This skill helps to acquire and reconstruct typical trade patterns so that anomalies marked by large reconstruction error may be found (Kim, 2016).

2.2 Empirical Studies

As the first cryptocurrency to emerge, Bitcoin has attracted considerable scholarly attention due to its higher market capital in cryptocurrency market. Kampers et al., (2022) uncovered bitcoin market manipulation using Local Outlier Factor (LOF) and K-Nearest Neighbourhood (KNN). Their study reported over 80% accuracy in anomaly detection when using high-frequency data, demonstrating the effectiveness of density-based methods in identifying manipulative strategies in the Bitcoin market. A similar study using price return analysis, by Shi et al., (2019) shows that anomalous return usually generates significant market volatility. The study implies that useful markers of market anomalies could be statistical characteristics of Bitcoin returns.

In a comparative study by Witayanont & Viyanon, (2025), Isolation Forest proved to outperform Histogram-based Outlier Score (HBOS) in spotting crashes in Bitcoin. The results reveal that tree-based methods might be sufficient to portray the non-linear relationship in the price of Bitcoin. Ethereum also draws the interest of structural anomaly experts because of its unusual anomaly pattern and smooth capabilities. Ehsan et al., (2024) demonstrates that since machine-learning classifiers can control aberrant transaction clusters related to smart execution, they assist anomaly identification in Ethereum. Expanding on this, Patel et al., (2020) developed a deep learning-based graph anomaly detection approach, using the network architecture of Ethereum transactions, for Ethereum block-chain networks and found that graph convolutional networks may perhaps successfully identify suspicious transaction patterns. Hisham et al., (2023) assert that the hybrid model Searching for Unrelated Variables (SULOV) is more accurate in identifying anomalies, including regional abnormalities and in effectively recognizing temporal dependencies in Ethereum price movement. Stellar has also gained attention from researchers due to its low-cost, fast-paced transaction since its launch in 2014. Bielecki (2023) notes that trade patterns are following stochastic process, requiring specialized anomaly detection techniques. According to his study, Stellar's faster settling time generates different temporal anomaly footprints than other well-known cryptocurrencies.

Other research on anomaly detection among multiple cryptocurrencies has examined by Yahia et al., (2024) holistically investigated machine learning methods for anomaly identification between two cryptocurrencies. Using three unsupervised machine-learning methods, they discovered that Local Outlier Factor (LOF) outperforms Isolation Forest (IF) in terms of predicting accuracy and is the most effective method for identifying abnormalities in bitcoin returns. Despite these advances, few studies have compared anomaly detection across six distinct cryptocurrencies using multiple time intervals and four or more unsupervised machine-learning models. Addressing this gap, the present study evaluates four unsupervised machine-learning models: Local Outlier Factor (LOF), One-Class Support Vector Machine (OC-SVM), Isolation Forest, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) across six different cryptocurrencies.

These include a blend of three highly volatile currencies and three consistent currencies over three different time intervals: daily, hourly, and 15 minutes. Through this comparative analysis, the study attempts to identify the most efficient strategies for detecting anomalies in high-frequency cryptocurrency data. The findings offer practical insights for analysts and investors for future investment.

3. Methodological Framework and Real Data

To detect anomalies, this study utilized four unsupervised machine-learning techniques. In the first stage, the relative performance of the four techniques was evaluated on the basis of Model Evaluation Metrics (Explained below) by generating artificial synthetic data. After confirmation of results, we use same four unsupervised machine learning techniques to real data of six cryptocurrencies to detect the anomalies.

3.1 Synthetic Data Generation

In order to generate synthetic data, we use Standard T-distribution as financial return data is heavily tailed (McLeay, 1986). Here T-distribution will be as follows.

$$\boldsymbol{f(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})}\ (1 + \frac{t^2}{v})^{-\frac{(v+1)}{2}}}$$

Here t denotes the value for which PDF (Probability Density Function) is evaluated ($-\infty < t < \infty$). $\boldsymbol{v}$ is the degree of freedom which is greater than 0. $\boldsymbol{\Gamma}$ is the gamma function.

The mean of standard T-distribution is 0 for all $\boldsymbol{v}$ greater than 1. Variance is $\frac{\boldsymbol{v}}{\boldsymbol{v-2}}$ for all v greater than 2.

After generating artificial synthetic data, we introduce 30 anomalies, which were uniformly distributed from -4 to +4, making a total of 2030 observations. Another similar study done by Agyemang, (2024) for generating artificial synthetic data, he used 1000 observation using normal distribution and added 10 anomalies.

3.2 Monte Carlo Simulation Design

The following steps were taken for model selection:

i. Generation of data by using Standard T-distribution and introducing anomalies in the data
ii. Application of anomaly detection methods (unsupervised machine learning models)
iii. Number of anomalies detected
iv. Confusion matrices
v. Repetition of steps (i) to (iv) 500 times
vi. Overall performance evaluation for 500 Monte Carlo simulations

3.3 Model Evaluation Metrics

To evaluate the performance of four unsupervised machine learning models, the following metrics were used: Accuracy, F1 score, Precision and Recall, with the final two being

specifically used for outliers (Kumar et al., 2021; Poutré et al., 2024; Rezapour Mashhadi, 2019).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = 2\ X \frac{Precision\ x\ Recall}{Precision + Recall}$$

Here:

TP (True Positives): The number of outliers correctly identified as outliers.

TN (True Negative): The number of normal observations correctly identified as normal.

FP (False Positive): The number of normal observations incorrectly identified as outliers.

FN (False Negative): The number of outliers incorrectly identified as normal observations.

3.4 Unsupervised Machine-Learning Techniques

3.4.1 Local Outlier factor

The LOF score of each point can be obtained by using the formula below:

$$LOF_K(P) = \frac{\sum_{i=0}^{Nk(P)} \frac{\rho k(0)}{\rho k(P)}}{|Nk(P)|}$$

Here *Nk(P)* is the transformer data point in the k-distance of Point *P* and $\rho k(P)$ is the local reachable density of point *P*. $LOF_K(P)$ is the Local outlier factor score of point *P*. If the value of LOF is less than, or closer to 1, it means that the density of point *P* is great than the density of neighboring points and vice versa, if the value is higher than one. Therefore, the higher value of *P* is suggesting it to be a outlier (Breunig et al., 2000b; Zou et al., 2023).

3.4.2 Isolation Forest

The Isolation Forest algorithm was employed as one of the principal unsupervised anomaly detection methods in this study due to its demonstrated efficacy in high-dimensional financial data (Cholevas et al., 2024; Liu et al., 2024). The algorithm's fundamental premise relies on the observation that anomalous data points are both infrequent and exhibit characteristics that differ markedly from normal observations, making them more susceptible to isolation through random partitioning.

The implementation utilized the following key components: i.e., for tree construction, a forest of t=100 isolation trees generated and each tree built using a subsample. Here the features were selected at random with uniform probability at each node. Therefore, split

values were also chosen at random between the observed minimum and maximum of the features. The anomaly score s(x, n) for observation *x* was computed as:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

Where: E[h(x)] represents the mean path length across all trees,

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

that serves as the normalization factor and H(n-1) denotes the harmonic number approximated by ln(n-1) + γ (γ ≈ 0.5772). Observations with s(x,n) > 0.5 were classified as anomalous (Agyemang, 2024; Liu et al., 2024).

3.4.3 One-Class Support Vector Machine (OC-SVM)

OC-SVM is a kernel-based method that defines a boundary, separating normal and anomalous instances. It is a function:

$$F: \partial \subset X \rightarrow \{+1, -1\}$$

Where +1 says that value is normal, and -1 otherwise. The classifier F given by:

$$F(x) = Sgn(\sum_{i=1}^{m} a_i K(x, xi) - b)$$

Where $b = \sum_{j=1}^{m} a_j K(xi, xj)$ for any xi ∈ ∂ and K(.,.) is a kernel function. The co-efficient a_i, i=1,2,…m are obtained as solution of the following optimization:

$$\underset{a1,..am}{Min} \sum_{i,j=1}^{m} a_i a_j K(xi, xj)$$

(Ananias & and Negri, 2021; Schölkopf et al., 2001; Yang et al., 2021).

3.4.4 Density Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN identifies anomalies as data points located in low density regions, making it effective for capturing localized density variations. The method finds the distance to the *k*th neighbor for each object and sorts the objects in ascendingly order of these distances. Each cluster has its own *Eps* value. The initial value of the *Eps* be the distance to the kth neighbor of the first object in the sorted dataset. The traditional DBSCAN was applied to the data using the present Eps value to get the first top dense cluster. The second top dense cluster used the smallest kth distance among the unclassified objects as a new value for Eps then the DBSCAN was applied again. This process continued until all objects clustered. This method produces several small clusters and could also deliver singleton clusters (Ester et al., 1996; Fahim, 2018).

3.5 Real Data Application

We used the Monte Carlo simulation to check the performance of the unsupervised machine learning models (Agyemang, 2024) and thereafter, used them on real datasets of six different cryptocurrencies.

3.6 Data and Source

To utilize unsupervised machine-learning techniques for detecting anomalies in the cryptocurrency market, this study used data obtained from the Kaggle database (https://www.kaggle.com/datasets). Given the availability of data at daily (732 observations), hourly (17544 observations), and 15 minutes (70176 observations) intervals, only data covering the period from April 2023 to March 2025 was utilized in this study. The study identified and analyzed six currencies: three highly volatile and three highly stable cryptocurrencies with significant market capitalization. This selection made to ensure the inclusion of diverse market behaviors. While the highly volatile currencies could possibly provide a deeper understanding into unpredictable fluctuations, the highly stable currencies might present a distinction with steady trends, enhancing the robustness of anomaly detection methods.

3.7 Commonly Detected Anomalies

This work presented the intersections of outlier sets found by every model using UpSet Plots. UpSet Plots allowed us to have combined anomalies in actual data sets. Horizontal bars reflect the overall deviations per model. Conversely, vertical bars show the overall count of anomalies that that combination shares (Blum & Gelfman, 2023).

