Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 13, No. 2, Desember 2025

ISSN 0216 – 0544
e-ISSN 2301– 6914

Comparative study of unsupervised anomaly detection methods on imbalanced time series data

Riza Aulia Hanifa[a], Aris Thobirin[b], Sugiyarto Surono[c]

[a,b,c]Department of Mathematics, Ahmad Dahlan University
St South Ringroad, Bantul, D.I Yogyakarta 55191
E-mail: riza2100015040@webmail.uad.ac.id, aris.thobi@math.uad.ac.id, sugiyarto@math.uad.ac.id

Abstract

Anomaly detection in time series data is essential, especially when dealing with imbalanced datasets such as air quality records. This study addresses the challenge of identifying point anomalies rare and extreme pollution levels within a highly imbalanced dataset. Failing to detect such anomalies may lead to delayed environmental interventions and poor public health responses. To solve this, we propose a comparative analysis of three unsupervised learning methods: K-means clustering, Isolation Forest (IForest), and Autoencoder (AE), including its LSTM variant. These algorithms are applied to monthly air quality data collected in 2023 from 2,110 cities across Asia. The models are evaluated using Area Under the Curve (AUC), Precision, Recall, and F1-score to assess their effectiveness in detecting anomalies. Results indicate that the Autoencoder and Autoencoder LSTM outperform the others with an AUC of 98.23%, followed by K-means (97.78%) and IForest (96.01%). The Autoencoder's reconstruction capability makes it highly effective for capturing complex temporal patterns. K-means and IForest also show strong results, offering efficient and interpretable solutions for structured data. This research highlights the potential of unsupervised anomaly detection techniques for environmental monitoring and provides practical insights into handling imbalanced time series data.

Key words: Anomaly detection, Autoencoder, Imbalance data, Isolation Forest, and K-means.

INTRODUCTION

Nowadays, technology has made breakthroughs in data collection in various research fields, allowing the collection of large amounts of data over time. Thus, this technology contributes to forming time series data[1]. In clustering issues, time series data is one of the most common data types. It is also frequently utilized in biology for gene expression data and in finance for stock market analysis[2]. This data reflects time that changes periodically and sequentially[3]. The most commonly used series are annual, quarterly, monthly, weekly, and daily frequencies[4]. Each observation in a time series presents information obtained from previous observations, thus enabling analysis of historical patterns and prediction of future values[5], [6]. A deep understanding of these data characteristics is helpful in analysis, but a big challenge arises when identifying anomalies among standard patterns.

Anomaly detection in time series data has become a significant challenge in modern data analysis, which focuses on identifying patterns or observations that deviate from the general characteristics of the majority of the data[7].

As values or observations differ significantly from the normal pattern in a dataset, anomalies are often identified as outliers that can affect the normality of the data and thus require special attention[8], [9]. This phenomenon can arise in a variety of contexts and is commonly caused by measurement errors, human error, mechanical errors, or technical problems[10], [11]. Moreover, anomalies can also indicate unusual events, such as operational disruptions, changes in environmental patterns, or unexpected incidents[12]. Anomaly identification is not only important for detecting deviations but also has the potential to provide critical insights into system understanding and development.

However, detecting anomalies is not simple, especially in real-world time series data, which often has imbalanced characteristics. In this condition, anomalous events only cover a small portion of the entire dataset[13]. This imbalance can cause the model to tend to ignore the minority patterns (anomalous data), thus reducing the model's ability to recognize anomalies effectively [14], [15].

In various fields such as environmental monitoring, anomalies like extreme spikes in air pollution levels are rare and often obscured by the dominant normal patterns. This anomaly is known as a point anomaly, when a single data point is very different from other data points within a certain time frame or its general pattern[16]. Point anomaly can indicate significant and rare environmental events, equipment failures, or data recording errors. If undetected, these anomalies can cause delays in environmental policy responses and public health interventions.

Unsupervised learning approaches have evolved as the leading solution to overcome these challenges[17]. These methods do not require label data, making them well-suited for situations where manual labeling is difficult or expensive[18]. Various algorithms, including K-means, Isolation Forest, and Autoencoder, have unique approaches to capturing patterns and identifying aberrant data. However, their effectiveness is highly dependent on the characteristics of the dataset and the underlying assumptions.

