Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 535-541, 2, 2024

Araştırma Makalesi

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 535-541, 2, 2024

Research Article

Digital Twin Model for Elevator Anomaly Detection: A LOF Approach

Elif Cesur*, Ayşe Balcı [2]

[1] *Industrial Engineering / Faculty of Engineering and Natural Sciences, Istanbul Medeniyet University, Turkey*
[2] *Industrial Engineering / Faculty of Engineering and Natural Sciences, Istanbul Medeniyet University, Turkey*

Email of the corresponding author: elif.karakaya@medeniyet.edu.tr

(Received: 12 March 2024, Accepted: 13 March 2024)

(4th International Artificial Intelligence and Data Science Congress ICADA 2024, March 14-15, 2024)

ATIF/REFERENCE: Cesur, E. & Balcı, A. (2024). Digital Twin Model for Elevator Anomaly Detection: A LOF Approach. *International Journal of Advanced Natural Sciences and Engineering Researches, 8(2), 535-541.*

Abstract – Digital twins provide the capability to transfer real-time data into a virtual environment through sensors, enabling the detection and prediction of abnormalities in production processes. A review of the literature reveals that digital twins play a significant role in improving efficiency in production processes and have become a crucial element in industrial competition. When anomalies are predicted and detected in advance, our ability to intervene increases. This allows for the prevention of potential problems, minimizing damage, and facilitating the implementation of predictive maintenance activities. Furthermore, it reduces costs resulting from unexpected failures and contributes to the reliable operation of systems.

In this study, data from three sensors installed in an elevator were collected to attempt to create a digital twin of the elevator. The aim was to detect anomalies in the collected data and improve effectiveness through predictive maintenance. Real-time data analysis and anomaly detection were facilitated using the Local Outlier Factor (LOF) algorithm, an anomaly detection algorithm. LOF evaluates the uniqueness of each event based on its distance to its k-nearest neighbors. It is an unsupervised anomaly detection method advantageous for cases where labeling large amounts of data is not feasible. In total, we collected 107,267 data points, of which 4,734 were identified as outliers, enabling us to comprehensively analyze the reasons behind their outlier status.

The intended contribution of this study is to demonstrate that the creation of digital twins in systems leads to the detection of anomalies in production processes, thereby increasing efficiency and reducing costs by minimizing unplanned downtime.

Keywords – Digital Twin, Anomaly Detection, Predictive Maintenance, Local Outlier Factor, Smart Manufacturing

I. INTRODUCTION

This research presents an investigation into the use of digital twins and anomaly detection with predictive maintenance to enhance system effectiveness, a result of the development of Industry 4.0 and the increasingly vital role of smart manufacturing systems. Industry 4.0 can be defined as the smartening of industry [1], and it can also be described as the integration of information and communication technologies in production [2]. Industry 4.0 offers the potential for faster response to customer demands, facilitating the adoption of new business models, production processes, and other innovations by increasing the flexibility, speed, efficiency, and quality of production processes [4]. The adaptation of businesses to Industry 4.0 technologies enables them to achieve more effective outcomes. Industry 4.0 represents an economic transition supported by information, data, and the Internet of Things (IoT). This transition influences the existing structures, markets, and business processes of the industrial age, leading to a new era of digitalization and the transformation of production systems into smarter networks. Recently, the addition of components such as mobile technologies, cloud computing, social media, and big data to the industrialization process has introduced a new perspective, guiding the market toward a new phase of competition and product differentiation [5].

Big data, real-time data processing, and connectivity are revolutionary distinctions brought by Industry 4.0. This significant difference also has important implications in the leadership context [6].

The concept of digital twins represents a virtual and dynamic model of services and processes that interact with real-time data, integrating the real world with the virtual world [7]. In other words, a digital twin is a digital replica that accurately reflects the current state of a physical object [8].

