

**Klasifikasi *Author Matching* pada Data Bibliografi  
menggunakan Metode *Recurrent Neural Networks***

**SKRIPSI**

**Diajukan Untuk Melengkapi Salah Satu Syarat**

**Memperoleh Gelar Sarjana Komputer**



**OLEH :**

**M Jorgi Ramadhan**

**09011381722125**

**JURUSAN SISTEM KOMPUTER**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2023**

**LEMBAR PENGESAHAN**

**Klasifikasi *Author Matching* pada Data Bibliografi menggunakan  
Metode *Recurrent Neural Networks***

**SKRIPSI**

**Diajukan Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sistem Komputer**

Oleh:

**M Jorgi Ramadhan**

**09011381722125**

**Palembang, <sup>26</sup> Mei 2023**


Mengetahui,

Pembimbing Tugas Akhir

Ketua Jurusan Sistem Komputer

  
**Dr. Firdaus, S.T., M.Kom.**

**NIP. 197801212008121003**

  
**Dr. Ir. H. Sukemi, M.T**

**NIP. 196612032006041001**



## HALAMAN PERSETUJUAN

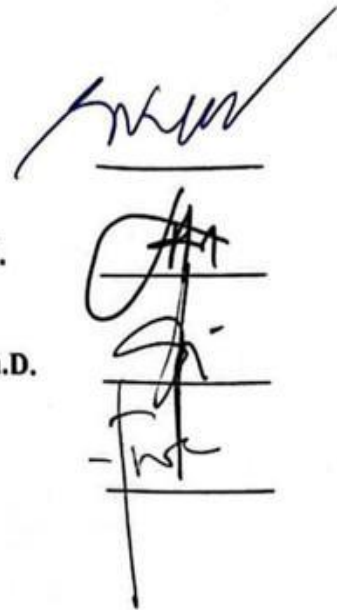
Telah diuji dan lulus pada :

Hari : Rabu

Tanggal : 1 Maret 2023

Tim Penguji :

1. Ketua : Dr. Ir. H. Sukemi, M.T.
2. Sekretaris : Abdurahman, S. Kom., M. HAN.
3. Penguji : Prof. Ir. Siti Nurmaini, M.T., Ph.D.
4. Pembimbing : Dr. Firdaus, S.T., M.Kom.



Mengetahui, 28/3/23

Ketua Jurusan Sistem Komputer



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

NIP. 196612032006041001

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : M Jorgi Ramadhan

NIM : 09011381722125

Judul : Klasifikasi *Author Matching* pada Data Bibliografi menggunakan Metode *Recurrent Neural Networks*

Hasil Pengecekan Software Turnitin : 14%

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, Mei 2023



M Jorgi Ramadhan

NIM. 09011381722125

## KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Puji syukur Alhamdulillah penulis panjatkan atas kehadiran Allah SWT yang telah memberikan karunia dan rahmat-Nya, sehingga penulis dapat menyelesaikan penulisan Proposal Tugas Akhir ini yang berjudul “**Klasifikasi Author Matching pada Data Bibliografi menggunakan Metode Recurrent Neural Networks**”.

Dalam laporan ini penulis menjelaskan mengenai pemodelan untuk identifikasi dan klasifikasi author terhadap suatu publikasi dengan disertai data-data yang diperoleh penulis saat melakukan penelitian dan pengujian data. Penulis berharap agar tulisan ini dapat bermanfaat bagi orang banyak.

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada beberapa pihak atas ide dan saran serta bantuannya dalam menyelesaikan penulisan Proposal Tugas Akhir ini. Oleh karena itu, penulis ingin mengucapkan rasa syukur kepada Allah SWT dan terimakasih kepada yang terhormat :

1. Allah SWT, yang telah memberikan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penulisan Proposal Tugas Akhir ini dengan baik dan lancar.

2. Orang tua saya tercinta yang telah membesarkan saya dengan penuh kasih sayang dan selalu mengajarkan saya dalam berbuat hal yang baik. Terimakasih untuk segala do'a, motivasi dan dukungannya baik moril, materil maupun spritual selama ini.

3. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.

4. Bapak Dr. Ir. H. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

5. Bapak Dr. Firdaus, M.Kom, selaku Dosen 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 Prof. Dr. Erwin, M.Si., selaku Pembimbing Akademik Jurusan Sistem Komputer.

