

OPTIMASI ALGORITMA STEMMING PORTER UNTUK PEMROSESAN TEKS DALAM BAHASA INDONESIA

Rajnaparamitha Kusumastuti*¹, Dewi Oktafiani², Yustikamasy Astica³

Prodi S1 Teknologi Informasi STMIK Amikom Surakarta^{1,2}, Magister Informatika Universitas
Amikom Yogyakarta

Korespodensi e-mail: rajna@dosen.amikomsolo.ac.id¹, dewioktafiani@dosen.amikomsolo.ac.id,
yustikamasy@students.amikom.ac.id³

Abstrak

Stemming adalah metode untuk menemukan kata dasar dengan menghapus imbuhan. Proses ini penting dalam pengolahan data karena mempengaruhi kinerja sistem. Ada banyak algoritma untuk stemming, termasuk algoritma Porter. Namun, algoritma ini sering gagal dalam proses stemming, sehingga perlu pengembangan aturan. Pengembangan yang diusulkan adalah dengan melakukan tiga percabangan, yang masing-masing memeriksa kata dalam kamus. Tujuannya adalah untuk meningkatkan kinerja dan mengurangi kesalahan. Penelitian ini menguji kegagalan stemming dari penelitian sebelumnya dengan algoritma yang telah dioptimalkan, yang memperbaiki proses dari library Porter. Optimasi ini menghilangkan tahap possessive pronoun dan selalu memeriksa kata yang telah distemming. Selain itu, pengujian dilakukan menggunakan data dari Twitter dengan kata kunci KOMINFO. Hasilnya menunjukkan akurasi 84% untuk alur stemming optimasi, lebih baik dari 80% pada algoritma Porter.

Kata kunci: KOMINFO, NLP, Porter, Stemming, Text Preprocessing

Abstract

Stemming is a method for finding root words by removing affixes. This process is crucial in data processing because it affects system performance. Numerous stemming algorithms exist, including the Porter algorithm. However, these algorithms often fail in the stemming process, necessitating rule development. The proposed development involves three branching paths, each of which checks the word in a dictionary. The goal is to improve performance and reduce errors. This study tested stemming failures from previous studies with an optimized algorithm that improves the process from the Porter library. This optimization eliminates the possessive pronoun stage and always checks for stemmed words. Furthermore, testing was conducted using data from Twitter with the keyword KOMINFO. The results showed an accuracy of 84% for the stemming optimization path, better than the 80% accuracy of the Porter algorithm.

Keywords: Communication and Information Technology, NLP, Porter, Stemming, Text Preprocessing

PENDAHULUAN

Dalam analisis teks, tahap pra-pemrosesan menjadi langkah krusial untuk memastikan data teks memiliki format yang terstruktur dan siap untuk diolah lebih lanjut. *Text preprocessing* merupakan proses mengubah teks agar memiliki struktur yang baik dengan cara menyeleksi teks. Salah satu tahapan pada *text preprocessing* adalah *stemming* [1]. *Stemming* merupakan metode untuk mencari kata dasar pada suatu kata dengan cara menghilangkan kata imbuhan [2]. Imbuhan yang dihilangkan dapat berupa *sufiks*, *konfiks*, dan *prefiks*. *Stemming* teks bahasa Indonesia mempunyai kesulitan yang cukup tinggi karena terdapat imbuhan yang beraneka ragam, dan juga bertingkat.

Stemming pada teks berbahasa Indonesia dapat dilakukan melalui dua pendekatan utama, yaitu dengan menggunakan *library* atau dengan menerapkan aturan mandiri tanpa bantuan *library* (*non-library*) [3]. Pendekatan berbasis *library* lebih umum digunakan karena lebih stabil dan cenderung memiliki tingkat kesalahan yang lebih rendah, mengingat *library* tersebut telah melalui tahap pengembangan dan pengujian yang matang. Di sisi lain, metode *non-library* menawarkan keunggulan dalam efisiensi waktu pemrosesan, namun rentan terhadap kesalahan, terutama jika aturan yang digunakan kurang lengkap atau tidak dirancang secara komprehensif [4].

