

DETEKSI ANOMALI UNSUPERVISED DATA TIME SERIES DENGAN K-MEANS, IFOREST, DAN AUTOENCODER

^aRiza Aulia Hanifa, ^bAris Thobirin, ^cSugiyarto Surono

^{a,b,c}Department of Mathematics, Ahmad Dahlan University
Jalan Ringroad Selatan, Bantul, D.I Yogyakarta

E-mail: riza2100015040@webmail.uad.ac.id, aris.thobi@math.uad.ac.id, sugiyarto@math.uad.ac.id

Abstract

Deteksi anomali dalam data *time series* adalah tugas penting dalam banyak domain, terutama ketika set data imbalance. Penelitian ini membandingkan kinerja tiga algoritma *unsupervised learning*: K-means, Isolation Forest, dan Autoencoder dalam mendeteksi anomali pada dataset kualitas udara yang imbalance di Asia pada tahun 2023. Dataset ini terdiri dari data kualitas udara bulanan untuk 2110 kota, menjadikannya contoh nyata dari tantangan analisis data lingkungan yang kompleks dan tidak seimbang. Evaluasi kinerja didasarkan pada metrik seperti AUC, *Precision*, *Recall*, dan *F1-score*. Hasilnya menunjukkan bahwa Autoencoder layer Dense dan layer LSTM mencapai kinerja terbaik secara keseluruhan dengan AUC 98,23%, diikuti K-means (97,78%) dan Isolation Forest (96,01%). Keunggulan Autoencoder terletak pada kemampuannya dalam merekonstruksi data dengan pola kompleks, sedangkan Isolation Forest menawarkan solusi yang efisien untuk deteksi anomali. Sementara itu, K-means menunjukkan performa yang baik pada dataset ini karena pola data yang memungkinkan pengelompokan yang efektif berdasarkan jarak antar kluster, menjadikannya alternatif yang layak untuk situasi dengan distribusi data yang lebih terstruktur. Studi ini memberikan wawasan tentang penerapan algoritma ini untuk deteksi anomali dalam dataset lingkungan dunia nyata, serta kontribusi signifikan untuk mendukung pengambilan keputusan berbasis data yang lebih baik dalam konteks pemantauan kualitas udara.

Kata kunci: Autoencoder, Data imbalance, Deteksi anomali, Isolation Forest, K-means

INTRODUCTION

Dewasa ini, teknologi telah menghasilkan terobosan besar dalam pengumpulan data di berbagai bidang penelitian, dimana memungkinkan pengumpulan sejumlah besar data dari waktu ke waktu. Dengan demikian, teknologi ini berkontribusi pada terbentuknya data deret waktu (*time series*)[1]. Data *time series* merupakan salah satu tipe data yang populer dalam masalah *clustering* dan digunakan secara luas dari gen data ekspresi dalam biologi hingga analisis pasar saham di bidang keuangan[2]. Data ini mencerminkan waktu yang terus berubah secara periodik dan teratur[3]. Seri yang paling umum digunakan adalah frekuensi tahunan, triwulanan, bulanan, mingguan, dan harian[4]. Setiap observasi dalam

time series mempresentasikan informasi yang diperoleh dari berbagai pengamatan sebelumnya, sehingga memungkinkan analisis pola historis dan prediksi nilai di masa depan[5][6]. Pemahaman mendalam tentang karakteristik data ini sangat membantu dalam analisis, tetapi tantangan besar muncul ketika harus mengidentifikasi anomali di antara pola-pola normal.

