Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, S. 436-443, Nisan 2020

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue, pp. 436-443, April 2020

Research Article

Anomaly Detection in Bitcoin Prices using DBSCAN Algorithm**

Ahmet Şakir Dokuz [1*], Mete Çelik[2], Alper Ecemiş[3]

[1] Niğde Ömer Halisdemir Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Niğde, Türkiye (ORCID: 0000-0002-1775-0954)
[2] Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye (ORCID: 0000-0002-1488-1502)
[3] Niğde Ömer Halisdemir Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Niğde, Türkiye (ORCID: 0000-0001-5455-0006)

(Conference Date: 5-7 March 2020)

(**DOI:** 10.31590/ejosat.araconf57)

ATIF/REFERENCE: Dokuz, A. Ş., Çelik, M. & Ecemiş, A. (2020). Anomaly Detection in Bitcoin Prices using DBSCAN Algorithm. *Avrupa Bilim ve Teknoloji Dergisi*, (Special Issue), 436-443.

Abstract

Blockchain is an emerging technology which is also behind the Bitcoin digital money. Daily bitcoin transactions are increasing due to the popular and widespread investments. The increase of Bitcoin related datasets and this increased big dataset requires novel approaches and methods to analyze using data mining techniques. In addition, fluctuations and anomalies in the bitcoin prices could mean a great deal to economists and discovering anomalies in bitcoin prices is important. In this study, anomaly detection in Bitcoin prices is performed based on the change of Bitcoin price difference and the change of Bitcoin price difference in percentage with respect to previous day using 8-years of Bitcoin price dataset of the period of 2012-2019. First, the dataset is pre-processed and unnecessary columns are deleted. Then, 2 different datasets are created by using daily bitcoin prices, i.e. bitcoin price difference dataset and bitcoin price difference in percentage dataset. After that, for detecting anomalous price changes, DBSCAN algorithm and statistical method are used, and the performance of the algorithms are evaluated. The results show that the DBSCAN algorithm and statistical method successfully detects anomalies in bitcoin prices for both of the datasets. However, the DBSCAN algorithm performs better than the statistical method which could detect anomalies even they are close to the normal daily price changes. Also, in this study, bitcoin price difference dataset and bitcoin price difference in percentage dataset are compared and the differences of the results for both datasets and their reasons are explained.

Keywords: Bitcoin price, Blockchain, Data mining, Anomaly detection, DBSCAN Algorithm, statistical approach

DBSCAN Algoritması Kullanarak Bitcoin Fiyatlarında Anormallik Tespiti

Öz

Blokzincir, bitcoin dijital para biriminin de alt yapısını oluşturan yeni bir teknolojidir. Popüler ve yaygın yatırımlar sayesinde günlük gerçekleştirilen bitcoin işlem sayısı gün geçtikçe artmaktadır. Bitcoin verisi her geçen gün artmakta ve dolayısıyla artan büyük bitcoin verisinin analizi ve madenciliği için yeni veri madenciliği yöntemlerine ihtiyaç duyulmaktadır. Buna ek olarak, Bitcoin fiyatındaki dalgalanmalar ve anormal fiyat değişimleri ve bu değişimlerdeki anormalliklerin keşfi ekonomistler için büyük önem taşımaktadır. Bu çalışmada, 2012-2019 yıllarına ait 8 yıllık bitcoin fiyat veri kümesi kullanılarak bitcoin fiyat farkı ve bitcoin fiyatı yüzdesel farkı olmak üzere iki farklı veri kümesi oluşturulup, anormallik tespiti gerçekleştirilmiştir. Öncelikle veri kümesi ön işlem aşamasından geçirilerek gereksiz sütunlar çıkarılmıştır ve daha sonra günlük fiyat farkları kullanılarak veri setleri oluşturulup, DBSCAN algoritması ile anormallik tespiti yapılmıştır. Ayrıca bu çalışmada DBSCAN aloritmasının sonuçları istatistiksel yöntemin sonuçları ile karşılaştırılıp, tartışılmıştır. Sonuçlar incelendiğinde, DBSCAN algoritması ve istatistiksel metodun bitcoin fiyatlarındaki anormallikleri her iki veri kümesinde de başarıyla tespit edebildiği görülmüştür. Bununla birlikte DBSCAN algoritması normal günlük fiyat değişimlerine yakın olan anormak fiyat değişimlerini de keşfedebildiği için istatistiksel metottan daha iyi performans