7. Mbak Renny selaku admin Jurusan Sistem Komputer yang telah membantu mengurus seluruh berkas.

8. Sahabat dekat dilingkungan saya yang selalu memberi motivasi yaitu pras,angga,dwiki,udin,wawan,ade dan yuka.

9. Teman-teman satu angkatan dan IsysRG yang selama masa perkuliahan membantu saya dalam proses mencari ilmu.

10. Dan semua pihak yang telah membantu.

Penulis menyadari bahwa laporan ini masih sangat jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun sangatlah diharapkan penulis. Akhir kata penulis berharap, semoga proposal tugas akhir ini bermanfaat dan berguna bagi khalayak.

Wassalamu'alaikum Wr. Wb.

Palembang, 1 Maret 2023

Penulis,



M Jorgi Ramadhan

NIM. 09011381722125

**Klasifikasi *Author Matching* pada Data Bibliografi menggunakan Metode  
*Recurrent Neural Networks***

**M Jorgi Ramadhan (09011381722125)**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : [mjorgi99@gmail.com](mailto:mjorgi99@gmail.com)

**ABSTRAK**

Dalam analisis data bibliografi, kecocokan penulis yang akurat sangat penting untuk berbagai aplikasi seperti analisis kutipan dan profil penelitian. Penelitian ini berfokus pada klasifikasi kecocokan penulis pada data bibliografi menggunakan metode RNNs. Dengan memanfaatkan kekuatan RNNs, tujuan penulis adalah meningkatkan akurasi klasifikasi kecocokan penulis pada dua skenario khusus yaitu kasus homonim dan sinonim. Penulis menggunakan parameter berupa *Author*, *Co Author*, Judul, Tahun serta *Venue*. Pada kasus homonim penulis berhasil mencapai tingkat akurasi yang sangat baik sebesar 96,19% menggunakan model terbaik. Tingkat akurasi yang tinggi ini menunjukkan efektivitas metode RNNs dalam membedakan antara penulis dengan nama yang mirip. Selain itu, penulis juga menyelidiki kasus sinonim, di mana penulis dapat memiliki nama yang berbeda tetapi merujuk pada entitas yang sama. Pada kasus ini, Penulis memperoleh tingkat akurasi yang baik sebesar 9,81%, menunjukkan potensi RNNs dalam mengidentifikasi penulis sinonim dalam data bibliografi. Secara keseluruhan, penelitian ini menyoroti peran penting RNNs dalam klasifikasi kecocokan penulis pada data bibliografi walau menggunakan data yang tidak tabular, hasil ini menunjukkan keefektifan metode ini dalam membedakan antara penulis dengan nama yang mirip, baik pada kasus homonim maupun sinonim.

**Keywords** : Data bibliografi, *Author Matching*, Klasifikasi, Homonim, Sinonim, *Recurrent Neural Network*.

***Author Matching Classification in Bibliographic Data using Recurrent Neural Networks Method.***

**M Jorgi Ramadhan (09011381722125)**

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

*Email : [mjorgi99@gmail.com](mailto:mjorgi99@gmail.com)*

***ABSTRACT***

*In bibliographic data analysis, accurate author matching is crucial for various applications such as citation analysis and research profiling. This research focuses on the classification of author matching in bibliographic data using RNNs method. By leveraging the power of RNNs, the authors aim to improve the accuracy of author matching classification in two specific scenarios: homonym and synonym cases. The authors utilize parameters such as Author, Co-Author, Title, Year, and Venue. In the homonym case, the authors achieved an excellent accuracy rate of 96.19% using the best-performing model. This high accuracy demonstrates the effectiveness of the RNNs method in distinguishing between authors with similar names. Furthermore, the authors also investigated the synonym case, where authors may have different names but refer to the same entity. In this scenario, they obtained a good accuracy rate of 9.81%, indicating the potential of RNNs in identifying synonymous authors in bibliographic data. Overall, this research highlights the important role of RNNs in author matching classification in bibliographic data. Despite using non-tabular data, the results demonstrate the effectiveness of this method in distinguishing between authors with similar names, both in homonym and synonym cases.*

