

Riset Jurnal Literatur : Morfologi Stemming Enhance Confix Stripping, Sastrawi Dan Tala Pada Bahasa Daerah Melayu-Ambon

Vinnesa Patricia Carolina¹, Ema Utami², Ainul Yaqin³

^{1,2} Magister Informatika, Universitas AMIKOM Yogyakarta, Sleman, Indonesia

³ Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta, Sleman, Indonesia

Article Info

Article history:

Diterima 19 April 2024

Revisi 21 April 2024

Diterbitkan 22 April 2024

Keywords:

Stemming Ambon

Literatur Review

Enhanced Confix Stripping

Sastrawi

Tala

ABSTRAK

Stemming dalam bahasa Ambon merupakan tantangan signifikan karena lexiconnya yang luas, mencakup sekitar 127.000 kata dasar seperti yang tercatat dalam Kamus Besar Bahasa Indonesia. Kompleksitas stemming timbul dari tugas untuk mengekstrak kata-kata dasar dari kata-kata yang memiliki imbuhan, memerlukan penghapusan berbagai imbuhan seperti awalan, sisipan, akhiran, dan kombinasinya. Pentingnya proses ini dalam memengaruhi kualitas hasil analisis sangat besar. Untuk mengatasi kompleksitas linguistik ini, beberapa algoritma stemming telah dikembangkan. Diantaranya, Enhanced Confix Stripping, Sastrawi, dan Tala, masing-masing menawarkan teknik unik untuk menangani kompleksitas stemming dalam bahasa Indonesia. Dalam penelitian yang dilakukan, Enhanced Confix Stripping menjadi metode yang paling banyak digunakan dengan 17 penelitian, menunjukkan popularitas dan efektivitasnya dalam konteks bahasa Ambon. Selanjutnya, Sastrawi juga cukup digunakan dengan 13 penelitian. Sementara itu, Tala hanya digunakan dalam 2 penelitian. Pemilihan algoritma yang tepat menjadi kunci dalam memastikan akurasi dan kehandalan proses stemming dalam kerangka analisis. Perkembangan dalam bidang ini tercermin dari tren publikasi penelitian yang mengalami perjalanan dinamis selama beberapa tahun terakhir. Meskipun terdapat penurunan signifikan pada tahun 2024, dengan hanya 2 penelitian yang berhasil dipublikasikan, penting untuk mempertimbangkan faktor-faktor seperti proyek penelitian yang sedang berlangsung, ketersediaan dana, dan situasi eksternal yang mungkin memengaruhi output penelitian. Kesimpulan ini menyoroti pentingnya eksplorasi terus-menerus dalam mengembangkan teknologi pemrosesan teks yang disesuaikan dengan karakteristik linguistik bahasa Ambon, serta peran penting pemilihan algoritma stemming yang tepat dalam memastikan kualitas analisis yang akurat.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Vinnesa Patricia Carolina,

Universitas AMIKOM Yogyakarta, Jl, Ring Road Utara, Sleman, 051024, Indonesia

Email: vinnesa.p.c@students.amikom.ac.id

1. PENDAHULUAN

Terletak di bagian timur Indonesia, Ambon memiliki banyak kebudayaan yang membanggakan seperti halnya kain tradisionalnya, etnis, dan juga bahasa daerah yang digunakan. Bahasa berperan sebagai pondasi budaya, dan beragam kelompok etnis di sana menggunakan bahasa masing-masing. Penting untuk dipahami bahwa bahasa Ambon digunakan oleh masyarakat Ambon. Namun, di dunia modern yang semakin terhubung, kemampuan berkomunikasi dalam bahasa minoritas seperti bahasa Ambon seringkali terbatas.

Bahasa Ambon memiliki nama lain bahasa Melayu Ambon, karena bahasa Ambon merupakan salah satu dialek yang dituturkan oleh warga wilayah Maluku [1]. Dalam prakteknya bahasa ini mengalami banyak serapan dari berbagai bahasa seperti bahasa Melayu Makassar dan bahasa Portugis akibat dari penjajahan. Selain itu, ketika

bahasa Indonesia mulai menjadi bahasa yang baku dan banyak digunakan oleh warga. Bahasa Ambon juga menyerap kata-kata dari bahasa tersebut, kosakata yang digunakan tidak jauh-jauh dari lingkungan sekitar tempat tinggal masyarakat, atau mencerminkan kebiasaan dan pola hidup masyarakat Maluku [2].

Stemming adalah proses pemisahan kata dasar dari kata berimbuhan dalam sebuah kalimat dengan memisahkan kata dasar dan imbuhan yang dapat terdiri dari awalan (*prefixes*), sisipan (*infixes*), dan akhiran (*suffixes*) [3]. Algoritma stemming akan berbeda untuk setiap bahasa yang digunakan [4]. Dalam stemming pada bahasa Indonesia ada beberapa metode stemming yang umumnya digunakan, diantaranya adalah *Enhanced Confix Stripping*, Sastrawi dan Tala.

Para peneliti telah mengeksplorasi berbagai algoritma stemming untuk mengatasi kompleksitas bahasa Ambon. Stemming, yang didefinisikan sebagai proses pengurangan infleksi atau derivasi menjadi bentuk dasarnya, mirip dengan mentransformasikan "menyapu" menjadi "sapu," umumnya digunakan untuk pra-pemrosesan dalam aplikasi berbasis teks [5]. Algoritma stemming biasanya dikategorikan menjadi dua jenis: berbasis statistik dan berbasis aturan. Stemmer berbasis statistik menggunakan algoritma tanpa supervisi yang memanfaatkan data pelatihan untuk membangun model untuk stemming, sedangkan stemmer berbasis aturan menggunakan aturan yang telah ditentukan sebelumnya untuk proses stemming. Penerapan stemming dapat digunakan untuk pengembangan mesin pencari dan untuk mengurangi plagiarisme [4]. Namun, tantangan seperti over stemming dan under stemming umum terjadi selama proses stemming. Penting untuk dicatat bahwa setiap bahasa memiliki karakteristik dan struktur unik, terutama dalam hal struktur afiks, sehingga memerlukan penyesuaian dalam metode stemming untuk sesuai dengan karakteristik bahasa dan metode stemming yang ada dalam setiap bahasa berbeda satu sama lain [6].

