

**PENINGKATAN AKURASI IMPUTASI DATA YANG HILANG
PADA DATA DERET WAKTU MULTIVARIAT
MENGUNAKAN *DEEP LEARNING***

SKRIPSI

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**



Oleh :

Yultriyen

09011282025075

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA**

2024

LEMBAR PENGESAHAN

PENINGKATAN AKURASI IMPUTASI DATA YANG HILANG PADA DATA DERET WAKTU MULTIVARIAT MENGUNAKAN *DEEP LEARNING*

SKRIPSI


Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh :
Yultriyen
09011282025075

Palembang, 21 Agustus 2024

Mengetahui,

Pembimbing I Tugas Akhir,



Prof. Dr. Ir. Bambang Tutuko, M.T.
NIP. 196001121989031002

Pembimbing II Tugas Akhir,



Dr. Firdaus, S.T., M.Kom.
NIP. 197801212008121003

Ketua Jurusan Sistem Komputer 21/8/24



Dr. Ir. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada:

Hari : Rabu

Tanggal : 10 Juli 2024

Tim Penguji :


1. Ketua : Sutarno, M.T.

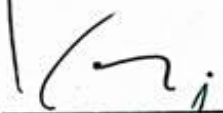
2. Sekretaris : Iman Saladin B. Azhar, M.MSI.

3. Penguji : Novi Yusliani, S. Kom., M.T.

4. Pembimbing 1 : Prof. Dr. Ir. Bambang Tutuko, M.T.


5. Pembimbing 2 : Dr. Firdaus, S.T., M.Kom.








15/8/2024



Mengetahui, 4/8/24

Ketua Jurusan Sistem Komputer




Dr. Ir. Sukemi, M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Yultriyen

NIM : 09011282025075

Judul : Peningkatan akurasi imputasi data yang hilang pada data deret waktu multivariat menggunakan *Deep Learning*.

Hasil Pengecekan Software Turnitin : 6%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Palembang, 2 Juli 2024

Mahasiswa,



Yultriyen

NIM.09011282025075

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas kasih karunia dan berkat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Peningkatan Akurasi Imputasi Data yang Hilang pada Data Deret Waktu Multivariat menggunakan *Deep Learning*”.

Selama pembuatan Tugas Akhir dan penulisan laporan Tugas Akhir ini, penulis banyak mendapatkan bantuan dan peran serta dari semua pihak, baik secara langsung maupun tak langsung. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Tuhan Yesus Kristus yang telah melimpahkan berkah serta nikmat kesehatan dan kesempatan kepada penulis dalam mengerjakan Tugas Akhir.
2. Kedua orang tua, saudara, dan keluarga besar yang telah mendoakan dan memberikan motivasi serta support.
3. Bapak Dr. Ir. Sukemi, M.T. Selaku Ketua Jurusan Sistem Komputer Universitas Sriwijaya.
4. Bapak Dr. Firdaus, S.T., M. Kom. Dan Prof. Dr. Ir. Bambang Tutuko, M.T. Selaku Pembimbing Tugas Akhir yang telah meluangkan waktu untuk membimbing, memberikan saran, dan motivasi serta ilmu yang pastinya berguna untuk penulis.
5. Ibu Prof. Dr. Ir. Siti Nurmaini, M.T., selaku Head of Intelligent System Research Group (ISysRG) yang telah memberikan kesempatan untuk menjadi bagian dari team research group ini.
6. Intelligent System Research Group (ISysRG) atas bantuan infrastruktur dalam menyelesaikan Tugas Akhir.
7. Bapak Abdurahman, S.Kom., M.Han., selaku Dosen Pembimbing Akademik Jurusan Sistem Komputer Universitas Sriwijaya.
8. Mbak Anggun Islami dalam membantu pembuatan Tugas Akhir ini.
9. Mba Ade Iriani Safitri, Mbak Annisa Darmawahyuni, dan kak Naufal Rachmatullah sebagai mentor di ISysRG.
10. Teman – teman di Serai Indah Indralaya dan Palembang dalam pengisi hari – hari penulis membuat Tugas Akhir.

11. Rekan seperjuangan di group ISysRG yang selalu membantu dalam pengerjaan Tugas Akhir.
12. Pak Angga selaku Administrasi Jurusan Sistem Komputer yang telah membantu dalam proses administrasi terkait Tugas Akhir.
13. Teman seperjuangan angkatan 2020 Sistem Komputer yang telah memberikan bantuan selama pengerjaan Tugas Akhir.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan dan keterbatasan. Oleh karena itu, penulis membutuhkan dukungan berupa kritik dan saran yang membangun agar penulis lebih baik di masa mendatang. Semoga skripsi ini dapat memberikan manfaat dan menjadi referensi serta tambahan informasi bagi pembaca.

