

**KLASIFIKASI AUTHOR PADA DATA BIBLIOGRAFI
MENGUNAKAN DEEP NEURAL NETWORK DAN
SUPPORT VECTOR MACHINE**

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



Oleh :

TIO ARTHA NUGRAHA

09011181520027

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

2019

HALAMAN PENGESAHAN

**KLASIFIKASI AUTHOR PADA DATA BIBLIOGRAFI
MENGUNAKAN DEEP NEURAL NETWORK DAN
SUPPORT VECTOR MACHINE**

TUGAS AKHIR

Sebagai salah satu syarat untuk menyelesaikan
Program Studi Sistem Komputer
Jenjang S1

Oleh :

TIO ARTHA NUGRAHA
09011181520027

Palembang, Desember 2019

Mengetahui,

Ketua Jurusan Sistem Komputer



Rossi Passarella, S.T., M.Eng.
NIP. 19780611 201012 1 004

Pembimbing Tugas Akhir

Firdaus, S.T., M.Kom.
NIP. 19780121 200812 1 003

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Rabu

Tanggal : 11 Desember 2019

Tim Penguji :

1. Ketua : Kemahyanto Exaudi, S.Kom., M.T.
2. Sekretaris : Firdaus, S. T., M. Kom.
3. Anggota I : Dr. Erwin, M.Si.
4. Anggota II : Huda Ubaya, S.T., M.T.



Mengetahui,

Ketua Jurusan Sistem Komputer



Rossi Passarella, S. T., M. Eng.

NIP. 19780611 201012 1 004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Tio Artha Nugraha

NIM : 09011181520027

Judul : Klasifikasi Author pada Data Bibliografi menggunakan
Deep Neural Network dan Support Vector Machine

Hasil pengecekan *Software Turnitin* : 3%

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 tidak dipaksakan.



Palembang, Desember 2019



Tio Artha Nugraha

KATA PENGANTAR



Assalamu'alaikum Wr.Wb.

Alhamdulillahirabbil'alamin, puji dan syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan nikmat, taufik, dan hidayah-Nya yang sangat besar dan tidak pernah berhenti kepada penulis sehingga penulis dapat menyelesaikan Tugas Akhir ini yang berjudul **“Klasifikasi Author pada Data Bibliografi menggunakan Deep Neural Network dan Support Vector Machine”**.

Dalam tugas akhir ini penulis menjelaskan mengenai pemodelan untuk melakukan identifikasi dan klasifikasi author pada suatu publikasi digital dengan disertai data-data yang diperoleh penulis saat melakukan penelitian. Penulis berharap agar tulisan ini dapat bermanfaat bagi orang banyak dan menjadi bahan bacaan bagi yang tertarik untuk meneliti tentang *Author Name Disambiguation* (AND).

Pada kesempatan ini, dengan segala kerendahan hati penulis mengucapkan terima kasih kepada semua pihak atas bantuan, bimbingan, dan saran yang telah diberikan dalam menyelesaikan Tugas Akhir ini, antara lain:

1. Orang tua saya tercinta yang telah membesarkan saya dengan penuh kasih sayang dan selalu mengajarkan saya dalam berbuat hal yang baik. Terimakasih untuk segala doa, motivasi dan dukungannya baik moril, materiil maupun spiritual selama ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rossi Passarella, S.T., M.Eng., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

4. Bapak Dr. Reza Firsandaya Malik, M.T., selaku Pembimbing Akademik Jurusan Sistem Komputer.
5. Bapak Firdaus, S.T., M.Kom., selaku Pembimbing Tugas Akhir yang telah berkenan meluangkan waktunya guna membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan Tugas Akhir ini.
6. Bapak Dr. Erwin, M.Si. dan Bapak Huda Ubaya, S.T., M.T., selaku Penguji dalam Sidang Tugas Akhir yang berkenan meluangkan waktunya guna menguji, memberikan arahan, saran serta nasihat untuk penulis.
7. Kak Naufal Rachmatullah, S.Kom., M.T. dan Mbak Annisa Darmawahyuni, M.Kom. yang sangat suportif dan memberikan arahan serta nasihat kepada penulis untuk menyelesaikan Tugas Akhir ini dengan baik.
8. Teman seperjuangan yang tergabung dalam grup penelitian yang sangat membantu penulis dalam menyelesaikan Tugas Akhir ini.
9. Teman-teman seperjuangan Sistem Komputer Angkatan 2015 serta semua pihak yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa Tugas Akhir ini masih sangat jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun sangatlah diharapkan penulis agar dapat segera diperbaiki sehingga Tugas Akhir ini dapat dijadikan sebagai masukan ide dan pemikiran yang bermanfaat bagi semua pihak dan menjadi tambahan bahan bacaan bagi yang tertarik dalam penelitian tentang *Author Name Disambiguation (AND)*.

Wassalamu'alaikum Wr. Wb.