Setiap bahasa memiliki algoritma stemming yang berbeda, yang membedakannya dari yang digunakan dalam bahasa-bahasa lain. Beberapa algoritma stemming telah dikembangkan secara khusus untuk bahasa Indonesia, termasuk Tala, Sastrawi, dan *Enhanced Confix Stripping*. Metode ini menggunakan aturan dan heuristik untuk melakukan stemming dengan menghilangkan awalan dan akhiran dari kata-kata. Algoritma Tala, yang dibuat oleh Atmaja dan Purwarianti, adalah metode stemming lain yang disesuaikan untuk bahasa Indonesia. Algoritma Tala menggunakan algoritma Porter dan mengikuti prinsip operasi berbasis aturan [8]. Algoritma ini menghasilkan akar kata menggunakan teknik berbasis kamus dan aturan linguistik. Algoritma Sastrawi juga umum digunakan untuk stemming dalam Bahasa Indonesia. Sastrawi adalah algoritma stemmer yang digunakan untuk mengatasi tantangan mengubah kata-kata menjadi kata-kata sederhana [9]. Algoritma ini menggunakan kombinasi pencarian kamus dan stemming berbasis aturan untuk menentukan bentuk dasar kata-kata. Algoritma *Enhanced Confix Stripping* adalah versi perbaikan dari *Confix Stripping* asli dengan penambahan modifikasi pada pemotongan kata-kata [10]. Algoritma ini dibuat khusus untuk Bahasa Indonesia dan menggunakan kumpulan aturan dan kamus untuk melakukan stemming, dengan fokus pada penghapusan afiks dan pencarian kamus.

Temuan penelitian tentang bahasa Jawa [11], Madura [12], Bali [13], dan Minangkabau [6] menyoroti baik keefektifan maupun keterbatasan algoritma stemming yang ada dalam konteks linguistik masing-masing. Studi-studi ini memberikan wawasan tentang kesesuaian algoritma-algoritma tersebut dan area di mana mereka mungkin memerlukan penyempurnaan. Namun, ketika diterapkan pada bahasa Ambon, muncul tantangan yang menunjukkan perlunya penelitian lebih lanjut dan adaptasi untuk meningkatkan efektivitas mereka dalam domain linguistik spesifik ini.

Berlanjut dari diskusi sebelumnya, tinjauan literatur ini bertujuan untuk memberikan pemeriksaan menyeluruh dan penilaian perbandingan berbagai algoritma stemming untuk pemrosesan teks dalam bahasa Ambon. Dengan menggabungkan penelitian yang sudah ada dan menyorot area-area kesenjangan pengetahuan, tinjauan ini bermaksud untuk mendorong kemajuan dalam teknologi pemrosesan teks yang disesuaikan dengan ciri-ciri linguistik yang khas dari bahasa Ambon. Melalui analisis dan sintesis yang cermat dari literatur yang relevan, tinjauan ini bertujuan untuk menguraikan kelebihan, kekurangan, dan aplikasi prospektif dari algoritma-algoritma ini dalam ranah pemrosesan teks bahasa Ambon. Pada akhirnya, tinjauan ini bertujuan untuk memandu upaya penelitian masa depan di bidang ini.

2. RUMUSAN MASALAH

2.1. Rencana Penelitian

Studi Literature Review Sistematis (SLR) berupaya mengidentifikasi studi-studi relevan yang penting, mengumpulkan data yang diperlukan, dan kemudian menilai serta menggabungkan temuan-temuan untuk mendapatkan pemahaman yang lebih dalam tentang subjek penelitian [14]. Terlepas dari subjek spesifik,

bidang akademik, atau sudut pandang teoritis tertentu, Literature Review Sistematis (SLR) [15], sebagaimana diuraikan di bawah ini.

Mengembangkan serangkaian pertanyaan penelitian (RQs) sangat penting saat menggunakan pendekatan literature review sistematis (SLR) [16]. Pertanyaan-pertanyaan yang disajikan dalam Tabel 1 memiliki peran krusial dalam membentuk kerangka kerja yang jelas, tujuan, dan efektif untuk proyek penelitian. Pendekatan komprehensif ini membantu dalam menyempurnakan dan memfokuskan proses penelitian, yang pada akhirnya meningkatkan efektivitasnya.



Gambar 1. Skema Penelitian

Pada Gambar 1, ditunjukkan proses yang dilakukan dalam penelitian ini. Proses dimulai dari adanya latar belakang mengapa penelitian akan dilakukan, hingga diakhiri dengan penulisan laporan dari hasil penelitian yang telah dilakukan. Pada Tabel 1, ditunjukkan rumusan masalah yang berisi pertanyaan yang menjadi fondasi mengapa penelitian ini dilakukan.

Tabel 1. Rumusan Masalah (*Research Question*)

TABLE I. RUMUSAN MASALAH	
ID	Rumusan Masalah
RQ1	Apa pendekatan pengumpulan data yang biasa digunakan oleh para peneliti dalam studi yang terkait dengan stemming?
RQ2	Metodologi apa yang diterapkan dalam bidang penelitian stemming?
RQ3	Apa kesimpulan yang dihasilkan dari analisis stemming dalam konteks penelitian?