Previous research has shown the effectiveness of unsupervised learning methods in detecting anomalies in industrial processes, particularly in screw-tightening data. The study by West et al. (2023)[19] used K-means clustering with Dynamic Time Warping (DTW) to detect anomalies in the unbalanced automotive industry assembly process, while Ribeiro et al. (2021)[20] compared the performance of Autoencoder and Isolation Forest in detecting anomaly in torque-angle pairs during the tightening process. Although both made significant contributions, their approaches are still limited to the manufacturing domain with mechanical sensor data, and they focus on a single type of method or a specific domain.

On the other hand, research by Wei et al. (2022)[21] applied an unsupervised deep learning-based approach using a combination of Long Short-Term Memory (LSTM) and Autoencoder to detect anomalies in indoor air quality data (CO2). This model is designed to capture long-term dependencies in time series data and calculate reconstruction errors to identify anomalies. The study demonstrated excellent performance and proved that unsupervised methods effectively identify abnormal values in air quality time series data, even without labeled data. These findings affirm the potential of Autoencoder methods and other unsupervised techniques in environmental monitoring and serve as an important foundation for the development of more adaptive alternative approaches.

In this study, the researchers addressed the issue of point anomaly detection in imbalanced time series data, with a real case study involving air quality data from 2110 cities in Asia in 2023. The data contains the average air quality values collected each month, thus forming a representation of a monthly scaled time series. This data not only reflects the challenges of analyzing complex environmental data but also provides a real example of the importance of anomaly detection in supporting sustainable air pollution monitoring systems.

To address this challenge, a comparative study was conducted on three unsupervised learning methods: K-means clustering, Isolation Forest (IForest), and Autoencoder (AE), including its LSTM variant. These three methods were chosen because of their ability to detect abnormal patterns without requiring labeled data, which is generally not available in environmental datasets. The evaluation was conducted using four main metrics, namely AUC, Precision, Recall, and F1-score. With this approach, this research is expected to make a significant contribution to supporting environmental data analysis and anomaly detection systems for better decision-making.

MATERIAL AND METHODS

This research uses the 2023 Asian air quality data, including monthly air quality averages in 2110 cities. The data consists of attributes such as rank, year, and monthly values from January to December. Because this research falls within the domain of environmental monitoring, its main objective is to identify abnormal pollution levels that may reflect real-world environmental anomalies.

This dataset reflects the characteristics of an imbalanced time series, where anomaly values are rare occurrences compared to the overall data. In this context, the type of anomaly detected is a point anomaly, which are individual values that significantly deviate from the general pattern. These anomalies can indicate extreme pollution events, data recording errors, or unusual local conditions.

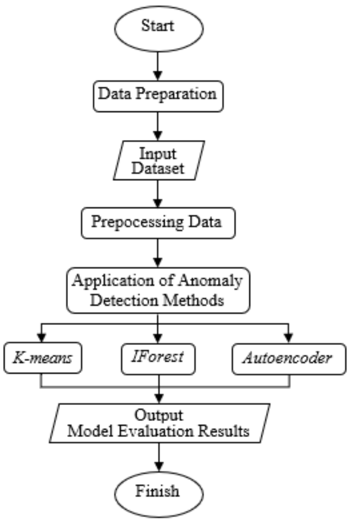

Fig. 1. Flowchart of the research

The process begins with preparing air quality data in Asia, which will be used as the research dataset. The dataset then went through a preprocessing stage, where data cleaning was done to address missing values and reduce simple outliers. In addition, data standardization was performed using RobustScaler, and the stationarity test process was conducted using the Augmented Dickey-Fuller (ADF) Test. Next, anomaly detection is performed by applying three main methods: K-means to cluster data based on feature similarity, Isolation Forest to identify easily isolated data as outliers, and Autoencoder, which utilizes neural networks to reconstruct standard data and detect anomalies through reconstruction errors. The results of these three methods are evaluated using metrics such as AUC, precision, recall, and F1-score for unsupervised learning to measure the performance of each method in detecting anomalies. Fig. 1. shows the research flow chart.