Palembang, 2 Juli 2024

Mahasiswa,



Yultriyen

NIM.09011282025075

**Peningkatan Akurasi Imputasi Data yang Hilang pada Data Deret Waktu
Multivariat menggunakan *Deep Learning***

YULTRIYEN (09011282025075)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : yultriyeneyes@gmail.com

ABSTRAK

Data hilang merupakan masalah umum dan kompleks dalam dunia industri, sehingga proses pengolahan data menjadi lebih sulit. Metode imputasi, baik secara konvensional maupun menggunakan neural network, digunakan untuk mengatasi masalah ini dengan memperkirakan atau menghitung nilai yang hilang. Deep learning dipilih karena kemampuannya dalam menggali informasi tersembunyi dalam data, yang secara signifikan meningkatkan proses imputasi data. Pada penelitian ini terdapat tiga metode deep learning yang digunakan, yaitu LL-CNN, EDR-CNN, dan MIRNet. Kinerja metode ini dievaluasi berdasarkan root mean squared error (RMSE), mean absolute error (MAE), dan R-squared (R^2) pada delapan data yang berbeda, diantaranya: MIMIC-IV, MIMIC III, Beijing Multi Site Air Quality, Air Quality Itali, Air Quality India, US Pollution, Beijing pm2.5, dan Guangzhou. Hasil penelitian menunjukkan bahwa EDR-CNN memberikan kinerja terbaik pada kedelapan data yang telah disebutkan.

Kata kunci : Imputasi data, Data yang hilang, LL-CNN, EDR-CNN, MIRNet

*Improving Accuracy in Missing Data Imputation for Multivariate Time Series
Using Deep Learning*

YULTRIYEN (09011282025075)

Computer System Department, Computer Science Faculty, Sriwijaya

Email : yultriyeneyes@gmail.com

ABSTRACT

Missing data is a common and complex issue in the industrial world, making data processing more challenging. Imputation methods, whether conventional or using neural networks, are employed to address this issue by estimating or computing the missing values. Deep learning is chosen for its ability to unearth hidden information within data, significantly enhancing the data imputation process. This study utilizes three deep learning methods: LL-CNN, EDR-CNN, and MIRNet. The performance of these methods is evaluated based on root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2) on eight different datasets: MIMIC-IV, MIMIC III, Beijing Multi-Site Air Quality, Air Quality Italy, Air Quality India, US Pollution, Beijing PM2.5, and Guangzhou. The results of the study show that EDR-CNN provides the best performance across all eight datasets.

Keywords : *Data imputation, Missing data, LL-CNN, EDR-CNN, MIRNet*

DAFTAR ISI

LEMBAR PENGESAHAN.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PERNYATAAN.....	iv
KATA PENGANTAR.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Tujuan Penelitian.....	3
1.3. Rumusan Masalah.....	3
1.4. Batasan Masalah.....	3
1.5. Metodologi Penelitian.....	3
1.5.1. Studi Pustaka dan Literatur.....	4
1.5.2. Metode Konsultasi.....	4
1.5.3. Metode Pembuatan Model.....	4
1.5.4. Metode Pengujian dan Validasi.....	4
1.5.5. Metode Hasil dan Analisa.....	4
1.5.6. Metode Penarikan Kesimpulan dan Saran.....	4
1.6. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. <i>Database</i>	6
2.2. <i>Data Time Series</i>	7
2.2.1. <i>Trend</i>	8
2.2.2. <i>Seasonality</i>	8
2.2.3. <i>Cycles</i>	8
2.2.4. <i>Irregular Variation</i>	9
2.2.5. <i>Autocorrelation</i>	9
2.2.6. <i>Stationarity</i>	9
2.3. Imputasi Data.....	9
2.4. Imputasi Multivariat.....	10
2.4.1. <i>Low Light CNN</i>	11
2.4.2. <i>Enhanced Deep Residual CNN</i>	12
2.4.3. <i>Multi-Scale Residual Block</i>	13
2.5. <i>Learning Rate</i>	14
2.6. Evaluasi Model.....	15
2.6.1. Mean Absolute Error.....	15
2.6.2. Mean Squared Error.....	15
2.6.3. Root mean Squared Error.....	16
2.6.4. <i>R-Squared</i>	16
BAB III METODOLOGI PENELITIAN.....	17
3.1. Kerangka Kerja.....	17

3.2. Akuisisi Data	18
3.2.1. MIMIC-III	19
3.2.2. MIMIC-IV	20
3.2.3 Beijing Multi-Site Air Quality	21
3.2.4. <i>Air Quality</i> Itali	22
3.2.5. <i>Air Quality</i> India	25
3.2.6. <i>US Pollution</i>	26
3.2.7. Beijing pm2.5	27
3.2.8. Guangzhou	28
3.3. Pra-Pengolahan Data	29
3.4. Pembersihan Data	30
3.5. Normalisasi Data	31
3.6. <i>Splitting Data</i>	34
3.7. Segmenasi Data	35
3.7.1. Pembuatan Nilai Hilang Buatan	36
3.7.1.1. Data Label	37
3.7.1.2 Data Not a Number	37
3.7.1.3 Data Mask	38
3.7.2. Imputasi Nilai yang Hilang	38
3.8. Data Imputasi Multivariat	39
3.8.1. LL-CNN	39
3.8.2. EDR-CNN	44
3.8.3 MIRNet	48
3.9. Evaluasi Kinerja Model	51
BAB IV HASIL DAN ANALISA	52
4.1. Hasil Ekstraksi Data	52
4.2. Hasil Imputasi pada Penelitian tanpa nilai NaN	52
4.2.1. <i>Vital Sign</i> MIMIC-IV	52
4.2.1.1. Hasil Imputasi	53
4.2.2. <i>Vital Sign</i> MIMIC-III	63
4.2.2.1. Hasil Imputasi	63
4.2.3. Beijing Multi-Site Air Quality	72
4.2.3.1. Hasil Imputasi	73
4.2.4. <i>Air Quality</i> Italia	77
4.2.4.1. Hasil Imputasi	77
4.2.5. <i>Air Quality</i> India	81
4.2.5.1. Hasil Imputasi	81
4.2.6. <i>US Pollution</i>	85
4.2.6.1. Hasil Imputasi	85
4.2.7. Beijing pm2.5	88
4.2.7.1. Hasil Imputasi	89
4.2.8. Guangzhou	92
4.2.8.1. Hasil Imputasi	93
4.3. Analisis Imputasi data berdasarkan epoch 200	96
4.3.1. Hasil dan analisis menggunakan data MIMIC-IV pada setiap model	96
4.3.2. Hasil dan analisis menggunakan data MIMIC-III pada setiap model	99

