

KLASIFIKASI *AUTHOR NAME DISAMBIGUATION*
MENGGUNAKAN ALGORITMA VARIATIONAL
AUTOENCODER DAN DEEP NEURAL NETWORK



TESIS

HAMID RAHMAN
ENTERPRISE SOFTWARE ENGINEERING
24242002P

PROGRAM STUDI TEKNIK INFORMATIKA – S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
TAHUN 2026

**KLASIFIKASI *AUTHOR NAME DISAMBIGUATION*
MENGUNAKAN ALGORITMA VARIATIONAL
AUTOENCODER DAN DEEP NEURAL NETWORK**



**Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar**

MAGISTER KOMPUTER

**HAMID RAHMAN
ENTERPRISE SOFTWARE ENGINEERING
24242002P**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2026**

Halaman Pengesahan Pembimbing Tesis

Judul Tesis: **KLASIFIKASI *AUTHOR NAME* *DISAMBIGUATION*
MENGUNAKAN ALGORITMA VARIATIONAL
AUTOENCODER DAN DEEP NEURAL NETWORK**

Oleh HAMID RAHMAN, NIM 24242002P, Tesis ini telah disetujui dan disahkan oleh Pembimbing Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE SOFTWARE ENGINEERING, Program Pascasarjana Universitas Bina Darma pada 26 Februari 2026 dan telah dinyatakan LULUS.

Palembang, 26 Februari 2026
Mengetahui,
Program Studi Teknik Informatika – S2
Universitas Bina Darma
Ketua,


Universitas Bina Darma
Magister Teknik Informatika

.....
Dr. Usman Ependi, S.Kom., M.Kom.

Pembimbing,

.....
Dr. Usman Ependi, S.Kom., M.Kom.

Halaman Pengesahan Penguji Tesis

Judul Tesis: **KLASIFIKASI *AUTHOR NAME* *DISAMBIGUATION*
MENGUNAKAN ALGORITMA VARIATIONAL
AUTOENCODER DAN DEEP NEURAL NETWORK**

Oleh HAMID RAHMAN , NIM 24242002P, Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE SOFTWARE ENGINEERING, Program Pascasarjana Universitas Bina Darma pada 26 Februari 2026 dan telah dinyatakan LULUS.

Palembang, 26 Februari 2026

Mengetahui,

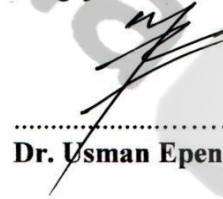
Program Pascasarjana
Universitas Bina Darma

Direktur,


PROGRAM PASCASARJANA

.....
Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Penguji I,



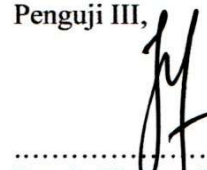
.....
Dr. Usman Ependi, M.Kom.

Penguji II,



.....
Prof. Dr. Edi Surya Negara, M.Kom.

Penguji III,



.....
Dr. A. Haidar Mirza, S.T., M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : HAMID RAHMAN
NIM : 24242002P

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (Tesis, Skripsi, Tugas Akhir) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (Magister, Sarjana, dan Ahli Madya) di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis/Skripsi/Tugas Akhir, yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 26 Februari 2026
Yang Membuat Pernyataan,