2.2. Strategi Studi

Peneliti dilakukan pencarian secara menyeluruh terhadap karya ilmiah di berbagai basis data terkemuka terkait penelitian atau jurnal ilmiah seperti ScienceDirect, IEEE Xplore, SpringerLink, Semantic Scholar, Google Scholar, dan Elsevier. Pencarian ini dilakukan menggunakan beberapa kata kunci, kata kunci yang digunakan mencakup istilah dalam bahasa Indonesia ataupun bahasa Inggris, untuk memastikan pengambilan literatur yang relevan secara komprehensif dan inklusif. Berikut ini adalah kata kunci yang digunakan untuk melakukan pencarian artikel, penelitian atau jurnal ilmiah dalam penelitian ini:

- “Indonesian Stemming Algorithm”
- “Algoritma Stemming Bahasa Indonesia”
- “Stemming Algorithm”

- “Algoritma Stemming”
- “Enhance Confix Stripping Algorithm”
- “Sastrawi Algorithm”
- “Tala Algorithm”

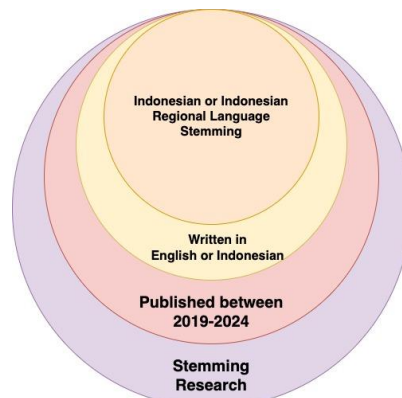
2.3. Seleksi Studi

Dalam melakukan penilaian terhadap naskah penelitian, penentuan kriteria merupakan langkah penting. Dengan menggunakan kriteria yang jelas, peneliti dapat menyaring penelitian, artikel ilmiah, atau jurnal lain yang relevan dengan penelitian tersebut. Biasanya, peneliti menggunakan dua jenis kriteria, yaitu kriteria inklusi dan eksklusi, untuk menyusun artikel. Kriteria inklusi adalah kriteria yang harus dipenuhi agar suatu penelitian dianggap relevan dan dimasukkan ke dalam penelitian ini. Berikut adalah kriteria inklusi dari penelitian ini:

- Artikel penelitian merupakan penelitian yang dilakukan dari tahun 2019 hingga 2024.
- Artikel penelitian yang dipilih ditulis dalam Bahasa Inggris atau Bahasa Indonesia.
- Topik utama dari penelitian harus berkaitan dengan metode stemming dalam bahasa Indonesia atau bahasa Daerah di Indonesia.

Kriteria eksklusi adalah kriteria yang menentukan apakah suatu penelitian, artikel ilmiah, atau jurnal dianggap tidak memenuhi syarat untuk dimasukkan ke dalam penelitian ini. Dalam konteks penelitian ini, kriteria eksklusi terdiri dari hal-hal berikut:

- Tidak adanya penjelasan yang jelas mengenai metode yang digunakan dalam penelitian.
- Penelitian tidak mencapai tujuan-tujuan yang telah ditetapkan sebelumnya.

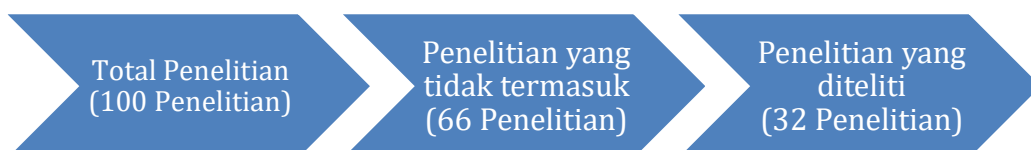


Gambar 2. Kriteria Inklusi Penelitian

2.4. Asesmen Studi

Untuk memperoleh pemahaman yang komprehensif mengenai kualitas keseluruhan dari penelitian, evaluasi kualitas yang cermat sangat penting. Tahap evaluasi ini memiliki peran krusial dalam menentukan relevansi dan kesesuaian data yang diidentifikasi untuk dimasukkan dalam penelitian. Dalam konteks penelitian ini, data yang terkumpul akan menjalani proses evaluasi yang ketat yang dipandu oleh serangkaian kriteria yang telah ditetapkan untuk mengukur kualitasnya. Penggunaan standar-standar ini untuk evaluasi kualitas memastikan tinjauan yang sistematis dan tidak bias, sehingga meningkatkan kekokohan dan keandalan penelitian. Setiap artikel yang dipilih untuk dimasukkan dalam penelitian ini harus memenuhi kriteria-kriteria berikut:

- Apakah penelitian diterbitkan antara tahun 2019 dan 2024?
- Apakah penelitian ditulis dalam bahasa Indonesia atau bahasa Inggris?
- Apakah topik utama penelitian berkaitan dengan stemming Bahasa Indonesia atau Bahasa Daerah Indonesia?



Gambar 2. Seleksi Penelitian

2.5. Pengumpulan Data

Pada tahap ini, data yang diekstrak dari naskah yang ditinjau mencakup elemen-elemen penting seperti tahun publikasi, dataset yang digunakan dalam studi yang sedang diteliti, metodologi yang digunakan untuk pengumpulan data, pendekatan stemming spesifik yang diadopsi dalam penelitian yang dianalisis, dan implikasi stemming terhadap studi tersebut. Selanjutnya, semua data yang relevan dicatat secara cermat dalam dokumen spreadsheet. Dokumentasi yang cermat ini menjadi dasar untuk analisis komprehensif atas data yang terkumpul. Organisasi sistematis informasi ini memfasilitasi penyelidikan yang efektif dan terstruktur yang sesuai dengan standar penelitian ilmiah yang telah ditetapkan.