4.3.3. Hasil dan analisis menggunakan data Beijing pada setiap model	102
4.3.4. Hasil dan analisis menggunakan data <i>Air Quality</i> Itali pada setiap model.....	103
4.3.5. Hasil dan analisis menggunakan data <i>Air Quality</i> India pada setiap model.....	105
4.3.6. Hasil dan analisis menggunakan data <i>US Pollution</i> pada setiap model	107
4.3.7. Hasil dan analisis menggunakan data Beijing pm2.5 pada setiap model	108
4.3.8. Hasil dan analisis menggunakan data Guangzhou pada setiap model.....	110
4.4. Analisis Imputasi data berdasarkan fitur pada arsitektur <i>Deep Learning</i>	111
4.4.1. MIMIC-IV	111
4.4.2. MIMIC-III	113
4.4.3. Beijing Multi-Site Air Quality.....	115
4.4.4. <i>Air Quality</i> Itali	116
4.4.5. <i>Air Quality</i> India	118
4.4.6. <i>US Pollution</i>	120
4.4.7. Beijing pm2.5	121
4.4.8. Guangzhou	123
4.5. Summary	125
BAB V KESIMPULAN.....	128
5.1. Kesimpulan	128
5.2. Saran	130
DAFTAR PUSTAKA.....	131

DAFTAR GAMBAR

Gambar 2.1. Arsitektur LL-CNN [28].....	11
Gambar 2.2. Arsitektur EDR-CNN [29].....	12
Gambar 2.3. Arsitektur MIRNet [30]	13
Gambar 3.1. Kerangka kerja	18
Gambar 3.2. Data Pre-processing.....	30
Gambar 3.3. Splitting Data.....	34
Gambar 3.4. Segmenasi Data	36
Gambar 3.5. Pembuatan Artificial NaN	37
Gambar 4.1. Menampilkan <i>bar chart</i> dari hasil penelitian sebelumnya [33].....	127
Gambar 4.2. Bar Chart hasil perbandingan penelitian terbaru	127

DAFTAR TABEL

Tabel 3.1. Data tanda vital pasien ICU MIMIC-III	20
Tabel 3.2. Data tanda vital pasien ICU MIMIC-IV	21
Tabel 3.3. Data Beijing Multi-Site Air Quality	22
Tabel 3.4. Data <i>Air Quality</i> Itali.....	24
Tabel 3.5. Data <i>Air Quality</i> India.....	25
Tabel 3.6. Data US Pollution	27
Tabel 3.7. Data Beijing pm2.5	27
Tabel 3.8. Data Guangzhou.....	29
Tabel 3.9. Data MIMIC-IV tanpa nilai NaN	31
Tabel 3.10. MinMax Scaler Data <i>Vital Sign</i> MIMIC-IV	32
Tabel 3.11. Robust Scaler Data Beijing pm2.5	33
Tabel 3.13. <i>Hyperparameter</i> yang akan digunakan pada data MIMIC-IV	40
Tabel 3.14. <i>Hyperparameter</i> yang akan digunakan pada data MIMIC-III	40
Tabel 3.15. <i>Hyperparameter</i> yang akan digunakan pada data Beijing <i>Multi-Site Air Quality</i>	41
Tabel 3.16. <i>Hyperparameter</i> yang akan digunakan pada data <i>Air Quality</i> Itali.....	41
Tabel 3.17. <i>Hyperparameter</i> yang akan digunakan pada data <i>Air Quality</i> India.....	42
Tabel 3.18. <i>Hyperparameter</i> yang akan digunakan pada data US <i>Pollution</i>	42
Tabel 3.19. <i>Hyperparameter</i> yang akan digunakan pada data Beijing pm2.5	43
Tabel 3.20. <i>Hyperparameter</i> yang akan digunakan pada data Guangzhou.....	43
Tabel 3.21. <i>Hyperparameter</i> yang akan digunakan pada data MIMIC-IV	44
Tabel 3.22. <i>Hyperparameter</i> yang akan digunakan pada data MIMIC-III	44
Tabel 3.23. <i>Hyperparameter</i> yang akan digunakan pada data Beijing <i>Multi-Site Air Quality</i>	45
Tabel 3.24. <i>Hyperparameter</i> yang akan digunakan pada data <i>Air Quality</i> Itali.....	45
Tabel 3.25. <i>Hyperparameter</i> yang akan digunakan pada data <i>Air Quality</i> India.....	46
Tabel 3.26. <i>Hyperparameter</i> yang akan digunakan pada data US <i>Pollution</i>	46
Tabel 3.27. <i>Hyperparameter</i> yang akan digunakan pada data Beijing pm2.5	47
Tabel 3.28. <i>Hyperparameter</i> yang akan digunakan pada data Guangzhou.....	47
Tabel 3.29. <i>Hyperparameter</i> yang akan digunakan pada data MIMIC-IV	48
Tabel 3.30. <i>Hyperparameter</i> yang akan digunakan pada data MIMIC-III	48
Tabel 3.31. <i>Hyperparameter</i> yang akan digunakan pada data Beijing <i>Multi-Site Air Quality</i>	49
Tabel 3.32. <i>Hyperparameter</i> yang akan digunakan pada data <i>Air Quality</i> Itali.....	49
Tabel 3.33. <i>Hyperparameter</i> yang akan digunakan pada data <i>Air Quality</i> India.....	50
Tabel 3.34. <i>Hyperparameter</i> yang akan digunakan pada data US <i>Pollution</i>	50
Tabel 3.35. <i>Hyperparameter</i> yang akan digunakan pada data Beijing pm2.5	50
Tabel 3.36. <i>Hyperparameter</i> yang akan digunakan pada data Guangzhou.....	51
Tabel 4.1 Data <i>Vital Sign</i> MIMIC-IV.....	53
Tabel 4.2 Hasil RMSE dari data MIMIC-IV	54
Tabel 4.3 Hasil MAE dari data MIMIC-IV	57
Tabel 4.4 Hasil R ² dari data MIMIC-IV	59
Tabel 4.5 Data <i>Vital Sign</i> MIMIC-III.....	63

