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/

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN ii
HALAMAN PERNYATAAN KARYA ILMIAH iii
PRAKATA iv
DAFTAR ISI v
DAFTAR GAMBAR viii
DAFTAR TABEL x
INTISARI xi
ABSTRACT xii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Batasan Penelitian 4
1.4 Tujuan Penelitian 5
1.5 Manfaat Penelitian 5
1.6 Kontribusi Penelitian 5
BAB II TINJAUAN PUSTAKA 7
BAB III METODOLOGI PENELITIAN 15
3.1 Analisis Permasalahan 15
3.2 Arsitektur Model Imputasi 15
3.3 Akuisisi Dataset dan Analisis Karakteristik Data 17
3.4 *Data Preprocessing* 18
3.4.1 *Data Partitioning* 18
3.4.2 Stabilisasi *Outlier* 19
3.4.3 *Quantile Transformer* 20
3.5 Simulasi *Missing Value* 21
3.5.1 *Missing Completely at Random* 22
3.5.2 *Missing at Random* 22
3.5.3 *Missing Not at Random* 24
3.6 MICE-*ElasticNet* Model 25
3.7 DDIM *Model* 30
3.7.1 Pembentukan Representasi Terpusat 32
3.7.2 Proses *Forward Diffusion* 32
3.7.3 Prediksi *Noise* 33
3.7.4 Fungsi *Loss* 33
3.7.5 *Partial Noising* dan *Reverse Denoising* 35

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/

3.7.6 *Observed-Lock* 36
3.8 *Hybrid* MICE-*ElasticNet* + DDIM 39
3.9 Pengujian dan Eksperimen 41
3.9.1 Skenario Eksperimen 42
3.9.2 *Hyperparameter Tuning* MICE-*ElasticNet* 43
3.9.3 *Hyperparameter Tuning* DDIM-*only* 43
3.9.4 *Hyperparameter Tuning Hybrid*-DDIM 44
3.9.5 *Final Multi-run Experiment* 45
3.10 Evaluasi 46
3.10.1 *Reconstruction Error* 47
3.10.2 *Distributional Fidelity* 49
3.10.3 *Correlation Preservation* 50
3.10.4 *Downstream Utility* 52
BAB IV IMPLEMENTASI 54
4.1 Lingkungan dan Konfigurasi Eksperimen 54
4.2 *Splitting Data* dan *Preprocessing Data* 54
4.2.1 Implementasi Stabilisasi *Outlier* 55
4.2.2 Implementasi *Quantile Transformer* 56
4.3 Pembentukan Simulasi *Missing Value* 57
4.3.1 Implementasi MCAR 59
4.3.2 Implementasi MAR 60
4.3.3 Implementasi MNAR 61
4.4 Implementasi Pembangunan Model MICE-*ElasticNet* 63
4.4.1 Implementasi *Hyperparameter Tuning* MICE-*ElasticNet* 68
4.4.2 Implementasi Imputasi Model MICE-*ElasticNet* 70
4.5 Implementasi Pembangunan Model DDIM 73
4.5.1 Implementasi Arsitektur DDIM *Imputer* 75
4.5.2 Implementasi Proses *Training* DDIM 76
4.5.3 Implementasi *Reverse Denoising* dan *Observed-Lock* 80
4.5.4 Implementasi *Hyperparameter Tuning* DDIM-*only* 83
4.5.5 Implementasi *Final Multi-run* DDIM-*only* 85
4.6 Implementasi Pembangunan Model *Hybrid* MICE-*ElasticNet* + DDIM 88
4.6.1 Implementasi Input *Hybrid* dari Hasil MICE 88
4.6.2 Implementasi *Refinement* dan *Final Multi-run* Hybrid 89

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/

4.7 Implementasi Evaluasi Empat Dimensi 91
4.7.1 Implementasi *Reconstruction Error* 92
4.7.2 Implementasi *Distributional Fidelity* 93
4.7.3 Implementasi *Correlation Preservation* 94
4.7.4 Implementasi *Downstream Utility* 95
BAB V HASIL PENELITIAN DAN PEMBAHASAN 98
5.1 Desain Eksperimen dan Evaluasi 98
5.2 Hasil Pemilihan *Hyperparameter* 99
5.3 Evaluasi Multidimensi 102
5.3.1 *Reconstruction Accuracy* 109
5.3.2 *Distributional Fidelity* 111
5.3.3 *Correlation Preservation* 118
5.3.4 *Downstream Utility* 121
5.4 Analisis Waktu Komputasi 124
5.5 Analisis Perilaku dan *Trade-off* Metode 125
5.6 *Trade-off* antara Rekonstruksi dan Distribusi 127
5.7 Pengaruh *Missing Rate* terhadap Kinerja Empat Dimensi 129
5.8 Pengaruh Mekanisme *Missing Value* terhadap Perilaku Metode 132
5.9 Kontribusi Estimasi Awal MICE pada Hybrid 136
5.10 Perbandingan Performa Model dengan Penelitian Sebelumnya 140
5.11 Keterbatasan Penelitian 144
BAB VI KESIMPULAN 146
6.1 Kesimpulan Penelitian 146
6.2 Saran 146
DAFTAR PUSTAKA 148