2.6. Sintesis Data

Pada tahap ini dalam proses penelitian, kami telah mengumpulkan sebuah kumpulan data yang substansial berisi 100 studi, dengan teliti menyaring judul-judul dan abstrak untuk relevansi. Penyaringan awal ini mengurangi pilihan kami sebanyak 66 artikel. Namun, pengamatan kami tidak berhenti di situ. Dengan menggunakan kriteria inklusi dan eksklusi yang ketat, kami secara cermat mengeksklusikan 66 artikel yang sejalan dengan tujuan penelitian kami. Jika suatu artikel memenuhi kriteria inklusi kami, maka artikel tersebut dimasukkan ke dalam tinjauan literatur kami; sebaliknya, artikel-artikel yang memenuhi kriteria eksklusi diabaikan. Proses pemilihan yang cermat ini menghasilkan satu set akhir berisi 32 artikel, yang kemudian menjalani tinjauan dan analisis yang detail. Data yang diperoleh dan temuan utama dari artikel-artikel ini menjalani pemeriksaan menyeluruh, dan sintesisnya disajikan secara metodis dalam Tabel 2. Pendekatan ini sesuai dengan standar penelitian ilmiah, yang menjamin eksplorasi literatur yang komprehensif dan terstruktur dengan baik.

Tabel 2. Jurnal yang di Review (*Reviewd Paper*)

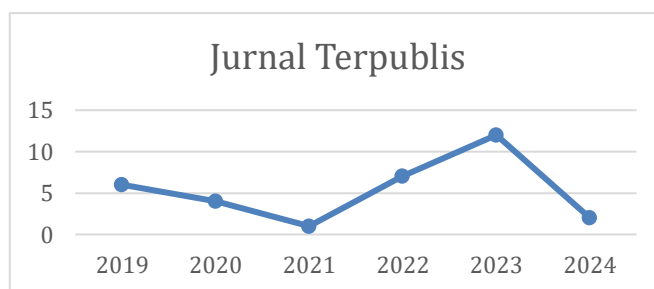
Penelitian	TABLE II. JURNAL YANG DI REVIEW		
	Tahun	Sumber Data	Metode
[17]	2024	Data from web	Sastrawi
[18]	2024	Data from web	Sastrawi
[19]	2023	Printed document	Sastrawi
[8]	2023	Data from web	Tala
[13]	2023	Corpus or dictionary	Tala
[11]	2023	Corpus or dictionary	ECS
[12]	2023	Corpus or dictionary	ECS
[20]	2023	Corpus or dictionary	Sastrawi
[21]	2023	Data from web	Sastrawi
[22]	2023	Printed document	ECS
[23]	2023	Data from web	Sastrawi
[24]	2023	Data from web	Sastrawi
[25]	2023	Printed document	Sastrawi
[26]	2023	Printed document	ECS
[3]	2022	Corpus or dictionary	ECS
[27]	2022	Printed document	ECS
[28]	2022	Corpus or dictionary	ECS
[6]	2022	Corpus or dictionary	ECS
[29]	2022	Printed document	ECS
[30]	2022	Data from web	Sastrawi
[31]	2022	Corpus or dictionary	ECS
[32]	2021	Data from web	Sastrawi
[33]	2020	Corpus or dictionary	Sastrawi
[34]	2020	Printed document	Sastrawi
[35]	2020	Data from web	Sastrawi
[36]	2020	Corpus or dictionary	ECS

Penelitian	Tahun	Sumber Data	Metode
[37]	2019	Printed document	ECS
[38]	2019	Printed document	ECS
[39]	2019	Printed document	ECS
[40]	2019	Printed document	ECS
[41]	2019	Printed document	ECS
[42]	2019	Printed document	ECS

3. HASIL DAN PEMBAHASAN

3.1. Tahun Penerbitan

Sebagai bagian dari analisis, naskah yang terkumpul akan dikelompokkan dan diklasifikasikan berdasarkan tahun publikasi mereka. Pendekatan kronologis ini akan memungkinkan penilaian yang lebih nuansa terhadap perkembangan tren penelitian dan kemajuan dalam pemrosesan teks dalam konteks bahasa Ambon atau Bahasa Daerah Indonesia lainnya. Setiap korpus literatur dari setiap tahun akan diperiksa secara cermat, dengan penekanan pada mengidentifikasi temuan utama, metodologi, dan implikasi untuk algoritma stemming dalam bahasa Ambon. Dengan mendekonstruksi literatur secara tahunan, studi ini berharap dapat memberikan pencerahan tentang tren perkembangan penelitian dalam pemrosesan teks untuk bahasa Ambon, memberikan informasi yang signifikan.



Gambar 3. Perkembangan dari Tahun ke Tahun

Pada tahun 2019, terjadi publikasi sebanyak 6 penelitian, ini menandai awal mulainya eksplorasi algoritma pemrosesan teks yang disesuaikan dengan nuansa linguistik bahasa Ambon. Awal yang sederhana ini membentuk dasar bagi kemajuan-kemajuan berikutnya dalam bidang ini.

Tahun 2020 terjadi penurunan dalam publikasi penelitian, pada tahun tersebut hanya 4 penelitian yang dipublikasikan. Hal ini mencerminkan menurunnya minat dalam mempelajari teknik-teknik pemrosesan teks dan aplikasinya dalam bahasa daerah khususnya bahasa Ambon.

Pada tahun 2021, menjadi titik terendah dari jurnal dalam dataset yang peneliti kumpulkan, karena pada tahun tersebut hanya ada 1 penelitian. Dari tahun ini ke tahun berikutnya mulai mengalami perkembangan dan terus naik secara bertahap.

Tahun 2022 menandai lonjakan signifikan, dengan 7 penelitian yang dipublikasikan. Lonjakan output penelitian ini menandakan kedewasaan bidang ini, seiring para peneliti menyelami aspek-aspek yang lebih rumit dalam pemrosesan teks dan berusaha mengatasi tantangan dan keterbatasan yang masih ada.