Library untuk *stemming* yang sering digunakan dalam bahasa Indonesia mencakup berbagai pilihan seperti Sastrawi, Nazief-Adriani, Porter, dan lainnya [5], [6], [7]. Masing-masing *library* memiliki keunggulan dan kelemahannya, tergantung pada kebutuhan dan kompleksitas teks yang diolah. Misalnya, *library* seperti Sastrawi dirancang khusus untuk bahasa Indonesia dan memiliki cakupan imbuhan yang lebih relevan, sementara Porter, yang awalnya dirancang untuk bahasa Inggris, telah dimodifikasi agar dapat digunakan pada bahasa Indonesia.

Berbagai penelitian telah dilakukan untuk mengembangkan algoritma stemmer, salah satunya adalah pengembangan Algoritma Porter dengan menambahkan kamus sebagai bagian dari proses [8], [9]. Penambahan ini bertujuan untuk meningkatkan akurasi hasil *stemming* sekaligus mengurangi ambiguitas yang sering terjadi dalam proses penghilangan imbuhan. Dalam penelitian tersebut, metode yang diterapkan meliputi pemeriksaan kata-kata hasil pemotongan imbuhan terhadap kamus, yang bertujuan memastikan bahwa kata yang dihasilkan merupakan kata dasar yang valid. Pendekatan ini berhasil meningkatkan akurasi *stemming* hingga sebesar 13,33%, menunjukkan potensi signifikan dari penambahan kamus dalam algoritma *stemming* untuk bahasa Indonesia.

Penelitian lain mengeksplorasi penggunaan Algoritma Porter yang diuji pada dokumen teks berbahasa Indonesia. Hasil pengujian menunjukkan bahwa modifikasi algoritma *stemming* menghasilkan *recall* sebesar 89,9% [4]. Namun, meskipun hasilnya cukup tinggi, penelitian ini masih mengidentifikasi beberapa kelemahan, seperti kata-kata yang mengalami *error* dalam proses *stemming* serta kata-kata yang gagal diproses sepenuhnya. Temuan ini menunjukkan bahwa meskipun algoritma Porter cukup andal, masih ada ruang untuk perbaikan guna meningkatkan akurasi dan mengurangi kegagalan dalam proses *stemming*.

Penelitian serupa juga pernah dilakukan oleh [10] yang berhasil melakukan *stemming* dengan model pembuatan aturan atau *non library*. Pada penelitiannya mengusulkan perubahan aturan stemmer dan penambahan kosakata. Penelitian ini menganalisis sentimen masyarakat terhadap calon presiden Indonesia 2024 di Twitter menggunakan metode *Rule-Based* dengan tingkat keberhasilan *preprocessing* dan analisis sebesar 90%, menunjukkan bahwa Ganjar Pranowo memiliki sentimen positif tertinggi dan negatif terendah, sementara Prabowo mendapat sentimen negatif tertinggi. Akan tetapi penggunaan *stemming* ini masih memiliki

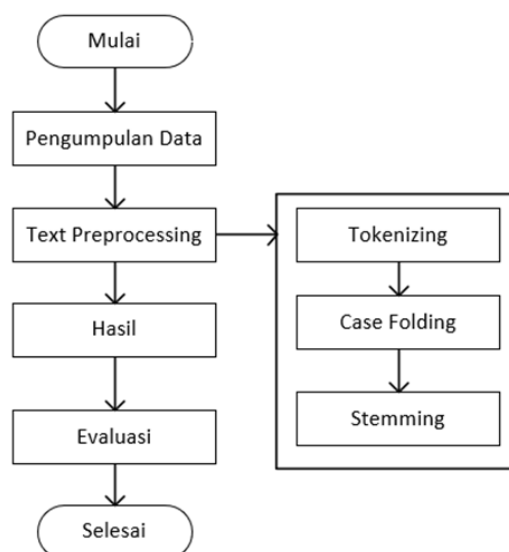
banyak kesalahan pada kata yang merupakan nama tempat atau nama orang, kata yang termasuk singkatan, imbuhan akhiran, *over stemming*, *error* pada kata yang salah ketik, dan *understemming*.