Preprocessing Data

Preprocessing steps were performed to ensure data quality as follows:

1 Handling missing values: Using the cubic interpolation method to replace missing values.
2 Differencing to make the data stationary
3 Data standardization with RobustScaler

Anomaly Detection Methods

This research uses three unsupervised learning algorithms to detect anomalies:

1. K-means Clustering: A clustering algorithm that divides data into k clusters based on the proximity of each object to the cluster center (centroid)[22]. The distance between data and k centroids is calculated using the Euclidean distance[23], formulated in Equation (1).

$$d_{i,j} = \sqrt{\sum_{k=1}^{n}\left(x_{ik} - c_{jk}\right)^2} \quad (1)$$

   In the equation, $d_{i,j}$ represents the Euclidean distance between object i and cluster center j, $x_{i,k}$ is the k attribute value of object i, $c_{j,k}$ is the k attribute value of cluster center j, and n is the total number of attributes. This distance becomes the basis for determining whether data is closer to a particular cluster center or further away from the center, which may indicate an anomaly.
2. Isolation Forest: A decision tree-based algorithm isolates data points using random division[24]. Faster isolated points are considered anomalies. The score is based on the average path length $h(x)$ to isolate data points in all isolation trees. The parameter $c(x)$ is used to normalise the path length based on the data size, so the anomaly score can be calculated using the function $s(x)$ in Equation (2).

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

3. Autoencoder: An artificial neural network designed to reconstruct data. Anomalies are identified based on high reconstruction errors. The two primary parts of an autoencoder are the encoder and the decoder[25]. The encoder reduces the dimension of the input data to produce a compressed representation of the data, while the decoder reconstructs the data from the compressed representation. The calculations for the encoder, decoder, and reconstruction error functions are described in equations (3) to (5):
   Encoder:

$$h = f(Wx + b) \quad (3)$$

Decoder:

$$\hat{x} = f'(W'h + b) \quad (4)$$

Reconstruction error:

$$\Delta = f_1(x, \hat{x}) \quad (5)$$

In this process, x represents the original input data, h is the output of the hidden layer, f and f' represent the encoding and decoding functions, respectively. At the same time, f_1 is used to calculate the reconstruction error. In addition, Δ represents the reconstruction error value, W and W' are weight matrices, b and b' are bias vectors, and $\hat{x}$ represents the reconstructed data

Performance Evaluation

The performance evaluation of the anomaly detection model is performed using the confusion matrix, which describes the model's performance in terms of correct and incorrect predictions in two classes: anomalous and normal. The confusion matrix consists of four main components:

- True Positives (TP), which indicates the number of anomalies that are detected as anomalies.
- False Positives (FP), which indicates the number of normal data misclassified as anomaly.
- False Negatives (FN), indicates the number of anomalies that were not detected and classified as normal
- True Negatives (TN), indicates the amount of normal data classified as normal.

After analyzing the confusion matrix, we continue with a discussion of the main performance metrics used to evaluate the model, namely:

- AUC (Area Under Curve): Measures the ability of the model to distinguish between positive and negative classes.
- Precision: Measures the ability of the model to detect specific anomalies.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

- Recall (Sensitivity): Measures the ability of the model to detect all anomalies.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

- F1-score: Harmony between precision and recall to provide an overall performance picture.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

RESULT AND DISCUSSION

Data Description

Exploratory data analysis revealed seasonal patterns in the air quality data, with anomalies indicating rare and extreme pollutant levels. Descriptive statistics and visualisations highlighted the dataset's imbalance (see Table 1).

Table 1. Average descriptive statistics table

	Rata-Rata
count	2164
mean	23.3052
std	16.1439
min	1.475
25%	10.0729
50%	19.1042
75%	31.4021
max	124.917

Table 1 shows that from the data, it can be seen that air pollution in cities has a significant variation, ranging from 1,475 to 124,917. There is a significant difference between the cities with the highest and lowest pollution. Next, we will conduct a stationarity test to see if this data has a trend or seasonal pattern.

Stationarity Test

The Augmented Dickey-Fuller (ADF) test was conducted to test the stationarity of the data. The ADF test results in Fig. 2. show that the data is not stationary, as the ADF Statistic (0.878) is more significant than all critical values (both at 1%, 5%, and 10%), and the p-value (0.993) is much greater than 0.05.

```
ADF Statistic: 0.8783632915177038
p-value: 0.9928013642123126
Critical Values:
        1%: -3.4334137212590194
        5%: -2.8628934347449033
        10%: -2.567490502788733
```

Fig. 2. ADF test results

Determining the Optimum Number of Clusters

Determining the optimum number of clusters is an important step in clustering methods to ensure representative results of the data being analyzed. In this study, two main methods, the Elbow Method and the Silhouette Score, were used to evaluate the most appropriate number of clusters. Fig. 3. shows the results of the Elbow Method, while Fig. 4. illustrates the evaluation results using Silhouette Score.