* *Corresponding Author*: Niğde Ömer Halisdemir Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendsiliği Bölümü, Niğde, Türkiye, ORCID: 0000-0002-1775-0954

This paper was presented at the ***International Conference on Access to Recent Advances in Engineering and Digitalization (ARACONF 2020)*.

göstermiştir. Ayrıca, bu çalışmada bitcoin fiyat farkı veri kümesi ve bitcoin fiyatı yüzdesel farlı veri kümesi karşılaştırılmış ve her bir veri kümesi için olan sonuçlar ve sebepleri tartışılmıştır.

Anahtar Kelimeler: Bitcoin fiyatı, Blokzincir, Veri madenciliği, Anormallik tespiti, DBSCAN algoritması, İstatistiksel yöntem

1. Giriş

Blockchain is a network protocol that provides transfer of assets online without requiring a mediator, such as banks (Melanie, 2017). Bitcoin is an electronic money which is based on blockchain technology and is proposed by Satoshi Nakamoto in 2008 (Nakamoto, 2008).

Bitcoin gained attention from financial and economics domains due to its ability to transform current payment and money systems (Böhme et al., 2015). The increase in bitcoin transactions led the increase of bitcoin related datasets. Analysis of such datasets requires data analysis and data mining techniques.

However, bitcoin has several challenges. First of all, due to the decentralized nature and anonymity of bitcoin, there is a challenge against illegal transactions and cyber attacks (Sas & Mara, 2017). In addition, the accumulation of data from day to day creates the problem of scalability. Also, fluctuations in bitcoin prices complicate the future bitcoin price prediction and create abnormal situations.

Anomaly detection is the process of discovering instances in the datasets which are different than the rest of the instances (Agrawal & Agrawal, 2015). Anomaly detection provides information about unexpected behaviors of people, devices, and data-related objects. In this study, anomaly detection has been used to get insight about the fluctuations of bitcoin prices. The prices of bitcoin could change dramatically because there is no stable and protected market for bitcoin. Because of this reason, bitcoin can be seen as illegal or treated as unreliable investment tool (Id et al., 2019).

In the literature, Id et al. performed anomaly detection in five biggest digital money platforms to observe abnormal activities and resulted that the anomalous activities could be linked with price manipulation and money laundering (Id et al., 2019). Dokuz et al. analyzed daily, hourly, and monthly aspects of bitcoin data (Dokuz et al., 2019). Thai and Lee detected suspected users and transactions in bitcoin dataset using machine learning methods (Lee & Edu, 2016). Monamo et al. detected fraud activities in bitcoin transactions using k-means 2clustering algorithm (Monamo et al., 2016). Bartoletti et al. used machine learning algorithms for detection of bitcoin addresses that are related to Ponzi schemes (Bartoletti et al., 2018). Sayadı et al. performed anomaly detection in bitcoin electronic transactions using machine learning algorithms (Sayadi, Rejeb, & Choukair, 2019). In the study, one-class SVM algorithm and k-means clustering algorithm are used for anomaly detection. Baek et al. detected DDOS attacks in 22 network using deep learning methods, and the authors claim that this method could be applied to other blockchain systems (Baek et al., 2019).

In this study, anomaly detection in bitcoin prices is performed using two methods, namely DBSCAN algorithm and statistical method. The selected bitcoin price dataset is 8-years dataset and belongs to the period of 2012-2019. First, the dataset is pre-processed and unnecessary columns are deleted. After that, for detecting anomalous daily changes, DBSCAN algorithm and statistical method are used, and then the performances of the algorithms are evaluated.

The rest of this study is organized as follows. Section 2 presents the dataset of this study, and introduces DBSCAN and statistical anomaly detection methods. Section 3 presents the results of the used methods. Section 4 presents the conclusions.

2. Material and Method

In this section, first bitcoin price dataset which is used in this study is explained, and then DBSCAN and statistical anomaly detection methods are introduced.

2.1. Bitcoin Price Dataset

In this study, 8-years of daily bitcoin price dataset which is belong to the period of January 2012 to December 2019 (Investing, 2020). The original dataset has 2921 instances and contains 7 columns, such as, date, now, open, high, low prices, volume, and change percentage. In this study, we only used daily price information and so other columns are removed from the dataset.