Deteksi anomali pada data *time series* telah menjadi tantangan signifikan dalam analisis data modern, dimana prosesnya berfokus pada identifikasi pola atau pengamatan yang menyimpang dari karakteristik umum mayoritas data[7]. Sebagai nilai atau pengamatan yang berbeda secara signifikan dari pola normal dalam dataset, anomali sering diidentifikasi sebagai outlier yang dapat mempengaruhi kenormalan

data sehingga memerlukan perhatian khusus[8][9]. Fenomena ini dapat muncul dalam berbagai konteks yang umumnya disebabkan oleh kesalahan pengukuran, faktor human error, kesalahan mekanis, atau permasalahan teknis[10][11]. Lebih dari itu, keberadaan anomali juga dapat mengindikasikan kejadian-kejadian tidak biasa, seperti gangguan operasional, perubahan pola lingkungan, atau insiden tak terduga[12]. Identifikasi anomali tidak hanya penting untuk mendeteksi penyimpangan, tetapi juga berpotensi memberikan wawasan kritis dalam pemahaman dan pengembangan sistem.

Namun, mendeteksi anomali tidaklah sederhana, terutama pada data *time series* di dunia nyata yang sering kali memiliki karakteristik tidak seimbang (*imbalanced*). Dalam kondisi ini, kejadian anomali hanya mencakup sebagian kecil dari keseluruhan dataset[13]. Ketidakseimbangan ini dapat menyebabkan model cenderung mengabaikan pola minoritas (data anomali), sehingga mengurangi kemampuan model untuk mengenali anomali secara efektif [14][15].

Pendekatan *unsupervised learning* telah berkembang sebagai solusi utama untuk mengatasi tantangan ini[16]. Metode ini tidak memerlukan data label, sehingga sangat cocok untuk situasi di mana pemberian label manual sulit atau mahal[17]. Beragam algoritma telah digunakan untuk mendeteksi anomali secara *unsupervised*, termasuk K-means, Isolation Forest, dan Autoencoder. Setiap algoritma memiliki pendekatan unik dalam menangkap pola dan mengidentifikasi data yang menyimpang, namun efektivitasnya sangat bergantung pada karakteristik dataset dan asumsi yang mendasarinya.

Dalam penelitian ini, peneliti menggunakan data kualitas udara di Asia tahun 2023 untuk mengevaluasi kinerja ketiga algoritma tersebut. Data ini terdiri atas nilai rata-rata tahunan kualitas udara yang dihimpun dari 2110 kota, dengan representasi deret waktu bulanan. Data ini tidak hanya mencerminkan tantangan analisis data lingkungan yang kompleks, tetapi juga memberikan contoh nyata tentang pentingnya deteksi anomali dalam mendukung pemantauan polusi udara secara berkelanjutan.

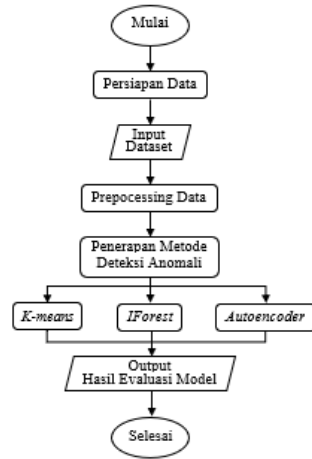
Penelitian ini berkontribusi dalam mengatasi tantangan deteksi anomali pada data deret waktu yang tidak seimbang dengan menganalisis performa tiga algoritma, yaitu K-means, Isolation Forest, dan Autoencoder. Evaluasi dilakukan secara komparatif menggunakan metrik untuk menentukan algoritma paling efektif. Hasil analisis ini diharapkan memberikan wawasan praktis tentang bagaimana algoritma *unsupervised learning* dapat diterapkan pada data lingkungan, seperti kualitas udara, yang sering kali bersifat tidak seimbang dan dinamis. Dengan pendekatan ini, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam mendukung analisis data lingkungan dan deteksi anomali untuk pengambilan keputusan berbasis data yang lebih baik.

MATERIAL AND METHODS

Penelitian ini menggunakan data kualitas udara Asia tahun 2023 yang mencakup rata-rata bulanan kualitas udara di 2110 kota. Data terdiri dari atribut seperti peringkat (*rank*), tahun, dan nilai bulanan dari Januari hingga Desember. Dataset ini mencerminkan karakteristik *time series* yang tidak seimbang, di mana nilai anomali merupakan kejadian yang jarang terjadi dibandingkan dengan keseluruhan data.