Palembang, Desember 2019
Penulis,

Tio Artha Nugraha

NIM. 09011181520027

*Author Classification in Bibliographic Data using
Deep Neural Network and Support Vector Machine*

Tio Artha Nugraha (09011181520027)

*Computer Engineering Department, Computer Science Faculty,
Sriwijaya University*

Email : tioarthanugraha@gmail.com

Abstract

Author Name Disambiguation (AND) is a problem of name ambiguity to the publication in a Digital Library (DL) database caused by the Homonymity and Synonymity of the author's name. The proposed method is a Deep Neural Network (DNN) and Support Vector Machine (SVM) to classify data. The DBLP Labeled Data dataset by Jinseok Kim, et. al. is used for the classification task. This study concerned with processing data with the techniques of normalization and transformation data to create an effective feature for classification. The performance evaluation of the research conducted is accuracy, precision, and recall. The parameters are important to evaluate the AND classification process, especially the identification of the author. For the result, DNN achieves accuracy, precision, and recall, which is 99.98%, 97.71%, and 97.83%, respectively. In addition, SVM produces accuracy, precision, and recall 99.98%, 95.33%, 95.09%, respectively. From the comparison of the two classification methods, DNN outperformed SVM for data classification and author identification.

Keywords : *Digital Library, Bibliographic Data, Author Name Disambiguation, Homonym, Synonym, Deep Neural Network, Support Vector Machine*

***Klasifikasi Author pada Data Bibliografi menggunakan
Deep Neural Network dan Support Vector Machine***

Tio Artha Nugraha (09011181520027)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : tioarthanugraha@gmail.com

Abstrak

Author Name Disambiguation (AND) merupakan permasalahan ambiguitas nama terhadap suatu publikasi pada database *Digital Library* (DL) yang disebabkan karena kondisi *Homonymity* dan *Synonymity* nama penulis (*author*). Metode yang diusulkan dalam tugas akhir ini adalah *Deep Neural Network* (DNN) dan *Support Vector Machine* (SVM) untuk melakukan klasifikasi data. Dataset bibliografi yang digunakan adalah *Dataset DBLP Labeled Data* oleh Jinseok Kim, dkk. Penelitian yang dilakukan berfokus dalam pengolahan data dengan berbagai macam teknik *Normalization* dan *Transformation Data* untuk menciptakan suatu fitur yang efektif untuk digunakan dalam klasifikasi. Parameter utama penelitian yang dilakukan adalah *accuracy*, *precision*, dan *recall* yang merupakan parameter penting untuk mengetahui tingkat keberhasilan metode yang dilakukan dalam mengatasi permasalahan AND khususnya identifikasi *author*. Klasifikasi DNN menghasilkan nilai *accuracy* 99,98%, *precision* 97,71%, dan *recall* 97,83%. Klasifikasi SVM menghasilkan nilai *accuracy* 99,98%, *precision* 95,33%, dan *recall* 95,09%. Dari perbandingan kedua metode klasifikasi tersebut, DNN menjadi metode terbaik untuk melakukan klasifikasi data dan identifikasi *author*.

Kata Kunci : *Digital Library, Bibliographic Data, Author Name Disambiguation, Homonym, Synonym, Deep Neural Network, Support Vector Machine*

DAFTAR ISI

	HALAMAN
HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN	iv
KATA PENGANTAR	v
Abstract	vii
Abstrak	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Tujuan dan Manfaat.....	3
1.2.1. Tujuan	3
1.2.2. Manfaat	3
1.3. Perumusan dan Batasan Masalah	3
1.3.1. Perumusan Masalah	3
1.3.2. Batasan Masalah	4
1.4. Metodologi Penelitian	4
1.4.1. Metode Studi Pustaka dan Literatur.....	4
1.4.2. Metode Konsultasi	5
1.4.3. Metode Pembuatan Model.....	5

1.4.4. Metode Pengujian dan Validasi	5
1.4.5. Metode Hasil dan Analisa	5
1.4.6. Metode Penarikan Kesimpulan dan Saran	6
1.5. Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	8
2.1. Author Name Disambiguation (AND)	8
2.1.1. Faktor Permasalahan AND	9
2.1.2. Kasus Permasalahan AND	12
2.1.3. Taksonomi Hierarki AND	13
2.1.4. Dataset AND	16
2.2. Text Preprocessing	20
2.3. Porter Stemmer	25
2.4. Lancaster Stemmer	31
2.5. Lemmatization	33
2.6. One Hot Encoder	34
2.7. Term Frequency (TF)	35
2.8. Document Frequency (DF)	36
2.9. TF-IDF	37
2.10. Deep Neural Network (DNN)	38
2.11. Support Vector Machine (SVM)	44
2.12. Principal Component Analysis (PCA)	49
2.13. Performance Measurements	51
BAB III METODOLOGI	53
3.1. Akuisisi Data	54
3.2. Persiapan Data	55
3.2.1. Cleaning Process Data	57