Tabel 4.6 Hasil RMSE dari data MIMIC-III	64
Tabel 4.7 Hasil MAE dari data MIMIC-III	66
Tabel 4.8 Hasil R ² dari data MIMIC-III.....	70
Tabel 4.9 Data Beijing <i>Multi-Site Air Quality</i>	73
Tabel 4.10 Hasil RMSE dari data Beijing	74
Tabel 4.11 Hasil MAE dari data Beijing	74
Tabel 4.12 Hasil R ² dari data Beijing.....	76
Tabel 4.13 Data <i>Air Quality Itali</i>	77
Tabel 4.14 Hasil RMSE dari data <i>Air Quality Itali</i>	77
Tabel 4.15 Hasil MAE dari data <i>Air Quality Itali</i>	79
Tabel 4.16 Hasil R ² dari data <i>Air Quality Itali</i>	80
Tabel 4.17 Data <i>Air Quality India</i>	81
Tabel 4.18 Hasil RMSE dari data <i>Air Quality India</i>	82
Tabel 4.19 Hasil MAE dari data <i>Air Quality India</i>	83
Tabel 4.20 Hasil R ² dari data <i>Air Quality India</i>	84
Tabel 4.21 Data <i>US Pollution</i>	85
Tabel 4.22 Hasil RMSE dari data <i>US Pollution</i>	86
Tabel 4.23 Hasil MAE dari data <i>US Pollution</i>	86
Tabel 4.24 Hasil R ² dari data <i>US Pollution</i>	87
Tabel 4.25 Data Beijing pm2.5	88
Tabel 4.26 Hasil RMSE dari data Beijing pm2.5	89
Tabel 4.27 Hasil MAE dari data Beijing pm2.5	90
Tabel 4.28 Hasil R ² dari data Beijing pm2.5.....	91
Tabel 4.29 Data Guangzhou.....	92
Tabel 4.30 Hasil RMSE dari data Guangzhou.....	93
Tabel 4.31 Hasil MAE dari data Guangzhou.....	94
Tabel 4.32 Hasil R ² dari data Guangzhou	95
Tabel 4.33 Hasil RMSE terbaik dari data MIMIC-IV	96
Tabel 4.34 Hasil MAE terbaik dari data MIMIC-IV	97
Tabel 4.35 Hasil R ² terbaik dari data MIMIC-IV.....	98
Tabel 4.36 Hasil RMSE terbaik dari data MIMIC-III	99
Tabel 4.37 Hasil MAE terbaik dari data MIMIC-III	100
Tabel 4.38 Hasil R ² terbaik dari data MIMIC-III.....	101
Tabel 4.39 Hasil RMSE terbaik dari data Beijing	102
Tabel 4.40 Hasil MAE terbaik dari data Beijing	102
Tabel 4.41 Hasil R ² terbaik dari data Beijing.....	103
Tabel 4.42 Hasil RMSE terbaik dari data <i>Air Quality Itali</i>	104
Tabel 4.43 Hasil MAE terbaik dari data <i>Air Quality Itali</i>	104
Tabel 4.44 Hasil R ² terbaik dari data <i>Air Quality Itali</i>	105
Tabel 4.45 Hasil RMSE terbaik dari data <i>Air Quality India</i>	105
Tabel 4.46 Hasil MAE terbaik dari data <i>Air Quality India</i>	106
Tabel 4.47 Hasil R ² terbaik dari data <i>Air Quality India</i>	106
Tabel 4.48 Hasil RMSE terbaik dari data <i>US Pollution</i>	107
Tabel 4.49 Hasil MAE terbaik dari data <i>US Pollution</i>	107
Tabel 4.50 Hasil R ² terbaik dari data <i>US Pollution</i>	108