4. Results and Discussion

This section, presents the process of selecting the best-performing model and the identification of anomalies in real data sets at different time intervals.

4.1 Monte Carlo Simulation Results

The synthetic dataset consisted of 2,000 observations including 30 deliberately added anomalies (Agyemang, 2024). Four unsupervised machine learning methods were evaluated: DBSCAN, Isolation Forest, Local Outlier Factor (LOF), and One-Class Support Vector Machine (OC-SVM). The experiment was conducted 500 times to develop resilience and the outcomes were compiled to evaluate consistency over runs. SVM was implemented using Python's sklearn.svm module. Following significant hyperparameter tuning, the best configuration was discovered to be nu = 0.1, which denotes the predicted proportion of outliers, and gamma = 0.1, which regulates the model's sensitivity to individual data points.

Assuring convergence, the model allowed a maximum of 2,000 iterations using an adaptive learning rate identified as optimal for the dynamic change of step sizes during training. Particularly in using high-dimensional datasets, the Isolation Forest approach in recursive partitioning was chosen for its efficiency in spotting abnormalities. The main settings were

n_estimators set to 100, supplying a sufficient number of base estimators for effective ensemble learning, and max_samples set to auto, thereby allowing the model to ascertain a reasonable subsampling rate. To obtain the outlier ratio straight away from the data, the contamination parameter was set to auto. By maximizing the balance between sensitivity and specificity in local density estimation, setting n_neighbors = 20 let the LOF technique to flourish. This value specified the neighborhood size used depending on relative density fluctuations in order to identify outliers. DBSCAN was used since it could find outliers as points discovered in low-density areas without depending on predefined cluster counts.

Separating noise from true clusters allowed experimentation to maximize the parameters eps = 0.5 (the greatest distance for neighbor consideration) and min_samples = 10 (the minimum number of points needed to form a dense region). All models were only taught on t-distributed synthetic data to probe the basic structure of the prevailing class. Every algorithm categorized data points in the evaluation stage as either normal or aberrant, according to its set decision limits. Ensuring constant performance measurements, the 500 repetitions of the experiment yielded a thorough assessment of algorithmic stability in several contexts.

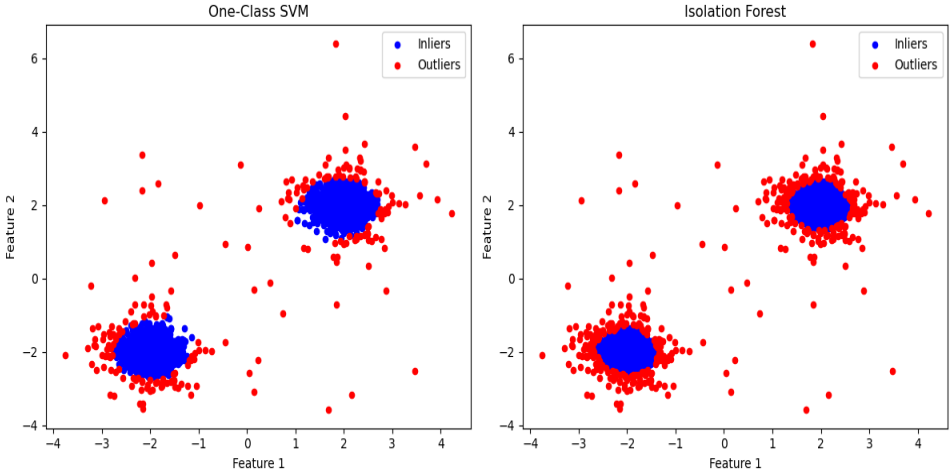


Figure 1: Anomaly detection with One-Class SVM and Isolation Forest on Synthetic Data Set

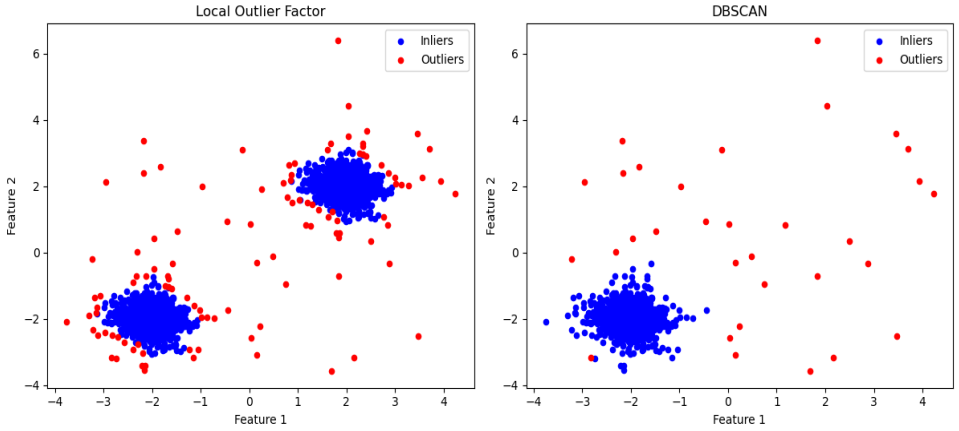


Figure 2: Anomaly Detection with Local Outlier Factor and DBSCAN on Synthetic Data Set

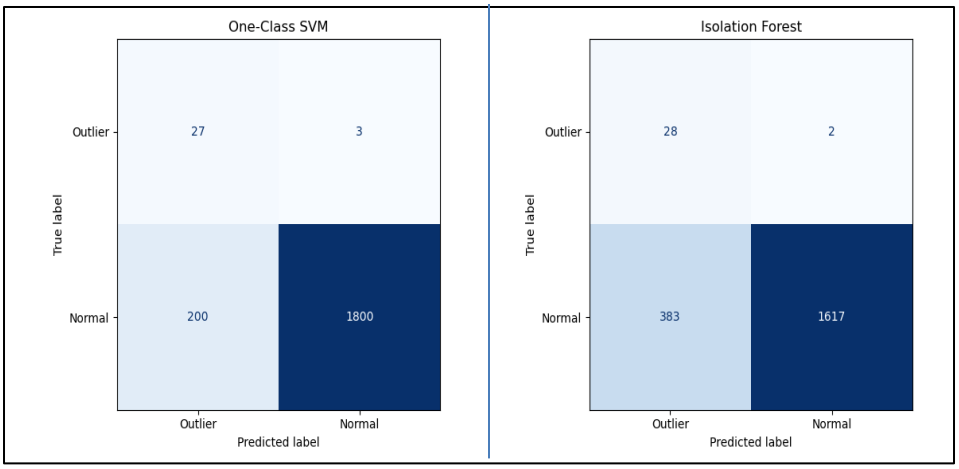


Figure 3: Confusion Matrices of One-Class SVM and Isolation Forest on Synthetic Data Set

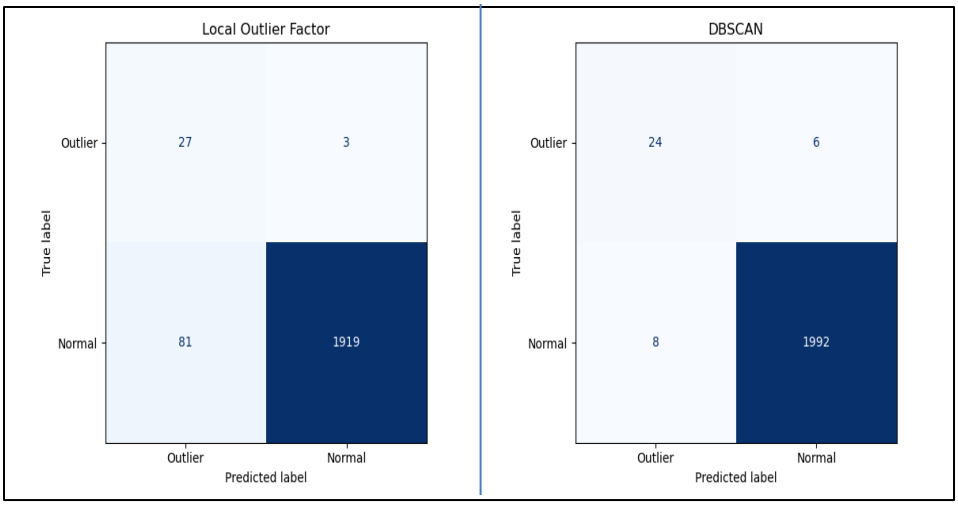


Figure 4: Confusion Matrices of Local Outlier Factor and DBSCAN on Synthetic Data Set

Table 1: Model Performance Metrics

Model	Accuracy	Precision (Outliers)	Recall (Outliers)	F1 Score (Outliers)
One-Class SVM	0.9000	0.1189	0.9000	0.2101
Isolation Forest	0.8103	0.0681	0.9333	0.1270
Local Outlier Factor	0.9586	0.2500	0.9000	0.3913
DBSCAN	0.9931	0.7500	0.8000	0.7742

The evaluation of four unsupervised learning algorithms reveals significant variations in their ability to detect anomalies, with each model demonstrating distinct strengths and limitations. These findings provide valuable insights for practitioners in selecting appropriate anomaly detection methods for real-world applications.

DBSCAN emerged (see Table 1) as the most effective model, demonstrating exceptional performance across all metrics. Exhibiting an optimal combination of Recall (80.00%) and Precision (75.00%), it demonstrated an accuracy of 99.31% and an impressive F1-score of 0.7742. The low false positive rate and ongoing identification of genuine abnormalities indicate that the density-based approach used by DBSCAN was particularly suitable for the data characteristics. This highlights its potential applications in areas such as network security systems and financial fraud detection, where operational efficiency and detection accuracy are paramount. The Local Outlier Factor (LOF) method demonstrated an

impressive accuracy of 95.86%. While its rate of Recall aligned with that of leading models (90.00%), the comparatively lower Precision (25.00%) indicates a propensity for false positives. In the absence of anomalies in medical diagnostics or critical infrastructure monitoring, this performance profile may be attainable, even if it incurs higher costs than evaluating false positives.