HAMID RAHMAN
NIM: 24242002P

ABSTRAK

Ambiguitas Nama Penulis atau *Author Name Disambiguation* (AND) merupakan tantangan fundamental yang memengaruhi integritas dan kualitas data dalam ekosistem perpustakaan digital. Pendekatan pencocokan berpasangan (*pairwise matching*) digunakan dalam menyelesaikan masalah tersebut, tetapi muncul kendala signifikan berupa ketidakseimbangan data (*imbalanced data*) yang ekstrem, di mana jumlah pasangan penulis berbeda (kelas mayoritas) sangat mendominasi pasangan penulis sama (kelas minoritas). Kondisi ini mengakibatkan model klasifikasi menjadi bias dan gagal mengidentifikasi penulis yang sebenarnya. Penelitian ini bertujuan mengembangkan solusi untuk mengatasi ketidakseimbangan data tersebut menggunakan algoritma *Variational Autoencoder* (VAE) sebagai teknik *generative oversampling*, serta mengintegrasikannya dengan *Deep Neural Network* (DNN) untuk proses klasifikasi. Metodologi penelitian memanfaatkan dataset DBLP yang melalui tahap pra-pemrosesan serta ekstraksi fitur kesamaan leksikal dan jarak temporal. Tiga skenario pengujian dirancang untuk membandingkan kinerja model: *baseline* (tanpa penanganan), metode VAE, dan metode SMOTE sebagai pembanding. Hasil eksperimen menunjukkan bahwa model *baseline* mengalami kegagalan total dalam mendeteksi kelas minoritas dengan nilai *F1-Score* 0.00. Sebaliknya, penerapan algoritma VAE terbukti paling efektif dengan mencapai *F1-Score* agregat sebesar 0.9997, sedikit lebih unggul dibandingkan metode SMOTE yang mencapai 0.9992. Penelitian ini menyimpulkan bahwa integrasi VAE dan DNN mampu menghasilkan data sintetis berkualitas tinggi yang menyeimbangkan distribusi kelas, sehingga secara signifikan meningkatkan akurasi dan sensitivitas sistem dalam menyelesaikan masalah disambiguasi nama penulis.

Kata Kunci: Ambiguitas Nama Penulis, *Imbalanced Data*, *Variational Autoencoder*, *Deep Neural Network*, *Oversampling*.

ABSTRACT

Author Name Disambiguation (AND) is a fundamental challenge affecting data integrity and quality within digital library ecosystems. In addressing this issue using a pairwise matching approach, a significant obstacle arises in the form of extreme data imbalance, where the number of different author pairs (majority class) vastly dominates the same author pairs (minority class). This condition causes classification models to become biased and fail to identify the actual authors. This study aims to develop a solution to overcome this data imbalance using the Variational Autoencoder (VAE) algorithm as a generative oversampling technique, integrating it with a Deep Neural Network (DNN) for the classification process. The research methodology utilizes the DBLP dataset, which undergoes pre-processing and feature extraction for lexical similarity and temporal distance. Three testing scenarios are designed to compare model performance: baseline (without data handling), the VAE method, and the SMOTE method as a benchmark. Experimental results indicate that the baseline model completely failed to detect the minority class with an F1-Score of 0.00. Conversely, the application of the VAE algorithm proved to be the most effective, achieving an aggregate F1-Score of 0.9997, slightly outperforming the SMOTE method which reached 0.9992. This study concludes that the integration of VAE and DNN successfully generates high-quality synthetic data that balances class distribution, thereby significantly improving the accuracy and sensitivity of the system in resolving author name disambiguation problems.

Kata Kunci: *Author Name Disambiguation, Imbalanced Data, Variational Autoencoder, Deep Neural Network, Oversampling.*

MOTTO DAN PERSEMBAHAN

"We cannot change what is done, so we must insist upon the truth and be loyal to what matters; in the end, we stand firm in our integrity and our effort."

Segala puji dan syukur penulis panjatkan ke hadirat Allah Swt. atas segala limpahan rahmat, hidayah, dan karunia-Nya. Berkat kemudahan dari-Nya, yang terwujud melalui dukungan berbagai pihak, penulis akhirnya dapat menyelesaikan penelitian Tesis ini.

Melalui kerja keras dan doa yang tak terputus, dengan segenap kerendahan hati dan rasa syukur, penulis mempersembahkan karya sederhana ini secara khusus kepada:

1. Kedua Orang Tua Tercinta, Ayahanda dan Ibunda. Yang telah mendidik dengan penuh kasih sayang, mengiringi dengan doa yang tak pernah putus, serta menjadi sumber kekuatan tak ternilai dalam setiap langkah penulis.
2. Keluarga dan Adik Penulis. Atas segenap hati, dukungan tulus, dan semangat yang senantiasa diberikan.
3. Para Rekan-rekan Seperjuangan. Terima kasih telah menjadi mitra diskusi dan saling memotivasi penulis dalam menyelesaikan pendidikan magister ini.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan penyusunan tesis yang berjudul “**Klasifikasi Author Name Disambiguation Menggunakan Algoritma Variational Autoencoder dan Deep Neural Network**”. Tesis ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi Magister Teknik Informatika di Universitas Bina Darma Palembang. Penulis menyadari bahwa kelancaran proses ini tidak terlepas dari bimbingan, arahan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Bapak Prof. Dr. Edi Surya Negara Harahap, M.Kom. Selaku Rektor Universitas Bina Darma Palembang dan Penguji yang telah memberikan bimbingan serta masukan berharga dalam proses penyelesaian tesis ini.
2. Bapak Prof. Dr. Ir. Achmad Syarifudin, M.Sc. Selaku Direktur Pascasarja Universitas Bina Darma Palembang
3. Bapak Dr. Usman Ependi, S.Kom., M.Kom., selaku Dosen Pembimbing dan Ketua Program Studi Magister Teknik Informatika, yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta motivasi kepada penulis dalam penulisan tesis ini.
4. Bapak Dr. A Haidar Mirza, S.T.,M.Kom. selaku penguji yang telah memberikan motivasi dan arahan dalam penulisan tesis ini.
5. Segenap Bapak dan Ibu Dosen Program Studi Magister Teknik Informatika atas bekal ilmu pengetahuan yang telah dibagikan.
6. Pihak Sekretariat Pascasarjana Universitas Bina Darma Palembang yang telah memberikan bimbingan pelayanan dengan baik.

Palembang, 26 Februari 2026

Hamid Rahman

DAFTAR ISI

HALAMAN DEPAN	ii
HALAMAN PENGESAHAN PEMBIMBING TESIS	iii
HALAMAN PENGESAHAN PENGUJI TESIS	iv
SURAT PERNYATAAN.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
MOTTO DAN PERSEMBAHAN.....	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	3
1.3. Perumusan Masalah	4
1.4. Tujuan Penelitian	5
1.5. Kebaruan (<i>Novelty</i>)	5
1.6. Manfaat Penelitian	6
1.7. Pembatasan Masalah	6
1.8. Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	9
2.1. Tinjauan Pustaka	9
2.2. Landasan Teori.....	12
2.2.1. <i>Machine Learning</i>	12
2.2.2. <i>Deep Learning</i>	12
2.2.3. <i>Similarity</i>	16
2.2.4. <i>Imbalanced Data</i>	17
2.2.5. <i>Text Normalization</i>	19
2.2.6. <i>Min-Max Scaller</i>	20

2.2.7.	<i>Confusion Matrix</i>	21
2.2.8.	<i>Receiver Operating Characteristic Curve (ROC)</i>	23
BAB III METODOLOGI PENELITIAN.....		26
3.1.	Waktu Dan Tempat Penelitian	26
3.2.	Prosedur Penelitian.....	26
3.2.1.	Studi Pustaka.....	28
3.2.2.	Persiapan Data.....	28
3.2.3.	Pra-Pemrosesan Data	30
3.2.4.	Ekstraksi Fitur	35
3.2.5.	Proses Penanganan Imbalanced Data.....	38
3.2.6.	Proses Klasifikasi	45
3.2.7.	Analisa Hasil	49
3.3.	Jadwal Penelitian.....	50
BAB IV HASIL DAN PEMBAHASAN		51
4.1.	Hasil	51
4.1.1.	Penanganan Data.....	51
4.1.2.	Penanganan Data Tidak Seimbang Menggunakan VAE	53
4.1.3.	Penanganan Data Tidak Seimbang Menggunakan SMOTE	58
4.1.4.	Klasifikasi Pada Data Tidak Seimbang.....	62
4.1.5.	Klasifikasi Pada Data Seimbang Menggunakan VAE	65
4.1.6.	Klasifikasi Pada Data Seimbang Menggunakan SMOTE.....	69
4.2.	Pembahasan.....	74
4.2.1.	Dampak Penyeimbangan Data terhadap Kinerja Model.....	75
4.2.2.	Perbandingan Kinerja VAE dan SMOTE pada Data Validasi.....	77
4.2.3.	Analisis Komparatif Terhadap Penelitian Terdahulu.....	79
BAB V KESIMPULAN DAN SARAN.....		81
5.1.	Kesimpulan	81
5.2.	Saran.....	82
DAFTAR PUSTAKA		83
DAFTAR RIWAYAT HIDUP.....		94