Proses dimulai dengan persiapan data kualitas udara di Asia, untuk kemudian dijadikan dataset penelitian. Dataset ini kemudian melalui tahap preprocessing, di mana dilakukan pembersihan data untuk mengatasi nilai hilang. Selain itu, Standarisasi data dilakukan menggunakan RobustScaler dan proses uji stasioneritas dengan Uji Augmented Dickey-Fuller (ADF). Selanjutnya, deteksi anomali dilakukan dengan menerapkan tiga metode utama: K-means untuk mengelompokkan data berdasarkan kesamaan fitur, Isolation Forest untuk mengidentifikasi data yang mudah diisolasi sebagai outlier, dan Autoencoder yang memanfaatkan jaringan saraf untuk merekonstruksi data normal dan mendeteksi anomali melalui kesalahan rekonstruksi. Hasil dari ketiga metode ini dievaluasi dengan menggunakan metrik seperti AUC, precision, recall, dan F1-score untuk *unsupervised learning* mengukur kinerja setiap

metode dalam mendeteksi anomali. Gambar 1 menunjukkan diagram alir penelitian.



Gambar 1. Diagram alir penelitian

Preprocessing Data

Langkah prapemrosesan dilakukan untuk memastikan kualitas data sebagai berikut:

1. Penanganan nilai hilang: Menggunakan metode interpolasi cubic untuk mengganti nilai yang hilang.
2. Differencing untuk membuat data stasioner
3. Standarisasi data dengan RobustScaler

Metode Deteksi Anomali

Penelitian ini menggunakan tiga algoritma *unsupervised learning* untuk mendeteksi anomali:

1. K-means Clustering: Algoritma pengelompokan yang membagi data ke dalam k kluster berdasarkan kedekatan jarak setiap objek dengan pusat cluster (centroid)[18]. Perhitungan jarak antara data dengan k centroid dilakukan menggunakan jarak Euclidean[19], yang dirumuskan dalam persamaan (1)

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

Dalam persamaan tersebut, $d_{i,j}$ merepresentasikan jarak Euclidean antara objek i dan pusat cluster j , $x_{i,k}$ adalah nilai atribut ke- k dari objek i , $c_{j,k}$ adalah nilai atribut ke- k dari pusat cluster j , dan n adalah jumlah total

atribut. Jarak ini menjadi dasar untuk menentukan apakah suatu data lebih dekat ke pusat cluster tertentu atau berada jauh dari pusat, yang dapat mengindikasikan anomali.

2. Isolation Forest: Algoritma berbasis pohon keputusan yang mengisolasi titik data menggunakan pembagian acak[20]. Titik yang lebih cepat terisolasi dianggap sebagai anomali. Skor ini didasarkan pada panjang jalur rata-rata $h(x)$ untuk mengisolasi titik data dalam semua pohon isolasi. Parameter $c(x)$ digunakan untuk menormalkan panjang jalur berdasarkan ukuran data, sehingga skor anomali dapat dihitung menggunakan fungsi $s(x)$ pada Persamaan (2)

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

3. Autoencoder: Jaringan saraf tiruan yang dirancang untuk merekonstruksi data. Anomali diidentifikasi berdasarkan error rekonstruksi yang tinggi. Autoencoder terdiri dari dua komponen utama, yaitu encoder dan decoder[21]. Encoder berfungsi untuk mengurangi dimensi data masukan sehingga menghasilkan representasi data yang terkompresi, sementara decoder bertugas merekonstruksi data dari representasi terkompresi tersebut. Perhitungan untuk fungsi encoder, decoder, dan kesalahan rekonstruksi dijelaskan dalam persamaan (3) hingga (5):

Encoder:

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

Decoder:

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

Rekonstruksi error:

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

Dalam proses ini, x merupakan data masukan asli, h adalah output dari layer tersembunyi, f dan f' masing-masing merepresentasikan fungsi encoding dan decoding, sedangkan f_1 digunakan untuk menghitung kesalahan rekonstruksi. Selain itu, Δ melambangkan nilai kesalahan

rekonstruksi, W dan W' adalah matriks bobot, b dan b' adalah vektor bias, serta \hat{x} merepresentasikan data yang telah direkonstruksi.