3.2.2. Pemilihan Fitur	58
3.2.3. Komposisi Data.....	58
3.3. Pra-pemrosesan Data	60
3.3.1. Pemrosesan Fitur.....	62
3.3.2. Penggabungan dan Reduksi Fitur	69
3.4. Konsep Tuning Percobaan.....	70
3.5. Klasifikasi.....	71
3.5.1. Klasifikasi DNN	71
3.5.2. Klasifikasi SVM	73
3.6. Evaluasi Model.....	75
BAB IV HASIL DAN PEMBAHASAN	80
4.1. Hasil Akuisisi Data.....	80
4.2. Hasil Persiapan Data	81
4.3. Hasil Pra-pemrosesan	83
4.4. Hasil Klasifikasi	84
4.4.1. Hasil Klasifikasi DNN.....	84
4.4.2. Hasil Klasifikasi SVM.....	88
4.5. Komparasi Hasil	91
BAB V KESIMPULAN DAN SARAN	95
5.1. Kesimpulan.....	95
5.2. Saran.....	96
DAFTAR PUSTAKA	98

DAFTAR GAMBAR

Gambar 2.1. Diagram Venn AND	12
Gambar 2.2. Taksonomi Hierarki AND oleh Ferreira et al. [10]	16
Gambar 2.3. Proses <i>Case Folding</i> pada <i>Text Preprocessing</i>	22
Gambar 2.4. Proses <i>Tokenizing</i> pada <i>Text Preprocessing</i>	22
Gambar 2.5. Proses <i>Filtering/Stopword Removal</i> pada <i>Text Preprocessing</i>	23
Gambar 2.6. Proses <i>Stemming</i> pada <i>Text Preprocessing</i>	24
Gambar 2.7. Algoritma <i>Porter Stemmer</i>	25
Gambar 2.8. Flowchart Proses <i>Porter Stemmer</i>	26
Gambar 2.9. Arsitektur <i>Neural Network</i>	39
Gambar 2.10. Arsitektur <i>Deep Neural Network (DNN)</i>	41
Gambar 2.11. <i>Support Vector Machine (SVM)</i>	46
Gambar 2.12. Langkah <i>Principal Component Analysis (PCA)</i>	51
Gambar 3.1. Metode Penelitian	53
Gambar 3.2. Flowchart Persiapan Data	56
Gambar 3.3. Tahapan Pra-pemrosesan Data	61
Gambar 3.4. Flowchart Pra-pemrosesan Data	62
Gambar 3.5. Flowchart Fitur dengan Teknik OHE	63
Gambar 3.6. Flowchart Fitur <i>Author List</i>	65
Gambar 3.7. Flowchart Fitur <i>Year</i>	66
Gambar 3.8. Flowchart Fitur <i>Venue</i>	67
Gambar 3.9. Flowchart Fitur <i>Title</i>	69
Gambar 3.10. Arsitektur DNN	72
Gambar 3.11. <i>Best Ratio Tuning Gamma dan C</i>	74

Gambar 4.1. Pie Chart Komposisi Data AND.....	82
Gambar 4.2. Grafik <i>Model Loss</i> Terbaik Klasifikasi DNN.....	85
Gambar 4.3. Grafik <i>Model Accuracy</i> Terbaik Klasifikasi DNN	85
Gambar 4.4. Grafik <i>Performance Measurements</i> Klasifikasi DNN.....	88
Gambar 4.5. Grafik <i>Performance Measurements</i> Klasifikasi SVM.....	91

DAFTAR TABEL

Tabel 1. Contoh kasus <i>Homonym</i> dan <i>Synonym</i>	11
Tabel 2. Step 1a – Menghilangkan Sufiks Kata Jamak.....	27
Tabel 3. Step 1b – Menghilangkan Infleksi Kata.....	28
Tabel 4. Step 1b1 – Menghilangkan Infleksi Kata dengan Sufiks ‘-ed’ dan ‘-ing’	28
Tabel 5. Step 1c – Mengubah Kata dengan Sufiks ‘y’ menjadi ‘i’	28
Tabel 6. Step 2 – Mengubah Kata yang memiliki Multi-Sufiks	29
Tabel 7. Step 3 – Mengubah Kata yang memiliki Multi-Sufiks untuk Kata Tertentu.....	29
Tabel 8. Step 4 – Menghilangkan Sufiks Terakhir.....	30
Tabel 9. Step 5a – Menghilangkan Sufiks Kata dengan akhiran ‘e’	31
Tabel 10. Step 5b – Mereduksi Kata	31
Tabel 11. Perbandingan hasil <i>Lancaster Stemmer</i> dan <i>Porter Stemmer</i>	32
Tabel 12. Perbandingan hasil <i>Stemming</i> dan <i>Lemmatization</i>	34
Tabel 13. Contoh <i>One Hot Encoder</i> (OHE)	35
Tabel 14. Detail Dataset <i>The GILES</i>	54
Tabel 15. Tabel Kebenaran Komposisi Data.....	60
Tabel 16. Detail <i>Tuning</i> Klasifikasi DNN.....	73
Tabel 17. Detail <i>Tuning</i> Klasifikasi SVM.....	74
Tabel 18. Tabel Kebenaran <i>Confusion Matrix</i>	75
Tabel 19. Perubahan Sebelum dan Sesudah <i>Cleaning Process Data</i>	82
Tabel 20. Detail Komposisi Data	82
Tabel 21. Detail Fitur	83

