

**KLASIFIKASI AUTHOR PADA DATA BIBLIOGRAFI
MENGGUNAKAN ARSITEKTUR CAPSULE NEURAL
NETWORKS (CAPSNETS)**

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer



OLEH :

WAIS AL QARNI

09011281722068

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

2021

HALAMAN PENGESAHAN

KLASIFIKASI AUTHOR PADA DATA BIBLIOGRAFI MENGGUNAKAN ARSITEKTUR CAPSULE NEURAL NETWORKS (CAPSNETS)

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh :

WAIS AL QARNI
09011281722068

Palembang, 25 Juni 2021

Mengetahui,

Ketua Jurusan Sistem Komputer,



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

Pembimbing Tugas Akhir,


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

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada:

Hari : Senin

Tanggal : 21 Juni 2021

Tim Penguji :

- 1. Ketua** : Sutarno, S.T., M.T.

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

3. Penguji : Dr. Ir. Bambang Tutuko, M.T.

4. Pembimbing : Firdaus, M.Kom.

Mengetahui,

Ketua Jurusan Sistem Komputer



NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Wais Al Qarni
NIM : 09011281722068
Judul : Klasifikasi *Author* Pada Data Bibliografi Menggunakan Arsitektur *Capsule Neural Networks* (*CapsNets*)

Hasil Penyecekan Software iThenticate/Turnitin : 4 %

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.



Indralaya, Juli 2021



Wais Al Qarni
NIM. 09011281722068

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamualaikum warahmatullahi wabarakatuh

Alhamdulilahirabbil'alamin, puji dan syukur penulis panjatkan kepada Allah *Subhanahu wata 'ala* 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 Arsitektur Capsule Neural Networks (CapsNets)**". Serta sholawat dan salam kepada Nabi Muhammad *Shallallahu 'alaihi wasallam* sebagai sosok idola dan teladan bagi penulis.

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, materil maupun spiritual selama ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Ir. H. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Huda Ubaya, S.T., M.T., selaku Pembimbing Akademik Penulis.

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. Ir. Bambang Tutuko, 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., Mbak Annisa Darmawahyuni, M.Kom., dan Mbak Ade Iriani Sapitri, M.Kom., yang memberikan arahan serta nasihat kepada penulis untuk menyelesaikan Tugas Akhir ini dengan baik.
8. Teman seperjuangan yang tergabung dalam grup riset IsyRG khususnya bidang Teks yang sangat membantu penulis dalam menyelesaikan Tugas Akhir ini.
9. Teman-teman seperjuangan Sistem Komputer Angkatan 2017, Teman-teman dari Organisasi BO Fasco dan LDF WIFI 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 masukkan 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 warahmatullahi wabarakatuh

Palembang, 26 Juli 2021
Penulis,

Wais Al Qarni
NIM. 09011281722068

AUTHOR CLASSIFICATION IN BIBLIOGRAPHIC DATA USING CAPSULE NEURAL NETWORKS (CAPSNETS) ARCHITECTURE

Wais Al Qarni (09011281722068)

Dept. of Computer Engineering, Faculty of Computer Science,

Sriwijaya University

Email : wais.alq99@gmail.com

ABSTRACT

The problem with Author Name Disambiguation is to determine whether the same name in the bibliographic archive is the same author or not. Currently, author identification on DBLP is triggered by an author request who finds his publication mixed with other people's writing. Name ambiguity leads to incorrect identification and attribution of credit to authors. Despite many studies in the last decade, the problem of ambiguity of the author's name is still largely unsolved. In this paper, the Capsule Networks (CapsNets) method is proposed to resolve the ambiguity of the author's name. The proposed method gets the best accuracy on four Name Disambiguation problems including homonyms, synonyms, homonyms- synonyms, and non-homonyms- synonyms, namely an average of 99% on training and testing data. Likewise, the overall data tested has an accuracy of 99.83% with a low error value. In addition, CapsNets were tested with Performance Measurements including Sensitivity, Precision, and F1-Score. CapsNets can identify authors in DBLP bibliographic data by using a number of attributes such as author name, co-author, venue, title, and year.