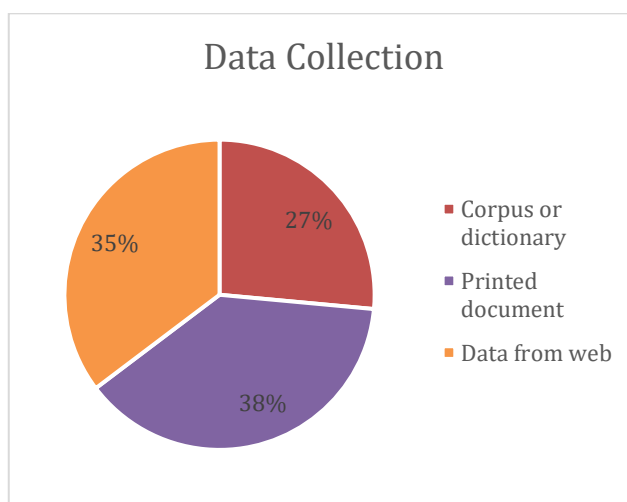
Pada tahun 2023, tren kenaikan berlanjut, dengan 21 penelitian yang dipublikasikan. Output yang konsisten ini menunjukkan minat dan investasi yang berkelanjutan dalam mengembangkan teknologi pemrosesan teks yang disesuaikan dengan karakteristik linguistik bahasa Ambon.

Namun, pada tahun 2024, jumlah penelitian yang dipublikasikan mengalami penurunan signifikan, hanya ada 2 penelitian yang berhasil dipublikasikan. Meskipun penurunan ini tampak mengkhawatirkan pada pandangan pertama, penting untuk mempertimbangkan faktor-faktor potensial seperti proyek penelitian yang sedang

berlangsung, ketersediaan dana, dan situasi eksternal yang mungkin telah memengaruhi output penelitian selama periode ini.

3.2. Sumber Data Penerbitan

Ketika memulai sebuah penelitian, para peneliti mengandalkan data untuk memandu penyelidikan mereka. Pengumpulan data melibatkan proses sistematis untuk memperoleh dan menganalisis data yang tepat dari berbagai sumber, dengan tujuan menjawab pertanyaan penelitian, mengungkap pola, mengeksplorasi kemungkinan, dan menilai hasil yang potensial. Selama proses ini, para peneliti harus menjelaskan sifat data tersebut, memastikan asal-usulnya, dan menjelaskan metodologi yang digunakan.



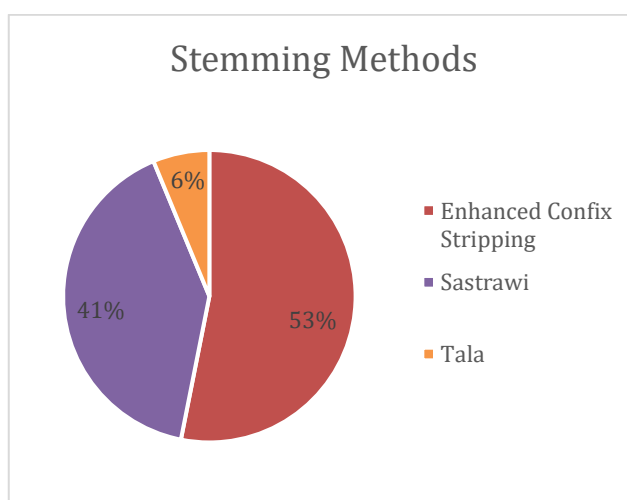
Gambar 4. Sumber Data Penelitian

Metode pengumpulan data yang digunakan sangat beragam, dengan strategi yang berbeda-beda tergantung pada konteksnya. Secara khusus, sektor ilmiah, komersial, dan pemerintahan sangat mengandalkan prosedur pengumpulan data yang efektif untuk memberi informasi pada usaha mereka masing-masing. Pendekatan komprehensif ini sejalan dengan standar ketat yang dipegang teguh dalam penelitian ilmiah, yang memastikan validitas dan reliabilitas temuan.

3.3. Metode yang Digunakan

Stemming merupakan tantangan yang besar dalam konteks bahasa Indonesia, karena kosakata yang luas, yang mencakup sekitar 127.000 kata dasar seperti yang terdokumentasi dalam Kamus Besar Bahasa Indonesia yang komprehensif. Kompleksitas stemming timbul dari tugas mengekstraksi kata-kata dasar dari kata-kata yang memiliki imbuhan, yang memerlukan penghapusan berbagai imbuhan seperti awalan, sisipan, akhiran, dan kombinasinya. Proses ini memiliki pentingnya yang besar karena sangat memengaruhi kualitas hasil analisis.

Untuk mengatasi kompleksitas linguistik ini, beberapa algoritma stemming telah dikembangkan. Ini termasuk algoritma Nazief-Adriani, Confix Stripping, Enhanced Confix Stripping, dan Sastrawi, masing-masing menawarkan teknik yang berbeda untuk menangani kerumitan stemming dalam bahasa Indonesia. Pemilihan algoritma yang tepat sangat penting untuk memastikan akurasi dan keandalan proses stemming dalam kerangka analisis dapat dilihat dalam Gambar 5.



Gambar 5. Metode Stemming Penelitian

Enhanced Confix Stripping adalah metode stemming yang menonjol dan digunakan dalam penelitian penelitian sebanyak 17 penelitian. Versi peningkatan dari algoritma Confix Stripping ini berfokus pada penghapusan imbuhan dan pencarian kamus untuk melakukan stemming, menawarkan solusi yang efektif untuk menangani kompleksitas stemming dalam bahasa Indonesia.