***Keywords:*** *Bibliographic data, Author Matching ,Classification, Homonym, Synonym, Recurrent Neural Network.*



## DAFTAR ISI

|  |             |
|--|-------------|
| <b>LEMBAR PENGESAHAN .....</b>                     | <b>ii</b>   |
| <b>HALAMAN PERSETUJUAN.....</b>                    | <b>iii</b>  |
| <b>HALAMAN PERNYATAAN.....</b>                     | <b>iv</b>   |
| <b>KATA PENGANTAR.....</b>                         | <b>v</b>    |
| <b>ABSTRAK .....</b>                               | <b>vii</b>  |
| <b>ABSTRACT .....</b>                              | <b>viii</b> |
| <b>DAFTAR ISI.....</b>                             | <b>ix</b>   |
| <b>DAFTAR GAMBAR.....</b>                          | <b>xii</b>  |
| <b>DAFTAR TABEL.....</b>                           | <b>xiv</b>  |
| <b>BAB I PENDAHULUAN.....</b>                      | <b>1</b>    |
| 1.1. Latar Belakang .....                          | 1           |
| 1.2. Tujuan .....                                  | 2           |
| 1.3. Perumusan dan Batasan Masalah.....            | 2           |
| 1.3.1. Perumusan Masalah.....                      | 2           |
| 1.3.2. Batasan Masalah.....                        | 2           |
| 1.4. Metodologi Penelitian.....                    | 3           |
| 1.4.1. Metode Studi Pustaka dan Literature .....   | 3           |
| 1.4.2. Metode Konsultasi.....                      | 3           |
| 1.4.3. Metode Pembuatan Model.....                 | 3           |
| 1.4.4. Metode Pengujian.....                       | 4           |
| 1.4.5. Metode Analisa dan Kesimpulan .....         | 4           |
| 1.4.6. Metode Penarikan Kesimpulan dan Saran ..... | 4           |
| 1.5. Sistematika Penelitian .....                  | 4           |
| <b>BAB II TINJAUAN PUSTAKA.....</b>                | <b>6</b>    |
| 2.1. Author Name Disambiguation.....               | 6           |
| 2.2. Taksanomi Author Name Disambiguation .....    | 6           |
| 2.3. Author Grouping .....                         | 7           |
| 2.4. Class Imbalance and Data Imbalace .....       | 8           |
| 2.3. <i>Text Pre-Procesing</i> .....               | 8           |
| 2.4. Normalisasi Teks.....                         | 8           |
| 2.4.1. Tokenizing.....                             | 8           |

|  |  |           |
|--|--|-----------|
| 2.4.2.                                     | Case Folding.....                                      | 9         |
| 2.4.3.                                     | <i>Punctuation</i> (Tanda Baca).....                   | 9         |
| 2.4.4.                                     | Text Filtering.....                                    | 10        |
| 2.4.5.                                     | Stemming .....   | 10        |
| 2.5.                                       | Kombinasi Data.....                                    | 11        |
| 2.6.                                       | Similarity measure and Jaro Winkler Similarity.....    | 12        |
| 2.7.                                       | Label Encoder .....                                    | 13        |
| 2.8.                                       | Min-Max Scaler .....                                   | 13        |
| 2.9.                                       | Recurrent Neural Network.....                          | 14        |
| 2.10.                                      | Performance measurement.....                           | 15        |
| <b>BAB III METODOLOGI PENELITIAN .....</b> |  | <b>16</b> |
| 3.1.                                       | Pendahuluan .....                                      | 16        |
| 3.2.                                       | Kerangka Kerja .....                                   | 16        |
| 3.3.                                       | Akuisisi Data.....                                     | 18        |
| 3.4.                                       | Komposisi Data.....                                    | 19        |
| 3.5.                                       | Pra-pemrosesan Data.....                               | 21        |
| 3.5.1.                                     | Pemrosesan Fitur .....                                 | 23        |
| 3.5.2.                                     | Penggabungan Fitur.....                                | 27        |
| 3.6.                                       | Percobaan Tunning.....                                 | 28        |
| 3.8.                                       | Klasifikasi .....                                      | 29        |
| 3.8.1.                                     | Klasifikasi Recurrent Neural Network.....              | 29        |
| 3.9.                                       | Evaluasi Model.....                                    | 29        |
| <b>BAB IV HASIL DAN PEMBAHASAN.....</b>    |  | <b>32</b> |
| 4.1.                                       | Hasil Akuisisi Data.....                               | 32        |
| 4.2.                                       | Hasil Kombinasi data .....                             | 33        |
| 4.3.                                       | Similarity Measurement .....                           | 35        |
| 4.3.1.                                     | Hasil Similarity Fitur Year .....                      | 35        |
| 4.3.2.                                     | Hasil Similarity Pada Data Variabel Kategorikal .....  | 36        |
| 4.4.                                       | Hasil Klafikasi.....                                   | 41        |
| 4.5.                                       | Confusion Matrix dan Evaluasi Model.....               | 46        |
| 4.7.                                       | Pengujian Sinonim dan Homonim.....                     | 48        |
| 4.7.1.                                     | Hasil Confusion Matrix Homonim dan Sinonim .....       | 48        |
| 4.7.2.                                     | Evaluasi Model terhadap kasus Homonim dan Sinonim..... | 50        |
| <b>BAB V KESIMPULAN DAN SARAN .....</b>    |  | <b>52</b> |
| 5.1.                                       | Kesimpulan .....                                       | 52        |
| 5.2.                                       | Saran .....  | 52        |