In recent studies, an increase in the use of digital twins has been observed for modeling real-world objects, detecting anomalies, and creating maintenance plans. The advantages offered by digital twins have contributed to their increasing adoption. According to Deng et al. (2021), digital twins are an inevitable goal of digital transformation.

Transforming data into meaningful insights and advancing decision-making and production processes based on this information leads to a higher-quality production process. Therefore, the use of digital twins is increasingly important today. Among the benefits of using digital twins is the ability for operators to test and optimize machine settings in the virtual world, thereby reducing machine setup times improving quality, and increasing efficiency through the use of machine learning methods. It is stated that this situation provides a significant competitive advantage for businesses [9]. Despite the increasing number of applications in the literature, there is still no high-quality application of digital twins [10].

Analyzing anomalies on data sets obtained with digital twins aims to identify deviations from the normal operation of the system as the observed object reflects its performance in the real world. Anomalies are outliers that stand out from other data points and do not conform to the general behavior of the data set. When anomalies are predicted and detected in advance, the opportunity to intervene in these situations increases. Preventing potential problems in advance minimizes potential damage and allows for the implementation of predictive maintenance activities.

In this study, data was obtained from sensors installed in elevators connected to digital twin technology. It was observed that the data saved as a CSV file was corrupted. Noisy data was corrected by preprocessing the data. Subsequently, a heatmap was generated using the Autoviz library to examine the correlation between columns. The data was visualized. Then, the Local Outlier Factor (LOF) algorithm, one of the machine learning algorithms, was run to detect anomalies. The anomaly detection in the study is performed on unsupervised data.

II. MATERIALS AND METHOD

Anomaly detection studies can be classified into supervised, unsupervised, and semi-supervised learning approaches. However, supervised learning models often require labeled anomaly data and are limited to defined anomalies, hence they may have disadvantages. In the future, the complexity of systems and the increase in data volume may limit the applicability of supervised learning methods. Therefore, anomaly studies based on unsupervised learning are preferred more due to reasons such as the difficulty of labeling by experts and the rarity of anomalous situations. Particularly, research on unsupervised time series anomaly detection continues to be an important subject.

Paradigms are suitable for anomaly detection roughly categorized into density estimation, clustering-based, reconstruction-based, and autoregression-based methods. Among density estimation methods, classical methods such as Local Outlier Factor (LOF) calculate the local density for outlier detection. The reason for choosing this algorithm is that it is more sensitive and flexible in detecting anomalies due to being based on local densities. The algorithm also does not have a threshold value, which needs to be set by the user, making LOF more flexible and easily applicable. Additionally, LOF can effectively work on multidimensional data sets.

The LOF algorithm evaluates the uniqueness of each event based on its distance to its k-nearest neighbors. Since LOF does not make any assumptions about data distribution, it can detect outliers effectively. The underlying idea of the algorithm is that the density around an outlier object significantly differs from the density around its neighbors.

LOF is an unsupervised anomaly detection method. The analyzed data are not labeled, which provides an advantage in cases where labeling is impractical due to large amounts of data.

LOF operates on the logic of the k-nearest neighbor algorithm and is generally distance-based rather than density-based. In the LOF algorithm, distance values are used to estimate density. Initially, it evaluates how isolated each data point is compared to its neighbors.

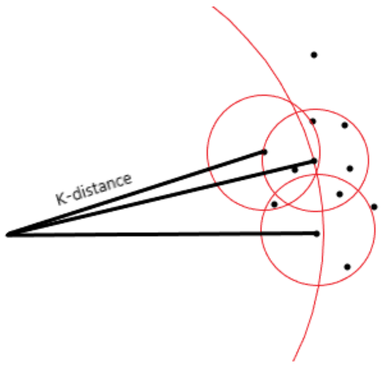

Fig 1. LOF Representation (Source: geeksforgeeks.org)

LOF, the density of points near the point being examined, and the average density of its k-nearest neighbors are compared. If the density of the point being examined is significantly lower than this average value, the point is considered an outlier. Data points with densities as large as their neighbors receive an approximate value of 1.