Sastrawi juga muncul dalam makalah-makalah tersebut, meskipun dengan frekuensi yang lebih tinggi, dengan 13 penelitian tercatat menggunakan metode tersebut. Algoritma stemming ini menggabungkan pencarian kamus dan stemming berbasis aturan untuk mendapatkan bentuk dasar kata-kata, memberikan peneliti dengan alat tambahan untuk analisis dan pemrosesan linguistik.

Terakhir, algoritma Tala, dengan 2 kasus tercatat, muncul dengan jumlah penggunaan yang paling sedikit dari penelitian-penelitian yang dimasukkan. Dikembangkan oleh Atmaja dan Purwarianti, Tala menggunakan pendekatan berbasis kamus yang dikombinasikan dengan aturan linguistik untuk menghasilkan akar kata, menawarkan perspektif unik dalam mengatasi tantangan stemming dalam bahasa Indonesia.

3.4. Kegunaan Stemming

Dari data yang disediakan, terlihat bahwa stemming secara dominan digunakan dalam konteks analisis sentimen, pemrosesan teks, dan klasifikasi dokumen. Area-area penelitian ini sering memerlukan pra-pemrosesan data teks untuk mengekstrak fitur-fitur yang bermakna dan meningkatkan efisiensi serta efektivitas tugas analisis atau klasifikasi selanjutnya.

Selain itu, stemming juga digunakan dalam tugas-tugas terjemahan, di mana normalisasi kata-kata melalui stemming membantu meningkatkan akurasi dan kelancaran teks terjemahan. Lebih lanjut, stemming juga diterapkan dalam sistem pengambilan informasi, di mana hal ini memfasilitasi pencocokan pertanyaan pengguna dengan dokumen-dokumen yang relevan dengan mengurangi variasi kata-kata menjadi bentuk akar yang umum.

Selain itu, data menunjukkan adanya fokus yang signifikan pada pengembangan algoritma stemming. Hal ini menunjukkan upaya yang terus berlanjut untuk meningkatkan kinerja dan adaptabilitas teknik stemming untuk berbagai bahasa dan konteks linguistik, termasuk bahasa Indonesia. Secara keseluruhan, penggunaan stemming di berbagai bidang penelitian menggarisbawahi pentingnya stemming sebagai langkah pra-pemrosesan yang fundamental dalam tugas analisis teks dan pengolahan informasi.

4. KESIMPULAN

Berdasarkan analisis data yang diberikan, terlihat bahwa eksplorasi algoritma pemrosesan teks yang disesuaikan dengan nuansa linguistik bahasa Ambon mengalami perjalanan yang dinamis selama beberapa tahun terakhir. Pada tahun 2019, dimulai dengan 6 penelitian, menandai awal dari perjalanan ini. Namun, minat dalam bidang ini mengalami penurunan pada tahun 2020, hanya dengan 4 penelitian yang dipublikasikan. Tren menurun ini mencerminkan kurangnya minat dalam mempelajari teknik-teknik pemrosesan teks dalam bahasa daerah, khususnya bahasa Ambon. Pada tahun 2021, jumlah penelitian mencapai titik terendah dengan hanya

1 penelitian yang dipublikasikan. Namun, sejak tahun tersebut, bidang ini mulai menunjukkan tanda-tanda pemulihan dan pertumbuhan yang bertahap. Tahun 2022 mencatat lonjakan signifikan dengan 7 penelitian yang berhasil dipublikasikan, menandai kedewasaan bidang ini. Tren kenaikan berlanjut pada tahun 2023, dengan 21 penelitian yang dipublikasikan, menunjukkan minat yang berkelanjutan dalam mengembangkan teknologi pemrosesan teks yang sesuai dengan bahasa Ambon. Meskipun demikian, pada tahun 2024, terjadi penurunan signifikan dengan hanya 2 penelitian yang berhasil dipublikasikan. Penurunan ini mungkin dipengaruhi oleh berbagai faktor seperti proyek penelitian yang sedang berlangsung, ketersediaan dana, dan situasi eksternal. Dari segi metode pemrosesan, metode Enhanced Confix Stripping menjadi yang paling dominan dengan 17 penelitian, diikuti oleh Sastrawi dengan 13 penelitian, dan Tala dengan hanya 2 penelitian. Data koleksi yang paling banyak berasal dari dokumen cetak dengan 13 data, diikuti oleh data dari web dengan 12 data, dan korpus atau kamus dengan 9 data.

UCAPAN TERIMA KASIH

Dari kami yang menulis, kami ingin mengucapkan terima kasih kepada Perguruan Tinggi kami Universitas AMIKOM Yogyakarta atas wawasan selama kami melaksanakan penelitian ini, serta dukungan, saran, dan motivasi yang diberikan selama proses studi ini.