Evaluasi Kinerja

Evaluasi kinerja model deteksi anomali dilakukan dengan menggunakan confusion matrix, yang menggambarkan kinerja model dalam hal prediksi yang benar dan salah pada dua kelas: anomali dan normal. Confusion matrix terdiri dari empat komponen utama:

- True Positives (TP), yang menunjukkan jumlah anomali yang benar-benar terdeteksi sebagai anomali
- False Positives (FP), yang menunjukkan jumlah data normal yang salah diklasifikasikan sebagai anomali
- False Negatives (FN), yang menunjukkan jumlah anomali yang tidak terdeteksi dan diklasifikasikan sebagai normal
- True Negatives (TN), yang menunjukkan jumlah data normal yang benar-benar diklasifikasikan sebagai normal.

Setelah menganalisis confusion matrix, dilanjutkan dengan pembahasan metrik performa utama yang digunakan untuk mengevaluasi model, yaitu

- AUC (Area Under Curve) mengukur kemampuan model untuk membedakan antara kelas positif (anomali) dan negatif (normal). Semakin tinggi nilai AUC, semakin baik kemampuan model dalam membedakan kelas.
- Presisi: Mengukur kemampuan model dalam mendeteksi anomali secara spesifik.

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

- Recall (Sensitivitas): Mengukur kemampuan model dalam mendeteksi semua anomali yang ada.

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

- F1-score: Harmoni antara presisi dan recall untuk memberikan gambaran performa keseluruhan.

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

RESULT AND DISCUSSION

Deskripsi Data

Analisis data eksplorasi mengungkapkan pola musiman dalam data kualitas udara, dengan anomali yang menunjukkan tingkat polutan yang langka dan ekstrem. Statistik deskriptif dan visualisasi menyoroti sifat ketidakseimbangan set data (lihat Tabel 1).

Table 1. Tabel Statistik deskriptif rata-rata

	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

Tabel 1 menunjukkan bahwa dari data tersebut, terlihat bahwa polusi udara di kota-kota memiliki variasi yang besar, mulai dari 1.475 hingga 124.917. Terdapat perbedaan yang signifikan antara kota dengan polusi tertinggi dan terendah.

Selanjutnya, Peneliti akan melakukan uji stasioneritas untuk melihat apakah data ini memiliki pola tren atau musiman.

Uji Stasioneritas

Uji Augmented Dickey-Fuller (ADF) dilakukan untuk menguji stasioneritas data.

Hasil uji ADF pada gambar 2 menunjukkan bahwa data tidak stasioner, karena ADF Statistic (0.878) lebih besar dari semua nilai kritis (baik di 1%, 5%, maupun 10%) dan p-value (0.993) jauh lebih besar dari 0.05.

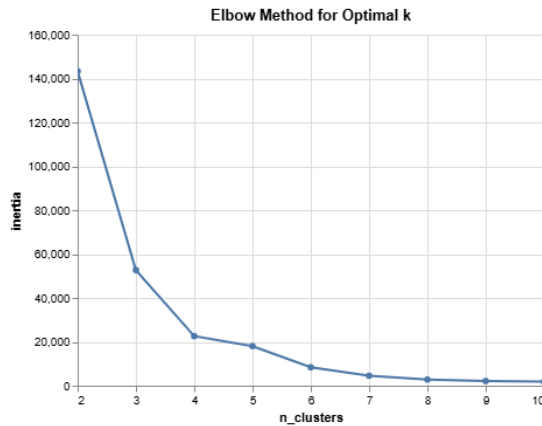
```
ADF Statistic: 0.8783632915177038
p-value: 0.9928013642123126
Critical Values:
1%: -3.4334137212590194
5%: -2.8628934347449033
10%: -2.567490502788733
```