Although demonstrating poor Accuracy (11.89%), the third model, One-Class SVM interestingly displayed 90.00% accuracy and recall. This trend implied that the model uses an inclusive detection approach, thereby spotting most anomalies but at a great expense of many false positives. Such performance may be acceptable in initial screening tests, where further verification rounds can eliminate false predictions. Isolation Forest finally demonstrated basic limitations in this regard. Although it displayed great Recall (93.33%), it's very low Accuracy (6.81%) makes it completely useless for most uses since most of its anomaly predictions would be false. This result emphasizes the need of context-specific algorithm evaluation since solutions that work in some fields would not be so in others.

These results carry several important implications for researchers as well as policymakers. First, they emphasize the importance of careful model selection depending on specific application requirements since no technique outsets another across all criteria (Poutré et al., 2024). Second, the data reveal how essentially distinct performance traits arise from different algorithmic approaches. Finally, the study emphasizes that higher accuracy by itself does not guarantee useful practical value as demonstrated by the precision limits of OC-SVM and Isolation Forest. These results suggest that DBSCAN offers institutions adopting anomaly detection systems the most balanced option where both detection rates and operational efficiency count. When maximizing anomaly detection is crucial, independent of false positive rates, LOF could be a good substitute. The low accuracy of OC-SVM and Isolation Forest in this study suggests they could need significant change or combination with other approaches to be successful in related applications.

Table 2: Average Model Performance Metrics across 500 Simulations

Model	Accuracy	Precision (Outliers)	Recall (Outliers)	F1 Score (Outliers)	Anomalies Detected
One-Class SVM	0.9006	0.1565	0.9278	0.2679	237.0
Isolation Forest	0.8044	0.0883	0.9575	0.1616	435.6
Local Outlier Factor	0.9610	0.3249	0.8726	0.4716	109.3
DBSCAN	0.9901	0.7970	0.6759	0.7271	34.3

Table 3: Average Confusion Matrix Stats (over 500 Simulations)

Model	TP	FN	FP	TN
One-Class SVM	900050	99950	1450	18550
Isolation Forest	801350	198650	850	19150
Local Outlier Factor	962800	37200	2550	17450
DBSCAN	996350	3650	6500	13500

In order to get robustness of the results, we did the 500-time simulation of the experiment. The results in table 2 confirms that DBSCAN outperform here as well by achieving exceptional accuracy (99.01%) and precision (79.70%), while maintaining reasonable recall (67.59%). Its high F1-score (0.7271) and low false positive rate demonstrate its reliability for practical applications where both detection accuracy and operational efficiency are crucial. The Local Outlier Factor (LOF) showed complementary strengths, with slightly lower accuracy (96.10%) but better recall (87.26%), making it a viable alternative when detecting more potential anomalies prioritized over precision. These results suggest that density-based approaches like DBSCAN and proximity-based methods like LOF offer the most robust solutions for real-data anomaly detection tasks. Our results are contrasted to the study conducted by Agyemang, (2024) where he compared unsupervised machine learning models on synthetic data and found OCSVM and Local Outlier Factor best fit. As the financial returns data is heavily tailed so we have to use t-distribution instead of normal distribution.

The remaining models exhibited more specialized performance profiles that may limit their standalone usefulness. As the One-Class SVM achieved high recall (92.78%) but very low precision (15.65%), indicating it captures most anomalies but generates numerous false alarms - a characteristic that might only be acceptable in multi-stage detection systems. As the Table 3 also shows that it shows 99950 false negative when we repeated the experiment 500 times. Isolation Forest demonstrated the extreme trade-off, with the highest recall (95.75%) but alarmingly poor precision (8.83%), flagging over four times as many anomalies as DBSCAN while maintaining the lowest F1-score (0.1616). These findings underscore that while some algorithms maximize anomaly detection, they do so at substantial operational costs. Practitioners must carefully consider their specific tolerance for false positives against the consequences of missed anomalies when selecting detection approaches, with DBSCAN representing the most generally applicable solution for balanced performance requirements. As DBSCAN detected only 3650 false negative values in 500 times repetition.

Therefore, table 2 and table 3 (which are repeated the experiment 500 times), confirm the results that are shown in table 1.

4.2 Real Data Results

After getting results from synthetic data, using Monte Carlo simulations and results shows DBSCAN is most suitable and now we will use these four unsupervised Machine Learning Models on six major cryptocurrencies across three different time intervals to check their performance and the results reveals several important insights about unsupervised anomaly detection in volatile markets. These results demonstrate how detection methodologies and sampling frequencies jointly influence anomaly identification in digital asset markets.

Table 4: Anomaly Detection of Cryptocurrencies Data on Different Time Interval

Method	Daily		Hourly		15 Minutes	
	No of Anomalies	Joint Anomalies with DBSCAN	No of Anomalies	Joint Anomalies with DBSCAN	No of Anomalies	Joint Anomalies with DBSCAN
Bitcoin						
OC-SVM	37	3	878	13	3509	22
Isolation Forest	37	3	878	13	3505	22
L.O.F	37	3	878	13	3509	22
DBSCAN	3	-	13	-	22	-
Dashcoin						
OC-SVM	37	0	878	7	3509	1
Isolation Forest	37	0	877	7	3502	1
L.O.F	37	0	878	7	3509	1
DBSCAN	0	-	7	-	1	-
Ethereum						
OC-SVM	37	6	878	15	2931	17
Isolation Forest	37	6	874	15	3507	17
L.O.F	37	6	878	15	3509	17
DBSCAN	6	-	15	-	17	-
Litecoin						
OC-SVM	36	2	819	14	3102	23
Isolation Forest	37	2	876	14	3296	23

L.O.F	36	2	869	14	894	23
DBSCAN	2	-	14	-	23	-
Stellar						
OC-SVM	37	13	878	20	3509	20
Isolation Forest	37	13	878	20	3505	20
L.O.F	37	13	878	20	3509	20
DBSCAN	13	-	20	-	20	-
TRON						
OC-SVM	37	7	878	28	3081	42
Isolation Forest	37	7	877	28	3478	42
L.O.F	37	7	878	28	3509	42
DBSCAN	7	-	28	-	42	-

From table 4, we observe that these four unsupervised machine-learning models displayed different kinds of anomaly detection over different time intervals. DBSCAN identified the least number of anomalies across all cryptocurrencies and timeframes, with daily detections ranging from 0 (Dashcoin) to 13 (Stellar). These results are aligned with the study of Yahia et al., (2024) for bitcoin and stellar using DBSCAN. These results reveal that DBSCAN's density-based technique detects the most statistically significant deviations. The other three models: OC-SVM, isolation forest, and LOF showed somewhat similar results, identifying the same or virtually the same number of anomalies for every asset and time period. As the results from table1 also says that other than DBSCAN, our three unsupervised machine-learning models are identifying false negative. Therefore, this convergence of multiple computational approaches suggests that they could be recognizing similar trends of typical price swings instead of true anomalies of occurrence.

However, the low overlap between the anomalies revealed by DBSCAN and those identified by other methods exposes important differences in how different algorithms interpret anomalies. For instance, in Bitcoin daily data, Isolation Forest, OCSVM and LOF found 37 anomalies, only 3 (8%) matched what DBSCAN found, suggesting that most events found by other models did not fulfill DBSCAN's stricter outlier criteria. Additionally, sampling frequency has obvious influences. The change from daily to 15-minute data yielded an increase in anomaly detection across all methods since 15-minute intervals usually generate 100–300 times more anomalies than daily data. This tendency most likely shows that, considering shorter timescale, high-frequency data includes more noise, which allows the algorithms to misinterpret normal daily ups and downs for any unexpected events. We discovered interesting differences in detection patterns, distinct to

every asset: Stellar showed the best concordance between DBSCAN and other approaches from 35% to 100% throughout several time periods.

Dashcoin produced the most varied findings, with no daily anomalies discovered by DBSCAN. From seven daily to 42 at 15-minute intervals, TRON revealed rising DBSCAN detections as frequency changed. The observed variations in volatility patterns, average deal size, and liquidity point to various market microstructure characteristics among cryptocurrencies. The results show that spotting unusual patterns in bitcoin markets requires careful review of the algorithms used and the chosen time periods since these choices greatly affect the observed events and count of them. Although DBSCAN's conservative findings are particularly successful at identifying truly amazing market events, competing methods seem more appropriate for thorough surveillance of unusual behavior.

4.2.1 Daily Data

From appendix “A-figure 1 to 6”, we observed that Dashcoin is very much volatile. Whereas, the high capital currency like Bitcoin and Ethereum are stable. The more the low capital currency the more they will be volatile. So, this behavior is also repeated in anomaly detection the currencies that are more volatile they have more anomalies. When we change the observations from daily (732 observations), hourly (17544 observations), and 15 minutes (70176 observations) intervals, the more anomalies detected.

While discussing the results of table 1 and table 2, we found a pattern that DBSCAN is detecting anomalies more accurately as compare to other three unsupervised machine learning models while using synthetic data. Now, while using real time daily data set of six different cryptocurrencies we observe that the anomalies detected by DBSCAN are lesser than other three models. So, in order to have robustness of results, we used the UpSets plot and found that in Bitcoin daily data the anomalies found by DBSCAN are detected by the other three methods as well. So, we can say that DBSCAN detect the less and accurate anomalies but we can consider the 13 other anomalies that are identify by the LOF, Isolation Forest and OCSVM jointly. This behavior is reported in the other 5 cryptocurrencies as well. In Dashcoin, DBSCAN did not detect an anomaly, so we can consider the 8 anomalies detected by all three models. Stellar has the greatest number of anomalies detected by DBSCAN but these are also detected by other three models.