Penelitian ini bertujuan untuk mengembangkan alur pada *library* Porter agar proses *stemming* dapat menangani kata-kata dengan struktur imbuhan yang lebih kompleks dan bertingkat. Dalam bahasa Indonesia, banyak kata memiliki imbuhan yang tidak hanya terdiri dari satu lapis, melainkan kombinasi prefiks, sufiks, dan konfiks. Hal ini seringkali menyebabkan kesalahan dalam proses *stemming* karena algoritma yang ada mungkin tidak dirancang untuk mengelola struktur yang kompleks tersebut. Pengembangan ini difokuskan pada penyesuaian dan penambahan aturan yang memungkinkan algoritma Porter untuk mengidentifikasi dan memproses imbuhan bertingkat secara lebih akurat. Selain itu, proses pengembangan ini juga bertujuan untuk mengurangi kemungkinan kesalahan, seperti pemotongan imbuhan yang tidak tepat atau kegagalan mengenali kata dasar. Salah satu pendekatan yang diusulkan adalah dengan menambahkan langkah pemeriksaan terhadap kamus kata dasar setelah proses penghilangan imbuhan, sehingga hasil *stemming* lebih terjamin keakuratannya.

METODE PENELITIAN

Pada penelitian ini, dilakukan beberapa tahapan sistematis untuk memastikan proses penelitian berjalan dengan baik dan hasil yang diperoleh dapat dipertanggungjawabkan. Alur penelitian secara umum dapat dilihat pada Gambar 1, yang dimulai dari tahap pengumpulan data hingga evaluasi hasil. Penelitian ini mengikuti alur sistematis yang dimulai dari pengumpulan data, di mana data relevan dikumpulkan untuk dianalisis.

Selanjutnya, data tersebut diproses melalui tahapan *text preprocessing*, yang meliputi *tokenizing* (memecah teks menjadi unit kecil), *case folding* (mengubah teks menjadi huruf kecil), dan *stemming* (menghilangkan imbuhan untuk mendapatkan kata dasar). Hasil dari proses ini kemudian dianalisis dan dievaluasi menggunakan metrik akurasi untuk mengukur keberhasilan metode yang digunakan. Tahap akhir adalah menyusun laporan dan memberikan rekomendasi pengembangan berdasarkan hasil evaluasi. Alur ini dirancang untuk menghasilkan penelitian yang terstruktur dan akurat.



Gambar 1. Alur Penelitian

Stemming

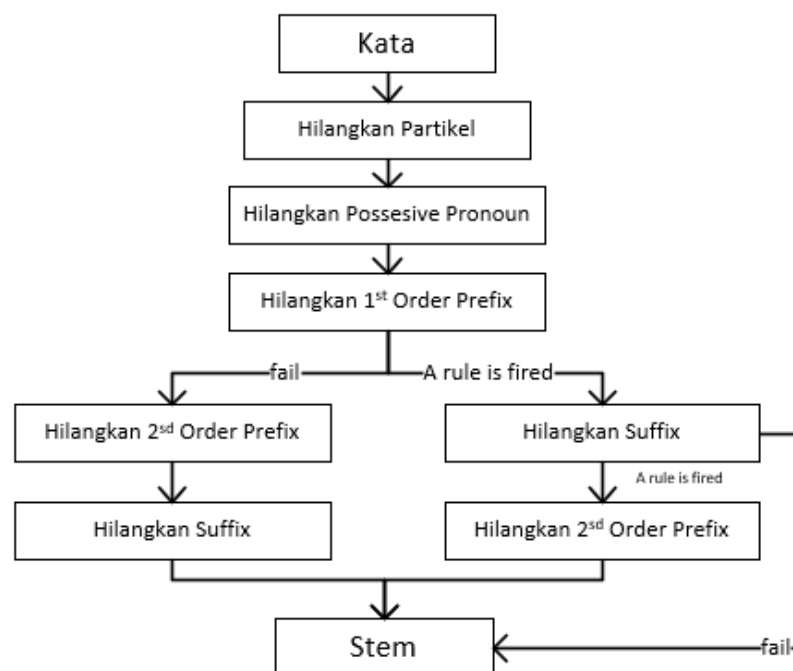
Stemming merupakan suatu proses mengekstraksi kata yang ada dari sufiksnya untuk mendapatkan akar kata. Hasil yang diperoleh dari proses *rooting* disebut stem. Penerapan proses pada setiap bahasa memiliki proses yang berbeda-beda sesuai dengan morfologi masing-masing bahasa. Untuk itu, tujuan dari proses *stemming* memiliki tujuan yang sama yaitu untuk mendapatkan kata dasar yang tetap, serta memerlukan algoritma yang berbeda untuk setiap bahasa[11]

Tahapan *stemming* ini merupakan tahapan paling utama dalam melakukan *text preprocessing* karena hasil yang dikeluarkan pada tahapan ini memiliki pengaruh besar terhadap kata dasar yang akan terbentuk. Kata dasar tersebut biasanya memiliki makna kata yang berubah sehingga maknanya menjadi berbeda dari isi dari teks aslinya. Sehingga *stemming* yang kurang baik dapat menyumbang bias yang cukup besar ke dalam proses yang akan dilakukan.