The daily price values of bitcoin between 2012-2019 is shown in Figure 1. As can be seen in the figure, bitcoin prices have fluctuations, especially in the years of 2018 and 2019. It has reached its highest values in 2018 and it is also incresed in 2019

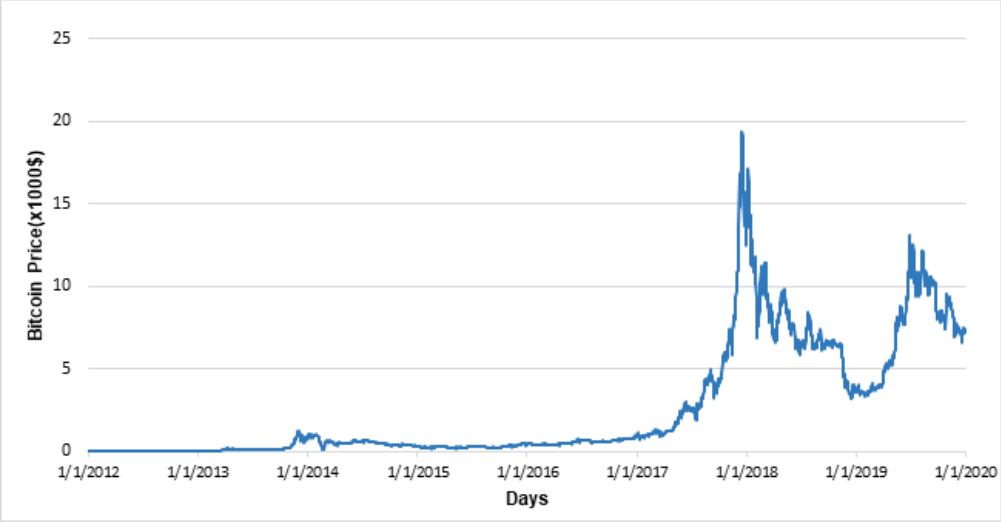

Fig. 1 Bitcoin price ($) between 2012-2019

In this study, bitcoin price data is prepared in 2 different ways, i.e. daily bitcoin price difference and daily bitcoin price in percentage difference, and analyses are performed. The daily bitcoin price difference dataset is prepared using the bitcoin price difference between current and previous day as shown in Equation (1). The daily bitcoin price in percentage difference dataset is prepared using the bitcoin price difference between current and previous day divided by bitcoin price for current day as shown in Equation (2).

$$z_i = x_i - x_{i-1} \tag{1}$$

$$q_i = \frac{x_i - x_{i-1}}{x_i} \tag{2}$$

In the equations, z_i presents daily bitcoin price difference on ith day, q_i presents daily bitcoin price difference in percentage for the ith day, x_i presents bitcoin price on the ith day and x_{i-1} presents the bitcoin price on the i-1th day.

2.2. DBSCAN Algorithm

DBSCAN is a density-based clustering algorithm (Ester et al., 1996). DBSCAN is a clustering algorithm. However, it successfully discovers anomalies based on the user given algortihm parameters of neighboring radius *Eps* and minimum number of points *minpts*. (Khan et al., 2014) (Çelik et al., 2011) (Ozekes et al., 2018).

There are three types of points in DBSCAN algorithm to perform clustering and to detect anomalies, i.e. core point, border point, and noise point (Khan et al., 2014). If point p satisfies the minimum minpts threshold within N_{Eps},(Equation 3), point p is called as core point. If point p do not satisfy the minimum minpts threshold within *Eps* distance, however point p directly density-reachable from a core point, then p is called as border point. If point p do not satisfy the minimum minpts threshold within *Eps* distance and if p is not a border point, then p is called as noise or anomaly point (Chen, Gao, & Li, 2010). For any point p, the number of neighboring points are calculated based on Equation 3: In the equation, D is dataset, *Eps* is radius, N_{Eps} is neighbourhood radius and p is an arbitrary point which is selected from the dataset. Dist is usually selected as Euclidean distance.

$$N_{Eps} = \ q \in D \rightarrow dist(p,q) < Eps \tag{3}$$

The anomaly detection method with DBSCAN algorithm which is used in the study is given in Algorithm 1.