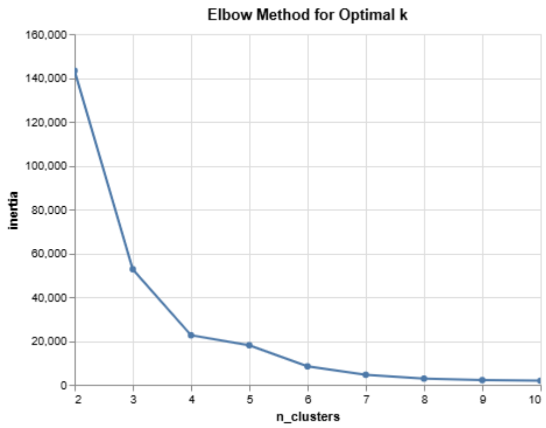

Fig. 3. Elbow method

In Fig. 3. the Elbow Method determines the optimum number of clusters based on the inertia value (sum of squared distances from points to cluster centers). The 'elbow' point on the graph shows a significant decrease in inertia before stabilizing, indicating the optimum number of clusters, $k = 3$.

Fig. 4. shows the evaluation results using the Silhouette Score, which measures how well the objects fit within their respective clusters. The highest Silhouette Score value is achieved at $k = 3$, indicating that the division of clusters at this point produces the most optimal cluster structure. Both methods consistently determine the optimum number of clusters, which is $k = 3$, so these results are used in further analyses.

Fig. 4. Silhouette score

Anomaly Detection

Anomaly detection is an important part of data analysis to identify values that deviate significantly from the general pattern. In this study, the number of anomalies detected was compared using three different methods: K-Means, Isolation Forest (Iforest), Autoencoder (AE) and Autoencoder LSTM (AE LSTM). Fig. 5. presents the comparison results of the number of anomalies detected by each method.

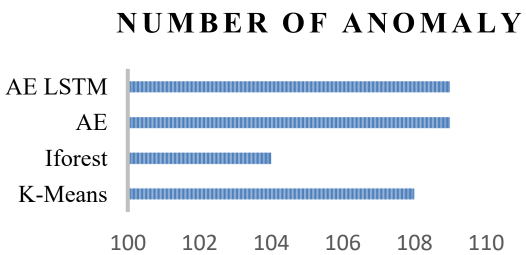

Fig. 5. Number of anomaly

Based on the visualisation in Fig. 5. the K-Means method detected 108 anomalies, while Iforest detected 104. On the other hand, AE and AE LSTM produced the same number of anomalies, 109. This difference in the number of anomalies detected by each method indicates that different approaches have varying sensitivity to patterns and irregularities in the data.

Model Performance

To provide a more in-depth look at the performance of each model, the confusion matrix in Fig. 6. shows the distribution of correct and incorrect predictions for the normal and anomaly classes.

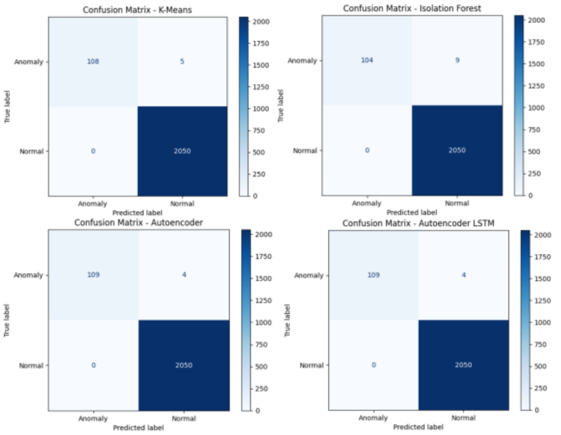

Fig. 6. Confusion matrix

After analyzing the confusion matrix, we proceed to discuss the main performance metrics used to evaluate the models, namely AUC, precision, recall, and F1-score. Table 2 summarises the performance metric values for each of the tested models. Autoencoder achieved the highest AUC and F1 score, followed by K-means and Isolation Forest.