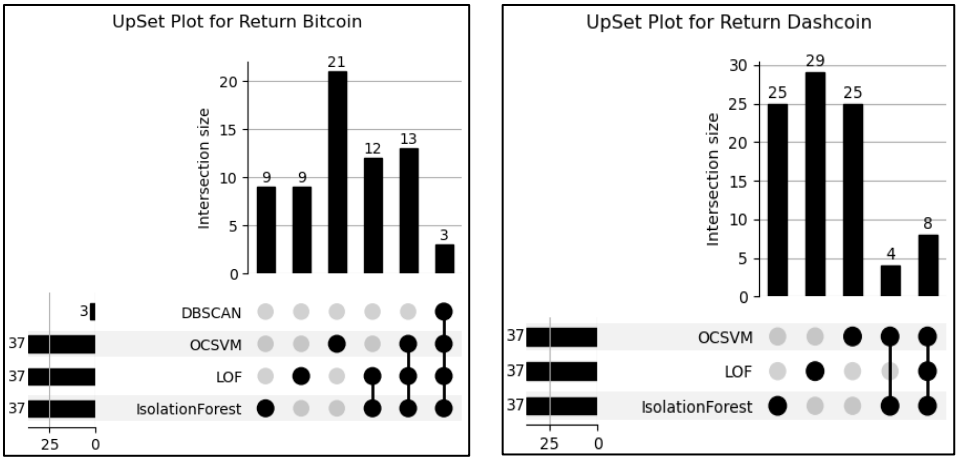


Figure 5: Joint Anomaly Detection on Bitcoin and Dashcoin Daily Data Return using UpSet Plot

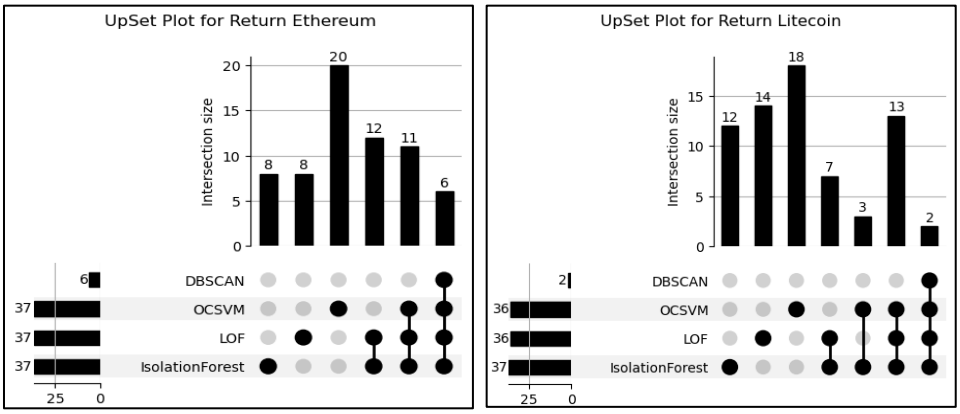


Figure 6: Joint Anomaly Detection on Ethereum and Litecoin Daily Data Return using UpSet Plot

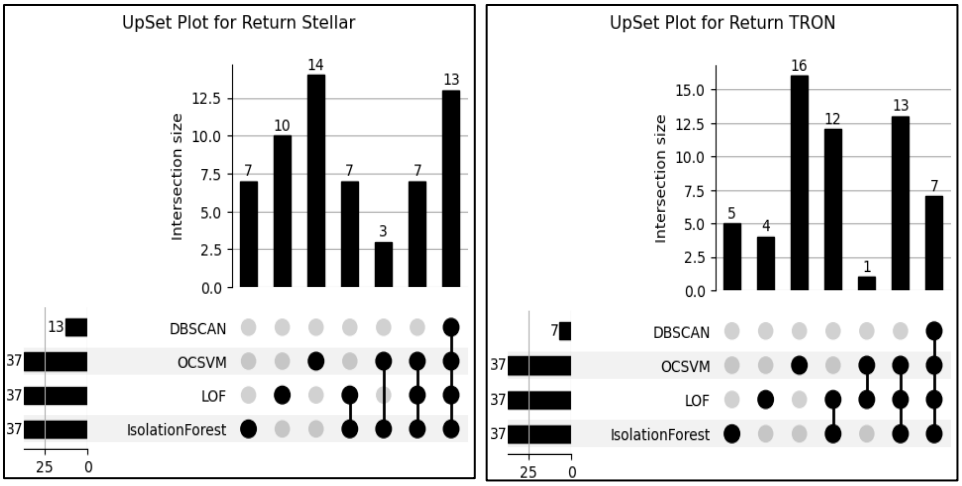


Figure 7: Joint Snomaly Detection on Stellar and TRON Daily Data Return using UpSet Plot

4.2.2 Hourly Data

From appendix "A-figure 7 to 12", we used hourly interval data of Bitcoin, Tron, Stellar, Dashcoin, Ethereum and Litecoin, Bitcoin and Dashcoin leads in anomalies with four of our unsupervised machine learning models. As hourly data of these currencies is more volatile as compare to other four cryptocurrencies.

From figure 8 to 10, we can have joint anomalies by using UpSet Plots. Again, we observed that the number of anomalies detected by DBSCAN are also caputered by the OCSVM, LOF and Isolation Forest. DBSCAN find 13 anomalies in Bitcoin, 7 anomalies in Dashcoin, 15 anomalies in Ethereum, 14 anomalies in Litecoin, 20 anomalies in Stellar and 28 anomalies in Tron. The number of anomalies is reduced in Dashcoin as compared to other five cryptocurrencies because it is more volatile.

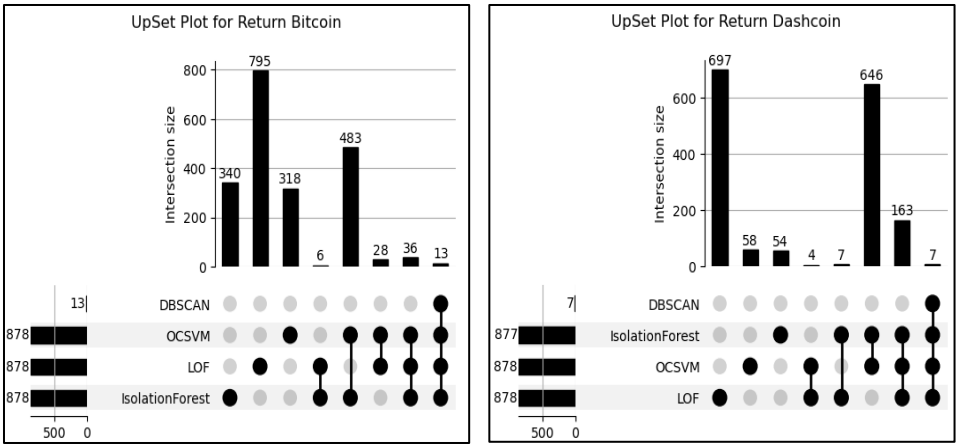


Figure 8: Joint Anomaly Detection on Bitcoin and Dashcoin Hourly Data Return using UpSet Plot

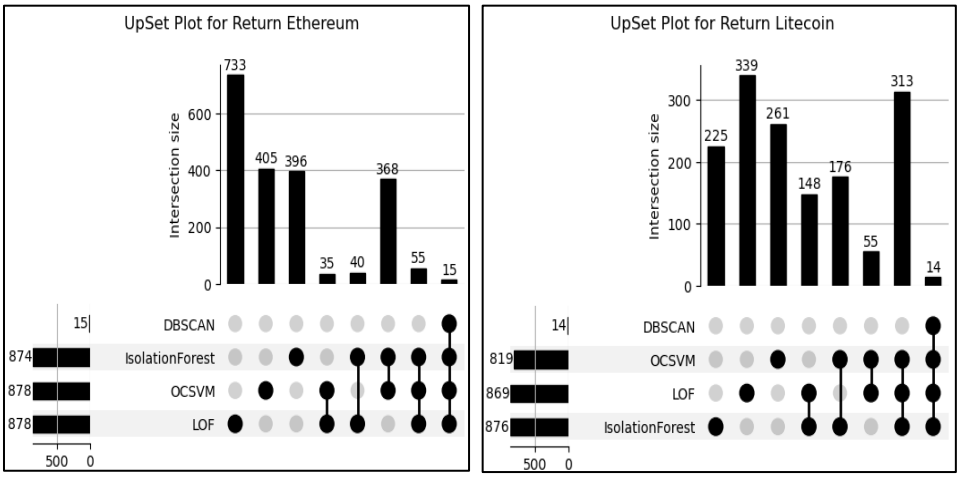


Figure 9: Joint Anomaly Detection on Ethereum and Litecoin Hourly Data Return using UpSet Plot

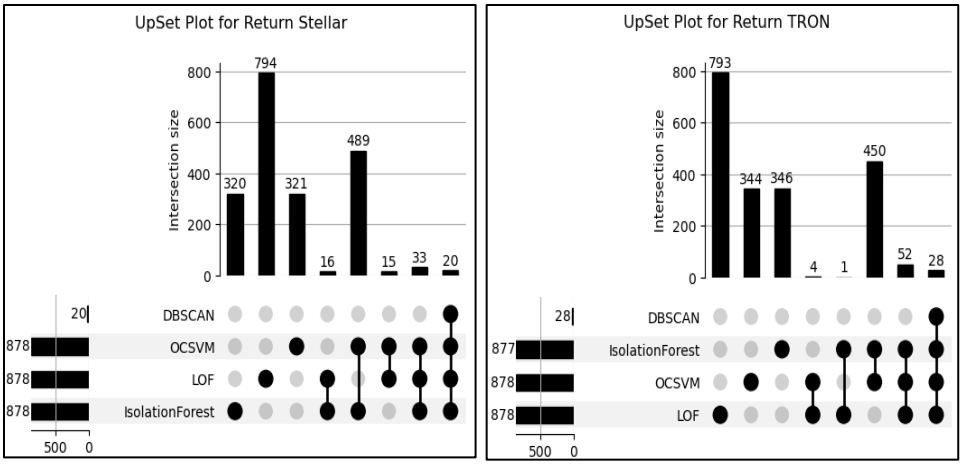


Figure 10: Joint Anomaly Detection on Stellar and TRON Hourly Data Return by using UpSet Plot

4.2.3 15 Minutes Data

In order to detect anomalies in six different cryptocurrencies data taken on 15 minutes time interval, we employed four unsupervised machine-learning models as can be seen in Appendix "A-Figures 13 to 18". As frequency of data is increasing, the number of anomalies is also increasing due to increase in volatility.

