

Analisis Hasil Imputasi Menggunakan Arsitektur Imputasi Autoencoder

^{1*} Arius Satoni Kurniawansyah

¹Dosen Tetap Program Studi Informatika Fakultas Ilmu Komputer Universitas Dehasen Bengkulu
Alamat Jalan Merawan No.20 Sawah Lebar, Kota Bengkulu.indonesia;

e-mail: ariussatoni@unived.ac.id

Program Studi Informatika Fakultas Ilmu Komputer Universitas Dehasen Bengkulu
Jl. Meranti Raya No. 32 Kota Bengkulu Kode Pos 38228 Telp. (0736) 22027, 26957 Fax. (0736) 341139

(Received: Nopember 2025, Revised: Februari 2026, Accepiend: April 2026)

Abstract— Missing values in multivariate time series data are a critical issue in many domains, especially in healthcare datasets such as MIMIC-IV. This study aims to analyze the performance of imputation results using an Autoencoder-based architecture. Autoencoder is a deep learning model capable of learning data representations and reconstructing missing values through latent feature extraction. The research methodology includes data preprocessing, missing value simulation, model training, and evaluation using metrics such as MAE, RMSE, and R^2 . The results show that Autoencoder-based imputation provides competitive performance in reconstructing missing values, particularly in nonlinear and complex patterns. However, the model's performance depends on the proportion of missing data and network architecture design. This study contributes to understanding the effectiveness of Autoencoder in multivariate time series imputation and provides a baseline for further development using hybrid models.

Keywords: Imputation, Autoencoder, Time Series, Missing Value, Deep Learning.

Intisari—Data deret waktu multivariat sering digunakan dalam berbagai bidang seperti kesehatan, keuangan, dan industri. Salah satu tantangan utama dalam pengolahan data tersebut adalah adanya missing value yang dapat mengurangi kualitas analisis dan prediksi. Pada dataset medis seperti MIMIC-IV, missing value sering terjadi akibat keterbatasan alat atau kesalahan pencatatan. Oleh karena itu, diperlukan metode imputasi yang mampu mengisi nilai yang hilang secara akurat. Metode tradisional seperti mean imputation dan KNN memiliki keterbatasan dalam menangkap pola kompleks. Oleh karena itu, pendekatan deep learning seperti Autoencoder menjadi alternatif yang menjanjikan karena mampu mempelajari representasi data secara nonlinier. Penelitian ini bertujuan untuk menganalisis performa imputasi menggunakan Autoencoder serta mengevaluasi keunggulannya dibandingkan pendekatan konvensional.

Kata Kunci: Imputasi, Autoencoder, Time Series, Missing Value, Deep Learning

I. PENDAHULUAN

Data deret waktu multivariat sering digunakan dalam berbagai bidang seperti kesehatan, keuangan, dan industri. Salah satu tantangan utama dalam pengolahan data tersebut adalah adanya missing value yang dapat mengurangi kualitas analisis dan prediksi. Pada dataset medis seperti MIMIC-IV, missing value sering terjadi akibat keterbatasan alat atau kesalahan pencatatan.

Oleh karena itu, diperlukan metode imputasi yang mampu mengisi nilai yang hilang secara akurat. Metode tradisional seperti mean imputation dan KNN memiliki keterbatasan dalam menangkap pola kompleks. Oleh karena itu, pendekatan deep learning seperti Autoencoder menjadi alternatif yang menjanjikan karena mampu mempelajari representasi data secara nonlinier.

Penelitian ini bertujuan untuk menganalisis performa imputasi menggunakan Autoencoder serta mengevaluasi keunggulannya dibandingkan pendekatan konvensional.

II. TINJAUAN PUSTAKA

A. Missing Value pada Time Series

Missing value merupakan kondisi di mana sebagian data tidak tersedia dalam dataset. Dalam time series, missing value dapat bersifat acak (MCAR), bergantung (MAR), atau tidak acak (MNAR).

B. Metode Imputasi

Beberapa metode imputasi yang umum digunakan antara lain:

- 1) Mean/Median Imputation
- 2) K-Nearest Neighbor (KNN)
- 3) Multiple Imputation (MICE)
- 4) Deep Learning-based Imputation

C. Autoencoder

Autoencoder adalah model neural network yang terdiri dari encoder dan decoder. Encoder memetakan input ke ruang laten, sedangkan decoder merekonstruksi kembali data.

Fungsi utama Autoencoder dapat dituliskan sebagai:

Encoder:

$$z = f(x)$$

Decoder:

$$\hat{x} = g(z)$$

Tujuan pelatihan adalah meminimalkan error rekonstruksi antara x dan \hat{x} .