Tabel 4.51 Hasil RMSE terbaik dari data Beijing pm2.5	108
Tabel 4.52 Hasil MAE terbaik dari data Beijing pm2.5	109
Tabel 4.53 Hasil R ² terbaik dari data Beijing pm2.5.....	109
Tabel 4.54 Hasil RMSE terbaik dari data Guangzhou.....	110
Tabel 4.55 Hasil MAE terbaik dari data Guangzhou.....	110
Tabel 4.56 Hasil R ² terbaik dari data Guangzhou	111
Tabel 4.57 Hasil variabel dari LL-CNN pada data <i>Vital Sign</i> MIMIC-IV	112
Tabel 4.58 Hasil Variabel dari EDR-CNN data <i>Vital Sign</i> MIMIC-IV	112
Tabel 4.59 Hasil variabel dari MIRNet pada data <i>Vital Sign</i> MIMIC-IV	113
Tabel 4.60 Hasil variabel dari LL-CNN pada data <i>Vital Sign</i> MIMIC-III.....	113
Tabel 4.61 Hasil Variabel dari EDR-CNN data <i>Vital Sign</i> MIMIC-III	114
Tabel 4.62 Hasil variabel dari MIRNet pada data <i>Vital Sign</i> MIMIC-III	114
Tabel 4.63 Hasil variabel dari LL-CNN pada data Beijing	115
Tabel 4.64 Hasil Variabel dari EDR-CNN pada data Beijing	115
Tabel 4.65 Hasil variabel dari MIRNet pada data Beijing.....	116
Tabel 4.66 Hasil variabel dari LL-CNN pada data <i>Air Quality</i> Itali.....	117
Tabel 4.67 Hasil Variabel dari EDR-CNN pada data <i>Air Quality</i> Itali.....	117
Tabel 4.68 Hasil variabel dari MIRNet pada data <i>Air Quality</i> Itali	118
Tabel 4.69 Hasil variabel dari LL-CNN pada data <i>Air Quality</i> India.....	118
Tabel 4.70 Hasil Variabel dari EDR-CNN pada data <i>Air Quality</i> India	119
Tabel 4.71 Hasil variabel dari MIRNet pada data <i>Air Quality</i> India	119
Tabel 4.72 Hasil variabel dari LL-CNN pada data <i>US Pollution</i>	120
Tabel 4.73 Hasil Variabel dari EDR-CNN pada data <i>US Pollution</i>	121
Tabel 4.74 Hasil variabel dari MIRNet pada data <i>US Pollution</i>	121
Tabel 4.75 Hasil variabel dari LL-CNN pada data Beijing pm2.5	122
Tabel 4.76 Hasil Variabel dari EDR-CNN pada data Beijing pm2.5.....	122
Tabel 4.77 Hasil variabel dari MIRNet pada data Beijing pm2.5.....	123
Tabel 4.78 Hasil variabel dari LL-CNN pada data Guangzhou.....	123
Tabel 4.79 Hasil Variabel dari EDR-CNN pada data Guangzhou.....	124
Tabel 4.80 Hasil variabel dari MIRNet pada data Guangzhou	124
Tabel 4.81 Hasil Metode terbaik pada data yang digunakan untuk penelitian.....	125

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam pengolahan data, sering terjadi kehilangan informasi akibat kelalaian, kegagalan sistem, atau penghapusan sengaja [1]. Kondisi tersebut bisa terjadi karena beberapa faktor, seperti kelalaian orang yang bertugas mendokumentasikan data, kegagalan sistem, atau penghapusan sengaja karena beberapa alasan. Oleh karena itu, peneliti menggunakan strategi imputasi untuk mengatasinya [1][2].

Saat ini data yang hilang adalah masalah yang kerap di temui, karena dapat menurunkan performa sebagian besar model *machine learning* [3]. Untuk mengisi nilai hilang pada data yang besar dapat menggunakan strategi teknik *Deep Learning* untuk mengkoordinasikan nilai yang hilang pada data, yang dikenal sebagai “metode imputasi” [4]. Metode ini dapat dihasilkan dengan beberapa cara yang berbeda, contoh nya dapat ditemukan pada literatur, seperti pada penelitian [3][5][6][7] yang memiliki masalah nilai hilang pada data dengan menggunakan metode *linear imputation* (LI), *predictive mean matching* (PMM), *random forest* (RF), *k-nearest neighbours* (k-NN), dan imputasi dari *kalman smoothing* pada *structural time series* (kalman) [5], Autoencoder[3], *generative adversarial network* (GAN) [7]. Tapi pada penelitian [8] memperkirakan nilai yang hilang dalam data deret waktu *Interne of Things* (IoT) menggunakan tiga algoritma interpolasi yaitu *radial basis functions*, *moving least squares* (MLS), dan *adaptive inverse distance weighted*.

Pada penelitian [9][10] dilakukan pengisian data nilai yang hilang dari database *mimic III* dan *mimic IV* dengan menggunakan *Machine Learning*, dari penelitian tersebut menghasilkan nilai RSME yang cukup baik. Metode PMM dan RF memiliki nilai RMSE yang lebih tinggi dibanding *linear imputation* dan *kalman* [5], pada penelitian [1], Autoencoder memiliki nilai RMSE yang tidak baik. Metode GAN [7] menunjukkan bahwa metode yang digunakan memiliki hasil kinerja yang lebih rendah dibanding dengan metode GAMIN. Pada penelitian ini saya menggunakan metode peningkatan citra untuk imputasi data karena metode ini menunjukkan akurasi tinggi pada citra. Dengan *transfer learning*, saya

menerapkan teknik ini pada data tabular. Sebelumnya, metode seperti U-Net, CNN, dan Autoencoder menghasilkan RMSE yang kurang baik.