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/

DAFTAR GAMBAR

Gambar 3. 1 Arsitektur model 17
Gambar 3. 2 *Multidimensional evaluations metrics* 46
Gambar 4. 1 Implementasi *data partitioning* 55
Gambar 4. 2 Implementasi fungsi stabilisasi *outlier* 56
Gambar 4. 3 Implementasi *Quantile Transformer* pada *final pipeline* 57
Gambar 4. 4 Implementasi fungsi statistik acuan simulasi *missing value* 58
Gambar 4. 5 Implementasi perhitungan jumlah sel *missing* 58
Gambar 4. 6 Implementasi pembentukan *df_miss* dan *mask_df* 59
Gambar 4. 7 Implementasi mekanisme MCAR 59
Gambar 4. 8 Implementasi pembentukan skor MAR 60
Gambar 4. 9 Implementasi kalibrasi probabilitas MAR 61
Gambar 4. 10 Implementasi pembentukan skor MNAR 62
Gambar 4. 11 Implementasi kalibrasi probabilitas MNAR 62
Gambar 4. 12 Implementasi fungsi regresi satu variabel pada MICE-*ElasticNet* 64
Gambar 4. 13 Implementasi *stochastic initialization* pada MICE-*ElasticNet* 65
Gambar 4. 14 Implementasi imputasi iteratif MICE-*ElasticNet* pada data *train* .. 67
Gambar 4. 15 Implementasi pemantauan konvergensi MICE-*ElasticNet* 67
Gambar 4. 16 Implementasi ruang pencarian hyperparameter MICE-*ElasticNet* 68
Gambar 4. 17 Implementasi evaluasi kombinasi *hyperparameter* MICE 69
Gambar 4. 18 Implementasi pemilihan konfigurasi terbaik MICE-*ElasticNet* 69
Gambar 4. 19 Implementasi imputasi final MICE-*ElasticNet* m=1 70
Gambar 4. 20 Implementasi cache MICE-*ElasticNet* pada jalur *tuning* 71
Gambar 4. 21 Implementasi konfigurasi MICE-m5 71
Gambar 4. 22 Implementasi eksekusi berulang MICE-m5 72
Gambar 4. 23 Implementasi agregasi rata-rata pada setiap sel MICE-m5 72
Gambar 4. 24 Implementasi imputasi final MICE-m5 73
Gambar 4. 25 Implementasi pembentukan *observed pool* DDIM-*only* 74
Gambar 4. 26 Implementasi *stochastic initialization* DDIM-*only* 74
Gambar 4. 27 Implementasi arsitektur utama DDIMImputer 75
Gambar 4. 28 Implementasi pembentukan *diffusion schedule* 76
Gambar 4. 29 Implementasi input fungsi *training* DDIM 77
Gambar 4. 30 Implementasi pembagian data *train* dan *validation* pada DDIM ... 77
Gambar 4. 31 Implementasi pembentukan residual pada proses *training* DDIM. 78
Gambar 4. 32 Implementasi *forward diffusion* menggunakan *q_sample()* 78
Gambar 4. 33 Implementasi pembentukan *noisy* residual dan prediksi *noise* 78
Gambar 4. 34 Implementasi prediksi *noise* pada jaringan DDIM 79
Gambar 4. 35 Implementasi loss pelatihan DDIM 79
Gambar 4. 36 Implementasi pembaruan parameter DDIM 79
Gambar 4. 37 Implementasi pemilihan model terbaik dan *early stopping* 80
Gambar 4. 38 Implementasi fungsi *denoise()* pada DDIM 80
Gambar 4. 39 Implementasi representasi terpusat dan *partial noising* 81
Gambar 4. 40 Implementasi pembentukan urutan *reverse timestep* 81
Gambar 4. 41 Implementasi pembaruan *timestep* pada DDIM 82
Gambar 4. 42 Implementasi reverse denoising dan observed-lock 82