Keywords: *Author Name Disambiguation, Author identification, Capsule Networks, DBLP, Homonyms, Performance Measurements, Synonyms*

***KLASIFIKASI AUTHOR PADA DATA BIBLIOGRAFI MENGGUNAKAN
ARSITEKTUR CAPSULE NEURAL NETWORKS (CAPSNETS)***

Wais Al Qarni (09011281722068)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : wais.alq99@gmail.com

ABSTRAK

Permasalahan Author Name Disambiguation adalah menentukan nama yang sama pada arsip bibliografi merupakan penulis yang sama atau tidak. Saat ini, identifikasi penulis pada DBLP dipicu oleh permintaan penulis yang menemukan terbitannya bercampur dengan tulisan orang lain. Ambiguitas nama menyebabkan kesalahan identifikasi dan atribusi kredit kepada penulis. Meskipun banyak penelitian dalam dekade terakhir, masalah ambiguitas nama penulis sebagian besar masih belum terpecahkan. Dalam makalah ini, metode Capsule Networks (CapsNets) diusulkan untuk menyelesaikan ambiguitas nama penulis. Metode yang diusulkan mendapatkan akurasi terbaik pada empat masalah Name Disambiguation diantaranya pada homonim, sinonim, homonim-sinonim, dan non-homonim-sinonim yaitu rata-rata 99% pada data training dan testing. Begitupun dengan keseluruhan data yang diuji mendapatkan akurasi 99.83% dengan nilai error yang rendah. Selain itu, CapsNets diuji dengan *Performance Measurements* diantaranya Sensitivity, Precision, dan F1-Score. CapsNets dapat mengidentifikasi author pada data bibliografi DBLP dengan menggunakan sejumlah atribut seperti author name, co-author, venue, title, dan year.

Kata Kunci : *Author Name Disambiguation, Capsule Networks, DBLP, Homonim, Identifikasi penulis, Performance Measurements, Sinonim.*

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN	iv
KATA PENGANTAR	v
ABSTRACT	vii
ABSTRAK.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN.....	xiv
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	4
1.4.3. Metode Pembuatan Model	5
1.4.4. Metode Pengujian dan Validasi.....	5
1.4.5. Metode Analisa Hasil.....	5
1.4.6. Metode Penarikan Kesimpulan dan Saran	5
1.5. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1. Author Name Disambiguation.....	7
2.1.1. Jenis-Jenis Author Name Disambiguation	9
2.2. Jenis Pendekatan Pada Author Name Disambiguation	11
2.3. Metode Author Assignment.....	13
2.4. Teknik Klasifikasi Author Name Disambiguation	14

2.4.1. Supervised AND Techniques.....	16
2.4.2. Unsupervised AND techniques.....	16
2.4.3. Semi-supervised AND techniques	17
2.5. Text Preprocessing	17
2.5.1. Text Cleaning dan Pre-processing	17
2.5.1.1 Tokenization	17
2.5.1.2 Stop Words.....	18
2.5.1.3 Capitalization.....	18
2.5.1.4 Stemming	18
2.5.1.5 Lemmatization.....	19
2.5.2 Syntactic Word Representation	20
2.5.2.1 N-Gram.....	20
2.5.3 Weighted Words	21
2.5.3.1 Bag of Words (BoW)	21
2.5.3.2 Term Frequency-Inverse Document Frequency	22
2.5.4 Word Embedding	23
2.5.4.1 Word2Vec.....	24
2.5.4.2 Global Vectors for Word Representation (GloVe)	25
2.6 Dimensionality Reduction	26
2.6.1 Principal Component Analysis (PCA)	26
2.7 Autoencoder	27
2.7.1 General Framework.....	27
2.8 Deep Neural Networks	28
2.9 Capsules Neural Networks	30
2.9.1 Convolutional layer	32
2.9.2 Primary capsule layer	32
2.9.3 Perhitungan matematis jaringan kapsul.....	35
2.9.4 Jaringan kapsul dengan <i>dynamic routing</i>	36
2.9.5 Jaringan kapsul dengan <i>static routing</i>	37
2.9.6 Perbedaan Arsitektur Neural Networks dan Arsitektur CapNets	37
2.10 Evaluation	39
2.10.1 Macro-Averaging and Micro-Averaging	40
2.10.2 F_β Score.....	41
2.10.3 Matthews Correlation Coefficient (MCC)	41
2.10.4 Area Under ROC Curve (AUC)	42
BAB III METODOLOGI.....	44