Ada beberapa pendekatan dari *Deep Learning* seperti pada penelitian [11][12] menggunakan *Low Light CNN* dapat mempelajari fitur – fitur penting yang ada pada data dengan menghasilkan imputasi yang efektif. Arsitektur LL-CNN digunakan untuk menangani data multidimensi, termasuk Tabel yang memiliki banyak atribut atau kolom. Keunggulan utama terletak pada kemampuan model untuk secara otomatis mengekstrak fitur-fitur yang relevan dari data tabular, memahami pola yang kompleks, dan menangani hubungan non-linear antar atribut.

Selain itu terdapat pendekatan *Deep Learning* yang mendukung untuk melakukan imputasi dengan data yang besar, seperti menggunakan metode EDR-CNN. Pada penelitian [13] EDR-CNN menunjukkan akurasi tertinggi dalam penelusuran sumber pencemaran air tanah. Arsitektur ini efektif menangani data yang hilang dengan mempelajari pola kompleks dan memanfaatkan blok residu untuk pembelajaran fitur relevan. Keunggulannya terletak pada kemampuannya memberikan solusi yang lebih baik daripada metode tradisional, terutama dalam dataset dengan pola atau ketergantungan kompleks. EDR-CNN dapat mengisi celah dalam data tabular dengan efektif, menjadikannya solusi andal untuk imputasi data deret waktu multivariat.

Selain ke-dua arsitektur tersebut, pendekatan *Deep Learning* lain yang dapat melakukan imputasi dengan data yang besar yaitu MIRNet. Pada penelitian [14] MIRNet, yang dirancang untuk pemulihan citra dua dimensi, menawarkan peningkatan signifikan dengan efisiensi komputasi tinggi dan keseimbangan antara waktu eksekusi dan kinerja. Arsitektur MIRNet dapat mengatasi data besar dengan efektivitas tinggi. Penelitian ini bertujuan untuk mengembangkan model yang lebih akurat dalam imputasi data. Berdasarkan penelitian sebelumnya yang menggunakan Autoencoder, CNN, dan U-Net, penelitian ini akan mengembangkan model imputasi data menggunakan LL-CNN, EDR-CNN, dan MIRNet untuk meningkatkan akurasi RMSE pada data umum. Tujuannya adalah meningkatkan kualitas imputasi data deret waktu multivariat dan kualitas data tanda vital dari berbagai sumber.

1.2. Tujuan Penelitian

Adapun tujuan dari Tugas Akhir ini, yaitu :

1. Meningkatkan evaluasi imputasi pada data umum dengan menggunakan tiga metode *Deep Learning*, yaitu LL-CNN, EDR-CNN, dan MIRNet untuk Imputasi Multivariat pada data tersebut.
2. Merancang dan menguji model imputasi pada data seperti MIMIC-IV, MIMIC-III, Beijing *Multi-Site Air Quality*, *Air Quality* Itali, *Air Quality* India, *US Pollution*, Beijing pm2.5, dan Guangzhou untuk menemukan model terbaik.

1.3. Rumusan Masalah

Penelitian ini mengatasi masalah nilai hilang pada data MIMIC IV, MIMIC III, Beijing Multi-Site Air Quality, Air Quality Itali, Air Quality India, US Pollution, Beijing pm2.5, dan Guangzhou dengan fokus pada imputasi multivariat. Metode yang digunakan adalah LL-CNN, EDR-CNN, dan MIRNet untuk mencapai hasil yang optimal.

1.4. Batasan Masalah

Berikut adalah batasan masalah dalam penelitian ini:

1. Pada penelitian ini terdapat 8 data yang digunakan diantaranya data: MIMIC IV, MIMIC, III, Beijing *Multi-Site Air Quality*, *Air Quality* Itali, *Air Quality* India, *US Pollution*, Beijing pm2.5, dan Guangzhou.
2. Data yang digunakan tidak memiliki nilai nan dengan tipe data deret waktu multivariat.
3. Metode imputasi yang digunakan melibatkan 3 arsitektur: LL-CNN, EDR-CNN, dan MIRNet.
4. Matriks evaluasi yang digunakan adalah MAE, R^2 , dan RMSE.

1.5. Metodologi Penelitian

Metode yang digunakan dalam penulisan tugas akhir dimulai dari studi pustaka sampai ke tahap penarikan kesimpulan dan saran. Setiap metode akan diuraikan secara rinci untuk memberikan Gambaran lengkap mengenai pendekatan yang diambil dalam penelitian ini.

1.5.1. Studi Pustaka dan Literatur

Bagian ini membahas metode yang digunakan untuk mencari dan mengumpulkan referensi data berupa literature yang terdapat pada buku, paper dan internet mengenai “Multivariate Imputation”.

1.5.2. Metode Konsultasi

Bagian ini menjelaskan bagaimana peneliti melakukan konsultasi bersama dengan para dosen pembimbing yang memiliki pengetahuan dan pengalaman dalam menangani permasalahan yang dihadapi selama penulisan tugas akhir.

1.5.3. Metode Pembuatan Model

Bagian ini menguraikan langkah-langkah dalam merancang dan membuat model untuk mendapatkan suatu perancangan pemodelan yang baik dengan menggunakan simulasi.

1.5.4. Metode Pengujian dan Validasi

Bagian ini menjelaskan cara menguji dan memvalidasi model terhadap imputasi yang telah dibuat, apakah metode tersebut dapat menghasilkan nilai yang baik atau tidak.