|                             |           |
|-----------------------------|-----------|
| <b>DAFTAR PUSTAKA .....</b> | <b>53</b> |
|-----------------------------|-----------|

## DAFTAR GAMBAR

|   | <b>Halaman</b> |
|---|----------------|
| <b>Gambar 2.1.</b> Taksonomi <i>Author Name Disambiguation</i> .....    | 7              |
| <b>Gambar 2.2.</b> Proses <i>Tokenizing</i> .....                       | 9              |
| <b>Gambar 2.3.</b> Proses <i>Case Folding</i> .....                     | 9              |
| <b>Gambar 2.4.</b> Proses <i>Punctuation</i> .....                      | 10             |
| <b>Gambar 2.5.</b> Proses <i>Filtering (Stopword Removal)</i> .....     | 10             |
| <b>Gambar 2.6.</b> Proses <i>Stemming</i> .....                         | 11             |
| <b>Gambar 2.7.</b> Contoh Kombinasi Data .....                          | 12             |
| <b>Gambar 2.7.</b> Contoh <i>Label Encoder</i> .....                    | 13             |
| <b>Gambar 2.8.</b> Proses RNNs.....                                     | 14             |
| <b>Gambar 3.1.</b> Diagram alir Penelitian.....                         | 18             |
| <b>Gambar 3.2.</b> Pra Pemrosesan Data Atribut Fitur.....               | 22             |
| <b>Gambar 3.3.</b> Pra Pemrosesan Data Atribut Label.....               | 22             |
| <b>Gambar 3.4.</b> <i>Flowchart</i> Pra Pemrosesan Data.....            | 23             |
| <b>Gambar 3.5.</b> <i>Flowchart</i> Pemrosesan Fitur <i>Year</i> .....  | 26             |
| <b>Gambar 3.6.</b> <i>Flowchart</i> Pemrosesan Fitur Label.....         | 27             |
| <b>Gambar 4.1.</b> Sample Data.....                                     | 31             |
| <b>Gambar 4.2.</b> Sample data Kombinasi fitur <i>Author Name</i> ..... | 32             |
| <b>Gambar 4.3.</b> Hasil Kombinasi fitur <i>Tahun</i> .....             | 33             |
| <b>Gambar 4.4.</b> Hasil Kombinasi fitur <i>Title</i> .....             | 33             |