3.1.	Akuisisi Data	46
3.1.1.	Pembersihan Data	49
3.1.2.	Pemilihan Fitur	51
3.1.3.	Komposisi Data.....	52
3.2.	Data Preprocessing	54
3.2.1.	Pemrosesan Fitur	56
3.2.2.	Penggabungan Fitur	62
3.3.	Ekstraksi Fitur.....	62
3.4.	Tuning Parameter	63
3.5.	Klasifikasi.....	64
3.6.	Evaluasi Model	65
BAB IV HASIL DAN PEMBAHASAN		69
4.1.	Hasil Akuisisi Data	69
4.2.	Hasil Pembersihan Data	69
4.3.	Hasil Pra-Pemrosesan Data	71
4.4.	Hasil Klasifikasi	73
4.4.1.	Hasil Pengujian Arsitektur CapsNets pada Klasifikasi Author	73
4.4.2.	Hasil Pengujian Pada 4 Model Ekstrasi Fitur.....	77
4.4.3.	Hasil Klasifikasi CapsNets pada Beberapa Kelas Data	79
4.4.4.	Perbandingan Hasil Klasifikasi Terhadap Model Lain.	82
4.4.5.	Perbandingan Hasil Klasifikasi CapsNets Terhadap Model Klasifikasi Lain Pada Beberapa Kelas Data.	84
BAB V KESIMPULAN DAN SARAN.....		86
5.1.	Kesimpulan.....	86
5.2.	Saran	87
DAFTAR PUSTAKA		88

DAFTAR GAMBAR

	Halaman
Gambar 2.1. Penyebab masalah AND	11
Gambar 2.2. Taksonomi Author Name Disambiguation	12
Gambar 2.3. Teknik Penyelesaian Author Name Disambiguation	15
Gambar 2.4. Global Vector for Word Representation	25
Gambar 2.5. Arsitektur sederhana Autoencoder	28
Gambar 2.6. Arsitektur Deep Neural Network	29
Gambar 2.7. Arsitektur Capsule Network pada Klasifikasi Teks	31
Gambar 2.8. Proses Klasifikasi CapsNets	34
Gambar 2.9. Perbedaan arsitektur Neural Networks dan Capsule Networks.....	38
Gambar 3.1. Alur Penelitian Klasifikasi Author	45
Gambar 3.2. Flowchart Data Cleaning	50
Gambar 3.3. Tahapan Preprocessing Data	55
Gambar 3.4. Flowchart Preprocessing Data	56
Gambar 3.5. Flowchart Pemrosesan fitur Author Name	57
Gambar 3.6. Flowchart Pemrosesan Fitur Co-Author Name	58
Gambar 3.7. Flowchart Pemrosesasn Fitur Venue	59
Gambar 3.8. Flowchart Pemrosesan Fitur Title	60
Gambar 3.9. Flowchart Pemrosesan Fitur Unique ID	61
Gambar 3.10. Ekstraksi Fitur	63
Gambar 3.11. Arsitektur Capsnets	64
Gambar 4.1. Pie Chart Komposisi Data	71
Gambar 4.2. Grafik <i>Model Accuracy</i> Terbaik Klasifikasi CapsNets	76
Gambar 4.3. Grafik <i>Model Loss</i> Terbaik Klasifikasi CapsNets	76

DAFTAR TABEL

	Halaman
Tabel 1. Penerapan Stemming pada kata	19
Tabel 2. Perbedaan Hasil Stemming dan Lemmatization	19
Tabel 3. Informasi Dataset <i>The GILES</i>	47
Tabel 4. Deskripsi Kesalahan Data Pada Dataset <i>The GILES</i>	48
Tabel 5. Informasi Dataset <i>The GILES</i> Setelah Perbaikan	49
Tabel 6. Tabel Kebenaran Komposisi Data	54
Tabel 7. Detail <i>Tuning</i> Klasifikasi CapsNets	65
Tabel 8. Tabel Kebenaran <i>Confusion Matrix</i>	66
Tabel 9. Hasil Sebelum dan Sesudah Proses Pembersihan Data	70
Tabel 10. Detail Komposisi Data	71
Tabel 11. Hasil Transformasi Fitur	72
Tabel 12. Hasil Transformasi <i>One Hot Encoder</i>	73
Tabel 13. Hasil klasifikasi model Capsule Networks	73
Tabel 14. Perbandingan Performa Matrik Pada 4 Model Fitur Ekstaksi	78
Tabel 15. <i>Performance Measurements</i> klasifikasi model Capsule Networks	79
Tabel 16. Perbandingan performa CapsNets dengan beberapa model klasifikasi	82
Tabel 17. Perbandingan performa CapsNets pada kategori data dengan beberapa model klasifikasi	84