Tabel 22. Detail Hasil Klasifikasi DNN.....	84
Tabel 23. <i>Performance Measurements</i> Klasifikasi DNN.....	87
Tabel 24. Detail Hasil Klasifikasi SVM.....	88
Tabel 25. <i>Performance Measurements</i> Klasifikasi SVM.....	90
Tabel 26. Perbandingan Hasil <i>Performance Measurements</i> DNN dan SVM	92

DAFTAR LAMPIRAN

Lampiran 1. Form Perbaikan

Lampiran 2. Cek Plagiat

BAB I

PENDAHULUAN

1.1. Latar Belakang

Suatu sistem dengan tugas dan fungsi yang baik untuk mengatasi disambiguasi kata secara tepat dan akurat masih menjadi satu permasalahan pokok pada penelitian *Word Sense Disambiguation* (WSD) dalam ruang lingkup *Natural Language Processing* (NLP) dan *Artificial Intelligence* (AI) [1], [2]. Pada WSD, suatu kata bisa memiliki arti dan makna yang berbeda karena banyak faktor sehingga sulit untuk mengidentifikasi kata yang dimaksud adalah sama atau tidak sama baik dalam arti, makna serta penulisan kata tersebut. Permasalahan ini juga berlaku untuk disambiguasi nama [3].

Author Name Disambiguation (AND) yang juga disebut *Personal Disambiguation* [3]–[5], merupakan sebuah proses untuk menyelesaikan konflik dan masalah ambiguitas yang terjadi ketika dua atau lebih publikasi memiliki data author yang sama dan sangat mirip. Ambiguitas yang terjadi disebabkan karena tingkat *Homonymity* dan *Synonymity* dari nama author tersebut. Penelitian AND sudah sangat banyak dilakukan oleh berbagai peneliti [6]–[9] dengan berbagai macam bentuk pendekatan dan penerapan metode serta teknik yang berbeda-beda. Teknik pendekatan yang paling banyak digunakan dalam penelitian AND adalah *Supervised Learning* dan *Unsupervised Learning* [2], [10]. Kedua buah teknik ini memiliki perbedaan pada tahap pemrosesan data dalam menentukan suatu klasifikasi dan prediksi terhadap suatu data. Penelitian dalam *Machine Learning* (ML) sudah banyak sekali dilakukan oleh para peneliti dengan metode dan pendekatan yang berbeda-beda dalam pengolahan datanya pada penelitian tentang AND.

AND merupakan sebuah sistem pengelompokan nama author terhadap suatu publikasi penelitian dalam suatu database berdasarkan dengan banyaknya

kesalahan dan konflik kesamaan serta kemiripan yang terjadi pada tiap data sehingga menyebabkan adanya pengambilan keputusan yang ambigu dalam menentukan dan membedakan author tersebut adalah orang yang sama atau tidak sama secara individual [5], [11]–[13]. Permasalahan disambiguasi nama author terutama pada data bibliografi sudah menjadi permasalahan utama dari berbagai macam penelitian sebelumnya bahkan sampai saat ini. Ambiguitas nama pada data bibliografi menjadi hal negatif yang mempengaruhi kualitas pencarian data yang dituju pada perpustakaan digital [7], [9], [14]. Faktanya, dua dari tiga orang author memiliki nama yang ambigu (sama/identik) dan satu dari lima orang author memiliki nama yang bervariasi pada publikasi penelitiannya [12].

Dasar penelitian yang akan dilakukan adalah lanjutan serta pengembangan dari penelitian-penelitian yang sudah ada sebelumnya dengan data dan metode yang saling mendukung untuk menemukan solusi terbaik dalam memecahkan permasalahan AND khususnya identifikasi author. Salah satu metode pendekatan dalam *Machine Learning* (ML) yang relatif stabil untuk mengolah varian data dalam melakukan klasifikasi dan pengambilan keputusan adalah *Deep Neural Network* (DNN) [8]. Selain itu juga ada metode pendekatan dalam *Machine Learning* (ML) yang melakukan klasifikasi dengan menggunakan fungsi-fungsi linier yang berdasarkan pada teori optimasi yaitu sistem pembelajaran *Support Vector Machine* (SVM).