Algoritma Porter

Algoritma porter merupakan algoritma yang berfungsi untuk melakukan *stemming*. Algoritma ini dikembangkan pada tahun 1992 oleh W.B. Frakes sebagai stemmer dalam bahasa inggris. Lalu, pada tahun 2003 Algoritma Porter dikembangkan oleh [12] untuk melakukan *stemming* dalam bahasa Indonesia. Algoritma Porter dalam melakukan *stemming* tidak bergantung pada database kamus. Tahapan proses *stemming* pada algoritma ini yaitu [4]:

1. Menghilangkan imbuhan atau partikel “-pun”, “-lah”, “-kah”, dan “-tah”
2. Menghilangkan kata ganti atau kata kepemilikan, misalnya “-ku”, “-nya”, dan “-mu”
3. Menghapus imbuhan awalan pada imbuhan pertama (jika ada awalan maka lanjut ke langkah berikutnya, yaitu langkah 4 namun jika tidak maka akan langsung ke langkah 5)
4. Lalu, menghapus imbuhan awalan yang kedua (selanjutnya ke langkah no 6)
5. Setelah itu, menghapus imbuhan akhiran (jika ada akhiran lanjut ke langkah no 7, namun jika tidak ada akhiran berarti dianggap sebagai kata dasar)
6. Kemudian menghapus imbuhan akhiran atau kata akhir (diartikan sebagai kata dasar)
7. Yang terakhir adalah dengan menghapus imbuhan awalan kedua dan kata dasar (diasumsikan sebagai *root word*)

Gambar 2. Alur *Stemming* Porter bahasa Indonesia

Library dari alur pada Gambar 2 menunjukkan bahwa proses yang dilakukan dalam melakukan *stemming* adalah dilakukan dengan menghilangkan *prefix* terlebih dahulu yang kemudian proses berikutnya dilakukan pengecekan secara berkala dengan pengecekan *suffix* pada setiap token kata hingga menghasilkan output akhir berupa kata dasar. Dalam penelitian sebelumnya[4] ditemukan kegagalan proses *stemming* terhadap 20 kata yang diantara dapat diperhatikan pada Tabel 1

Tabel 1. Sampel Kata Yang Gagal *Stemming*

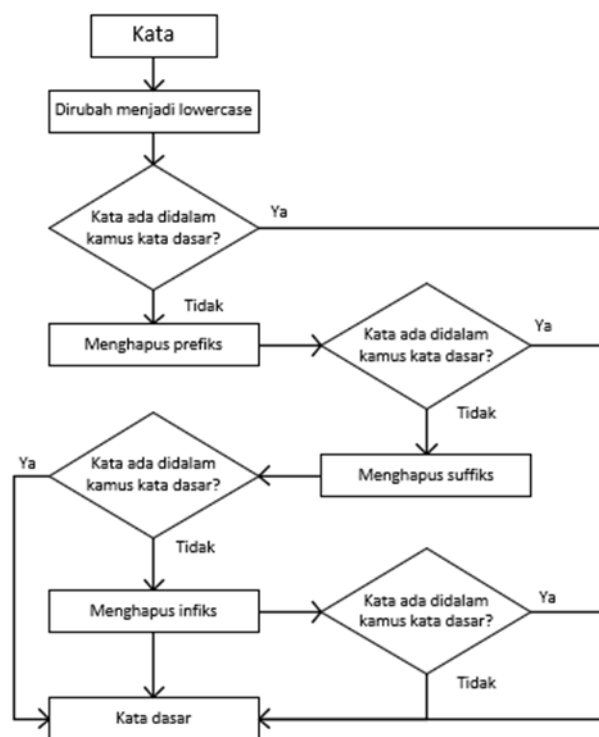
Token	Stemming Gagal	Hasil Seharusnya
Asupan	Asupan	Asup
Bertahun	Bertahun	Tahun
Bekerjasama	Bekerjasama	Kerjasama
Beratnya	Rat	Berat
Berepresi	Berepresi	Ekspresi
Berlaku	Berla	Laku
Berolah	Bero	Olah
Berpengalaman	Berpengalaman	Alam
Bersalah	Bursa	Salah
Bersekolah	Seko	Sekolah
Bertanggungjawab	Bertanggungjawab	Tanggungjawab
Bertanya	Berta	Tanya
Bertopologi	Bertopologi	Topologi
Berupa	Upa	Rupa
Bukanlah	Bu	Bukan
Dariku	Dari	Dari
Dengannya	Dengan	Dengan
Diadakan	Adakan	Ada
Dianjurkan	Dianjurkan	Anjur