DAFTAR LAMPIRAN

Lampiran 1. Form Perbaikan

Lampiran 2. Cek Plagiat

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perpustakaan digital ilmiah (*Digital Library*), seperti DBLP, CiteSeer, PubMed dan BDBComp menyediakan fitur dan layanan yang memfasilitasi penelitian dan penemuan literatur ilmiah [1]. Menurut Dongwon Lee dkk [2], yang menjadi tantangan *Digital Library* untuk memiliki konten berkualitas tinggi umumnya adalah masalah ambiguitas pada nama penulis. Tantangan utama pada masalah tersebut adalah menentukan dua nama yang sama atau identik pada catatan arsip bibliografi dengan mengacu pada penulis yang sama atau tidak. Kondisi tersebut dipersulit oleh dua karakteristik nama penulis, yaitu penulis yang berbeda memiliki dan menerbitkan nama yang sama pada catatan kepenggarangan sehingga nama penulis persis sama, tetapi entitas penulis berbeda. Demikian juga, seorang penulis terkadang menggunakan nama yang berbeda seperti menyingkat nama depan atau tengah sehingga menghasilkan catatan kepenggarangan dengan nama penulis yang berbeda, tetapi merujuk kepada entitas penulis yang sama. Pendekatan awal untuk menyelesaikan permasalahan ini sebagian besar mencakup disambiguasi secara manual. Namun, pertumbuhan jumlah peneliti yang semakin meningkat pesat pada perpustakaan digital membuat metode disambiguasi secara manual menjadi tidak praktis [3]. Oleh karena itu, beragam metode terbaru telah diusulkan untuk mengidentifikasi penulis secara otomatis [4].

Secara umum, solusi yang telah dilakukan untuk mengatasi masalah tersebut dapat dibagi menjadi dua pendekatan yaitu *author grouping* dan *author assignment*. Metode *author grouping* menerapkan fungsi kesamaan (*similarity function*) terhadap atribut-atribut referensi penulis untuk memutuskan apakah kumpulan referensi tersebut merujuk terhadap entitas penulis yang sama atau tidak. Fungsi kesamaan dapat dikelompokkan (*clustering*) menggunakan teknik *supervised machine learning* [5]-[6], atau dilakukan representasi korelasi antara

penulis dan rekan penulis dalam bentuk grafik [7]-[8]. Secara umum tidak semua fungsi yang digunakan dalam metode ini bersifat transitif dan biasanya membutuhkan banyak *sample* dan fitur memadai agar berfungsi dengan baik, yang biasanya sangat mahal untuk didapatkan. Selain berdasarkan teknik tersebut, terdapat pendekatan yang juga memberikan hasil terbaik dalam melakukan identifikasi author yaitu teknik *author assignment*.

Metode *author assignment* secara langsung menetapkan setiap referensi ke penulis tertentu dengan membangun model yang mewakili penulis (misalnya, kemungkinan seorang penulis menerbitkan artikel dengan (rekan) penulis lain, di tempat tertentu dan menggunakan daftar istilah tertentu di judul publikasi) menggunakan teknik *supervised machine learning* [9][10]. Metode ini sangat efektif ketika dihadapkan pada sejumlah besar contoh kutipan untuk setiap penulis. Penelitian Han dkk [9] mengusulkan dua model klasifikasi *supervised* yaitu naïve Bayes (NB) dan Support Vector Machine (SVM) dengan sekumpulan atribut seperti co-author names, judul dan venue publikasi untuk mengidentifikasi author pada data DBLP. Hasil penelitian menunjukkan bahwa hasil berdasarkan model SVM lebih akurat yaitu 95.6% sedangkan NB mendapatkan akurasi yaitu 91.3%. Meskipun hasil yang didapatkan cukup baik, tetapi model tersebut masih belum mampu menghasilkan performa maksimal dalam klasifikasi multiclass[11]. Sehingga pada penelitian ini penulis mengusulkan model klasifikasi deep learning yaitu Capsule Neural Networks (CapsNets) [8]-[13] yang telah menunjukkan performa yang efektif dalam penanganan kasus *multiclass classification* [14].