As we are detecting anomalies by using four different unsupervised machine learning models. DBSCAN as our best anomaly detection model on the basis on Monte Carlo simulation. So, in order to have comprehensive anomalies, we use the criteria that at least 3 of our models should detect anomaly at that point. We use UpSet Plot for this purpose and got 22 anomalies in Bitcoin, 1 in Dashcoin, 17 in Ethereum, 23 in Litecoin, 20 in Stellar and 39 in Tron. As Tron is very much volatile that's why shows maximum joint anomalies. When we increase the number of observations by reducing the time frame from daily to 15 minutes, the number of anomalies is decreased in dashcoin.

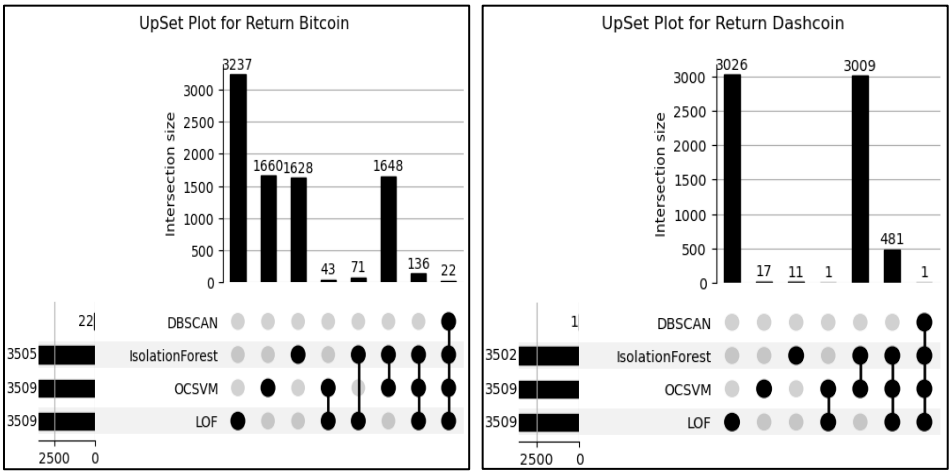


Figure 11: Joint Anomaly Detection on Bitcoin and Dashcoin 15-Minutes Interval Data Return using UpSet Plot

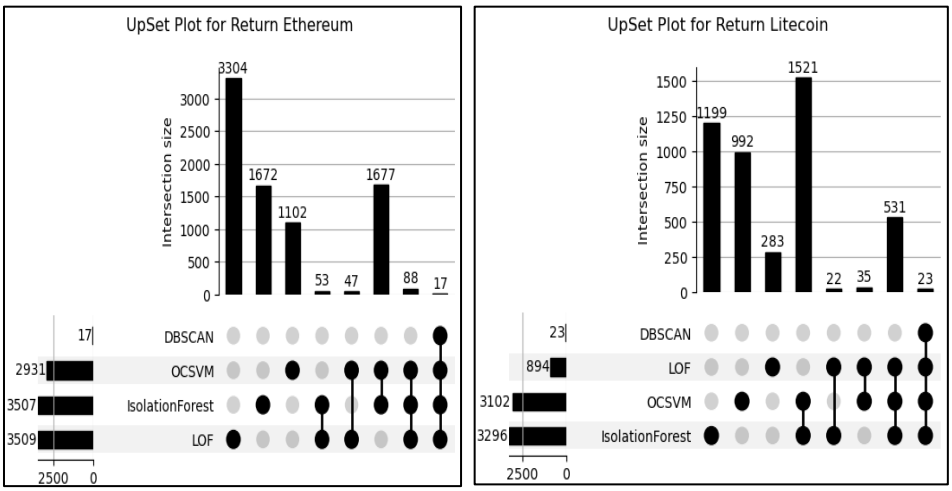


Figure 12: Joint Anomaly Detection on Ethereum and Litecoin 15-Minutes Interval Data Return using UpSet Plot

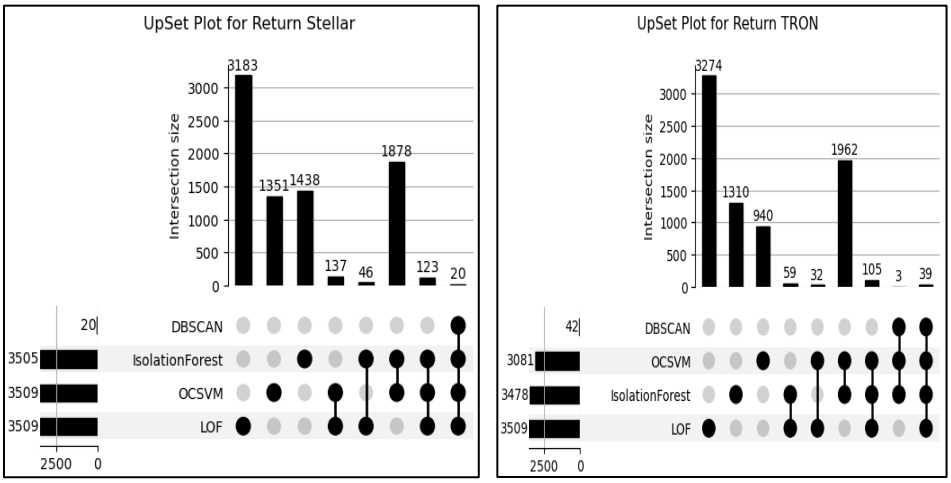


Figure 13: Joint Anomaly Detection on Stellar and TRON 15-Minutes Interval Data Return using UpSet Plot

5. Conclusion

This comprehensive evaluation is calculated as done by Cholevas et al., (2024) of our four unsupervised anomaly detection methods but also across six major cryptocurrencies and three-time resolutions, yields several critical insights. Results shows that cryptocurrencies market is not following efficient market hypothesis (EMH) as market information is not fully available, as a result anomalies occurs. The analysis reveals that detection outcomes are highly sensitive to both algorithmic choice and temporal granularity, with DBSCAN consistently demonstrating the most conservative detection profile. The striking divergence between DBSCAN outputs and those from other methods suggests fundamental differences in how these algorithms define anomalous behavior in cryptocurrency markets. The exponential increase in detected anomalies at higher frequencies (from daily to 15-minute intervals) underscores the challenge of distinguishing meaningful market events from normal volatility in high-frequency trading environments. The intraday traders can invest in Tron or Litecoin, as they are more frequent. Whereas Stellar and Litecoin are good for day-to-day trading.

For surveillance systems prioritizing high-confidence alerts, DBSCAN is recommended as the primary detector due to its conservative approach. For more comprehensive monitoring, DBSCAN may be supplemented with LOF or OC-SVM, through their outputs should be verified using secondary verification. It is suggested to avoid using Isolation Forest as a standalone detector given its excessive false positive rate. Based on the analysis using three

different time dimensions, it is suggested that daily data appears most suitable for identifying significant market events where as hourly data may offer optimal balance between sensitivity and specificity for routine monitoring. However, 15-minute data should be used cautiously with robust filtering mechanisms due to noise susceptibility. A tiered detection method can be developed for day traders and investors in the cryptocurrency market by combining DBSCAN's precision with the recall offered by other method and also by incorporating market context (e.g., news events, volume spikes) to validate detected anomalies for policy makers. There is great need for surveillance on Tron, Litecoin and Stellar as they have more anomalies as compare to other cryptocurrencies.

Future research should investigate hybrid approaches combining the strengths of different algorithms and explore adaptive time-window strategies that adjust to market volatility regimes, which will help to develop cryptocurrency-specific anomaly detection as well as investment strategies, which formulated on anomalies detected. Practitioners should match detection strategies to specific use cases, recognizing that algorithmic choices fundamentally shape surveillance outcomes in this complex, evolving markets.

Research Funding

The authors received no research grant or support for this research study.

REFERENCES

Agyemang, E. F. (2024). Anomaly detection using unsupervised machine learning algorithms: A simulation study. *Scientific African*, 26, e02386. https://doi.org/10.1016/j.sciaf.2024.e02386

Ananias, P. H. M., & and Negri, R. G. (2021). Anomalous behaviour detection using one-class support vector machine and remote sensing images: A case study of algal bloom occurrence in inland waters. *International Journal of Digital Earth, 14*(7), 921-942. https://doi.org/10.1080/17538947.2021.1907462

Aysan, A. F., Demir, E., Gozgor, G., & Lau, C. K. M. (2019). Effects of the geopolitical risks on Bitcoin returns and volatility. *Research in International Business and Finance, 47*, 511-518. https://doi.org/10.1016/j.ribaf.2018.09.011

Başçı, S., & Khan, A. U. I. (2023). Detecting Unknown Change Points for Heteroskedastic Data. *Dokuz Eylül Üniversitesi İşletme Fakültesi Dergisi, 24*(2), 81-98. https://doi.org/10.24889/ifede.1300907

Bhatia, P., & Jain, L. (2025). Lawful Sequence of Events and Cryptocurrency Anomalies: An Empirical Investigation. *FIIB Business Review, 14*(1), 71-88. https://doi.org/10.1177/23197145211042438

Bielecki, C. (2023). Anomaly Detection for Galactic Archaeology: Unveiling Stellar Streams with Machine Learning [Université de Montréal]. https://umontreal.scholaris.ca/server/api/core/bitstreams/b908a503-

Blum, M., & Gelfman, L. P. (2023). Visualizing Multimorbidity in Chronically Ill Populations Using UpSet Plots. *Journal of Pain and Symptom Management, 65*(4), e397-e398. https://doi.org/10.1016/j.jpainsymman.2022.12.003

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000a). LOF: Identifying density-based local outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 93-104. https://doi.org/10.1145/342009.335388

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000b). LOF: Identifying density-based local outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 93-104. https://doi.org/10.1145/342009.335388

Cholevas, C., Angeli, E., Sereti, Z., Mavrikos, E., & Tsekouras, G. E. (2024). Anomaly Detection in Blockchain Networks Using Unsupervised Learning: A Survey. *Algorithms,* 17(5), 201. https://doi.org/10.3390/a17050201

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis, 62*, 182-199. https://doi.org/10.1016/j.irfa.2018.09.003

Ehsan, A., Iqbal, Z., Abuowaida, S., Aljaidi, M., Usman Zia, H., Alshdaifat, N., & Khalf Alshammry, N. (2024). Enhanced Anomaly Detection in Ethereum: Unveiling and Classifying Threats with Machine Learning. *IEEE Access, 12*, 176440-176456. https://doi.org/10.1109/ACCESS.2024.3504300

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. (2002). A Geometric Framework for Unsupervised Anomaly Detection. In D. Barbará & S. Jajodia (Eds.), Applications of Data Mining in Computer Security (pp. 77-101). Springer US. https://doi.org/10.1007/978-1-4615-0953-0_4

Ester, M., Kriegel, H.-P., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *In Kdd Proceeding, 96*(34), 226-231.