Algorithm 1. DBSCAN Anomaly Detection Method

Inputs:
Eps: Distance threshold
MinPts: Minimum neighbor count threshold

Output: Anomalous price differences

1. data = readData()
2. dataset = preprocessingData(data)
3. anomalies = applyDBSCAN(Eps, MinPts, dataset)
4. **return** anomalies

When the steps of the Algorithm 1 are examined, the dataset is read at step 1. At step 2, the raw data is preprocessed using the *preprocessingData* function and unnecessary columns are deleted. At step 3, DBSCAN algorithm is applied on the prepared dataset with *Eps* and *MinPts* threshold parameters and anomalous price differences are detected. At step 4, detected anomalous prices are returned as the algorithm output.

2.3. Statistical Method

The statistical method is based on the assumption that the data are normally distributed. The normal distribution depends on 2 different variables called average (μ) and standard deviation (σ)(Tan et al., 2005). While performing anomaly detection with normal distribution, it is common practice to accept values between $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ as normal and other cases as abnormal (Çelik et al., 2011).

The anomaly detection with statistical method which is used in the study is given in Algorithm 2.

Algorithm 2. Statistical Anomaly Detection Method

Inputs:
μ: Mean
σ: Standard deviation
β: σ multiplier coefficient

Output: Anomalous price differences

```
1. data = readData()
2. dataset = preprocessingData(data)
3. upperlimit = μ + β σ,  lowerlimit = μ − β σ
4. for x:each instance from dataset
5.     if ( lowerlimit > x || x > upperlimit )
6.         anomalies.add(x)
7. return anomalies
```

When the steps of Algorithm 2 are examined, the dataset is read at step 1. At step 2, the raw data is preprocessed using the *preprocessingData* function and unnecessary columns are deleted. At step 3, upper and lower limits of the dataset are determined for anomaly detection. In steps 4, 5 and 6, each data in the dataset is compared with the upper and lower limits, and the values that are not between these two limits are detected abnormally. At step 7, detected anomalous prices are returned as the algorithm output.

3. Results and Discussion

This section presents the experimental results of this study. Experiments are carried out using 2 different datasets as presented in Section 2.1. These datasets are evaluated by statistical method and DBSCAN algorithm, and the results are evaluated.

3.1. Experiments on Bitcoin Price Difference in Percentage

The purpose of the bitcoin price percentage difference experiments is to evaluate the changes in bitcoin prices in percentage and to discover anomalies in the dataset. The average (μ) is calculated as 0.00449 and the standard deviation (σ) is calculated as 0.082298 for the statistical method. In addition, $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ are selected for the detection of anomalous price differences in percentage and the anomalies are presented in Figure 3.

When the results in Figure 3 are analyzed, it is observed that as the standard deviation coefficient increases, there is a decrease in the number of abnormal points detected. In addition, when looking at the results of $\mu \pm \sigma$, it is observed that the values that should be normal are detected abnormally. In $\mu \pm 3\sigma$ results, it is seen that abnormal points are overlooked and detected normally. As a result, it can be said that the most efficient abnormality detection for analytical experiments is carried out in the range of $\mu \pm 2\sigma$.

When the daily bitcoin price in percentage difference data is analyzed, it is observed that the largest percentage change is realized on 24.02.2014 with a daily difference of 3.36% and the second biggest percentage change is on 20.02.2014 with a rate of 1.29%. In this case, the closeness of the dates with high percentage change draws attention. In this context, when the fluctuations of bitcoin price data in Figure 1 in February 2014 are analyzed, it is seen that there is a significant decrease in bitcoin price. When the reason for the price decrease is analyzed, it is known that Mt.Gox company, which dominated the entire bitcoin market at that time, was hacked and the statement made accordingly (Blockonomi news, 2019).