III. METODOLOGI PENELITIAN

A. Dataset

Dataset yang digunakan adalah MIMIC-IV Vital Signs yang terdiri dari:

- a) Heart Rate
- b) SBP
- c) DBP
- d) Respiratory Rate
- e) Temperature
- f) SpO2
- g) Glucose

B. Preprocessing Data

Tahapan preprocessing meliputi:

1. Normalisasi data
2. Penanganan outlier
3. Simulasi missing value (10%, 20%, 30%)
4. Pembagian data train dan test

C. Arsitektur Autoencoder

Arsitektur yang digunakan terdiri dari:

- 1) Input Layer
- 2) Hidden Layer (Encoder)
- 3) Bottleneck Layer
- 4) Hidden Layer (Decoder)
- 5) Output Layer

D. Evaluasi Model

Evaluasi dilakukan menggunakan:

- 1) Mean Absolute Error (MAE)
- 2) Root Mean Square Error (RMSE)
- 3) Coefficient of Determination (R^2)

IV. HASIL DAN PEMBAHASAN

A. Hasil

Hasil menunjukkan bahwa Autoencoder mampu merekonstruksi data dengan cukup baik pada missing rate rendah ($\leq 20\%$). Pada missing rate tinggi ($> 30\%$), performa mulai menurun.

Tabel 1. Hasil Evaluasi Imputasi Autoencoder

Missing Rate	MAE	RMSE	R ²
10%	0.021	0.034	0.982
20%	0.035	0.052	0.961
30%	0.058	0.079	0.923

B. Pembahasan

Analisis Performa

- 1) MAE menunjukkan error absolut yang relatif kecil
- 2) RMSE lebih sensitif terhadap outlier
- 3) R² menunjukkan kemampuan model menjelaskan variansi data

Hasil eksperimen menunjukkan bahwa performa Autoencoder mengalami penurunan seiring meningkatnya missing rate.

1. **MAE (Mean Absolute Error)**
Nilai MAE meningkat dari 0.021 menjadi 0.058, menunjukkan bahwa kesalahan prediksi semakin besar ketika data yang hilang meningkat.
2. **RMSE (Root Mean Square Error)**
RMSE meningkat lebih signifikan dibanding MAE, yang menunjukkan sensitivitas terhadap outlier pada data imputasi.
3. **R² (Koefisien Determinasi).**
Nilai R² tetap tinggi (>0.9), yang menunjukkan bahwa model masih mampu menjelaskan variansi data dengan baik.

Analisis Berdasarkan Variabel

Tabel 2. MAE per Variabel (Missing Rate 20%)

Variabel	MAE
Heart Rate	0.028
SBP	0.041
DBP	0.039
Resp Rate	0.033
Temperature	0.021
SpO2	0.018
Glucose	0.052

Kelebihan Autoencoder

1. Mampu menangkap pola nonlinier
2. Cocok untuk data multivariat
3. Fleksibel dalam arsitektur

Eksperimen dilakukan menggunakan dataset **MIMIC-IV Vital Sign** dengan variabel: heart_rate, sbp, dbp, resp_rate, temperature, spo2, dan glucose.

Simulasi missing value dilakukan dengan skenario:

- a) 10% (rendah)
- b) 20% (sedang)
- c) 30% (tinggi)

Model Autoencoder dilatih dengan konfigurasi:

- a) Hidden layer: 64 – 32 – 64
- b) Aktivasi: ReLU

- c) Optimizer: Adam
- d) Loss function: Mean Squared Error (MSE)
- e) Epoch: 100

Visualisasi Hasil Imputasi

Secara umum, hasil rekonstruksi menunjukkan bahwa:

- a) Pola data asli dan hasil imputasi hampir identik pada missing rate rendah
- b) Pada missing rate tinggi, mulai terlihat deviasi terutama pada nilai ekstrem
- c) Autoencoder mampu menjaga trend utama data time series

Perbandingan dengan Metode Lain (Baseline)

Tabel III. Perbandingan Metode Imputasi (Missing Rate 20%)

Metode	MAE	RMSE	R ²
Mean	0.087	0.121	0.812
KNN	0.061	0.089	0.884
MICE	0.049	0.072	0.912
Autoencoder	0.035	0.052	0.961

Dari Tabel tersebut didapatkan hasil:

- 1) Autoencoder outperform semua metode tradisional
- 2) Peningkatan signifikan pada R² (~5–15%)
- 3) Cocok untuk data kompleks multivariat

Jika dikaitkan dengan riset yang lain:

1. Posisi Autoencoder

Sebagai baseline deep learning dan Tidak eksplisit menangkap temporal dependency.

2. Dibanding STING

STING unggul karena: Ada GRU (temporal learning) dan Ada Self-Attention, sedangkan Autoencoder hanya rekonstruksi statis.

3. Dibanding Kernel

Kernel unggul pada Klasifikasi dan Struktur data non-linear berbasis similarity.

V. PENUTUP

A. Kesimpulan

Autoencoder merupakan metode yang efektif untuk imputasi data deret waktu multivariat, terutama dalam menangani pola data yang kompleks. Hasil penelitian menunjukkan performa yang baik pada tingkat missing value rendah hingga sedang.

B. Saran

Penelitian selanjutnya dapat mengembangkan sebuah kolaborasi Arsitektur Imputasi antar Hybrid Autoencoder dan Attention

DAFTAR PUSTAKA

- [1] Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016.
- [2] Vincent, P., et al., "Stacked Denoising Autoencoders," JMLR, 2010.

- [3] Che, Z., et al., “Recurrent Neural Networks for Multivariate Time Series,” *Scientific Reports*, 2018.
- [4] Yoon, J., et al., “GAIN: Missing Data Imputation using GAN,” *ICML*, 2018.
- [5] Little, R. J. A., & Rubin, D. B., *Statistical Analysis with Missing Data*, Wiley, 2019.