Gambar 2. Hasil Uji ADF

Penentuan Jumlah Cluster yang Optimum

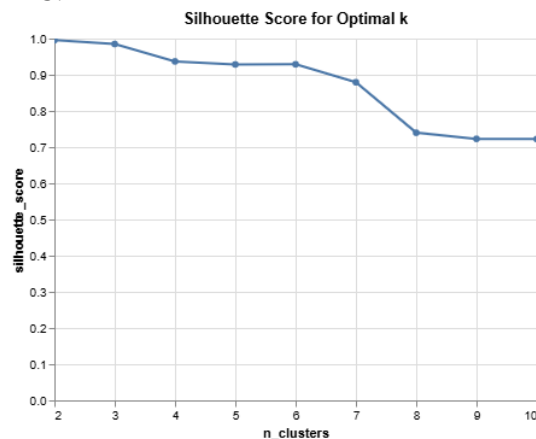
Penentuan jumlah cluster yang optimum merupakan langkah penting dalam metode clustering untuk memastikan hasil yang

representatif terhadap data yang dianalisis. Pada penelitian ini, digunakan dua metode utama, yaitu Metode Elbow dan Silhouette Score, untuk mengevaluasi jumlah cluster yang paling sesuai. Gambar 3 menunjukkan hasil dari Metode Elbow, sedangkan Gambar 4 menggambarkan hasil dari evaluasi menggunakan Silhouette Score.



Gambar 3. Metode Elbow

Pada Gambar 3, Metode Elbow digunakan untuk menentukan jumlah cluster yang optimum berdasarkan nilai inertia (jumlah jarak kuadrat dari titik ke pusat cluster). Titik "siku" pada grafik menunjukkan penurunan inertia yang signifikan sebelum akhirnya stabil, yang mengindikasikan jumlah cluster optimum, yaitu $k = 3$.



Gambar 4. Silhouette Score

Gambar 4 menunjukkan hasil evaluasi menggunakan Silhouette Score, yang mengukur seberapa baik objek berada dalam cluster masing-masing. Nilai Silhouette Score tertinggi dicapai pada $k = 3$, menunjukkan bahwa pembagian cluster pada titik tersebut

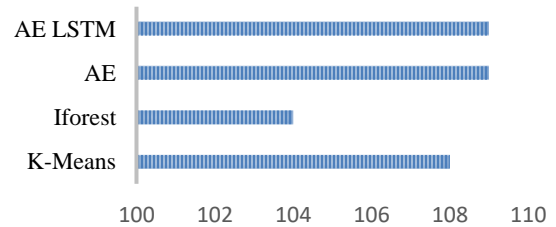
menghasilkan struktur cluster yang paling optimal.

Kedua metode ini menunjukkan konsistensi dalam menentukan jumlah cluster yang optimum, yaitu $k = 3$, sehingga hasil ini digunakan dalam analisis lebih lanjut.

Deteksi Anomali

Deteksi anomali merupakan bagian penting dari analisis data untuk mengidentifikasi nilai-nilai yang menyimpang secara signifikan dari pola umum. Dalam penelitian ini, jumlah anomali yang terdeteksi dibandingkan menggunakan empat metode berbeda: K-Means, Isolation Forest (Iforest), Autoencoder (AE), dan Autoencoder LSTM (AE LSTM). Gambar 5 menyajikan hasil perbandingan jumlah anomali yang terdeteksi oleh masing-masing metode.