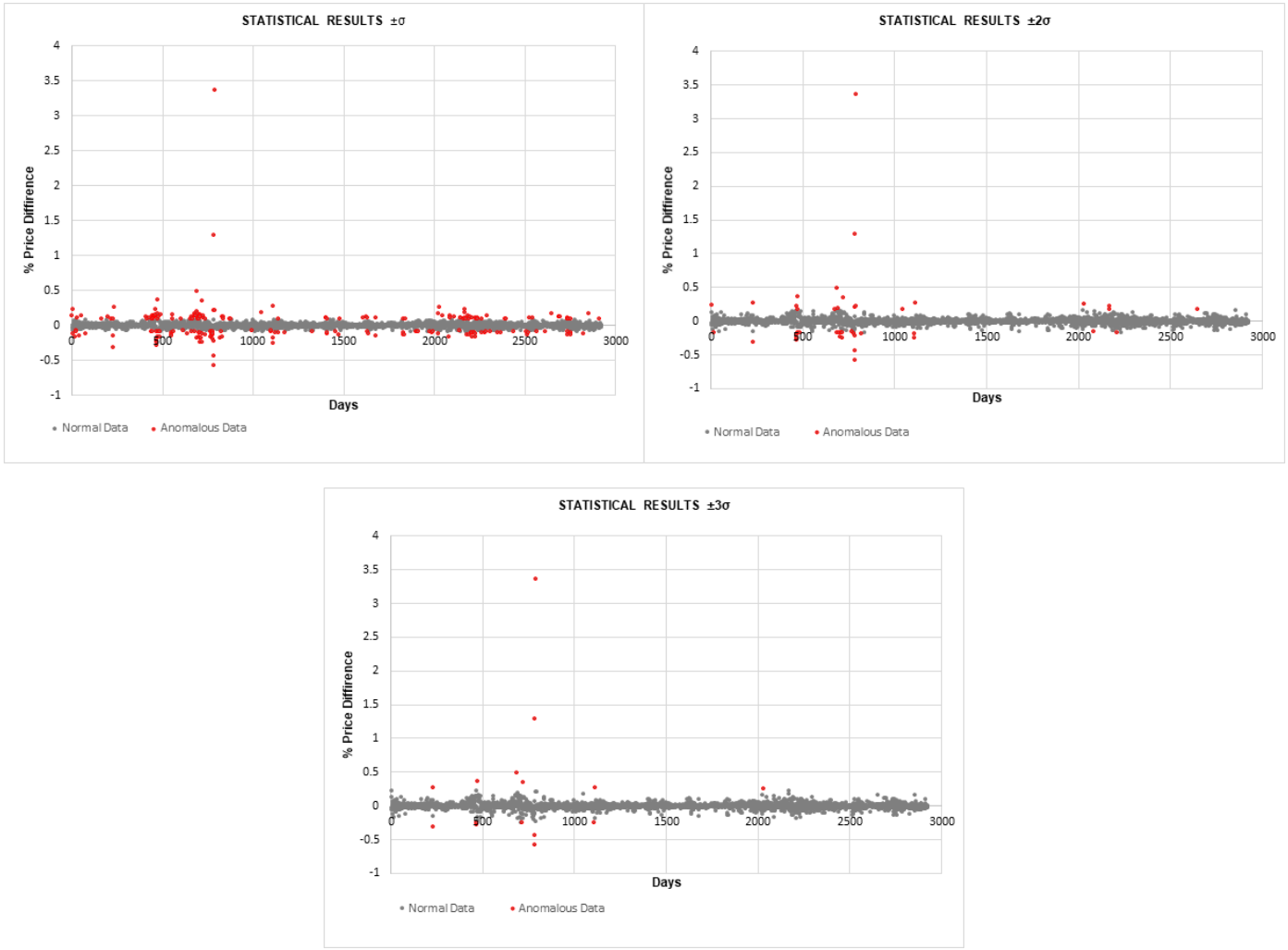

Fig. 3 Statistical results with daily bitcoin price in percentage difference dataset for ± σ, 2σ and 3σ (Red color represent anomalies)

In the experiment carried out using the DBSCAN algorithm with the daily bitcoin price in percentage difference dataset, it is determined that the most efficient result is obtained when the parameters are *Eps*=0.16 and *MinPts*=20, and the test results are given in Figure 4. As can be seen in Figure 4, the DBSCAN algorithm has successfully detected abnormal points.

When DBSCAN results are compared with μ ± 2σ results where the statistical method is the most successful, it is seen that the statistical method overlooked some points such as the point on 23.10.2019. The reason for this can be said that the statistical method distinguishes points in a linear range and the DBSCAN algorithm in a density-based range.