Table 2. Matrix evaluation

Metode	AUC	Precision	Recall	F-1 Score
K-Means	0,9778	1	0,9557	0,9773
IForest	0,9601	1	0,9203	0,9585
Autoencoder	0,9823	1	0, 9646	0,9819
Autoencoder LSTM	0,9823	1	0, 9646	0,9819

Each model has outstanding performance, as shown in Table 2. The autoencoder has the highest AUC value of 0.9823, demonstrating its best ability to capture complex temporal patterns and non-linear relationships in air quality data thanks to its layered architecture. K-means performed well, showing its effectiveness with simpler distance-metrics-based methodologies with an AUC of 0.9778. Isolation Forest effectively managed high-dimensional time series data by achieving a competitive AUC of 0.9601. With no false positives, all models accurately detected the correct abnormalities, achieving a perfect precision value of 1.0.

CONCLUSION

This study compares the performance of K-means, Isolation Forest, and Autoencoder in detecting anomalies in unbalanced air quality data. The results show that K-Means can distinguish normal and abnormal data with an accuracy of up to 97.78%, IForest with an accuracy of 96.01%, and AE and AE LSTM with an accuracy of up to 98.23%. These methods showed excellent results, with all anomaly detection accuracies above 90%, confirming the effectiveness of unsupervised learning methods in detecting anomalies in environmental datasets.

Autoencoder and Autoencoder LSTM are the best-performing models for this task, as they have relatively high AUC and F1-score values. Autoencoder outperformed the other methods, demonstrating its robustness in handling complex patterns. These findings provide a basis for selecting appropriate anomaly detection methods for real-world applications, especially environmental monitoring. Future research can explore hybrid approaches to improve detection accuracy further.

REFERENCES

[1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A Review on Outlier/Anomaly Detection in Time Series Data," ACM Comput. Surv., vol. 54, no. 3, pp. 1–33, Apr. 2022, doi: 10.1145/3444690. https://doi.org/10.1145/3444690

[2] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – A decade review," Inf. Syst., vol. 53, pp. 16–38, Oct. 2015, doi: 10.1016/j.is.2015.04.007. https://doi.org/10.1016/j.is.2015.04.007

[3] H. Rahadian, S. Bandong, A. Widyotriatmo, and E. Joelianto, "Image encoding selection based on Pearson correlation coefficient for time series anomaly detection," Alex. Eng. J., vol. 82, pp. 304–322, Nov. 2023, https://doi.org/10.1016/j.aej.2023.09.070 .

[4] M. B. Shrestha and G. R. Bhatta, "Selecting appropriate methodological framework for time series data analysis," J. Finance Data Sci., vol. 4, no. 2, pp. 71–89, Jun. 2018,

https://doi.org/10.1016/j.jfds.2017.11.001

[5] R. J. Hyndman and B. Rostami-Tabar, "Forecasting interrupted time series," J. Oper. Res. Soc., pp. 1–14, Sep. 2024, https://doi.org/10.1080/01605682.2024.2395315 .

[6] R. Kablaoui, I. Ahmad, S. Abed, and M. Awad, "Network traffic prediction by learning time series as images," Eng. Sci. Technol. Int. J., vol. 55, p. 101754, Jul. 2024, https://doi.org/10.1016/j.jestch.2024.101754 .

[7] R. Rahman and H. Nurul, "Keamanan Jaringan Kecerdasan Buatan dan Implementasi Solusi Keaman," Technol. Sci. Insights J., vol. 1, no. 1, pp. 33–36.

[8] E. A. Fadlilah, "Identifikasi Anomali Data Akademik Menggunakan Dbscan Outlier Detection," Pros. Sains Nas. Dan Teknol., vol. 12, no. 1, pp. 336–342, Nov. 2022, https://doi.org/10.36499/psnst.v12i1.7012 .

[9] G. R. Baihaqi and Mulaab, "LONG SHORT-TERM MEMORY FOR PREDICTION OF WAVE HEIGHT AND WIND SPEED USING PROPHET FOR OUTLIERS," J. Ilm. Kursor, vol. 12, no. 2, pp. 59–68, Dec. 2023, doi: https://doi.org/10.21107/kursor.v12i2.351 .

[10] M. N. K. Sikder and F. A. Batarseh, "Outlier detection using AI: a survey," in AI Assurance, Elsevier, 2023, pp. 231–291. https://doi.org/10.1016/B978-0-32-391919-7.00020-2 .

[11] G. Nassreddine, J. Younis, and T. Falahi, "Detecting Data Outliers with Machine Learning," Al-Salam J. Eng. Technol., vol. 2, no. 2, pp. 152–164, May 2023, https://doi.org/10.55145/ajest.2023.02.02.018 .