|  |    |
|--|----|
| <b>Gambar 4.5.</b> Hasil Kombinasi fitur Venue.....                            | 34 |
| <b>Gambar 4.6.</b> Sampel data untuk Hasil fitur Year .....                    | 35 |
| <b>Gambar 4.7.</b> Sampel data hasil similarity Author Name.....               | 36 |
| <b>Gambar 4.8.</b> Hasil Similarity Pada fitur Kategorikal.....                | 37 |
| <b>Gambar 4.9.</b> Visualisasi Hasil Similarity.....                           | 38 |
| <b>Gambar 4.10.</b> Hasil Similarity pada Labels.....                          | 39 |
| <b>Gambar 4.11.</b> Pie Char Labels.....                                       | 39 |
| <b>Gambar 4.12.</b> Grafik akurasi Model 0.....                                | 42 |
| <b>Gambar 4.13.</b> Grafik loss Model 0.....                                   | 43 |
| <b>Gambar 4.14.</b> Grafik akurasi Model 1.....                                | 43 |
| <b>Gambar 4.15.</b> Grafik loss Model 1.....                                   | 44 |
| <b>Gambar 4.16.</b> Grafik akurasi Model 2.....                                | 44 |
| <b>Gambar 4.17.</b> Grafik loss Model .....                                    | 45 |
| <b>Gambar 4.18.</b> Hasil Confusion Matrix Model 5 terhadap kasus Sinonim..... | 49 |
| <b>Gambar 4.19.</b> Hasil Confusion Matrix Model 3 terhadap kasus Homonim..... | 49 |

## DAFTAR TABEL

|   | <b>Halaman</b> |
|---|----------------|
| <b>Tabel 3.1.</b> Deskripsi Dataset DBLP .....                                  | 19             |
| <b>Tabel 3.2.</b> Komposisi Data.....   | 21             |
| <b>Tabel 3.3.</b> Tabel Kebenaran Confusion Matrix .....                        | 29             |
| <b>Tabel 4.1.</b> Rata-rata ekstraksi tiap fitur.....                           | 40             |
| <b>Tabel 4.2.</b> Rincian Data Latih dan Test Keseluruhan Klasifikasi .....     | 41             |
| <b>Tabel 4.3.</b> Hasil Akurasi 5 Model Pertama.....                            | 41             |
| <b>Tabel 4.4.</b> Hasil Performance Measurement.....                            | 46             |
| <b>Tabel 4.7.</b> Hasil uji semua model terhadap kasus homonim dan sinonim..... | 50             |

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Dengan meningkatnya catatan digital di berbagai domain, ambiguitas nama di informasi pada jurnal menimbulkan tantangan besar dalam beberapa tugas seperti mencari catatan penulis tertentu dan menilai kualitas pekerjaan mereka [1]. *Author Name Disambiguation* (AND) adalah masalah umum di domain perpustakaan digital di mana pemetaan catatan bibliografi penulis adalah masalah utama. Misalkan, dua penulis berbeda yang memiliki nama yang sama (contoh, Jorgi Ramadhan), yang dapat berbeda antara sebagai satu orang yang sama jika mengidentifikasi author menggunakan namanya, dapat disebut dengan “penggabungan entitas”. Kasus ambigu lainnya ialah satu author yang menggunakan berbagai nama yang berbeda di seluruh artikelnya (misal, Jorgi R, Ramadhan Jorgi), yang membuat karya penulis tersebut terhubung dengan beberapa penulis lain, yang dikenal sebagai “pemisahan entitas” [2].

Dalam pengelolaan perpustakaan digital, nama penulis dalam data bibliografi bersifat ambigu karena banyak penulis memiliki nama yang sama (homonim) atau nama varian yang dicatat untuk penulis yang sama (sinonim) [3]. Ambiguitas nama muncul ketika satu set catatan kutipan berisi nama penulis yang ambigu dan mungkin muncul dalam dua bentuk yang berbeda, dalam bentuk pertama nama penulis yang sama dapat muncul di bawah nama yang berbeda yang disebut sinonim, dan dalam bentuk kedua nama penulis yang berbeda mungkin memiliki nama yang mirip. disebut sebagai homonim [4]. Pada beberapa penelitian sebelumnya terhadap data dalam jumlah banyak dan *imbalance* pada kasus *Author Matching* telah dilakukan menggunakan machine learning dengan metode *Deep Neural network* sering didapatkan hasil yang sangat baik yaitu level akurasi mencapai 98% , namun hasil akurasi tidak valid karena data dengan label yang benar dan minoritas tidak representatif sama sekali, atau jika dijelaskan dalam konfusi matrix , maka data True Positive bernilai 0 [5].