Ditaati

Ditaati

Taati

Kegagalan *token* yang tidak ter-*stemming* jika diperhatikan memiliki bentuk kata berjenjang yang dan hal ini bertentangan dengan proses alur pada gambar 2 seharusnya dapat ter-*stemming*. Kemudian bentuk kata *possessive pronoun* berhasil ter-*stemming* tetapi kata tersebut masuk ke dalam tabel *error result stemming*, sehingga alur dari Gambar 2 yang digunakan pada penelitian [13] belum berhasil sepenuhnya melakukan pembentukan kata dasar. Untuk mengatasi kegagalan kata yang tidak berhasil ter-*stemming* pada penelitian yang dilakukan oleh peneliti pada Tabel 1 maka percobaan akan dilakukan untuk melakukan *stemming* pada *token* Tabel 1 dengan menggunakan alur *purposed stemming* seperti pada Gambar 3.

Gambar 3. Alur *Purposed Stemming*

Optimasi alur *stemming* yang kami gunakan adalah dengan cara membagi 3 kali percabangan, di setiap percabangan dilakukan pengecekan kata apakah kata tersebut sudah ada didalam kamus yang sudah dikumpulkan dari KBBI. Untuk percabangan pertama adalah menghapus *prefiks*, *prefiks* yang kami gunakan ada 13 *prefiks*, yaitu meng, per, be, ter, di, ke, se, peng, be, pem, men, mem, dan me. Untuk *sufiks* kami menggunakan 4 *sufiks* yaitu kan, an, i, dan nya. Lalu yang terakhir, untuk *infiks* kami menggunakan tiga kata, yaitu el, er, dan em. Setelah semua proses tersebut dilakukan dan tidak memiliki *prefiks*, *sufiks*, dan *infiks* akan dianggap kata tidak baku atau sebagai kata dasar.

Berdasarkan pengujian menggunakan alur *stemming* optimasi, hasil kata yang belum berhasil ter-*stemming* dapat berhasil ter-*stemming* menjadi kata dasar yang memiliki hasil sesuai. Hasil percobaan menggunakan alur pengembangan yang di usulkan, dapat dilihat hasilnya ditunjukkan oleh Tabel 2.

Tabel 2. Hasil Percobaan Menggunakan Alur *Stemming* Optimasi

Token	Stemming Gagal	Stemming Optimasi	Hasil Seharusnya
Asupan	Asupan	Asup	Asup
Bertahun	Bertahun	Tahun	Tahun
Bekerjasama	Bekerjasama	Kerjasama	Kerjasama
Beratnya	Rat	Berat	Berat
Berekpresi	Berekpresi	Ekspresi	Ekspresi
Berlaku	Berla	Laku	Laku
Berolah	Bero	Olah	Olah
Berpengalaman	Berpengalaman	Berpengalaman	Alama
Bersalah	Bersa	Salah	Salah
Bersekolah	Seko	Sekolah	Sekolah
Bertanggungjawab	Bertanggungjawab b	Tanggungjawab	Tanggungjawab
Bertanya	Berta	Tanya	Tanya
Bertopologi	Bertopologi	Topologi	Topologi
Berupa	Upa	Upa	Rupa
Bukanlah	Bu	Bukanlah	Bukan
Dariku	Dari	Dariku	Dari
Dengannya	Dengan	Dengan	Dengan
Diadakan	Adakan	Adakan	Ada

HASIL DAN PEMBAHASAN

Guna memperkuat dan mengetahui apakah alur pengembangan *stemming* dapat digunakan lebih baik untuk melakukan proses *stemming* pada *case* yang berbeda, berikut dilakukan percobaan menggunakan *dataset* yang sudah diperoleh melalui *twitter* mengenai topik Kementerian Komunikasi dan Informatika (KOMINFO) yang sedang hangat menjadi topik perbincangan perihal status pemblokiran akun terhadap beberapa platform seperti paypal, linkedin, github dll. Jumlah data uji yang digunakan adalah pengguna *twitter* yang sudah terverifikasi sebagai pengguna asli dan bukan merupakan robot. Data *twitter* hasil *crawling* ditunjukkan pada Tabel 3.