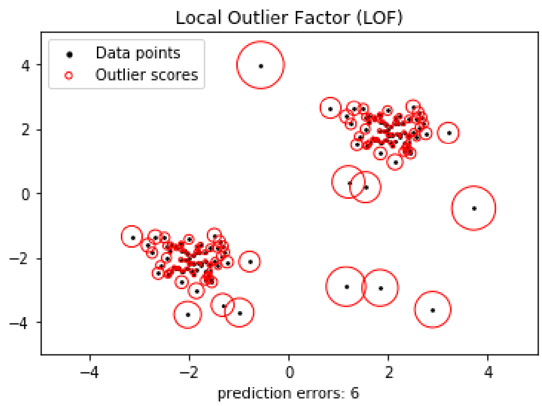

Fig 2. LOF Structure (Source: geeksforgeeks.org)

The LOF score serves to determine whether a point is anomalous. While other algorithms typically produce binary classifications for outlier status, LOF can also indicate how much of an outlier a point is.

One advantage of the LOF algorithm is its ability to accommodate uneven data distributions and different data densities.

For anomaly detection, we will use the Pycaret library. The reason for choosing this library is its ability to quickly analyze datasets, transform categorical variables, and provide convenience in data preprocessing steps. Pycaret also offers the capability to automate hyperparameter tuning by comparing multiple models.
.

III. CASE STUDY

In this study, we analyzed a dataset consisting of 112,002 rows and 3 columns. The columns are ball-bearing, humidity, and vibration. Data preprocessing steps were initially performed concerning the data. Irregular and noisy data were identified and organized through data preprocessing steps. A correlation map, indicated in Figure 1, was generated. A correlation map illustrates the correlation between two variables, measuring the strength of their relationship. In this correlation map, it was observed that humidity, ball-bearing, and vibration are lowly correlated with each other, as seen in Figure 1.

Histograms, boxplots, and scatterplots were drawn to observe the behaviors of the three columns we worked with in the data preprocessing step. In Figure 2, it can be observed that humidity values tend to concentrate around 74, which is also reflected in the distribution of the boxplot around the center of 74. In Figure 3, it is observed that vibration values are concentrated between 0 and 20, with a limited amount of data between 80 and 100.

In Figure 4, when examining the ball-bearing data, accumulation is observed at values of 20, 40, and 60 and the boxplot indicates the presence of means at these values. It is noted that the obtained scatter plots resemble each other.