Fahim, A. (2018). Homogeneous Densities Clustering Algorithm. *International Journal of Information Technology and Computer Science, 10*(10), 1-10. https://doi.org/10.5815/ijitcs.2018.10.01

Fama, E. F. (1970). Efficient capital markets. *Journal of Finance, 25*(2), 383-417. https://doi.org/10.2307/2325486

Fan, F., Carmine, V., Michail, B., Kanthan, L., Martinez-Rego, D., Wu, F., & Li, L. (2022). Cryptocurrency trading: A comprehensive survey. *Financial Innovation, 8*(1). https://doi.org/10.1186/s40854-021-00321-6

Golnari, A., Komeili, M. H., & Azizi, Z. (2024). Probabilistic deep learning and transfer learning for robust cryptocurrency price prediction. *Expert Systems with Applications, 255*, 124404. https://doi.org/10.1016/j.eswa.2024.124404

Hisham, S., Makhtar, M., & Aziz, A. A. (2023). Anomaly detection in smart contracts based on optimal relevance hybrid features analysis in the Ethereum blockchain employing ensemble learning. *International Journal of Advanced Technology and Engineering Exploration, 10*(109), 1552. https://doi.org/10.19101/IJATEE.2023.10102216

Kaleem, M., Jusoh, H. bin, Raza, H., Sadiq, M., & Hamzah, A. H. bin. (2024). A Machine Learning Approach to Predict Bankruptcy in Chinese Companies with ESG Integration. *Pakistan Journal of Commerce and Social Sciences, 18*(2), 335-357. https://doi.org/10.64534/Commer.2024.043

Kampers, O., Qahtan, A., Mathur, S., & Velegrakis, Y. (2022). Manipulation detection in cryptocurrency markets: An anomaly and change detection based approach. Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, 326-329. https://doi.org/10.1145/3477314.3507185

Khan, A. U. I., Ozcan, R., & Ibrahim, M. M. (2024). Unravelling Crash Risk Transmission: Cryptocurrency Impact on Stock Markets in G-7 and China. *Pakistan Journal of Commerce and Social Sciences, 18*(4), 848-871. https://doi.org/10.64534/Comm.2025.019

Kim, K. G. (2016). Book Review: Deep Learning. *Healthcare Informatics Research, 22*(4), 351. https://doi.org/10.4258/hir.2016.22.4.351

Kumar, A., Kumar, A., Bashir, A. K., Rashid, M., Kumar, V. D. A., & Kharel, R. (2021). Distance Based Pattern Driven Mining for Outlier Detection in High Dimensional Big Dataset. *ACM Transactions on Management Information System, 13*(1), 8:1-8:17. https://doi.org/10.1145/3469891

Kyriazis, N. A. (2021). A Survey on Volatility Fluctuations in the Decentralized Cryptocurrency Financial Assets. *Journal of Risk and Financial Management, 14*(7), 293. https://doi.org/10.3390/jrfm14070293

Latif, M. N., Kaplan, M., & Khan, A. U. I. (2025). Analyzing Anomalies for Financial Fraud Detection: A Case Study of Selected Insurance Companies Listed in Borsa Istanbul. *FWU Journal of Social Sciences, 19*(2). http://doi.org/10.51709/19951272

Liu, H., Zhao, B., Guo, J., Zhang, K., & Liu, P. (2024). A lightweight unsupervised adversarial detector based on autoencoder and isolation forest. *Pattern Recognition, 147*, 110127. https://doi.org/10.1016/j.patcog.2023.110127

McLeay, S. (1986). Student's T and the Distribution of Financial Ratios. *Journal of Business Finance & Accounting, 13*(2), 209-222. https://doi.org/10.1111/j.1468-5957.1986.tb00091.x

Naz, F., Sayyed, M., Rehman, R.-U.-, Naseem, M. A., Abdullah, S. N., & Ahmad, M. I. (2023). Calendar anomalies and market volatility in selected cryptocurrencies. *Cogent*

Business & Management, 10(1). https://doi.org/10.1080/23311975.2023.2171992

Patel, V., Pan, L., & Rajasegarar, S. (2020). Graph Deep Learning Based Anomaly Detection in Ethereum Blockchain Network. In M. Kutyłowski, J. Zhang, & C. Chen (Eds.), Network and System Security (pp. 132-148). Springer International Publishing. https://doi.org/10.1007/978-3-030-65745-1_8

Pérez-Cano, V., & Jurado, F. (2025). Fraud Detection in Cryptocurrency Networks-An Exploration Using Anomaly Detection and Heterogeneous Graph Transformers. *Future Internet, 17*(1), 44. https://doi.org/10.3390/fi17010044

Poutré, C., Chételat, D., & Morales, M. (2024). Deep unsupervised anomaly detection in high-frequency markets. *The Journal of Finance and Data Science, 10*, 100129. https://doi.org/10.1016/j.jfds.2024.100129

Rezapour Mashhadi, M. M. (2019). Anomaly Detection using Unsupervised Methods: Credit Card Fraud Case Study. *International Journal of Advanced Computer Science and Applications, 10*(11). https://doi.org/10.14569/IJACSA.2019.0101101

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the Support of a High-Dimensional Distribution. *Neural Computation, 13*(7), 1443-1471. https://doi.org/10.1162/089976601750264965

Shi, F.-B., Sun, X.-Q., Gao, J.-H., Xu, L., Shen, H.-W., & Cheng, X.-Q. (2019). Anomaly detection in Bitcoin market via price return analysis. *PLOS ONE*, 14(6), e0218341. https://doi.org/10.1371/journal.pone.0218341

Urquhart, A., & Yarovaya, L. (2024). Cryptocurrency research: Future directions. *The European Journal of Finance, 30*(16), 1849-1854. https://doi.org/10.1080/1351847X.2023.2284186

Wang, H., Zheng, J., Carvajal-Roca, I. E., Chen, L., & Bai, M. (2023). Financial Fraud Detection Based on Deep Learning: Towards Large-Scale Pre-training Transformer Models. In H. Wang, X. Han, M. Liu, G. Cheng, Y. Liu, & N. Zhang (Eds.), Knowledge Graph and Semantic Computing: Knowledge Graph Empowers Artificial General Intelligence (pp. 163-177). Springer Nature. https://doi.org/10.1007/978-981-99-7224-1_13

Witayanont, Y., & Viyanon, W. (2025). Anomaly Detection in Bitcoin Network: Using Distance-based and Tree-based Unsupervised Learning Methods. Proceedings of the 6th ACM International Symposium on Blockchain and Secure Critical Infrastructure, 1-7. https://doi.org/10.1145/3659463.3660022

Yahia, A., Mouhssine, Y., El Alaoui, A., & El Alaoui, S. O. (2024). Exploring machine learning-based methods for anomalies detection: Evidence from cryptocurrencies. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-024-00703-w

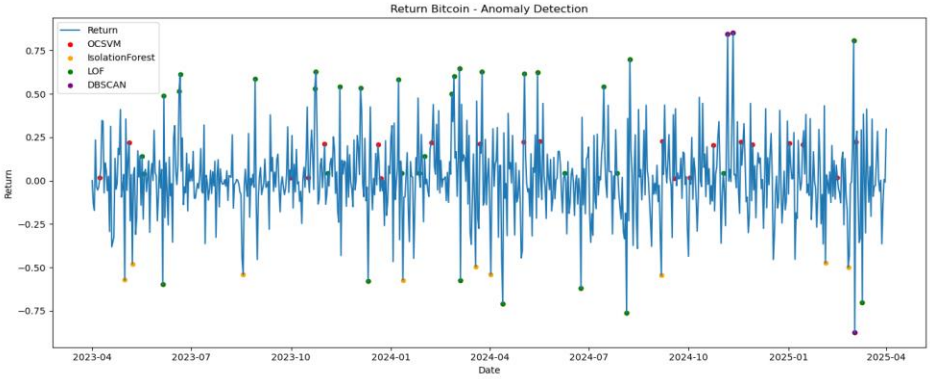
Yahia, A., Mouhssine, Y., Elalaoui, A., & Ouatik Elalaoui, S. (2024). Leveraging Machine Learning for Anomaly Detection Methods in Cryptocurrency: A Data-Driven Study. 2024 10th International Conference on Optimization and Applications (ICOA), 1-5. https://doi.org/10.1109/ICOA62581.2024.10754457

Yang, K., Kpotufe, S., & Feamster, N. (2021). An Efficient One-Class SVM for Anomaly Detection in the Internet of Things (No. arXiv:2104.11146). arXiv.