REFERENSI

- [1] Theresia Meturan, Laraswati Laraswati, and Lusi Nur Triani, "Bahasa Ambon dan Bahasa Indonesia: Analisis Fonologi," *Sintaksis*, vol. 1, no. 5, pp. 54–64, Sep. 2023, doi: 10.61132/sintaksis.v1i5.261.
- [2] L. F. Pesiwarissa, "CIGULU-CIGULU (TEKA-TEKI) MASYARAKAT TUTUR BAHASA MELAYU AMBON (KAJIAN ETNOSEMANTIK: SUATU PENDEKATAN AWAL)," *kolita*, vol. 21, no. 21, pp. 208–214, Oct. 2023, doi: 10.25170/kolita.21.4851.
- [3] S. H. Wibowo, R. Toyib, M. Muntahanah, and Y. Darnita, "Time complexity in rejang language stemming," *J.INFOTEL*, vol. 14, no. 3, pp. 174–179, Aug. 2022, doi: 10.20895/infotel.v14i3.764.
- [4] S. Tuhpatussania, E. Utami, and A. D. Hartanto, "COMPARISON OF PORTERS STEMMING ALGORITHM AND NAZIEF & ADRIANI'S STEMMING ALGORITHM IN DETERMINING INDONESIAN LANGUAGE LEARNING MODULES," *pilar*, vol. 18, no. 2, pp. 203–210, Sep. 2022, doi: 10.33480/pilar.v18i2.3940.
- [5] A. Sinaga and S. P. Nainggolan, "ANALISIS PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA STEMMING ARIFIN-SETIONO DAN NAZIEF-ADRIANI PADA DOKUMEN TEKS BAHASA INDONESIA," *Sebatik*, vol. 27, no. 1, pp. 63–69, Jun. 2023, doi: 10.46984/sebatik.v27i1.2072.
- [6] R. Sovia, S. Defit, and Yuhandri, "Development of the Minangkabau Local Language Translation Machine Based on Stemming," in *2022 International Symposium on Information Technology and Digital Innovation (ISITDI)*, Padang, Indonesia: IEEE, Jul. 2022, pp. 195–198. doi: 10.1109/ISITDI55734.2022.9944457.
- [7] J. Jumadi, D. S. Maylawati, L. D. Pratiwi, and M. A. Ramdhani, "Comparison of Nazief-Adriani and Paice-Husk algorithm for Indonesian text stemming process," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1098, no. 3, p. 032044, Mar. 2021, doi: 10.1088/1757-899X/1098/3/032044.
- [8] N. Pamungkas *et al.*, "Comparison of Stemming Test Results of Tala Algorithms with Nazief Adriani in Abstract Documents and National News," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 8, no. 1, pp. 33–41, Jan. 2023, doi: 10.25139/inform.v8i1.5569.
- [9] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 874, no. 1, p. 012017, Jun. 2020, doi: 10.1088/1757-899X/874/1/012017.
- [10] A. Juhari, I. O. Suzanti, Y. D. Pramudita, Husni, and N. P. W. Diantisari, "Enhanced Confix Stripping Stemmer And Cosine Similarity For Search Engine in The Holy Qur'an Translation," in *2020 6th Information Technology International Seminar (ITIS)*, Surabaya, Indonesia: IEEE, Oct. 2020, pp. 207–212. doi: 10.1109/ITIS50118.2020.9321041.

- [11] S. I. Melia, J. Sholihah, D. Nisak, I. S. Juniaristha, and A. T. Ni'mah, "The Ngoko Javanese Stemmer uses the Enhanced Confix Stripping Stemmer Method," *Rekayasa*, vol. 16, no. 1, pp. 107–112, Apr. 2023, doi: 10.21107/rekayasa.v16i1.19308.
- [12] E. Lindrawati, E. Utami, and A. Yaqin, "Comparison of Modified Nazief&Adriani and Modified Enhanced Confix Stripping algorithms for Madurese Language Stemming," *intensif*, vol. 7, no. 2, pp. 276–289, Aug. 2023, doi: 10.29407/intensif.v7i2.20103.
- [13] G. N. M. Nata, "Pengembangan Algoritma Stemmer Bilingual Bali-Indonesia Dengan Rule-Base," 2023.
- [14] M. Wahyu Ade Saputra, E. Utami, and A. Yaqin, "Unlocking Insights: A Literature Review on Enhanced Confix Stripping and Nazief & Adriani Algorithm Modifications for Makassar Language Text Stemming," *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 603–610, Mar. 2024, doi: 10.38124/ijisrt/IJISRT24MAR437.
- [15] Y. Karuniawati, E. Utami, and A. Yaqin, "A Systematic Literature Review of Stemming in Non-Formal Indonesian Language," vol. 8, no. 1, 2023.
- [16] Prema Adhitya Dharma Kusumah, Kusrini Kusrini, and Kusnawi Kusnawi, "Optimizing Data Security: A Literature Review on the Implementation of Beaufort Cipher for Vigenère Affine Cipher," Feb. 2024, doi: 10.5281/ZENODO.10685974.
- [17] J. Z. H. Yong, J. Y. Koh, J. X. Liew, and C. W. Tan, "Linguistic Harmony in Diversity: Lemmatizing Rojak Malay for Global Communication," in *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, Kuala Lumpur, Malaysia: IEEE, Jan. 2024, pp. 6–10. doi: 10.1109/ICDXA61007.2024.10470819.
- [18] L. Cahyaningrum, A. Luthfiarta, and M. Rahayu, "Sentiment Analysis on the Impact of MBKM on Student Organizations Using Supervised Learning with Smote to Handle Data Imbalance," 2024.
- [19] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, Y. Fauziah, and N. H. Cahyana, "Automated Text Annotation Using a Semi-Supervised Approach with Meta Vectorizer and Machine Learning Algorithms for Hate Speech Detection," *Applied Sciences*, vol. 14, no. 3, p. 1078, Jan. 2024, doi: 10.3390/app14031078.
- [20] Y. Purwati, F. S. Utomo, N. Trinarsih, and H. Hidayatulloh, "Feature Selection Technique to Improve the Instances Classification Framework Performance for Quran Ontology," *JOIV: Int. J. Inform. Visualization*, vol. 7, no. 2, p. 615, Jul. 2023, doi: 10.30630/joiv.7.2.1195.
- [21] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *J. Inf. Syst. Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.
- [22] N. P. Yunita, "Aplikasi Pencarian Hadis Menggunakan Vector Space Model Dengan Pembobotan TF-IDF Dan Confix-Stripping Stemmer," *JTIK*, vol. 10, no. 3, pp. 665–676, Jul. 2023, doi: 10.25126/jtiik.20231036736.
- [23] A. Xu, T. Tiffany, M. E. Phanie, and A. Simarmata, "Sentiment Analysis On Twitter Posts About The Russia and Ukraine War With Long Short-Term Memory," *Sinkron*, vol. 8, no. 2, pp. 789–797, Apr. 2023, doi: 10.33395/sinkron.v8i2.12235.
- [24] D. S. Maylawati, Y. J. Kumar, and F. B. Kasmin, "Combination of Graph-based Approach and Sequential Pattern Mining for Extractive Text Summarization with Indonesian Language," vol. 9, no. 2, 2023.
- [25] Setyobudi, R. (2023). Utilization of tds sensors for water quality monitoring and water filtering of carp pools using IoT. *EUREKA: Physics and Engineering*, (6), 69-77.
- [26] A. M. Billah, D. A. R. Wulandari, and Y. A. Auliya, "Rancang Bangun Chatbot Pengaduan Kekerasan Perempuan Anak Dengan Metode Fuzzy String Matching Dan Enhanced Confix Stripping Stemmer," *INFORMAL*, vol. 8, no. 2, p. 101, Aug. 2023, doi: 10.19184/isj.v8i2.42310.
- [27] I. O. Suzanti and A. Jauhari, "COMPARISON OF STEMMING AND SIMILARITY ALGORITHMS IN INDONESIAN TRANSLATED AL-QUR'AN TEXT SEARCH," *cursor*, vol. 11, no. 2, p. 91, Jan. 2022, doi: 10.21107/kursor.v11i2.280.
- [28] S. Suyanto, A. Sunyoto, R. N. Ismail, E. Rachmawati, and W. Maharani, "Stemmer and phonotactic rules to improve n-gram tagger-based indonesian phonemicization," *Journal of King Saud University* -