Pada penelitian kali ini penulis mencoba untuk mengklasifikasikan teks pada kasus *Author Name Disambiguation*. Klasifikasi menggunakan Machine Learning dengan metode *Recurrent Neural Network* (RNN), RNN dengan kemampuan pemodelan data dengan panjang variable sekuensial , telah banyak diterapkan untuk memecahkan masalah klasifikasi teks [6]. Namun data DBLP pada kasus *Author Name Disambiguation* adalah suatu data tabular *non-sequential* , hal tersebut menjadi salah satu permasalahan utama pada penelitian ini untuk mengetahui apakah pada kasus seperti AND metode klasifikasi menggunakan *Recurrent Neural network* akan memberikan hasil yang bagus atau tidak, dan permasalahan lain pada penelitian ini adalah untuk mengetahui apakah RNN mampu memberikan solusi yang signifikan pada kasus Homonim dan Sinonim. Berdasarkan latar belakang tersebut penulis akan melakukan penelitian **Klasifikasi Author Matching pada Data Bibliografi menggunakan Metode Recurent Neural Network** .

## **1.2. Tujuan**

Tujuan dari penulisan tugas akhir ini adalah memberikan solusi pada kasus *Author Name Disambiguation* (AND), lebih tepatnya pada Klasifikasi untuk *Author Matching* terutama pada kasus Homonim dan Sinonim menggunakan metode *Recurrent Neural Network* (RNN).

## **1.3. Perumusan dan Batasan Masalah**

### **1.3.1. Perumusan Masalah**

AND disebabkan oleh keterbatasan pada saat memasukan nama penulis pada jurnal yang menyebabkan adanya kemiripan antar nama penulis. Dengan menggunakan metode RNN diharapkan mendapat hasil terbaik untuk permasalahan AND terutama pada kesamaan nama penulis (*Author matching*) pada kasus Homonim dan Sinonim.

### **1.3.2. Batasan Masalah**

Berikut batasan masalah Tugas Akhir ini, adalah :



1. Penelitian hanya terbatas pada kasus AND dan *Author matching*
2. Bahasa program yang digunakan adalah *Python 3*.
3. Material dataset yang digunakan dalam penelitian ini adalah DBLP Labeled Data Dataset, hasil penelitian Jinseok Kim et al. [7] Kumpulan data berasal dari situs web dblp.org dan telah disanitasi dan dibersihkan (*Cleaning proces*).
4. Dalam Penelitian ini hanya digunakan Metode RNN Sebagai metode utama dalam klasifikasi.

Hasil penelitian ini dibatasi pada sensitivitas, spesifisitas, presisi, skor F1, tingkat kesalahan (*Error Rate*) dan akurasi, dan nilai kebenaran persentase yang dimaksudkan untuk dijadikan ukuran kesamaan penulis.

#### **1.4. Metodologi Penelitian**

Pada penelitian ini menggunakan metodologi sebagai berikut :

##### **1.4.1. Metode Studi Pustaka dan Literature**

Dalam metode ini, kami mencari dan mengumpulkan dokumen dari buku dan Internet dengan tema “Pencocokan penulis”.

##### **1.4.2. Metode Konsultasi**

Metode ini berkonsultasi dengan pihak-pihak yang memiliki pengetahuan dan wawasan yang cukup untuk mengatasi permasalahan yang dihadapi dalam penulisan Tugas Akhir “Pencocokan Penulis”.

##### **1.4.3. Metode Pembuatan Model**

Metode ini menggunakan simulasi untuk membuat desain pemodelan.

#### **1.4.4. Metode Pengujian**

Metode ini menguji simulasi yang dijalankan untuk mengetahui apakah simulasi dapat menghasilkan nilai akurasi yang baik.

#### **1.4.5. Metode Analisa dan Kesimpulan**

Hasil pengujian pada tugas akhir ini akan dianalisis kelemahannya untuk digunakan pada penelitian selanjutnya.

#### **1.4.6. Metode Penarikan Kesimpulan dan Saran**

Metode terakhir adalah menarik kesimpulan dengan meringkas secara singkat, padat dan jelas untuk memberikan tidak hanya informasi yang terstruktur dengan baik tetapi juga saran dan ide untuk penelitian masa depan.