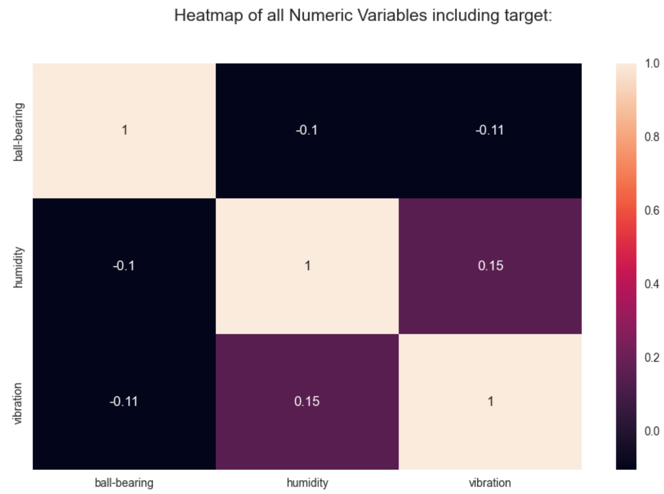

Fig 3. Correlation Map

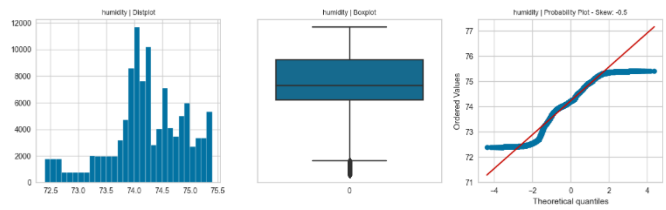

Fig 4. Graphs of the Humidity Column

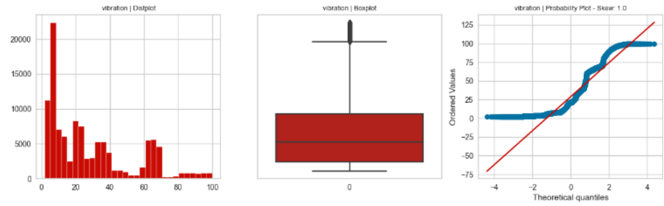

Fig 5. Graphs of the Vibration Column

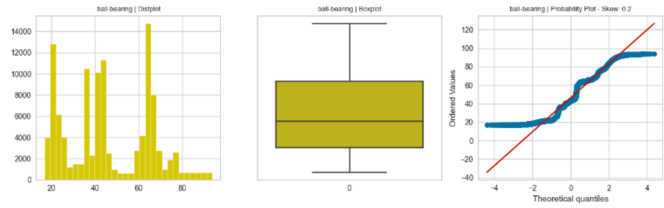

Figure 6. Graphs of the Ball-Bearing Column

As a result of the analyses conducted, the algorithm normalized the data, assigning anomaly values as 1 and normal values as 0. The algorithm also determined the anomaly scores, and the first 10 rows of the scores are as follows:

	ball-bearing	humidity	vibration	Anomaly	Anomaly_Score
0	2.492865	-0.328812	-0.439433	1	1.285632
1	2.492655	-0.328812	-0.439393	1	1.279927
2	2.492445	-0.330272	-0.439352	1	1.181194
3	2.492235	-0.330272	-0.439311	1	1.176689
4	2.492077	-0.330272	-0.439270	1	1.173800
5	2.491867	-0.331733	-0.439230	1	1.108143
6	2.491657	-0.331733	-0.439189	1	1.105375
7	2.491447	-0.331733	-0.439148	1	1.103484
8	2.491237	-0.333193	-0.439107	1	1.071248
9	2.491027	-0.333193	-0.439066	1	1.070626
10	2.490817	-0.333193	-0.439026	1	1.070626

Fig 7. Result of the LOF Algorithm

Among 112,002 data points, there are 4,734 anomaly values and 107,267 normal values.

When examining the relationship of abnormalities with columns, we obtain the following correlation graph.

Table 1. Relationship of Abnormalities with Columns

	Ball-Bearing	***Humidity***	***Vibration***	***Lof Anomaly***
Ball-Bearing	1			
Humidity	-0,10148207	1		
Vibration	-0,10507968	0,153545388	1	
Lof Anomaly	-0,02420057	0,040492921	0,012279363	1

IV. DISCUSSION

Based on the results obtained from the graphical analysis in Figure 8, the relationship between anomalies and the columns from which they originated was examined. While no significant change was observed in the humidity column, it was determined that anomalies occur when there is a decreasing trend in the ball-bearing column.

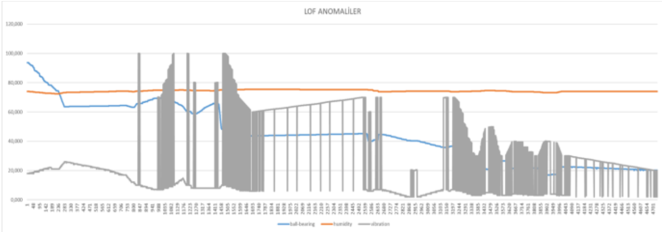

Fig 8. Graphical Representation of Anomalies

The decrease in the ball-bearing column may indicate an increase in mechanical issues within the system or the onset of wear and tear. This suggests that the ball-bearing column is an important indicator in detecting anomalies.