- Computer and Information Sciences*, vol. 34, no. 6, pp. 3807–3814, Jun. 2022, doi: 10.1016/j.jksuci.2021.01.006.
- [29] S. I. G. Situmeang, “Impact of Text Preprocessing on Named Entity Recognition Based on Conditional Random Field in Indonesian Text,” vol. 6, no. 36, 2022.
- [30] T. H. Jaya Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, “Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier,” *Procedia Computer Science*, vol. 197, pp. 660–667, 2022, doi: 10.1016/j.procs.2021.12.187.
- [31] H. Dwiharyono and S. Suyanto, “Stemming for Better Indonesian Text-to-Phoneme,” *Ampersand*, vol. 9, p. 100083, 2022, doi: 10.1016/j.amper.2022.100083.
- [32] B. Siswanto and Y. Dani, “Sentiment Analysis about Oximeter as Covid-19 Detection Tools on Twitter Using Sastrawi Library,” in *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, Semarang, Indonesia: IEEE, Sep. 2021, pp. 161–164. doi: 10.1109/ICITACEE53184.2021.9617216.
- [33] R. Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, “Improving the Accuracy of Text Classification using Stemming Method, A Case of Informal Indonesian Conversation.” Aug. 17, 2020. doi: 10.21203/rs.3.rs-41431/v1.
- [34] M. D. Purbolaksono, F. D. Reskyadita, - Adiwijaya, A. A. Suryani, and A. F. Huda, “Indonesian Text Classification using Back Propagation and Sastrawi Stemming Analysis with Information Gain for Selection Feature,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 1, pp. 234–238, Feb. 2020, doi: 10.18517/ijaseit.10.1.8858.
- [35] S. Fahmi, L. Purnamawati, G. F. Shidik, M. Muljono, and A. Z. Fanani, “Sentiment Analysis of Student Review in Learning Management System Based on Sastrawi Stemmer and SVM-PSO,” in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia: IEEE, Sep. 2020, pp. 643–648. doi: 10.1109/iSemantic50169.2020.9234291.
- [36] N. W. Wardani and P. G. S. C. Nugraha, “Stemming Teks Bahasa Bali dengan Algoritma Enhanced Confix Stripping,” *IJNSE*, vol. 4, no. 3, pp. 103–113, Dec. 2020, doi: 10.23887/ijnse.v4i3.30309.
- [37] T. Yusnitasari, I. Humaini, L. Wulandari, and D. Ikasari, “Information Retrieval for Popular Words in Bahasa Translation of Al Quran and Hadith Bukhori Using Enhance Confix Stripping (ECS) Stemming,” *AJSEA*, vol. 8, no. 1, p. 18, 2019, doi: 10.11648/j.ajsea.20190801.13.
- [38] W. Rifai and E. Winarko, “Modification of Stemming Algorithm Using A Non Deterministic Approach To Indonesian Text,” *Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, p. 379, Oct. 2019, doi: 10.22146/ijccs.49072.
- [39] M. A. Muchtar *et al.*, “Separation of Basic Words in Angkola Batak Text Documents using Enhanced Confix Stripping Stemmer Case: Mandailing Ethnic,” *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 648, no. 1, p. 012024, Oct. 2019, doi: 10.1088/1757-899X/648/1/012024.
- [40] K. M. Suryaningrum, “Implementasi Algoritma Confix Stripping untuk Pendeteksian Kesalahan pada Tenses,” *AITI*, vol. 16, no. 1, pp. 88–98, Dec. 2019, doi: 10.24246/aiti.v16i1.88-98.
- [41] A. S. Rizki, A. Tjahyanto, and R. Trialih, “Comparison of stemming algorithms on Indonesian text processing,” *TELKOMNIKA*, vol. 17, no. 1, p. 95, Feb. 2019, doi: 10.12928/telkomnika.v17i1.10183.
- [42] A. T. Ni'mah, D. A. Suryaningrum, and A. Z. Arifin, “Autonomy Stemmer Algorithm for Legal and Illegal Affix Detection use Finite-State Automata Method,” *EPI-IJE*, vol. 2, no. 1, pp. 46–55, Jun. 2019, doi: 10.25042/epi-ije.022019.09.