Capsules dengan transformasi matriks memungkinkan jaringan mempelajari hubungan tiap fitur secara keseluruhan, meningkatkan keragaman fitur, dan menangkap fitur pluralistik kalimat lokal yang dapat mengekspresikan fitur teks secara lebih komprehensif [15]. Penelitian ini, secara khusus menggunakan database DBLP untuk melakukan identifikasi penulis berdasarkan sekumpulan data atribut bibliografi seperti *author id*, *author name*, *title of paper*, *venue*, *year*, dan *author list*. Hasil penelitian ini dianalisis melalui *Performance measurement* dengan beberapa indikator seperti Akurasi, Error rate, Presisi, F1-Score, dan Recall.

1.2. Tujuan dan Manfaat

1.2.1. Tujuan

Adapun tujuan dari penulisan Tugas Akhir ini, yaitu:

1. Melakukan identifikasi author pada data bibliografi DBLP dengan membangun model klasifikasi *Capsule Neural Networks* (CapsNets).
2. Mengevaluasi penerapan model *Capsule Neural Networks* (CapsNets) dalam penyelesaian kasus identifikasi author pada data Bibliografi DBLP.

1.2.2. Manfaat

Adapun manfaat dari penulisan Tugas Akhir ini, yaitu:

1. Dapat menjadi solusi pemecahan masalah Ambiguitas Author dengan mengidentifikasi entitas author pada data bibliografi menggunakan metode *Capsule Neural Networks*.
2. Dapat menjadi referensi pada penelitian selanjutnya agar dapat dikembangkan lebih lanjut untuk menemukan solusi terbaik pemecahan masalah Ambiguitas Author.

1.3. Perumusan dan Batasan Masalah

1.3.1. Perumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah bagaimana menerapkan arsitektur *Capsule Neural Networks* (CapsNets) untuk mengklasifikasi ambiguitas author pada data bibliografi DBLP? Untuk dapat menyelesaikan masalah ini maka diterapkan metode CapsNets untuk melakukan klasifikasi.

1.3.2. Batasan Masalah

Berikut batasan masalah pada Tugas Akhir ini, yaitu :

1. Penelitian dilakukan dengan mencakup pada permasalahan *Author Name Disambiguation* (AND).
2. Penelitian ini mencakup program simulasi pada jupyter notebook dengan bahasa pemrograman python tanpa masuk kedalam pengaplikasian model kedalam sistem nyata.
3. Dataset yang digunakan pada penelitian ini berupa dataset bibliografi DBLP [16] yang telah melalui proses pembersihan data (*Cleaning Process*) tanpa dilakukan perbandingan dengan data bibliografi sejenis.
4. Keluaran yang dihasilkan dalam penelitian identifikasi author ini berupa hasil pengukuran nilai akurasi, presisi, f1-score, specificity, recall, dan nilai error.

1.4. Metodologi Penelitian

Penelitian ini menerapkan sejumlah metode diantaranya:

1.4.1. Metode Studi Pustaka dan Literatur

Pada metode ini, penulis melakukan pendalaman materi dengan mencari sejumlah referensi terkait kasus yang diangkat khususnya masalah *Auhor Name Disambiguation*, melakukan analisis terhadap paper terkait dan juga semua teori-teori yang berhubungan dengan masalah dan metode yang diangkat yang bersumber dari buku, e-jurnal, dan laporan tugas akhir.

1.4.2. Metode Konsultasi

Pada metode ini, penulis melakukan konsultasi dengan dosen pembimbing dan beberapa pihak yang memiliki pengetahuan dan riset dibidang ambiguitas author. Konsultasi dilakukan untuk mengetahui lebih dalam mengenai riset tersebut, masalah yang belum terpecahkan pada riset sebelumnya dan konsultasi

mengenai algoritma dan model yang perlu untuk diuji dalam penyelesaian kasus ambiguitas auhor.