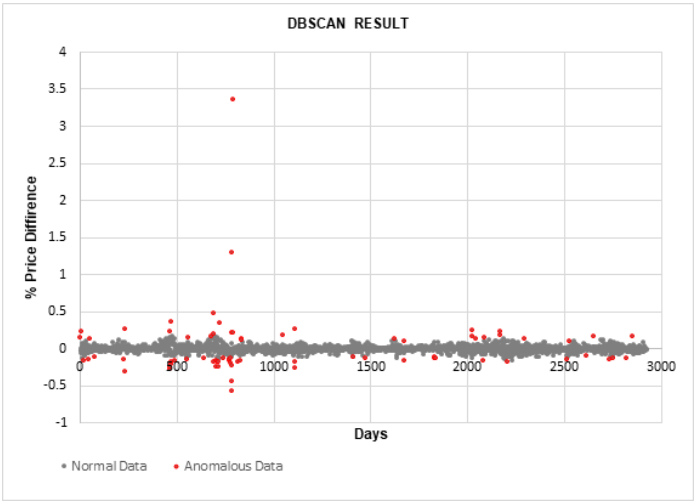

Fig. 4 DBSCAN result with daily bitcoin price in percentage difference dataset

3.2. Fiyat Farkı Deneyleri (Experiments on Bitcoin Price Difference)

The purpose of bitcoin price difference experiments is to determine the daily price changes that will be considered anomalies in bitcoin prices. The μ is calculated as 2557.87 and the σ is calculated as 3611.95 for statistical method. For the detection of abnormalities, $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ parameters are selected and the detected anomalies are presented in Figure 5.

When the results in Figure 5 are analyzed, it is seen that as the standard deviation coefficient increases, there is a decrease in the number of abnormal points detected. However, the most successful result for statistical method appears to be within $\mu \pm 2\sigma$.

In the experiment carried out using the DBSCAN algorithm with the Bitcoin price difference data set, it is determined that the most efficient result is detected when the parameters are Eps 0.03 and MinPts 30, and the test results are given in Figure 6. When the results in Figure 6 are examined, it is possible to say that the DBSCAN algorithm has successfully detected anomalies.

When DBSCAN results are compared with $\mu \pm 2\sigma$ results where the statistical method is the most successful, it is seen that the statistical method overlooked some abnormal points such as points on 09.01.2019, 23.02.2019 and 10.04.2019. The reason for this can be said that the statistical method distinguishes points in a linear range and the DBSCAN algorithm in a density-based range.