1.5.5 Metode Hasil dan Analisa

Bagian ini menyajikan hasil pengujian dan menganalisis percobaan dengan menggunakan metode yang digunakan pada penelitian, untuk mengetahui kelebihan dan kekurangan dari metode tersebut, sehingga dapat digunakan dan diimprovisasi untuk penelitian selanjutnya.

1.5.6. Metode Penarikan Kesimpulan dan Saran

Peneliti akan menyimpulkan hasil dari penelitian yang telah dilakukan dan memberikan saran yang dapat dijadikan referensi untuk penelitian selanjutnya.

1.6. Sistematika Penulisan

Adapun sistematika penulisan dalam tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Bab 1 memberikan uraian tentang awal dari suatu penulisan, meliputi latar belakang, perumusan dan batasan masalah, tujuan dan manfaat, metodologi penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab 2 membicarakan tentang fondasi teori yang digunakan untuk mengatasi permasalahan yang dihadapi dalam penelitian.

BAB III METODOLOGI

Bab 3 berisi penjelasan detail mengenai teknik, metode, serta alur proses yang digunakan dalam penelitian.

BAB IV Hasil dan Pembahasan

Bab 4 berisi analisis hasil pengujian yang mencakup evaluasi terhadap kelebihan dan kekurangan dari penelitian yang telah dilakukan.

BAB V PENUTUP

Bab 5 berisi tentang ringkasan kesimpulan yang diambil dari hasil penelitian sebelumnya dan rekomendasi untuk perbaikan dan pengembangan penelitian di masa depan.

DAFTAR PUSTAKA

Bab 6 berisi daftar referensi dari sumber – sumber informasi yang digunakan dalam metode literatur.

LAMPIRAN

Lampiran mencakup beberapa berkas berupa formulir perbaikan dan juga pemeriksaan tingkat kemiripan karya dengan sumber lain.

DAFTAR PUSTAKA

- [1] A. F. Costa, M. S. Santos, J. P. Soares, and P. H. Abreu, “Missing data imputation via denoising autoencoders: The untold story,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11191 LNCS, no. November, pp. 87–98, 2018, doi: 10.1007/978-3-030-01768-2_8.
- [2] X. Zhang, C. Yan, C. Gao, B. A. Malin, and Y. Chen, “Predicting Missing Values in Medical Data Via XGBoost Regression,” *J. Healthc. Informatics Res.*, vol. 4, no. 4, pp. 383–394, 2020, doi: 10.1007/s41666-020-00077-1.
- [3] R. C. Pereira, M. S. Santos, P. P. Rodrigues, and P. H. Abreu, “Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes,” *J. Artif. Intell. Res.*, vol. 69, pp. 1255–1285, 2020, doi: 10.1613/JAIR.1.12312.
- [4] M. Chen, H. Zhu, Y. Chen, and Y. Wang, “A Novel Missing Data Imputation Approach for Time Series Air Quality Data Based on Logistic Regression,” *Atmosphere (Basel)*, vol. 13, no. 7, 2022, doi: 10.3390/atmos13071044.
- [5] L. A. Menéndez García *et al.*, “A Method of Pruning and Random Replacing of Known Values for Comparing Missing Data Imputation Models for Incomplete Air Quality Time Series,” *Appl. Sci.*, vol. 12, no. 13, p. 6465, 2022, doi: 10.3390/app12136465.
- [6] J. C. Kim and K. Chung, “Recurrent Neural Network-Based Multimodal Deep Learning for Estimating Missing Values in Healthcare,” *Appl. Sci.*, vol. 12, no. 15, 2022, doi: 10.3390/app12157477.
- [7] S. Yoon and S. Sull, “Gamin: Generative adversarial multiple imputation network for highly missing data,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8453–8461, 2020, doi: 10.1109/CVPR42600.2020.00848.
- [8] Z. Ding, G. Mei, S. Cuomo, Y. Li, and N. Xu, “Comparison of Estimating Missing Values in IoT Time Series Data Using Different Interpolation Algorithms,” *Int. J. Parallel Program.*, vol. 48, no. 3, pp. 534–548, 2020,

- doi: 10.1007/s10766-018-0595-5.
- [9] O. T. Abdala and M. Saeed, “Estimation of missing values in clinical laboratory measurements of ICU patients using a weighted K-nearest neighbors algorithm,” in *Computers in Cardiology, 2004*, 2004, pp. 693–696. doi: 10.1109/CIC.2004.1443033.
- [10] X. Zhang, C. Yan, C. Gao, B. A. Malin, and Y. Chen, “Predicting Missing Values in Medical Data Via XGBoost Regression,” *J. Healthc. Informatics Res.*, vol. 4, no. 4, pp. 383–394, 2020, doi: 10.1007/s41666-020-00077-1.
- [11] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, “LLCNN: A convolutional neural network for low-light image enhancement,” *2017 IEEE Vis. Commun. Image Process. VCIP 2017*, vol. 2018-Janua, no. 2013, pp. 1–4, 2017, doi: 10.1109/VCIP.2017.8305143.
- [12] Y. Qu, C. liu, and Y. Ou, “LEUGAN:Low-Light Image Enhancement by Unsupervised Generative Attentional Networks,” vol. 14, no. 8, pp. 1–5, 2020, [Online]. Available: <http://arxiv.org/abs/2012.13322>
- [13] Q. Sun *et al.*, “Pr ep rin ot pe er re v Pr ep rin ot pe er ed,” pp. 1–8, 2023.
- [14] N. Li, Y. Zhao, Q. Pan, S. G. Kong, and J. C. W. Chan, *Full-Time Monocular Road Detection Using Zero-Distribution Prior of Angle of Polarization*, vol. 12370 LNCS. 2020. doi: 10.1007/978-3-030-58595-2_28.
- [15] R. K. Akiki, R. S. Anand, M. Borrelli, I. N. Sarkar, P. Y. Liu, and E. S. Chen, “Predicting open wound mortality in the ICU using machine learning,” *J. Emerg. Crit. Care Med.*, vol. 5, no. April, 2021, doi: 10.21037/jeccm-20-154.
- [16] M. Gupta, B. Gallamoza, N. Cutrona, P. Dhakal, R. Poulain, and R. Beheshti, “An extensive data processing pipeline for MIMIC-IV,” *Proc. Mach. Learn. Res.*, vol. 193, pp. 311–325, 2022.
- [17] S. Min Kim and J. Young Yoon, “환자의 활력 징후를 이용한 후향적 데이터의 분석과 연구를 위한 데이터 가공 및 시각화 방법 Data Processing and Visualization Method for Retrospective Data Analysis and Research Using Patient Vital Signs,” *J. Biomed. Eng. Res.*, vol. 42, pp. 175–185, 2021, [Online]. Available:

<http://dx.doi.org/10.9718/JBER.2021.42.4.175>

- [18] D. Li, J. Liu, and Y. Zhao, "Prediction of Multi-Site PM_{2.5} Concentrations in Beijing Using," pp. 1–19, 2022.
- [19] N. Agarwal, N. Brukhim, E. Hazan, and Z. Lu, "Boosting for control of dynamical systems," *37th Int. Conf. Mach. Learn. ICML 2020*, vol. PartF16814, pp. 84–91, 2020.
- [20] X. Wang and C. Wang, "Time Series Data Cleaning: A Survey," *IEEE Access*, vol. 8, pp. 1866–1881, 2020, doi: 10.1109/ACCESS.2019.2962152.
- [21] L. Maysofa and K. Umam Syaliman, "Implementasi Forecasting Pada Penjualan Inaura Hair Care Dengan Metode Single Exponential Smoothing Forecasting Implementation in Inaura Hair Care Sales With Single Exponential Smoothing Method," *J. Test. dan Implementasi Sist. Inf.*, vol. 1, no. 2, pp. 1–10, 2023.
- [22] M. Motamedi, N. Li, D. G. Down, and N. M. Heddle, "Demand Forecasting for Platelet Usage: from Univariate Time Series to Multivariate Models," 2021, [Online]. Available: <http://arxiv.org/abs/2101.02305>
- [23] A. T. Jebb, L. Tay, W. Wang, and Q. Huang, "Time series analysis for psychological research: Examining and forecasting change," *Front. Psychol.*, vol. 6, no. JUN, pp. 1–24, 2015, doi: 10.3389/fpsyg.2015.00727.
- [24] S. Darji, J., Biswas, N., Jones, L. D., & Ashili, "Handling missing data in the time-series data from wearables," 2023, [Online]. Available: <https://www.intechopen.com/online-first/1154877>
- [25] S. Jäger, A. Allhorn, and F. Bießmann, "A Benchmark for Data Imputation Methods," *Front. Big Data*, vol. 4, no. July, pp. 1–16, 2021, doi: 10.3389/fdata.2021.693674.
- [26] A. Jadhav, D. Pramod, and K. Ramanathan, "Comparison of Performance of Data Imputation Methods for Numeric Dataset," *Appl. Artif. Intell.*, vol. 33, no. 10, pp. 913–933, 2019, doi: 10.1080/08839514.2019.1637138.
- [27] I. Eldiyana, E. Nurlaelah, and N. Herrhyanto, "Estimasi Missing Data dengan Metode Multivariate Imputation by Chained Equations (MICE) untuk Membentuk Persamaan Regresi Linier Berganda," *J. EurekaMatika*, vol. 9, no. 1, pp. 95–106, 2021, doi: 10.17509/jem.v8i1.25750.

- [28] E. Ha, H. Lim, S. Yu, and J. Paik, “Low-light Image Enhancement Using Dual Convolutional Neural Networks for Vehicular Imaging Systems,” in *2020 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, Jan. 2020, pp. 1–2. doi: 10.1109/ICCE46568.2020.9043035.
- [29] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced Deep Residual Networks for Single Image Super-Resolution,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2017-July, pp. 1132–1140, 2017, doi: 10.1109/CVPRW.2017.151.
- [30] S. W. Zamir *et al.*, “Learning Enriched Features for Fast Image Restoration and Enhancement,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1934–1948, 2023, doi: 10.1109/TPAMI.2022.3167175.
- [31] J. Qi, J. Du, S. M. Siniscalchi, X. Ma, and C. H. Lee, “On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression,” *IEEE Signal Process. Lett.*, vol. 27, no. c, pp. 1485–1489, 2020, doi: 10.1109/LSP.2020.3016837.
- [32] R. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, “MIMIC-IV (version 2.0). PhysioNet.,” 2022, [Online]. Available: <https://doi.org/10.13026/7vcr-e114>.
- [33] M. I. KARIM, “IMPUTASI DATA YANG HILANG PADA DATA DERET WAKTU MULTIVARIAT MENGGUNAKAN ARSITEKTUR U-NET”.