Kedua metode klasifikasi tersebut sangat sering digunakan dalam penelitian AND. Dalam klasifikasi DNN maupun SVM, nilai *Accuracy* menjadi penentu performa klasifikasi tersebut bagus atau tidak terhadap dataset. Khususnya dalam permasalahan AND, nilai *Precision* dan *Recall* menjadi parameter yang penting untuk mengetahui performa klasifikasi yang dilakukan tepat sasaran terhadap permasalahan AND yang dihadapi yaitu identifikasi author [9], [15], [16]. Berdasarkan hal tersebut, maka penelitian ini akan mengklasifikasikan author pada data bibliografi dengan menggunakan *Deep Neural Network* (DNN) dan *Support Vector Machine* (SVM).

1.2. Tujuan dan Manfaat

1.2.1. Tujuan

Tujuan dari penulisan Tugas Akhir ini, yaitu :

1. Menyelesaikan permasalahan *Author Name Disambiguation* (AND) khususnya identifikasi author dengan menggunakan *Deep Neural Network* (DNN) dan *Support Vector Machine* (SVM).
2. Menentukan metode dan pendekatan terbaik dalam menyelesaikan permasalahan *Author Name Disambiguation* (AND) khususnya identifikasi author.

1.2.2. Manfaat

Manfaat dari penulisan Tugas Akhir ini, yaitu :

1. Membantu memecahkan permasalahan *Author Name Disambiguation* (AND) khususnya identifikasi author dengan menggunakan pendekatan *Supervised Learning* yaitu *Deep Neural Network* (DNN) dan *Support Vector Machine* (SVM).
2. Sebagai bahan referensi penelitian selanjutnya untuk dapat terus dikembangkan dalam menciptakan solusi terbaik mengatasi permasalahan *Author Name Disambiguation* (AND).

1.3. Perumusan dan Batasan Masalah

1.3.1. Perumusan Masalah

Bagaimana menentukan metode dan pendekatan terbaik dalam menyelesaikan permasalahan *Author Name Disambiguation* (AND) khususnya identifikasi author pada data bibliografi dengan menggunakan *Deep Neural*

Network (DNN) dan *Support Vector Machine* (SVM) untuk mendapatkan hasil yang akurat.

1.3.2. Batasan Masalah

Berikut batasan masalah pada Tugas Akhir ini, yaitu :

1. Penelitian dilakukan dengan mencakup permasalahan *Author Name Disambiguation* (AND) khususnya identifikasi author.
2. Penelitian ini menggunakan basis bahasa pemrograman *Python*.
3. Dataset bibliografi yang digunakan sebagai bahan penelitian adalah *Dataset DBLP Labeled Data* dari hasil penelitian yang dilakukan oleh Jinseok Kim et al. [17] yaitu dataset yang bersumber dari *dblp.org* dan sudah dilakukan *Cleaning Process*.
4. Hasil penelitian yang dilakukan dengan metode DNN dan SVM akan dibandingkan untuk menentukan *Machine Learning* (ML) terbaik dalam pengolahan permasalahan AND.
5. Output yang dihasilkan dari penelitian ini hanya berupa nilai *Sensitivity*, *Specificity*, *Precision*, *F1-Score*, *Error Rate* dan *Accuracy* sebagai tolak ukur untuk melihat tingkat kecocokan author dengan label.

1.4. Metodologi Penelitian

Pada Tugas Akhir ini, metodologi yang digunakan adalah sebagai berikut :

1.4.1. Metode Studi Pustaka dan Literatur

Pada metode ini, dilakukan pencarian dan pengumpulan referensi berupa literatur yang terdapat pada buku dan internet mengenai klasifikasi *Author Name*

Disambiguation (AND) dengan menggunakan *Deep Neural Network* (DNN) dan atau *Support Vector Machine* (SVM).

1.4.2. Metode Konsultasi

Dalam metode ini penulis melakukan konsultasi secara langsung dan atau tidak langsung kepada semua pihak narasumber yang memiliki pengetahuan serta wawasan yang baik dalam mengatasi permasalahan yang ditemui pada penulisan Tugas Akhir ini tentang klasifikasi *Author Name Disambiguation* (AND) dengan menggunakan *Deep Neural Network* (DNN) dan atau *Support Vector Machine* (SVM).

1.4.3. Metode Pembuatan Model

Metode selanjutnya adalah membuat suatu perancangan pemodelan dengan menggunakan program.

1.4.4. Metode Pengujian dan Validasi

Pengujian terhadap sistem yang telah dibuat perlu dilakukan untuk melihat batasan-batasan kinerja sistem tersebut dapat menghasilkan nilai akurasi yang baik atau sebaliknya.

1.4.5. Metode Hasil dan Analisa

Hasil dari pengujian pada Tugas Akhir ini akan dianalisa seluruh kelebihan serta kekurangannya, sehingga diharapkan dapat digunakan sebagai referensi yang baik untuk penelitian selanjutnya.