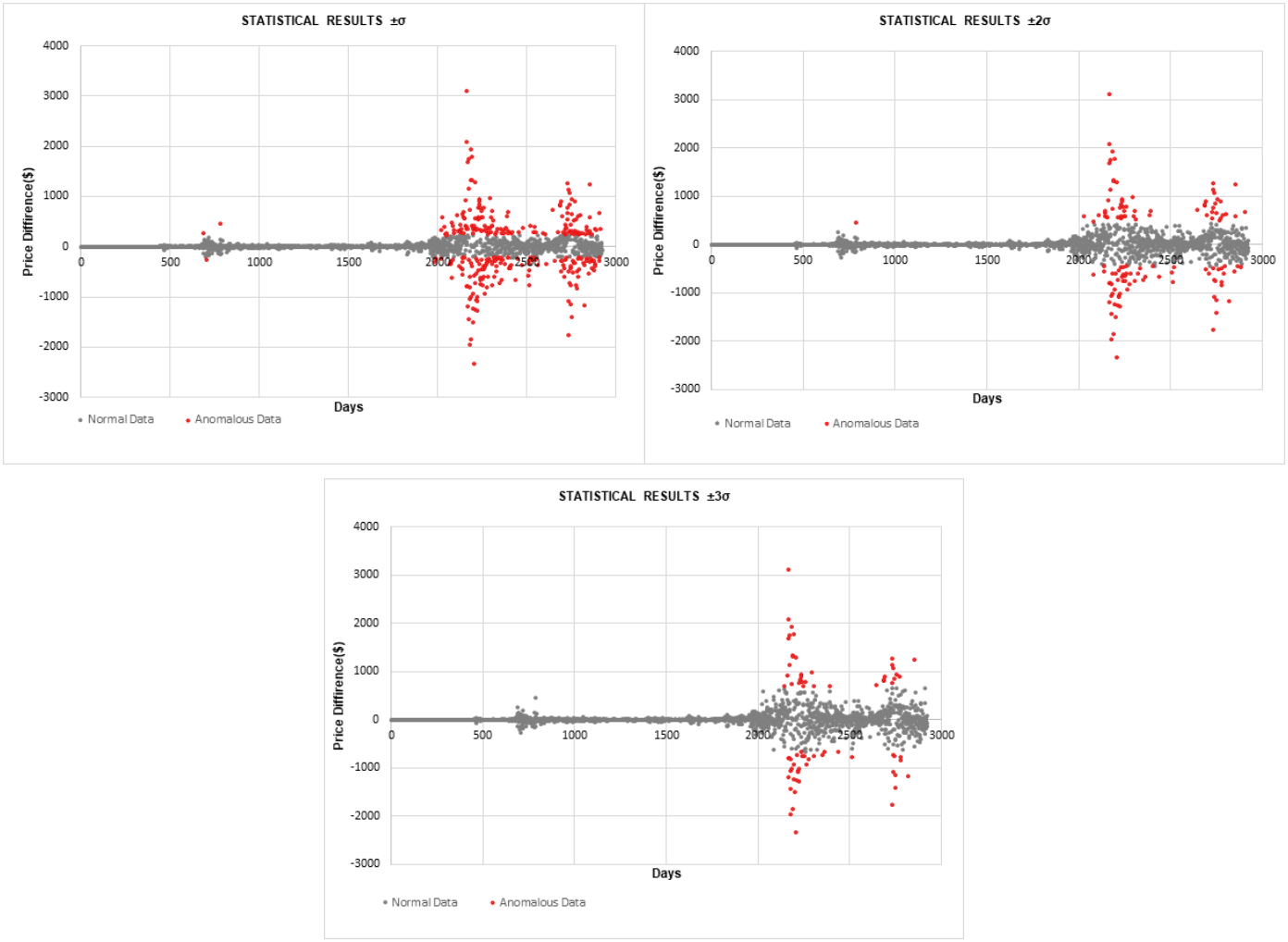

Fig. 5 Statistical results with bitcoin price difference dataset for ± σ, 2σ and 3σ (Red color represent anomalies)

When the price difference dataset characteristic is analyzed, it is seen that the highest daily positive price change is realized with $ 3100.7 on 06.12.2017 and the highest daily negative price change is realized with $ -2335.5 on 15.01.2018. The fact that the highest positive and negative variation is in the close range of dates shows the severity of the bitcoin price fluctuation in this period. It can also be said that the period with the highest fluctuation in prices is between 2018-2019.

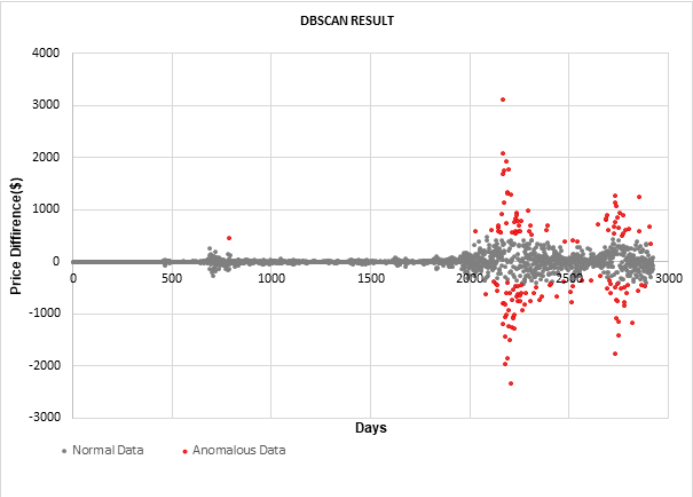

Fig. 6 DBSCAN result with bitcoin price difference dataset

4. Conclusions

Bitcoin is one of the popular digital coins which also have many suspects and manipulations in its prices. The changes in bitcoin prices could be sudden and detection of these anomalous changes could benefit many insights to investors and financial experts. In this study, anomaly detection in bitcoin prices is performed based on the changes of bitcoin prices with respect to previous day using 8-years of bitcoin price dataset of the period of 2012-2019. DBSCAN algorithm and statistical method are used for anomaly detection and the performance of the algorithm is evaluated. The results show that using one-day price difference is beneficial for extracting anomaly detection in bitcoin prices. DBSCAN algorithm had a good performance on detecting anomalous price fluctuations.

