UNIVERSITAS
GADJAH MADA
Model Hybrid MICE-ElasticNet dan Denoising Diffusion Implicit Model untuk Imputasi Data Tabular Numerik
RIKA FEBRIANTI, Prof. Dr. Ing. Mhd. Reza MI Pulungan, S.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

MODEL *HYBRID MICE-ELASTICNET* DAN *DENOISING DIFFUSION IMPLICIT MODEL* UNTUK IMPUTASI DATA TABULAR NUMERIK

Oleh

Rika Febrianti
24/546602/PPA/06847

Metode berbasis difusi untuk imputasi data tabular numerik menghadapi *trade-off* antara ketepatan rekonstruksi nilai hilang dan preservasi distribusi. Penelitian ini mengusulkan kerangka *Hybrid* MICE-*ElasticNet* + *Denoising Diffusion Implicit Model* (DDIM), dengan hasil MICE-*ElasticNet* sebagai estimasi awal dan panduan kondisional selama *reverse denoising*. Evaluasi dilakukan pada lima dataset *benchmark* melalui 45 skenario MCAR, MAR, dan MNAR dengan *missing rate* 20%, 40%, dan 60%. Kinerja diukur menggunakan RMSE, MAE, R^2, *Wasserstein distance*, CorrFrob, dan MSEDS untuk menilai rekonstruksi, distribusi, korelasi, dan utilitas prediktif. MICE-m5 memberikan hasil terbaik untuk rekonstruksi dan utilitas, dengan RMSE terendah pada 35 skenario dan MSEDS terendah pada 36 skenario. DDIM-*only* memperoleh *Wasserstein distance* terendah pada 17 skenario. *Hybrid* menurunkan rata-rata RMSE DDIM-*only* dari 2,4103 menjadi 1,3374 dan mengunggulinya pada 44 dari 45 skenario RMSE, seluruh skenario CorrFrob, serta 43 skenario MSEDS. Temuan ini menunjukkan bahwa kesamaan distribusi tidak selalu berhubungan langsung dengan akurasi rekonstruksi, sedangkan integrasi estimasi berbasis regresi dengan model difusi dapat meningkatkan keseimbangan antara rekonstruksi nilai dan preservasi struktur data.

Kata kunci: imputasi data, *diffusion model*, MICE, evaluasi multi-dimensi

UNIVERSITAS
GADJAH MADA
Model Hybrid MICE-ElasticNet dan Denoising Diffusion Implicit Model untuk Imputasi Data Tabular Numerik
RIKA FEBRIANTI, Prof. Dr. Ing. Mhd. Reza MI Pulungan, S.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

A HYBRID MICE-*ELASTICNET* AND DENOISING DIFFUSION IMPLICIT MODEL FOR TABULAR DATA IMPUTATION

Oleh

Rika Febrianti
24/546602/PPA/06847

Diffusion-based methods for numerical tabular data imputation face a trade-off between the accurate reconstruction of missing values and the preservation of data distributions. This study proposes a Hybrid MICE-ElasticNet + Denoising Diffusion Implicit Model (DDIM) framework, in which MICE-ElasticNet imputation results serve as both the initial estimates and conditional guidance during reverse denoising. The framework was evaluated on five benchmark datasets across 45 scenarios covering MCAR, MAR, and MNAR mechanisms with missing rates of 20%, 40%, and 60%. Performance was assessed using RMSE, MAE, R^2, Wasserstein distance, CorrFrob, and MSEDS to evaluate reconstruction accuracy, distribution preservation, correlation preservation, and predictive utility. MICE-m5 achieved the best overall performance in reconstruction and predictive utility, obtaining the lowest RMSE in 35 scenarios and the lowest MSEDS in 36 scenarios. DDIM-only achieved the lowest Wasserstein distance in 17 scenarios. The Hybrid model reduced the average RMSE of DDIM-only from 2.4103 to 1.3374 and outperformed it in 44 of 45 RMSE scenarios, all CorrFrob scenarios, and 43 MSEDS scenarios. These findings indicate that distributional similarity is not always directly associated with reconstruction accuracy, while integrating regression-based estimates with a diffusion model can improve the balance between value reconstruction and the preservation of data structure.

Keywords: missing values, imputation, diffusion model, MICE, multidimensional evaluation