1.4.6. Metode Penarikan Kesimpulan dan Saran

Metode ini merupakan tahap akhir dari Metodologi Penelitian, berdasarkan hasil dan analisis penelitian yang dilakukan maka akan didapatkan kesimpulan dan saran untuk penelitian selanjutnya.

1.5. Sistematika Penulisan

Dalam mempermudah penyusunan Tugas Akhir ini dan juga membuat isi dari setiap bab yang ada pada Tugas Akhir ini lebih jelas, maka dibuat sistematika penulisan sebagai berikut :

BAB I – PENDAHULUAN

Sebagai pondasi penelitian, bab ini membahas tentang Latar Belakang Masalah, Tujuan dan Manfaat, Perumusan dan Batasan Masalah, Metode Penelitian, dan Sistematika Penulisan dari penelitian yang dilakukan.

BAB II – TINJAUAN PUSTAKA

Bab selanjutnya merupakan penjelasan Dasar Teori, Konsep dan Prinsip Dasar yang dibutuhkan untuk memecahkan masalah dalam penelitian yang dilakukan.

BAB III – METODOLOGI

Metodologi yang digunakan akan dibahas secara rinci tentang teknik, metode, dan alur proses yang dilakukan dalam penelitian.

BAB IV – HASIL DAN PEMBAHASAN

Bab empat merupakan hasil pengujian dan analisis yang diperoleh dari penelitian serta pembahasan terhadap hasil yang telah dicapai meliputi kelebihan dan kekurangan dari penelitian yang telah dilakukan.

BAB V – KESIMPULAN DAN SARAN

Pada bab terakhir berisi kesimpulan yang bersumber dari hasil penelitian yang dilakukan beserta saran untuk penelitian selanjutnya khususnya tentang *Author Name Disambiguation (AND)*.

DAFTAR PUSTAKA

- [1] I. Iacobacci, M. T. Pilehvar, and R. Navigli, “Embeddings for word sense disambiguation: An evaluation study,” in *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 2016, vol. 2, no. 2003, pp. 897–907.
- [2] N. R. Smalheiser and V. I. Torvik, “Author name disambiguation,” *Annual Review of Information Science and Technology*, vol. 43, no. 1. pp. 287–313, 2009.
- [3] G. S. Mann and D. Yarowsky, “Unsupervised Personal Name Disambiguation,” in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, 2003, pp. 33–40.
- [4] Y. Chen and J. Martin, “Towards Robust Unsupervised Personal Name Disambiguation,” in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007, pp. 190–198.
- [5] D. Bollegala, Y. Matsuo, and M. Ishizuka, “Automatic discovery of personal name aliases from the web,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 6, pp. 831–844, 2011.
- [6] H. Han, W. Xu, H. Zha, and C. L. Giles, “A Hierarchical Naive Bayes Mixture Model for Name Disambiguation in Author Citations,” in *Proceedings of the 2005 ACM Symposium on Applied Computing*, 2005, pp. 1065–1069.
- [7] H. Han, L. Giles, H. Zha, C. Li, and K. Tsioutsoulouklis, “Two supervised learning approaches for name disambiguation in author citations,” in *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, 2004, pp. 296–305.

- [8] H. N. Tran, T. Huynh, and T. Do, “Author name disambiguation by using deep neural network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8397 LNAI, no. PART 1, pp. 123–132, 2014.
- [9] A. A. Ferreira, A. Veloso, M. A. Gonçalves, and A. H. F. Laender, “Effective Self-training Author Name Disambiguation in Scholarly Digital Libraries,” in *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, 2010, pp. 39–48.
- [10] A. A. Ferreira, M. A. Gonçalves, and A. H. F. Laender, “A brief survey of automatic methods for author name disambiguation,” *SIGMOD Record*, vol. 41, no. 2, pp. 15–26, 2012.
- [11] K. Kim, A. Sefid, B. A. Weinberg, and C. L. Giles, “A Web Service for Author Name Disambiguation in Scholarly Databases,” in *2018 IEEE International Conference on Web Services (ICWS)*, 2018, pp. 265–273.
- [12] V. I. Torvik and N. R. Smalheiser, “Author name disambiguation in MEDLINE,” *ACM Trans. Knowl. Discov. Data*, vol. 3, no. 3, pp. 1–29, 2009.
- [13] J. Zhu, Y. Yang, Q. Xie, L. Wang, and S. U. Hassan, “Robust hybrid name disambiguation framework for large databases,” *Scientometrics*, vol. 98, no. 3, pp. 2255–2274, 2014.
- [14] Z. Yu and B. Yang, “Researcher Name Disambiguation: Feature Learning and Affinity Propagation Clustering,” in *Foundations of Intelligent Systems*, 2018, pp. 225–235.
- [15] A. F. Santana, M. A. Gonçalves, A. H. F. Laender, and A. A. Ferreira, “On the combination of domain-specific heuristics for author name disambiguation: the nearest cluster method,” *Int. J. Digit. Libr.*, vol. 16, no. 3–4, pp. 229–246, 2015.
- [16] R. G. Cota, A. A. Ferreira, C. Nascimento, M. A. Gonçalves, and A. H. F.