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/

Gambar 4. 43 Implementasi pembentukan hasil akhir DDIM 83
Gambar 4. 44 Implementasi ruang pencarian *hyperparameter* DDIM-*only* 83
Gambar 4. 45 Implementasi *train* DDIM-*only* pada kombinasi *hyperparameter* 84
Gambar 4. 46 Implementasi pemilihan eta pada data validation 85
Gambar 4. 47 Implementasi pemilihan konfigurasi terbaik DDIM-*only* 85
Gambar 4. 48 Implementasi training final DDIM-only pada setiap run 86
Gambar 4. 49 Implementasi denoise final DDIM-only pada data train dan test . 87
Gambar 4. 50 Implementasi agregasi multi-run DDIM-*only* 87
Gambar 4. 51 Implementasi masukan Hybrid pada jalur tuning 88
Gambar 4. 52 Implementasi input final Hybrid dari hasil MICE 89
Gambar 4. 53 Implementasi *training* final *Hybrid* pada setiap *run* 90
Gambar 4. 54 Implementasi *denoise Hybrid* pada data *train* dan *test* 90
Gambar 4. 55 Implementasi agregasi multi-run Hybrid 91
Gambar 4. 56 Implementasi fungsi evaluasi satu metode dan satu skenario 92
Gambar 4. 57 Implementasi RMSE dan MAE pada posisi *missing* 93
Gambar 4. 58 Implementasi R^2 per fitur pada posisi *missing* 93
Gambar 4. 59 Implementasi *Wasserstein distance* data *test* 94
Gambar 4. 60 Implementasi *correlation preservation* menggunakan CorrFrob... 95
Gambar 4. 61 Implementasi skema *round-robin* pada *downstream utility* 96
Gambar 4. 62 Implementasi pembentukan matriks data *train* dan data *test* 96
Gambar 4. 63 Implementasi Ridge Regression dan MSE setiap target 96
Gambar 4. 64 Implementasi perhitungan rata-rata MSE 97
Gambar 5. 1 KDE gabungan hasil imputasi. (a) Dataset *Credit Card* 115
Gambar 5. 2 (b) Dataset CPU *Small* 115
Gambar 5. 3 (c) Dataset *Parkinsons* 116
Gambar 5. 4 (d) Dataset *Satellite* 116
Gambar 5. 5 (d) Dataset *Wine Quality* 117
Gambar 5. 6 Perubahan CorrFrob skenario. (a) Dataset *Credit Card* 119
Gambar 5. 7 (b) Dataset *CPU Small* 119
Gambar 5. 8 (c) Dataset *Parkinsons* 119
Gambar 5. 9 (d) Dataset *Satellite* 119
Gambar 5. 10 (e) Dataset *Wine Quality* 119

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/

DAFTAR TABEL

Tabel 2. 1 Studi *diffusion model* pada imputasi data tabular numerik 8
Tabel 2. 2 Lanjutan 9
Tabel 2. 3 Studi *diffusion model* pada imputasi data *time series* dan sensor 10
Tabel 2. 4 Studi *diffusion model* pada data *spatiotemporal* 11
Tabel 2. 5 Studi *diffusion model* pada aplikasi klinis dan *multi task* 12
Tabel 2. 6 Studi metode imputasi non diffusion 13
Tabel 3. 1 Dataset penelitian 18
Tabel 3. 2 *Hyperparameter tuning* MICE-*ElasticNet* 43
Tabel 3. 3 *Hyperparameter tuning* DDIM-*only* 44
Tabel 3. 4 *Hyperparameter tuning* DDIM 45
Tabel 5. 1 Dimensi evaluasi metode imputasi 99
Tabel 5. 2 Konfigurasi hyperparameter akhir pada setiap dataset 100
Tabel 5. 3 Hasil evaluasi multidimensi pada dataset *Credit Card* 103
Tabel 5. 4 Hasil evaluasi multidimensi pada dataset *CPU Small* 104
Tabel 5. 5 Hasil evaluasi multidimensi pada dataset *Parkinsons* 105
Tabel 5. 6 Hasil evaluasi multidimensi pada dataset *Satellite* 106
Tabel 5. 7 Hasil evaluasi multidimensi pada dataset *Wine Quality* 107
Tabel 5. 8 Rata-rata waktu komputasi eksperimen pada lima dataset 124
Tabel 5. 9 Contoh empiris hubungan antara RMSE dan *Wasserstein distance* .. 129
Tabel 5. 10 Perubahan rata-rata metrik dari *missing rate* 20% menjadi 60%..... 130
Tabel 5. 11 Rata-rata metrik berdasarkan mekanisme *missing value* 133
Tabel 5. 12 Perbandingan MICE, DDIM-*only*, dan *Hybrid* pada 45 skenario ... 137
Tabel 5. 13 Penurunan RMSE metode usulan terhadap *baseline* model difusi .. 141