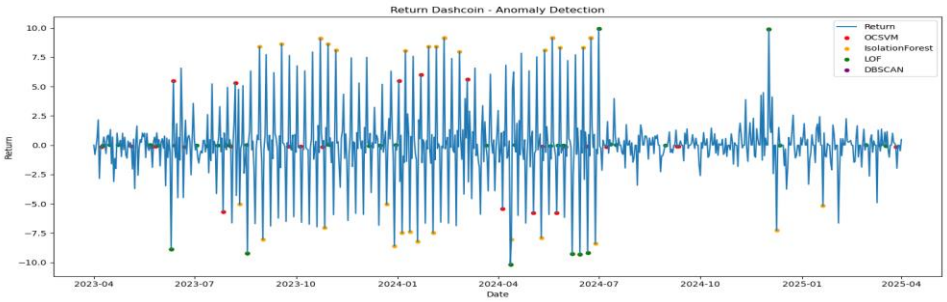
Zhu, Y., Dickinson, D., & Li, J. (2017). Analysis on the influence factors of Bitcoin's price based on VEC model. *Financial Innovation, 3*(1), 3. https://doi.org/10.1186/s40854-017-0054-0

Zou, D., Xiang, Y., Zhou, T., Peng, Q., Dai, W., Hong, Z., Shi, Y., Wang, S., Yin, J., & Quan, H. (2023). Outlier detection and data filling based on KNN and LOF for power transformer operation data classification. *Energy Reports, 9*, 698-711. https://doi.org/10.1016/j.egyr.2023.04.094

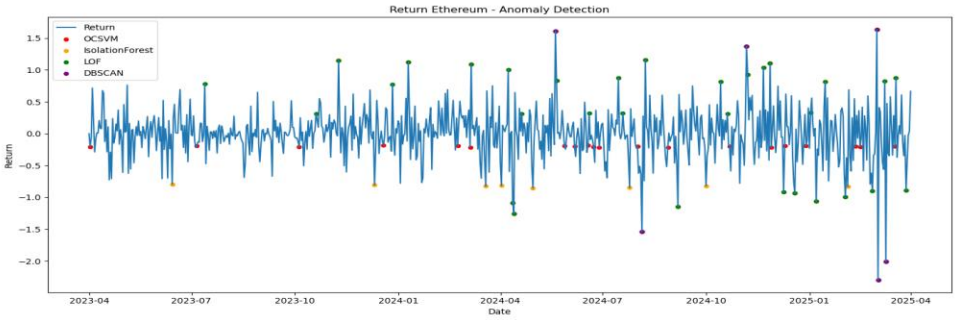
APPENDIX



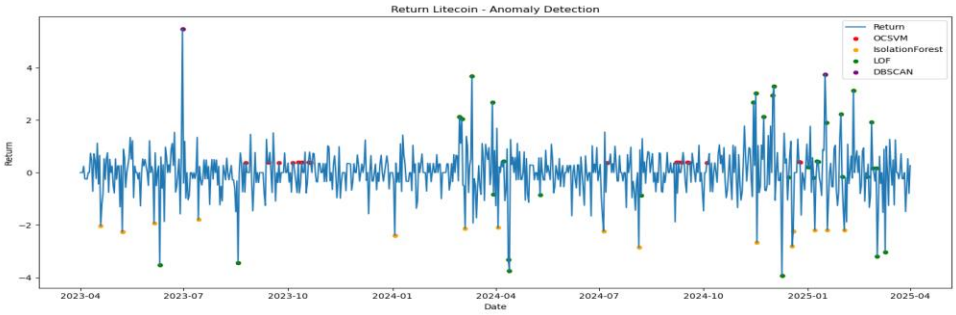
A-Figure 1: Anomaly detection on Bitcoin daily data return using four Unsupervised Machine-Learning Models



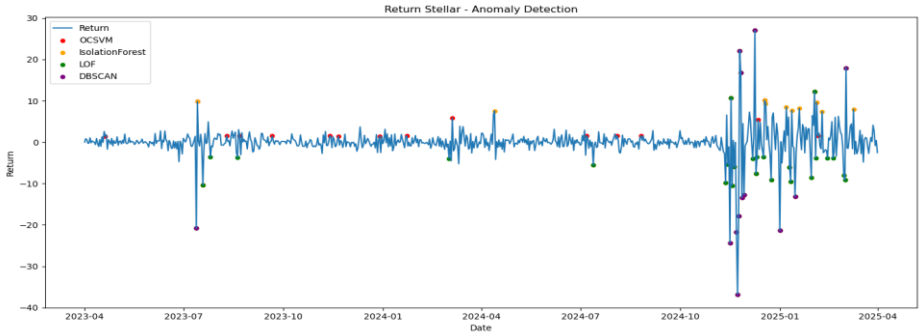
A-Figure 2: Anomaly detection on Dashcoin daily data return by using four different unsupervised machine-learning models



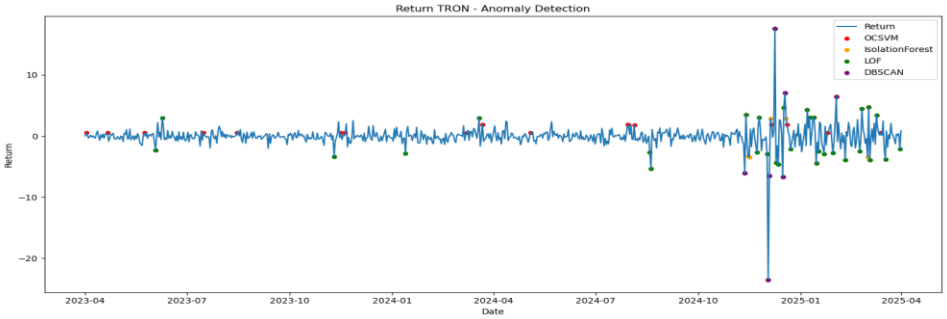
A-Figure 3: Anomaly detection on Ethereum daily data return by using four different unsupervised machine-learning models



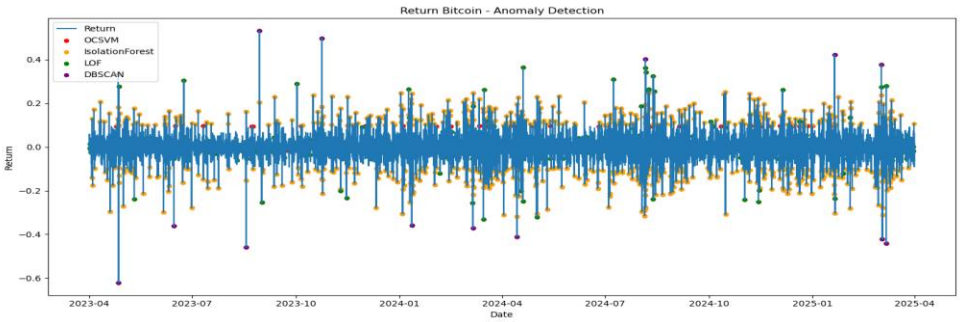
A-Figure 4: Anomaly detection on Litecoin daily data return by using four different unsupervised machine-learning models



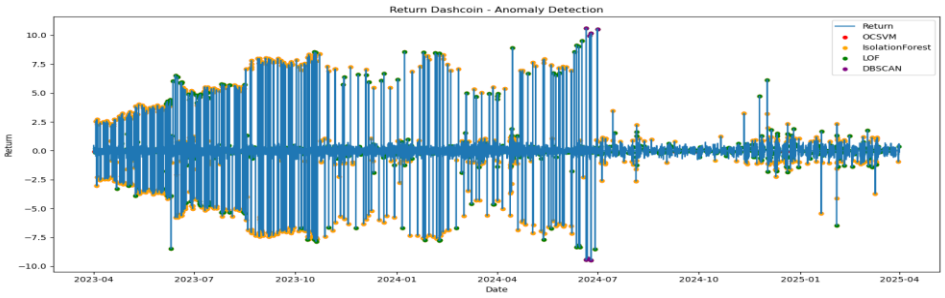
A-Figure 5: Anomaly detection on Stellar daily data return by using four different unsupervised machine-learning models



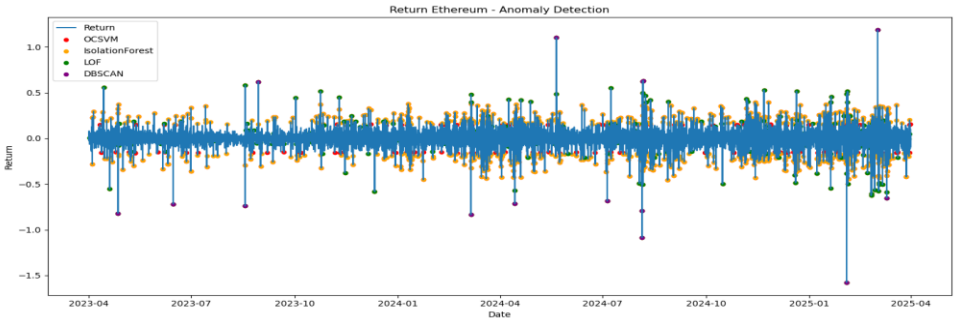
A-Figure 6: Anomaly detection on Tron daily data return by using four different unsupervised machine-learning models



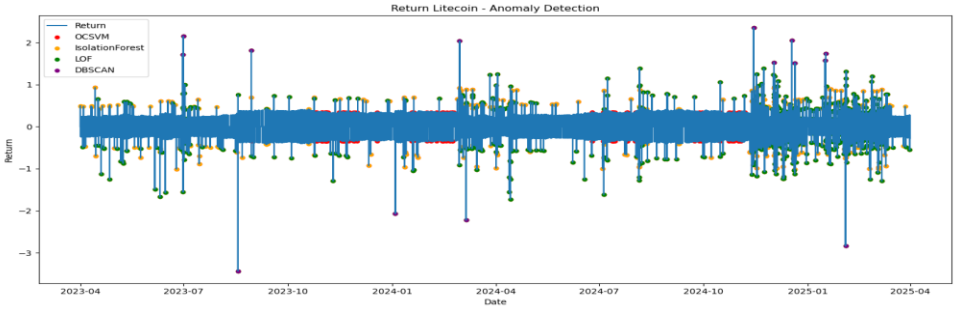
A-Figure 7: Anomaly detection on Bitcoin hourly data return by using four different unsupervised machine-learning models