### **1.5. Sistematika Penelitian**

Untuk memperjelas Penelitian, sistem deskripsi disusun sebagai berikut.:

#### **BAB I PENDAHULUAN**

Bab ini memberikan gambaran secara rinci dan jelas tentang latar belakang masalah, tujuan, kelebihan, rumusan, keterbatasan, serta metodologi dan sistem penelitian.

#### **BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan dasar-dasar teoritis dan bahan untuk pertanyaan penelitian dan solusi yang digunakan untuk membuat tugas akhir ini..

#### **BAB III METODOLOGI PENELITIAN**

Metodologi yang digunakan dirinci berkenaan dengan teknik, metode dan proses yang dilakukan dalam penelitian.

#### **BAB IV HASIL DAN ANALISIS**

Bagian berikut menjelaskan hasil studi yang dilakukan, termasuk kekuatan dan kelemahan metode yang digunakan.

## **BAB V KESIMPULAN**

Dalam bab terakhir ini, hasil penelitian disajikan secara singkat, padat, dan jelas, serta saran untuk penelitian selanjutnya, terutama yang berkaitan dengan tugas akhir yang diselesaikan..

## DAFTAR PUSTAKA

- [1] K. M. Pooja, S. Mondal, and J. Chandra, "Online author name disambiguation in evolving digital library," *Neurocomputing*, vol. 493, pp. 1–14, 2022, doi: 10.1016/j.neucom.2021.07.104.
- [2] J. Kim, "Scale-free collaboration networks: An author name disambiguation perspective," *J. Assoc. Inf. Sci. Technol.*, vol. 70, no. 7, pp. 685–700, 2019.
- [3] J. Kim, J. Kim, and J. Owen-Smith, "Ethnicity-based name partitioning for author name disambiguation using supervised machine learning," *J. Assoc. Inf. Sci. Technol.*, 2021, doi: 10.1002/asi.24459.
- [4] I. Hussain and S. Asghar, "A survey of author name disambiguation techniques: 2010-2016," *Knowl. Eng. Rev.*, vol. 32, 2017, doi: 10.1017/S0269888917000182.
- [5] Z. Yamani, S. Nurmaini, and D. P. Rini, "Author Matching Classification with Anomaly Detection Approach for Bibliometric Repository Data," *Comput. Eng. Appl. J.*, vol. 9, no. 2, pp. 79–92, 2020.
- [6] Y. Zhao, Y. Shen, and J. Yao, "Recurrent neural network for text classification with hierarchical multiscale dense connections," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2019-Augus, pp. 5450–5456, 2019, doi: 10.24963/ijcai.2019/757.
- [7] 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.
- [8] I. Hussain and S. Asghar, "A survey of author name disambiguation techniques: 2010-2016," *Knowl. Eng. Rev.*, vol. 32, pp. 1–24, 2017, doi: 10.1017/S0269888917000182.

- [9] 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, doi: 10.1007/978-3-319-05476-6\_13.
- [10] M. Shoaib, A. Daud, and T. Amjad, “Author Name Disambiguation in Bibliographic Databases: A Survey,” *arXiv Prepr. arXiv2004.06391*, 2020.
- [11] Y. Ma, Y. Wu, and C. Lu, “A graph-based author name disambiguation method and analysis via information theory,” *Entropy*, vol. 22, no. 4, 2020, doi: 10.3390/E22040416.
- [12] A. A. Ferreira, M. A. Gonçalves, and A. H. F. Laender, “A brief survey of automatic methods for author name disambiguation,” *Acm Sigmod Rec.*, vol. 41, no. 2, pp. 15–26, 2012.
- [13] F. Momeni and P. Mayr, “Evaluating co-authorship networks in author name disambiguation for common names,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9819 LNCS, pp. 386–391, 2016, doi: 10.1007/978-3-319-43997-6\_31.
- [14] F. Momeni and P. Mayr, “Using co-authorship networks for author name disambiguation,” in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 2016, vol. 2016-Septe, pp. 261–262, doi: 10.1145/2910896.2925461.
- [15] L. A. Bugnon, C. Yones, D. H. Milone, and G. Stegmayer, “Deep neural architectures for highly imbalanced data in bioinformatics,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 31, no. 8, pp. 2857–2867, 2020, doi: 10.1109/TNNLS.2019.2914471.
- [16] B. Liu, M. Zhang, W. Ma, X. Li, Y. Liu, and S. Ma, “A two-step information accumulation strategy for learning from highly imbalanced data,” *Int. Conf. Inf. Knowl. Manag. Proc.*, vol. Part F1318, pp. 1289–1298, 2017, doi: 10.1145/3132847.3132940.