On the other hand, no significant trend related to anomalies was observed in the humidity column. These results indicate that humidity levels generally do not affect the occurrence of anomalies in the elevator system.

The vibration column exhibits a continuously changing trend. Vibration data can indicate the presence of mechanical faults in the machine or system.

V. CONCLUSION

Within the scope of the study, anomalies were identified, and their anomaly scores were printed. Upon examining the relationship between the obtained anomaly scores and the columns, it is observed that the humidity column has a positive correlation of 4%. Upon overall inspection of the graph, it can be said that the parameters of the dataset have minimal influence on each other.

The pre-determination of abnormal conditions in the digital twin model created with elevator data contributes to energy efficiency and effective utilization of resources. Additionally, it assists in reducing costs incurred from unexpected malfunctions. The early detection of abnormal conditions contributes to the reliable operation of systems. In this way, more efficient utilization is achieved. As a result of averting breakdowns, a more financially profitable model is established.

The study aims to contribute to the literature on how the system can be utilized more effectively and efficiently by detecting outliers within the scope of predictive maintenance.

REFERENCES

[1] Faller, C., & Feldmüller, D. (2015). Industry 4.0 learning factory for regional SMEs. Procedia Cirp, 32, 88-91.
[2] Basl, J. (2016). The pilot survey of the industry 4.0 principles penetration in the selected Czech and Polish companies. Journal of Systems Integration (1804-2724), 7(4).
[3] Zhu, Z., Xi, X., Xu, X. ve Cai, Y. (2021). “Digital Twin-Driven Machining Process for Thin-Walled Part Manufacturing”, Journal of Manufacturing Systems, 59, 453-466.
[4] Rüßmann, M., Lorenz, M., & Gerbert, P. (2015, Nisan). Industry 4.0 The Future of Productivity and Growth in Manufacturing Industries. The Boston Consulting Group, s. 1-20.
[5] Morrar, R., Arman, H. and Mousa S. (2017). The Fourth Industrial Revolution (Industry 4.0): A Social Innovation Perspective. Technology Innovation Management Review, 7(11).
[6] Öztürk, E., & Koç, K. H. (2017). Endüstri 4.0 Ve Mobilya Endüstrisi. İleri Teknoloji Bilimleri Dergisi, 6(3), 786-794.
[7] Çaliş Duman, M. (2022). İşletmeler İçin Yeni Bir Verimlilik Teknolojisi: Dijital İkiz. Verimlilik Dergisi189-206.
[8] Yukcu, S., & Aydin, Ö. (2020). Maliyet Düşürme Yöntemi Olarak Dijital İkiz. Muhasebe Bilim Dünyası Dergisi, 22(3), 563-579.
[9] Apillioğulları, L. (2019). “Dijital Dönüşüm: Akıllı Fabrikalar”, Aura Kitapları, İstanbul
[10] Scime, L., Singh, A. ve Paquit, V. (2021). “A Scalable Digital Platform for the Use of Digital Twins in Additive Manufacturing”, Manufacturing Letters, DOI: 10.1016/j.mfglet.2021.05.007.

[11] O.P. Popoola, K. Wang (2012). "Video-Based Abnormal Human Behavior Recognition: A Review," Syst. Man, Cybern. Part C Appl. Rev. IEEE Trans., 42(6), 865–878.
[12] S. Thomas, P. Seebck, S.M. Waldstein, U. SchmidtErfurth, G. Langs “Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery”, Lecture Notes in Computer Science, vol. 10265, pp. 146–157, 2017.
[13] K. Choi, J. Yi, C. Park, and S. Yoon, “Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines,” IEEE Access. 2021
[14] Wang, X., Wang, D., Zhang, Y., Jin, L., & Song, M. (2019, July). Unsupervised learning for log data analysis based on behavior and attribute features. In Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science (pp. 510-518).