[12] M. Čampulová, J. Michálek, P. Mikuška, and D. Bokal, "Nonparametric algorithm for identification of outliers in environmental data," J. Chemom., vol. 32, no. 5, p. e2997, May 2018, https://doi.org/10.1002/cem.2997 .

[13] D. Liang, J. Wang, W. Zhang, Y. Liu, L. Wang, and X. Zhao, "Tabular Data Anomaly Detection Based on Density Peak Clustering Algorithm," in 2022 International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China: IEEE, Jan. 2022, pp. 16–21. https://doi.org/10.1109/BDICN55575.2022.00011 .

[14] [R. Baidya and H. Jeong, "Anomaly Detection in Time Series Data Using Reversible Instance Normalized Anomaly Transformer," Sensors, vol. 23, no. 22, p. 9272, Nov. 2023, https://doi.org/10.3390/s23229272 .

[15] Q. Liu, P. Boniol, T. Palpanas, and J. Paparrizos, "Time-Series Anomaly Detection: Overview and New Trends," Proc. VLDB Endow., vol. 17, no. 12, pp. 4229–4232, Aug. 2024, https://doi.org/10.14778/3685800.3685842 .

[16] J. Liu, D. Yang, K. Zhang, H. Gao, and J. Li, "Anomaly and change point detection for time series with concept drift," World Wide Web, vol. 26, no. 5, pp. 3229–3252, Sep. 2023, https://doi.org/10.1007/s11280-023-01181-z .

[17] Z. Z. Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," ACM Comput. Surv., vol. 57, no. 1, pp. 1–42, Jan. 2025, https://doi.org/10.1145/3691338 .

[18] A. Priarone, U. Albertin, C. Cena, M. Martini, and M. Chiaberge, "Unsupervised Novelty Detection Methods Benchmarking with Wavelet Decomposition," Sep. 11, 2024, arXiv: arXiv:2409.07135. https://doi.org/10.1109/ICSRS63046.2024.10927428 .

[19] N. West, T. Schlegl, and J. Deuse, "Unsupervised anomaly detection in unbalanced time series data from screw driving processes using k-means clustering," Procedia CIRP, vol. 120, pp. 1185–1190, 2023, https://doi.org/10.1016/j.procir.2023.09.146 .

[20] D. Ribeiro, L. M. Matos, P. Cortez, G. Moreira, and A. Pilastri, “A Comparison of Anomaly Detection Methods for Industrial Screw Tightening,” in Computational Science and Its Applications – ICCSA 2021, vol. 12950, O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, B. O. Apduhan, A. M. A. C. Rocha, E. Tarantino, and C. M. Torre, Eds., in Lecture Notes in Computer Science, vol. 12950. , Cham: Springer International Publishing, 2021, pp. 485–500. https://doi.org/10.1016/j.aej.2023.09.070 .

[21] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, “LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data,” Apr. 14, 2022, arXiv: arXiv:2204.06701. https://doi.org/10.1109/JSEN.2022.3230361 .

[22] Y. Pratama, E. Sulistianingsih, N. N. Debataraja, and N. Imro’ah, “K-Means Clustering dan Mean Variance Efficient Portfolio dalam Portofolio Saham,” Jambura J. Probab. Stat., vol. 5, no. 1, pp. 24–30, Jun. 2024, https://doi.org/10.37905/jjps.v5i1.20298 .

[23] I. M. S. Bimantara and I. M. Widiartha, “Optimization of k-means clustering using particle swarm optimization algorithm for grouping traveler reviews data on tripadvisor siteS,” J. Ilm. Kursor, vol. 12, no. 1, pp. 1–10, Jun. 2023, https://doi.org/10.21107/kursor.v12i01.269 .

[24] S. Wang, “Isolation Forest Anomaly Detection Algorithm Based On Multi-level Sub-subspace Partition,” Int. J. Comput. Sci. Inf. Technol., vol. 4, no. 2, pp. 149–159, Oct. 2024, https://doi.org/10.62051/ijcsit.v4n2.20 .

[25] W. Skaf and T. Horváth, “Denoising Architecture for Unsupervised Anomaly Detection in Time-Series,” in New Trends in Database and Information Systems, vol. 1652, S. Chiusano, T. Cerquitelli, R. Wrembel, K. Nørvåg, B. Catania, G. Vargas-Solar, and E. Zumpano, Eds., in Communications in Computer and Information Science, vol. 1652. , Cham: Springer International Publishing, 2022, pp. 178–187. https://doi.org/10.1007/978-3-031-15743-1_17 .