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Terkait Author Name Disambiguation.....	11
Tabel 2.2 Matrix 2x2 Confusion Matrix.....	21
Tabel 3.1 Deskripsi DBLP Dataset	29
Tabel 3.2 Hasil Proses Kombinasi	34
Tabel 3.3 Hasil Penghitungan Kesamaan Data	37
Tabel 3.4 Proses Year Difference.....	38
Tabel 3.5 Jadwal Penelitian.....	50
Tabel 4.1 Distribusi Kelas Pairwise Matching pada Dataset	51
Tabel 4.2 Kombinasi Hyperparamater Tuning VAE.....	54
Tabel 4.3 Distribusi Kelas Setelah Penyeimbangan Data dengan VAE	56
Tabel 4.4 Kombinasi Hyperparameter Tuning SMOTE	59
Tabel 4.5 Distribusi Kelas Setelah Penyeimbangan Data Dengan SMOTE	60
Tabel 4.6 Classification Report DNN Data tidak seimbang.....	62
Tabel 4.7 Verifikasi Data Sampling VAE+DNN.....	65
Tabel 4.8 Classification Report Data Seimbang VAE dan DNN.....	68
Tabel 4.9 Verifikasi Data Sampling SMOTE+DNN.....	69
Tabel 4.10 Classification Report Data Seimbang SMOTE dan DNN.....	71
Tabel 4.11 Perbandingan Metrik Kinerja Agregat Antar Skenario.....	74
Tabel 4.12 Perbandingan Komparatif dengan Penelitian Terdahulu.....	79

DAFTAR GAMBAR

Gambar 2.1 Taxonomy Author Name Disambiguation Method	10
Gambar 2.2 Arsitektur Autoencoder	14
Gambar 2.3 Arsitektur DNN (Altin Karataş & Biberici, 2023)	16
Gambar 2.4 Ilustrasi ROC Curve	24
Gambar 2.5 Ilustrasi AUC Curve	25
Gambar 3.1 Prosedur Penelitian	27
Gambar 3.2 Pra-Pemrosesan Data	31
Gambar 3.3 Penghapusan Kolom	32
Gambar 3.4 Text Cleaning	33
Gambar 3.5 Proses Pembuatan Labeling Pasangan Data	35
Gambar 3.6 Diagram Alir Ekstraksi Fitur	36
Gambar 3.7 Alur Penanganan Data Tidak Seimbang Menggunakan VAE.....	39
Gambar 3.8 VAE Encoder	40
Gambar 3.9 VAE Decoder	41
Gambar 3.10 VAE Arsitektur	41
Gambar 3.11 Arsitektur DNN	46
Gambar 3.12 Workflow Arsitektur DNN	49
Gambar 4.1 Distribusi Kelas Asli	52
Gambar 4.2 Distribusi Kelas Setelah Pemisahan Data	53
Gambar 4.3 Kombinasi ke-1 Performa Terbaik Hyperparameter Tuning VAE..	55
Gambar 4.4 Distribusi Kelas Setelah Penyeimbangan Data	56
Gambar 4.5 Visualisasi t-SNE dari sampel data asli dan sintetis	57
Gambar 4.6 Distribusi Kelas Setelah Penyeimbangan Data Dengan SMOTE....	60
Gambar 4.7 Visualisasi t-Sne SMOTE	61
Gambar 4.8 ROC - DNN Data Tidak Seimbang	63
Gambar 4.9 Confusion Matrix - DNN Data Tidak Seimbang	64
Gambar 4.10 Grafik Model Loss (VAE+DNN)	66
Gambar 4.11 Grafik Model Accuracy (VAE+DNN)	66
Gambar 4.12 Confusion Matrix (VAE+DNN)	67

Gambar 4.13 ROC Curve (VAE+DNN)	69
Gambar 4.14 Grafik Model Loss (SMOTE+DNN).....	70
Gambar 4.15 Grafik Model Accuracy (SMOTE+DNN).....	71
Gambar 4.16 Confusion Matrix (SMOTE+DNN).....	72
Gambar 4.17 ROC Curve (SMOTE+DNN)	73
Gambar 4.18 Perbandingan Metrik Kinerja (same_author)	75
Gambar 4.19 Perbandingan Metrik Kinerja Keseluruhan	76
Gambar 4.20 Perbandingan Kinerja VAE+DNN dan SMOTE+DNN	77



DAFTAR LAMPIRAN

SK PEMBIMBING	95
LEMBAR KONSULTASI BIMBINGAN TESIS	96
LEMBAR PERBAIKAN TESIS	99
SURAT KETERANGAN LULUS	102