JUMLAH ANOMALI

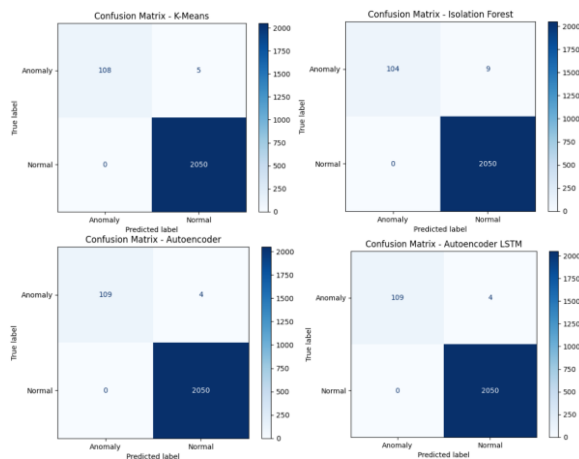


Gambar 5. Jumlah Anomali

Berdasarkan visualisasi gambar 5, metode K-Means mendeteksi 108 anomali, sementara Iforest mendeteksi 104 anomali. Di sisi lain, AE dan AE LSTM menghasilkan jumlah anomali yang sama, yaitu 109. Perbedaan jumlah anomali yang terdeteksi oleh setiap metode ini menunjukkan bahwa pendekatan yang berbeda memiliki sensitivitas yang bervariasi terhadap pola dan penyimpangan dalam data.

Performa Model

Untuk memberikan gambaran lebih mendalam tentang kinerja masing-masing model, confusion matrix pada gambar 6 menunjukkan distribusi prediksi yang benar dan salah untuk kelas normal dan anomali.



Gambar 6. confusion matrix

Setelah menganalisis confusion matrix, dilanjutkan dengan membahas metrik performa utama yang digunakan untuk mengevaluasi model, yaitu AUC, precision, recall, dan F1-score. Tabel 2 merangkum nilai metrik performa untuk masing-masing model yang telah diuji. Autoencoder mencapai skor AUC dan F1 score tertinggi, diikuti oleh K-means dan Isolation Forest.

Table 2. Tabel evaluasi matrik

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

Tabel 2 menunjukkan kinerja yang luar biasa di semua model. Autoencoder mencapai skor AUC tertinggi sebesar 0,9823 yang menunjukkan kemampuannya yang unggul dalam menangkap pola temporal yang kompleks dan hubungan non-linear dalam data kualitas udara melalui arsitektur multi-layernya. K-means menunjukkan kinerja yang sangat kuat

BIBLIOGRAPHY

[1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A Review on Outlier/Anomaly Detection in Time Series Data," Jun. 01, 2021, *Association for Computing Machinery*. doi: 10.1145/3444690.

dengan AUC 0,9778 menunjukkan keefektifannya meskipun pendekatannya lebih sederhana berdasarkan metrik jarak. Isolation Forest menunjukkan kinerja yang kompetitif dengan AUC sebesar 0,9601 memvalidasi efisiensinya dalam menangani data deret waktu berdimensi tinggi. Semua model mencapai nilai presisi sempurna 1.0, menunjukkan akurasi yang luar biasa dalam mengidentifikasi anomali yang benar tanpa *false positives*.

CONCLUSION

Penelitian ini membandingkan kinerja K-means, Isolation Forest, dan Autoencoder dalam mendeteksi anomali pada data kualitas udara yang tidak seimbang. Hasil penelitian menunjukkan bahwa K-Means mampu membedakan data normal dan abnormal dengan akurasi mencapai 97,78%, IForest dengan akurasi 96,01%, AE dan AE LSTM dengan akurasi hingga 98,23%. Metode-metode ini menunjukkan hasil yang sangat baik, dengan semua akurasi deteksi anomali yang di atas 90% pada dataset lingkungan.

Autoencoder dan Autoencoder LSTM tampaknya menjadi model dengan performa terbaik secara keseluruhan untuk tugas ini, karena memiliki nilai AUC dan F1-score yang tergolong tinggi. Autoencoder mengungguli metode lainnya, menunjukkan ketangguhannya dalam menangani pola yang kompleks. Temuan ini memberikan dasar untuk memilih metode deteksi anomali yang tepat untuk aplikasi dunia nyata, terutama dalam pemantauan lingkungan. Penelitian di masa depan dapat mengeksplorasi pendekatan hibrida untuk meningkatkan akurasi deteksi lebih lanjut.