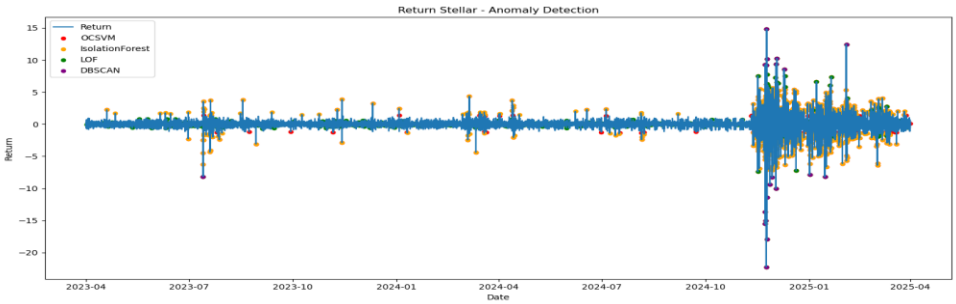
A-Figure 8: Anomaly detection on Dashcoin hourly data return by using four different unsupervised machine-learning models



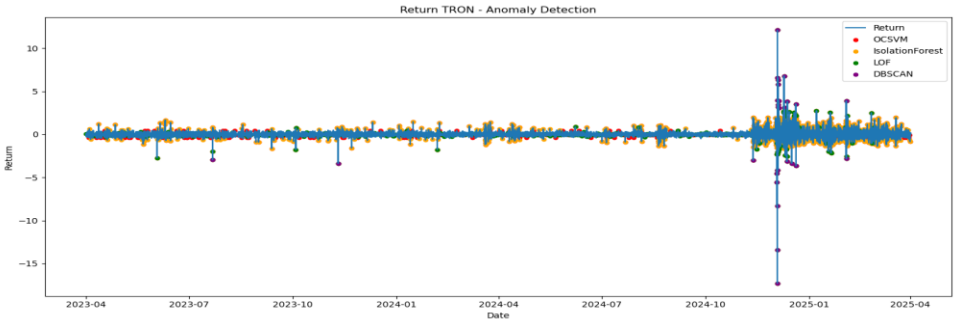
A-Figure 9: Anomaly detection on Ethereum hourly data return by using four different unsupervised machine-learning models



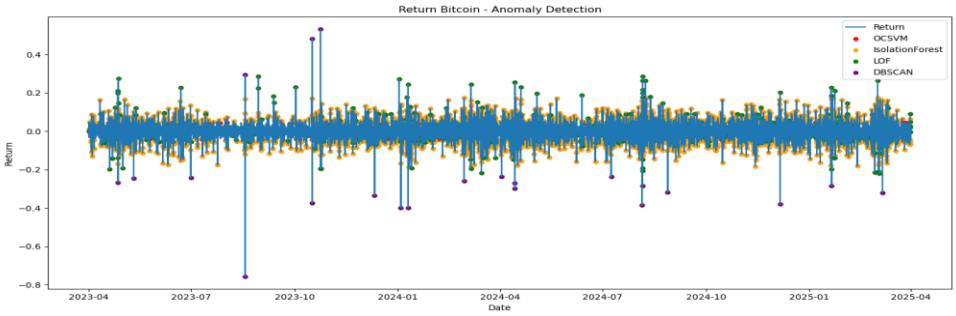
A-Figure 10: Anomaly detection on Litecoin hourly data return by using four different unsupervised machine-learning models



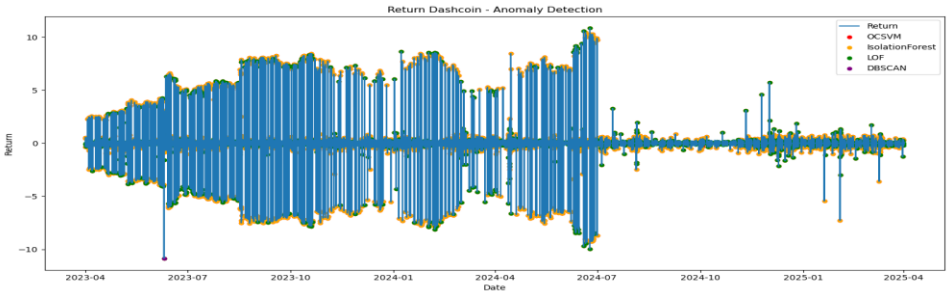
A-Figure 11: Anomaly detection on Stellar hourly data return by using four different unsupervised machine-learning models



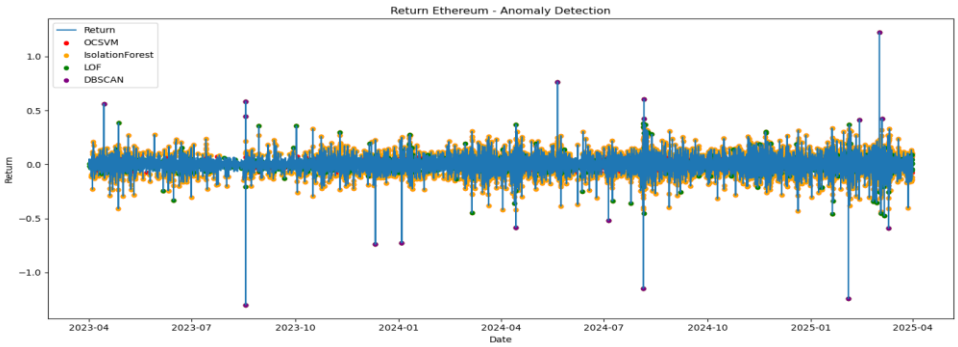
A-Figure 12: Anomaly detection on TRON hourly data return by using four different unsupervised machine-learning models



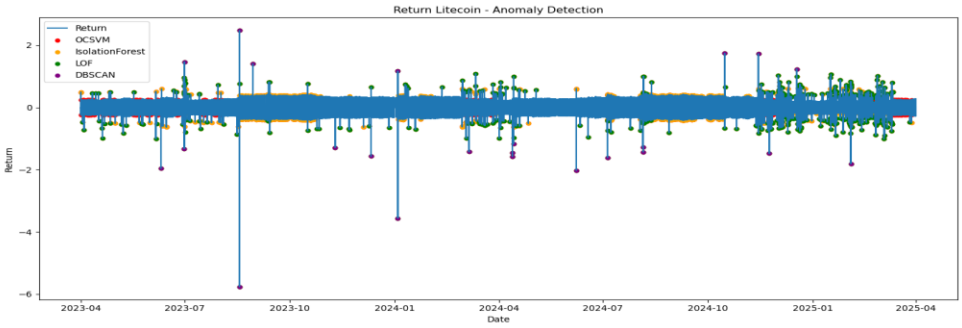
A-Figure 13: Anomaly detection on Bitcoin 15-minutes interval data return by using four different unsupervised machine-learning models



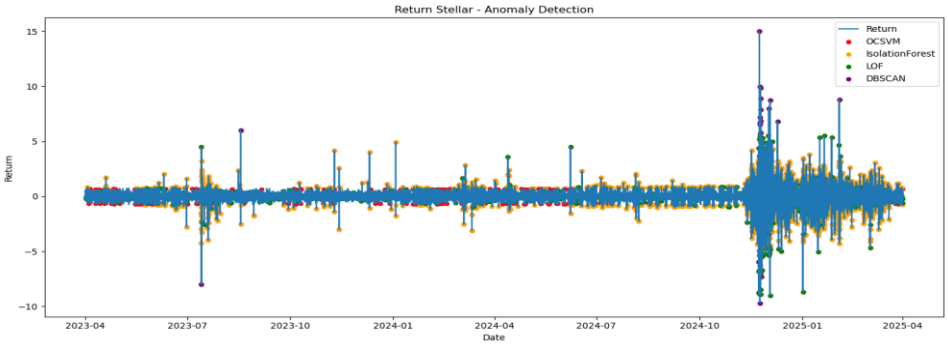
A-Figure 14: Anomaly detection on Dashcoin 15-minutes interval data return by using four different unsupervised machine-learning models



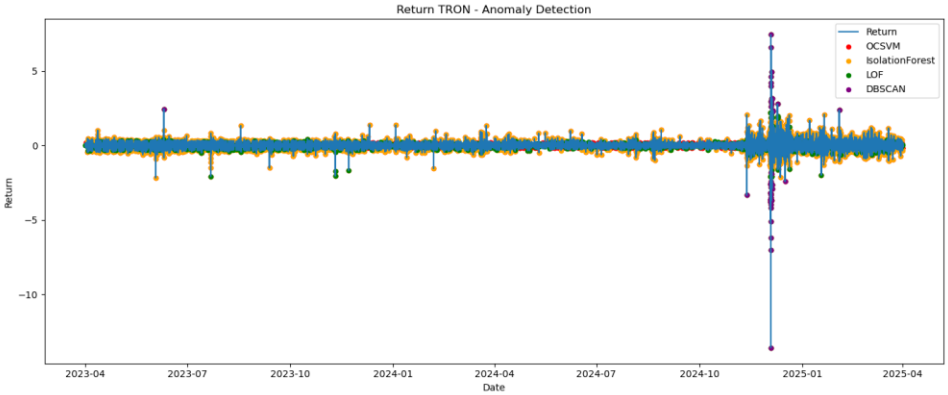
A-Figure 15: Anomaly detection on Ethereum 15-minutes interval data return by using four different unsupervised machine-learning models



A-Figure 16: Anomaly detection on Litecoin 15-minutes interval data return by using four different unsupervised machine-learning models



A-Figure 17: Anomaly detection on Stellar 15-minutes interval data return by using four different unsupervised machine-learning models



A-Figure 18: Anomaly detection on TRON 15-minutes interval data return by using four different unsupervised machine-learning models