When the bitcoin price difference in precentage dataset and the bitcoin price difference dataset are compared, it is seen that abnormal points are concentrated in the bitcoin price difference in precentage in February 2014, and in the bitcoin price difference dataset between 2018 and 2019. Although both datasets are produced from bitcoin price dataset, abnormalities in 2018-2019 cannot be detected properly in the bitcoin price difference in precentage dataset. The reason for this is that the daily price change increases with the increase in bitcoin price value and this is not reflected in the percentage change.

Reference

Agrawal, S., & Agrawal, J. (2015). Survey on Anomaly Detection using Data Mining Techniques. *Procedia - Procedia Computer Science*, *60*, 708–713. https://doi.org/10.1016/j.procs.2015.08.220

Baek, U.-J., Lee, M., Park, J., & Kim, M. (2019). DDoS Attack Detection on Bitcoin Ecosystem using. *Ecosystem Using Deep-Learning. In 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 1–4.

Bartoletti, M., Pes, B., & Serusi, S. (2018). Data mining for detecting Bitcoin Ponzi schemes. *Crypto Valley Conference on Blockchain Technology (CVCBT)*, 75–84.

Blockonomi news. (2019). Retrieved February 1, 2020, from https://blockonomi.com/mt-gox-hack/

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). *Bitcoin: Economics, Technology, and Governance*. *29*(2), 213–238.

Çelik, M., Dadaşer-Çelik, F., & Dokuz, A. Ş. (2011). Anomaly detection in temperature data using DBSCAN algorithm. *INISTA 2011 - 2011 International Symposium on INnovations in Intelligent SysTems and Applications*, 91–95. https://doi.org/10.1109/INISTA.2011.5946052

Chen, M., Gao, X. D., & Li, H. F. (2010). Parallel DBSCAN with Priority R-tree. *ICIME 2010 - 2010 2nd IEEE International Conference on Information Management and Engineering*, *3*, 508–511. https://doi.org/10.1109/ICIME.2010.5477926

Dokuz, A. Ş., Ecemiş, A., & Celik, M. (2019). Hourly , Daily , and Monthly Analysis of Big Dataset of Bitcoin Blocks. *International Conference on Engineering Technologies (ICENTE'19)*.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, *96*(34), 226–231.

Id, F. S., Sun, X., Gao, J., Xu, L., Shen, H., & Cheng, X. (2019). *Anomaly detection in Bitcoin market via price return analysis*. *6*(14).

Investing. (2020). Retrieved February 1, 2020, from www.investing.com

Khan, K., Rehman, S. U., Aziz, K., Fong, S., Sarasvady, S., & Vishwa, A. (2014). DBSCAN: Past, present and future. *5th International Conference on the Applications of Digital Information and Web Technologies, ICADIWT 2014*, 232–238. https://doi.org/10.1109/ICADIWT.2014.6814687

Lee, S., & Edu, T. S. (2016). Anomaly Detection in Bitcoin Network Using Unsupervised Learning Methods. *ArXiv Preprint ArXiv:1611.03941*.

Melanie, S. (2017). Anticipating the Economic Benefits of Blockchain. *Technology Innovation Management Review*, *7*(10), 6–13. https://doi.org/10.22215/timreview/1107

Monamo, P., Marivate, V., & Twala, B. (2016). Unsupervised Learning for Robust Bitcoin Fraud Detection. *Information Security for South Africa (ISSA)*, 129–134.

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. https://doi.org/10.1007/s10838-008-9062-0

Ozekes, A., Celik, M., Ozkok, F. O., Komuscu, A. U., & Dadaser-celik, F. (2018). AutoVDBSCAN : An Automatic and Level-Wise Varied-Density Based Anomaly Detection Algorithm. *7th International Conference on Advanced Technologies (ICAT'18)*.

Sas, C., & Mara, U. T. (2017). Design for trust: An exploration of the challenges and opportunities of bitcoin users. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6499–6510.

Sayadi, S., Rejeb, S. Ben, & Choukair, Z. (2019). Anomaly Detection Model Over Blockchain Electronic Transactions. *15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 895–900.

Tan, P.-N. T., Steinbach, M., Anuj, K., & Kumar, V. (2005). *Introduction to Data Mining*.