- Laender, “An unsupervised heuristic-based hierarchical method for name disambiguation in bibliographic citations,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, no. 9, pp. 1853–1870, 2010.
- [17] J. Kim, “Evaluating author name disambiguation for digital libraries: a case of DBLP,” *Scientometrics*, vol. 116, no. 3, pp. 1867–1886, 2018.
- [18] T. K. Saha, B. Zhang, and M. Al Hasan, “Name disambiguation from link data in a collaboration graph using temporal and topological features,” *Soc. Netw. Anal. Min.*, vol. 5, no. 1, pp. 1–14, 2015.
- [19] Y. Qian, Q. Zheng, T. Sakai, J. Ye, and J. Liu, “Dynamic author name disambiguation for growing digital libraries,” *Inf. Retr. J.*, vol. 18, no. 5, pp. 379–412, 2015.
- [20] J. Kim, “Evaluating author name disambiguation for digital libraries: a case of DBLP,” *Scientometrics*, vol. 116, no. 3, pp. 1867–1886, 2018.
- [21] T. Backes, “Effective Unsupervised Author Disambiguation with Relative Frequencies,” in *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, 2018, pp. 203–212.
- [22] W. Liu *et al.*, “Author Name Disambiguation for PubMed,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 765–781, 2014.
- [23] I. Hussain and S. Asghar, “LUCID: Author name disambiguation using graph Structural Clustering,” in *2017 Intelligent Systems Conference (IntelliSys)*, 2017, pp. 406–413.
- [24] D. Shin, T. Kim, J. Choi, and J. Kim, “Author name disambiguation using a graph model with node splitting and merging based on bibliographic information,” *Scientometrics*, vol. 100, no. 1, pp. 15–50, 2014.
- [25] G. Louppe, H. T. Al-Natsheh, M. Susik, and E. J. Maguire, “Ethnicity Sensitive Author Disambiguation Using Semi-supervised Learning,” in *Knowledge Engineering and Semantic Web*, 2016, pp. 272–287.

- [26] B. Zhang and M. Al Hasan, “Name disambiguation in anonymized graphs using network embedding,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1239–1248.
- [27] J. Zhu, X. Wu, X. Lin, C. Huang, G. P. C. Fung, and Y. Tang, “A novel multiple layers name disambiguation framework for digital libraries using dynamic clustering,” *Scientometrics*, vol. 114, no. 3, pp. 781–794, 2018.
- [28] C. Chen, H. Junfeng, and W. Houfeng, “Clustering Technique in Multi-document Personal Name Disambiguation,” in *Proceedings of the ACL-IJCNLP 2009 Student Research Workshop*, 2009, pp. 88–95.
- [29] J. Kim, J. Kim, and J. Owen-Smith, “Generating automatically labeled data for author name disambiguation: an iterative clustering method,” *Scientometrics*, vol. 118, no. 1, pp. 253–280, Jan. 2019.
- [30] B. Zhang, M. Dundar, and M. Al Hasan, “Bayesian Non-Exhaustive Classification A Case Study: Online Name Disambiguation Using Temporal Record Streams,” in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 1341–1350.
- [31] M. Levin, S. Krawczyk, S. Bethard, and D. Jurafsky, “Citation-Based Bootstrapping for Large-Scale Author Disambiguation,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 5, pp. 1030–1047, 2012.
- [32] J. Tang, A. C. M. Fong, B. Wang, and J. Zhang, “A unified probabilistic framework for name disambiguation in digital library,” *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 975–987, 2012.
- [33] D. Zhang, J. Tang, J. Li, and K. Wang, “A Constraint-based Probabilistic Framework for Name Disambiguation,” in *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, 2007, pp. 1019–1022.
- [34] M. Khabsa, P. Treeratpituk, and C. L. Giles, “Online Person Name