[2] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering - A decade review," *Inf Syst*, vol. 53, pp. 16–38, May 2015, doi: 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," *Alexandria Engineering Journal*, vol. 82, pp. 304–

- 322, Nov. 2023, doi: 10.1016/j.aej.2023.09.070.
- [4] M. B. Shrestha and G. R. Bhatta, "Selecting appropriate methodological framework for time series data analysis," *Journal of Finance and Data Science*, vol. 4, no. 2, pp. 71–89, Jun. 2018, doi: 10.1016/j.jfds.2017.11.001.
- [5] R. J. Hyndman and B. Rostami-Tabar, "Forecasting interrupted time series," *Journal of the Operational Research Society*, 2024, doi: 10.1080/01605682.2024.2395315.
- [6] R. Kablaoui, I. Ahmad, S. Abed, and M. Awad, "Network traffic prediction by learning time series as images," *Engineering Science and Technology, an International Journal*, vol. 55, Jul. 2024, doi: 10.1016/j.jestch.2024.101754.
- [7] R. Rahman and Nurul Hikma, "Keamanan Jaringan Kecerdasan Buatan dan Implementasi Solusi Keamanan," *Technology Sciences Insights Journal*, vol. 1(1), pp. 33–36, 2024.
- [8] E. A. Fadlilah, "Identifikasi Anomali Data Akademik Menggunakan Dbscan Outlier Detection," *Prosiding Sains Nasional dan Teknologi*, vol. 12, no. 1, p. 336, Nov. 2022, doi: 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," *Jurnal Ilmiah Kursor*, vol. 12, no. 2, pp. 59–68, Dec. 2023, doi: 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. doi: 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 Journal for Engineering and Technology*, vol. 2, no. 2, pp. 152–164, May 2023, doi: 10.55145/ajest.2023.02.02.018.
- [12] M. Čampulová, J. Michálek, P. Mikuška, and D. Bokál, "Nonparametric algorithm for identification of outliers in environmental data," *J Chemom*, vol. 32, no. 5, May 2018, doi: 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)*, IEEE, Jan. 2022, pp. 16–21. doi: 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, Nov. 2023, doi: 10.3390/s23229272.
- [15] Q. Liu, P. Boniol, T. Palpanas, and J. Paparrizos, "Time-Series Anomaly Detection: Overview and New Trends," *Time-Series Anomaly Detection: Overview and New Trends. PVLDB*, vol. 17, no. 12, pp. 4229–4232, 2024, doi: 10.14778/3685800.3685842.
- [16] Z. Z. Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," Nov. 2022, doi: 10.1145/3691338.
- [17] A. Priarone, U. Albertin, C. Cena, M. Martini, and M. Chiaberge, "Unsupervised Novelty Detection Methods Benchmarking with Wavelet Decomposition," Sep. 2024, [Online]. Available: <http://arxiv.org/abs/2409.07135>
- [18] Y. Pratama, E. Sulistianingsih, N. N. Debatara, and N. Imro'ah, "K-Means Clustering dan Mean Variance Efficient Portfolio dalam Portofolio Saham," *Jambura Journal of Probability and Statistics*, vol. 5, no. 1, pp. 24–30, Jun. 2024, doi: 10.37905/jjps.v5i1.20298.
- [19] 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," *Jurnal Ilmiah Kursor*, vol. 12, no. 1, pp. 1–10, Jun. 2023, doi: 10.21107/kursor.v12i01.269.
- [20] S. Wang, "Isolation Forest Anomaly Detection Algorithm Based On Multi-level Sub-subspace Partition," *International Journal of Computer Science and Information Technology*, vol. 4, no. 2, pp. 149–159, Oct. 2024, doi: 10.62051/ijcsit.v4n2.20.
- [21] W. Skaf and T. Horváth, "Denoising Architecture for Unsupervised Anomaly Detection in Time-Series," 2022, pp. 178–187. doi: 10.1007/978-3-031-15743-1_17.