1.4.3. Metode Pembuatan Model

Pada metode ini, Penulis melakukan pengujian model dengan membangun arsitektur model deep learning untuk klasifikasi author, kemudian membandingkannya dengan model klasifikasi lain.

1.4.4. Metode Pengujian dan Validasi

Pada metode ini, Penulis melakukan pengujian dan validasi terhadap model yang diterapkan untuk melihat performa yang dihasilkan agar dapat dilakukan perbaikan pada arsitektur yang dibangun.

1.4.5. Metode Analisa Hasil

Pada metode ini, penulis melakukan analisa terhadap hasil yang diperoleh dari klasifikasi terhadap model yang dibangun untuk mengetahui model yang dibangun sesuai dengan tujuan penelitian.

1.4.6. Metode Penarikan Kesimpulan dan Saran

Pada metode ini, Penulis mengambil kesimpulan berdasarkan hasil penelitian, yang kemudian dapat digunakan sebagai referensi untuk penelitian selanjutnya.

1.5. Sistematika Penulisan

Untuk mempermudah dalam penyusunan Tugas akhir ini, maka dibuat sistematika penulisan sehingga isi dari setiap Bab lebih jelas. Adapun sistematika penulisan pada penelitian ini yaitu sebagai berikut:

DAFTAR PUSTAKA

- [1] S. A. Hussain, Ijaz, “A survey of author name disambiguation techniques;,” *Knowl. Eng. Rev.*, vol. 20, no. 2, pp. 117–125, 2017, doi: 10.1017/S0000000000000000.
- [2] B.-W. O. Dongwon Lee, Jaewoo Kang, Prasenjit Mitra, C. Lee Giles, “Are your citations clean?,” *Commun. ACM*, vol. 50, no. 12 (2007), pp. 33–38, 2017, doi: 10.1039/9781782628361-00169.
- [3] Y. Qian, Q. Zheng, T. Sakai, J. Ye, and J. Liu, “Dynamic author name disambiguation for growing digital libraries,” *Inf. Retr. Boston.*, vol. 18, no. 5, pp. 379–412, 2015, doi: 10.1007/s10791-015-9261-3.
- [4] Q. Shen, T. Wu, H. Yang, Y. Wu, H. Qu, and W. Cui, “NameClarifier: A Visual Analytics System for Author Name Disambiguation,” *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 141–150, 2017, doi: 10.1109/TVCG.2016.2598465.
- [5] B. W. On and D. Lee, “Scalable name disambiguation using multi-level graph partition,” in *Proceedings of the 7th SIAM International Conference on Data Mining*, 2007, pp. 575–580, doi: 10.1137/1.9781611972771.64.
- [6] A. Culotta, P. Kanani, R. Hall, M. Wick, and A. McCallum, “Author disambiguation using error-driven machine learning with a ranking loss function,” *AAAI Work. - Tech. Rep.*, vol. WS-07-14, pp. 32–37, 2007.
- [7] B. On, E. Elmacioglu, and D. Lee, “Improving Grouped-Entity Resolution using Quasi-Cliques,” in *Sixth International Conference on Data Mining (ICDM’06)*, 2006, pp. 1008–1015.
- [8] X. Fan, J. Wang, X. Pu, L. Zhou, and B. Lv, “On graph-based name disambiguation,” *J. Data Inf. Qual.*, vol. 2, no. 2, pp. 1–23, 2011, doi:

- 10.1145/1891879.1891883.
- [9] H. Han, L. Giles, H. Zha, C. Li, and K. Tsoutsouliklis, “Two supervised learning approaches for name disambiguation in author citations,” in *Proceedings of the ACM IEEE International Conference on Digital Libraries, JCDL 2004*, 2004, pp. 296–305, doi: 10.1145/996350.996419.
 - [10] 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 ACM International Conference on Digital Libraries*, 2010, pp. 39–48, doi: 10.1145/1816123.1816130.
 - [11] D. Tomar and S. Agarwal, “A comparison on multi-class classification methods based on least squares twin support vector machine,” *Knowledge-Based Syst.*, vol. 81, no. February, pp. 131–147, 2015, doi: 10.1016/j.knosys.2015.02.009.
 - [12] L. Xiao, H. Zhang, W. Chen, Y. Wang, and Y. Jin, “MCapsNet: Capsule network for text with multi-task learning,” *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 4565–4574, 2020, doi: 10.18653/v1/d18-1486.
 - [13] H. Peng *et al.*, “Hierarchical Taxonomy-Aware and Attentional Graph Capsule RCNNs for Large-Scale Multi-Label Text Classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 4347, no. c, pp. 1–14, 2019, doi: 10.1109/TKDE.2019.2959991.
 - [14] D. Pan, Y. Lu, and P. Kang, *A Deep Learning Model for Multi-label Classification Using Capsule Networks*, vol. 11643 LNCS. Springer International Publishing, 2019.
 - [15] L. L. D. L. M. Chen, C. He, “Text Classification Based on A New Joint Network,” in *5th International Conference on Control, Robotics and Cybernetics (CRC)*, 2020, pp. 13–18, doi: 10.1016/j.ins.2018.10.006.
 - [16] J. Kim, “Evaluating author name disambiguation for digital libraries: a case

- of DBLP,” *Scientometrics*, vol. 116, no. 3, pp. 1867–1886, 2018, doi: 10.1007/s11192-018-2824-5.
- [17] 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, doi: 10.1007/s11192-017-2363-5.
 - [18] J. Kim and J. Kim, “Effect of forename string on author name disambiguation,” *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 7, pp. 839–855, 2020, doi: 10.1002/asi.24298.
 - [19] J. M. B. Silva and F. Silva, “Feature extraction for the author name disambiguation problem in a bibliographic database,” *Proc. ACM Symp. Appl. Comput.*, vol. Part F1280, pp. 783–789, 2017, doi: 10.1145/3019612.3019663.
 - [20] A. A. Ferreira, M. A. Gonçalves, and A. H. F. Laender, “A brief survey of automatic methods for author name disambiguation,” *SIGMOD Rec.*, vol. 41, no. 2, pp. 15–26, 2012, doi: 10.1145/2350036.2350040.
 - [21] K. Kowsari, K. J. Meimandi, M. Heidarysafa, and S. Mendu, “Text Classification Algorithms : A Survey,” *Information*, vol. 18, p. 150, 2019, doi: 10.3390/info10040150.
 - [22] G. Gupta and S. Malhotra, “Text Document Tokenization for Word Frequency Count using Rapid Miner (Taking Resume as an Example),” *Int. Conf. Adv. Eng. Technol.*, no. Icaet, pp. 24–26, 2015.
 - [23] T. Verma, R. Renu, and D. Gaur, “Tokenization and Filtering Process in RapidMiner,” *Int. J. Appl. Inf. Syst.*, vol. 7, no. 2, pp. 16–18, 2014, doi: 10.5120/ijais14-451139.
 - [24] H. Saif, M. Fernandez, Y. He, and H. Alani, “On stopwords, filtering and data sparsity for sentiment analysis of twitter,” *Proc. 9th Int. Conf. Lang. Resour. Eval. Lr. 2014*, pp. 810–817, 2014.

- [25] V. Gupta and G. S. Lehal, “A Survey of Text Mining Techniques and Applications - Volume 1, No. 1, August 2009 - JETWI,” *J. Emerg. Technol. Web Intell.*, vol. 1, no. 1, pp. 60–76, 2009.
- [26] M. K. Dalal and M. A. Zaveri, “Automatic Text Classification: A Technical Review,” *Int. J. Comput. Appl.*, vol. 28, no. 2, pp. 37–40, 2011, doi: 10.5120/3358-4633.
- [27] H. A. Almuzaini and A. M. Azmi, “Impact of Stemming and Word Embedding on Deep Learning-Based Arabic Text Categorization,” *IEEE Access*, vol. 8, pp. 127913–127928, 2020, doi: 10.1109/ACCESS.2020.3009217.
- [28] M. Schmitt and M. Constant, “Neural Lemmatization of Multiword Expressions,” in *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, 2019, pp. 142–148, doi: 10.18653/v1/w19-5117.
- [29] Y. Zhang, M. Lease, and B. C. Wallace, “Active discriminative text representation learning,” in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017, pp. 3386–3392.
- [30] J. G. Flores and L. Gillard, “Bag-of-senses versus bag-of-words: comparing semantic and lexical approaches on sentence extraction,” in *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2018, pp. 672–675.
- [31] S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents,” *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.
- [32] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *J. Innov. Eng. Educ.*, vol. 3, no. March, pp. 71–77, 2020.
- [33] C. D. Pennington, Jeffrey and Socher, Richard and Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 conference on*

- empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [34] C. R. García-Alonso, L. M. Pérez-Naranjo, and J. C. Fernández-Caballero, “Visualizing Data using t-SNE Laurens,” *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008, doi: 10.1007/s10479-011-0841-3.
 - [35] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.
 - [36] W. Wang, Y. Huang, Y. Wang, and L. Wang, “Generalized autoencoder: A neural network framework for dimensionality reduction,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 490–497.
 - [37] H. Liang, X. Sun, Y. Sun, and Y. Gao, “Text feature extraction based on deep learning: a review,” *Eurasip J. Wirel. Commun. Netw.*, vol. 2017, no. 1, pp. 1–12, 2017, doi: 10.1186/s13638-017-0993-1.
 - [38] C. A. P. Sarath *et al.*, “An autoencoder approach to learning bilingual word representations,” *Adv. Neural Inf. Process. Syst.*, vol. 3, no. January, pp. 1853–1861, 2014.
 - [39] S. Nurmaini, R. F. Malik, A. Darmawahyuni, and M. Naufal, “Author identification in bibliographic data using deep neural networks,” *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 19, no. 3, pp. 911–919, 2021, doi: 10.12928/TELKOMNIKA.v19i3.18877.
 - [40] H. D. Aykut Çayır, Ugur Ünal, “Random CapsNet for est model for imbalanced malware type classification task ~ Ünal , Hasan Da g,” *Comput. Secur.*, vol. 102, p. 102133, 2021, doi: 10.1016/j.cose.2020.102133.
 - [41] M. Yang, W. Zhao, L. Chen, Q. Qu, Z. Zhao, and Y. Shen, “Investigating the transferring capability of capsule networks for text classification,” *Neural Networks*, vol. 118, pp. 247–261, 2019, doi:

- 10.1016/j.neunet.2019.06.014.
- [42] R. Katarya and Y. Arora, “Study on Text Classification using Capsule Networks,” *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 501–505, 2019, doi: 10.1109/ICACCS.2019.8728394.
 - [43] N. Hinton, Geoffrey E. Sabour, Sara. Frosst, “MATRIX CAPSULES WITH EM ROUTING,” in *conference paper at ICLR 2018*, 2018, vol. 41, no. 1, pp. 43–50, doi: 10.2514/1.562.
 - [44] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao, “Investigating capsule networks with dynamic routing for text classification,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 2020, pp. 3110–3119, doi: 10.18653/v1/d18-1350.
 - [45] J. Kim, S. Jang, E. Park, and S. Choi, “Text classification using capsules,” *Neurocomputing*, vol. 376, pp. 214–221, 2020, doi: 10.1016/j.neucom.2019.10.033.
 - [46] J. Kim and J. Kim, “The impact of imbalanced training data on machine learning for author name disambiguation,” *Scientometrics*, vol. 9, 2018, doi: 10.1007/s11192-018-2865-9.
 - [47] J. Kim, “Evaluating author name disambiguation for digital libraries : a case of DBLP,” *Scientometrics*, vol. 116, pp. 1867–1886, 2018, doi: 10.1007/s11192-018-2824-5.
 - [48] 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, pp. 15–50, 2014, doi: 10.1007/s11192-014-1289-4.
 - [49] A. Filipe, S. Marcos, 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, pp.

229–246, 2015, doi: 10.1007/s00799-015-0158-y.

- [50] 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, doi: 10.1007/s11192-018-2865-9.
- [51] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, “Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods,” *Appl. Soft Comput. J.*, vol. 86, p. 105836, 2020, doi: 10.1016/j.asoc.2019.105836.
- [52] A. Subasi, *Chapter 2 - Data preprocessing*. Elsevier Inc., 2020.
- [53] A. Nurdin, B. A. S. Aji, A. Bustamin, and Z. Abidin, “Perbandingan Kinerja Word Embedding Word2Vec , Glove ,” *J. TEKNOKOMPAK*, vol. 14, no. 2, pp. 74–79, 2020.
- [54] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 3857–3867, 2017.