- [17] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.
- [18] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.
- [19] A. I. Kadhim, "An Evaluation of Preprocessing Techniques for Text Classification," *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 6, pp. 22–32, 2018.
- [20] O. T. Tran and V. T. Bui, "Neural Text Normalization in Speech-to-Text Systems with Rich Features," *Appl. Artif. Intell.*, vol. 35, no. 3, pp. 193–205, 2021, doi: 10.1080/08839514.2020.1842108.
- [21] C. Zhang, T. Baldwin, B. Kimelfeld, and Y. Li, "Adaptive Parser-Centric Text Normalization," pp. 1159–1168, 2013.
- [22] L. A. Mullen, K. Benoit, O. Keyes, D. Selivanov, and J. Arnold, "Fast , Consistent Tokenization of Natural Language Text," vol. 2017, pp. 2017–2019, 2018, doi: 10.21105/joss.00655.
- [23] D. Oard and G. Marchionini, "A conceptual framework for text filtering," *Coll. Park. MD, Univ. Maryl.*, p. 96-25, 1996.
- [24] V. Jha, N. Manjunath, P. D. Shenoy, and K. R. Venugopal, "HSRA: Hindi stopword removal algorithm," in *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, 2016, pp. 1–5, doi: 10.1109/MicroCom.2016.7522593.
- [25] J. Singh and V. Gupta, "A systematic review of text stemming techniques," *Artif. Intell. Rev.*, vol. 48, no. 2, pp. 157–217, 2017, doi: 10.1007/s10462-016-9498-2.
- [26] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian

- Networks: The Combination of Knowledge and Statistical Data,” *Mach. Learn.*, vol. 20, no. 3, pp. 197–243, 1995, doi: 10.1023/A:1022623210503.
- [27] Z. YAMANI, S. NURMAINI, FIRDAUS, M. Naufal R, and W. K. SARI, “Author Matching Using String Similarities and Deep Neural Networks,” vol. 172, no. Siconian 2019, pp. 474–479, 2020, doi: 10.2991/aisr.k.200424.073.
- [28] S. Y. Yuliani, S. Y. Yuliani, S. Sahib, M. F. Abdollah, Y. S. Wijaya, and N. H. M. Yusoff, “Hoax news validation using similarity algorithms,” *J. Phys. Conf. Ser.*, vol. 1524, no. 1, 2020, doi: 10.1088/1742-6596/1524/1/012035.
- [29] J. M. Keil, “Efficient bounded Jaro-winkler similarity based search,” *Lect. Notes Informatics (LNI), Proc. - Ser. Gesellschaft fur Inform.*, vol. P-289, pp. 205–214, 2019, doi: 10.18420/btw2019-13.
- [30] T. Phaladisailoed and T. Numnonda, “Machine learning models comparison for bitcoin price prediction,” *Proc. 2018 10th Int. Conf. Inf. Technol. Electr. Eng. Smart Technol. Better Soc. ICITEE 2018*, pp. 506–511, 2018, doi: 10.1109/ICITEED.2018.8534911.
- [31] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, “Recurrent neural network wave functions,” *Phys. Rev. Res.*, vol. 2, no. 2, pp. 1–17, 2020, doi: 10.1103/PhysRevResearch.2.023358.
- [32] W. Zaremba, I. Sutskever, and O. Vinyals, “Recurrent Neural Network Regularization,” no. 2013, pp. 1–8, 2014.
- [33] P. Liu, X. Qiu, and H. Xuanjing, “Recurrent neural network for text classification with multi-task learning,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2016-Janua, pp. 2873–2879, 2016.
- [34] C. Du and L. Huang, “Text classification research with attention-based recurrent neural networks,” *Int. J. Comput. Commun. Control*, vol. 13, no. 1, pp. 50–61, 2018, doi: 10.15837/ijccc.2018.1.3142.

- [35] 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.
- [36] 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.