- Disambiguation with Constraints,” in *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2015, pp. 37–46.
- [35] V. I. Torvik, M. Weeber, D. R. Swanson, and N. R. Smalheiser, “A probabilistic similarity metric for Medline records: a model for author name disambiguation.,” *AMIA Annu. Symp. Proc.*, vol. 56, no. 2, p. 1033, 2003.
- [36] I. Hussain and S. Asghar, “Author Name Disambiguation by Exploiting Graph Structural Clustering and Hybrid Similarity,” *Arab. J. Sci. Eng.*, vol. 43, no. 12, pp. 7421–7437, 2018.
- [37] Y. Zhang, F. Zhang, P. Yao, and J. Tang, “Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop.,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1002–1011.
- [38] P. Treeratpituk and C. L. Giles, “Disambiguating Authors in Academic Publications Using Random Forests,” in *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2009, pp. 39–48.
- [39] M. Khabsa, P. Treeratpituk, and C. L. Giles, “Large scale author name disambiguation in digital libraries,” in *2014 IEEE International Conference on Big Data (Big Data)*, 2014, pp. 41–42.
- [40] J. Tang, L. Yao, D. Zhang, and J. Zhang, “A combination approach to web user profiling,” *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 1, 2010.
- [41] I. S. Kang, P. Kim, S. Lee, H. Jung, and B. J. You, “Construction of a large-scale test set for author disambiguation,” *Inf. Process. Manag.*, vol. 47, no. 3, pp. 452–465, 2011.
- [42] B. Malley, D. Ramazzotti, and J. T. Wu, “Data Pre-processing,” in *Secondary Analysis of Electronic Health Records*, Cham: Springer International Publishing, 2016, pp. 115–141.
- [43] V. Balakrishnan and E. Lloyd-Yemoh, “Stemming and lemmatization: A

- comparison of retrieval performances,” in *Proceedings of SCEI Seoul Conferences*, 2015, pp. 174–179.
- [44] M. A. El Affendi and K. H. S. Al Rajhi, “Text encoding for deep learning neural networks: A reversible base 64 (Tetrsexagesimal) Integer Transformation (RIT64) alternative to one hot encoding with applications to Arabic morphology,” *6th Int. Conf. Digit. Information, Networking, Wirel. Commun. DINWC 2018*, vol. 64, pp. 70–74, 2018.
- [45] J. D. Prusa and T. M. Khoshgoftaar, “Designing a Better Data Representation for Deep Neural Networks and Text Classification,” in *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, 2016, pp. 411–416.
- [46] J. D. Prusa and T. M. Khoshgoftaar, “Improving deep neural network design with new text data representations,” *J. Big Data*, vol. 4, no. 1, 2017.
- [47] N. Azam and J. Yao, “Comparison of term frequency and document frequency based feature selection metrics in text categorization,” *Expert Syst. Appl.*, vol. 39, no. 5, pp. 4760–4768, 2012.
- [48] L. Adhianto *et al.*, “HPCTOOLKIT: Tools for performance analysis of optimized parallel programs,” *Concurr. Comput. Pract. Exp.*, vol. 22, no. 6, pp. 685–701, 2010.
- [49] S. Shanmuganathan, “Artificial Neural Network Modelling: An Introduction,” in *Artificial Neural Network Modelling*, S. Shanmuganathan and S. Samarasinghe, Eds. Cham: Springer International Publishing, 2016, pp. 1–14.
- [50] M. Shah and R. Kapdi, “Object detection using deep neural networks,” in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2017, pp. 787–790.
- [51] S. Nurmaini, P. R. Umi, R. M. Naufal, and A. Gani, “Cardiac arrhythmias classification using Deep Neural Networks and principle component analysis

- algorithm,” *Int. J. Adv. Soft Comput. its Appl.*, vol. 10, no. 2, pp. 14–32, 2018.
- [52] D. Yu, M. L. Seltzer, J. Li, J.-T. Huang, and F. Seide, “Feature Learning in Deep Neural Networks - Studies on Speech Recognition Tasks,” *arXiv e-prints*, p. arXiv:1301.3605, Jan. 2013.
- [53] A. Ben-Hur and J. Weston, “A User’s Guide to Support Vector Machines,” in *Data Mining Techniques for the Life Sciences*, O. Carugo and F. Eisenhaber, Eds. Totowa, NJ: Humana Press, 2010, pp. 223–239.
- [54] P. J. Dendek, L. Bolikowski, and M. Lukasik, “Evaluation of Features for Author Name Disambiguation Using Linear Support Vector Machines,” in *2012 10th IAPR International Workshop on Document Analysis Systems*, 2012, pp. 440–444.
- [55] A. Al-Anazi and I. D. Gates, “A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs,” *Eng. Geol.*, vol. 114, no. 3–4, pp. 267–277, 2010.
- [56] J. Kim and J. Kim, “The impact of imbalanced training data on machine learning for author name disambiguation,” *Scientometrics*, vol. 117, no. 1, pp. 511–526, 2018.
- [57] M.-C. Müller, F. Reitz, and N. Roy, “Data sets for author name disambiguation: an empirical analysis and a new resource,” *Scientometrics*, vol. 111, no. 3, pp. 1467–1500, 2017.
- [58] R. Sangeetha and D. B. Kalpana, “Identifying Efficient Kernel Function in Multiclass Support Vector Machines,” *Int. J. Comput. Appl.*, vol. 28, no. 8, pp. 18–23, 2011.
- [59] G. S. Cho, N. Gantulga, and Y. W. Choi, “A comparative study on multi-class SVM & kernel function for land cover classification in a KOMPSAT-2 image,” *KSCE J. Civ. Eng.*, vol. 21, no. 5, pp. 1894–1904, 2017.