Tabel 3. Dataset *Twitter*

No	<i>Dataset Tweet</i>
1	Nasib gamers di Indonesia disorot media-media asing setelah Steam diblokir Kominfo
2	Tema tersebut sengaja diambil karena saat ini sedang beredar isu Kominfo akan menutup beberapa sosial media. #publisherstory
3	Ramainya penyebaran kabar hoax, disebut Kominfo baru-baru ini untuk mengeluarkan ancaman pemblokiran bagi sejumlah aplikasi media sosial. Namun, hal ini jauh dari fakta. Cek kebenarannya berikut ini!
4	Kominfo Tambah Waktu Lima Hari bagi PSE yang Belum Mendaftar
5	Saya setuju @kemkominfo memblokir aplikasi digital yg: - bandel tidak mau mendaftarkan aplikasinya; - tidak jelas kontribusinya terhadap perekonomian negara; - merusak mental bangsa; - tidak mau membayar pajak. Semoga @kemkominfo tidak ragu untuk bertindak.
6	Hasil survei Katadata Insight Center (KIC) bersama Kementerian Komunikasi dan Informatika (Kominfo) menyebutkan bahwa mayoritas masyarakat Indonesia mengakses berita melalui media sosial lho, Teman ANTARA

Dari data twitter yang digunakan terbentuk 26 token yang kemudian akan dilakukan proses stemming menggunakan alur pengembangan stemming. Hasil dari percobaan dapat diperhatikan pada Tabel 4.

Tabel 4. Hasil *Stemming* Dataset *Twitter*

Token	Sebelum Stemming	Sesudah Stemming Porter	Sesudah Optimasi Stemming Porter	Hasil Seharusnya
Diblokir	Diblokir	Blokir	Blokir	Blokir
Disorot	Disorot	Sorot	Sorot	Sorot
Setelah	Setelah	Telah	Sete	Telah
Diambil	Diambil	Ambil	Ambil	Ambil
Beredar	Beredar	Edar	Edar	Edar
Menutup	Menutup	Menutup	Tutup	Tutup
Beberapa	Beberapa	Beberapa	Berapa	Beberapa
Ramainya	Ramainya	Ramai	Rama	Ramai
Penyebaran	Penyebaran	Penyebaran	Sebar	Penyebaran
Disebut	Disebut	Sebut	Sebut	Sebut
Mengeluarkan	Mengeluarkan	Uar	Eluar	Keluar
Pemblokiran	Pemblokiran	blokir	Blokir	Blokir
Kebenarannya	Kebenarannya	kebenarannya	Benar	Benar
Mendaftar	Mendaftar	daftar	Daftar	Daftar
Melalui	Melalui	lalu	Lalu	Lalu
Aplikasinya	Aplikasinya	aplikasi	Aplikasi	Aplikasi
Kontribusinya	Kontribusinya	kontribusi	Kontribus	Kontribusi
Perekonomian	Perekonomian	ekonomi	Ekonomi	Ekonomi
Merusak	Merusak	rusak	Rusak	Rusak
Membayar	Membayar	bayar	Bayar	Bayar
Bertindak	Bertindak	tindak	Tindak	Tindak
Bersama	Bersama	sama	Sama	Sama
Kementerian	Kementerian	menteri	Menteri	Menteri
Menyebutkan	Menyebutkan	menyebutkan	Sebut	Sebut
Mengakses	Mengakses	akses	Akses	Akses
Terhadap	Terhadap	hadap	Hadap	Hadap

Hasil pengujian alur *stemming* optimasi yang sudah dilakukan, diperoleh hasil bahwa kata yang berhasil ter-*stemming* masih terdapat 4 kata yang belum berhasil ter-*stemming* dengan baik. Namun dalam kondisi ini, alur *stemming* optimasi yang digunakan berhasil melakukan *stemming* lebih baik dibandingkan pada alur *stemming* ketika hanya menggunakan alur *stemming* porter.

Untuk mengetahui performa dari hasil percobaan yang sudah dilakukan untuk melakukan *stemming* kata dari *twitter* maka perlu dilakukan evaluasi untuk mengetahui nilai akurasi dari proses yang sudah dilakukan. Nilai akurasi digunakan untuk mengetahui tingkat keakuratan. Hasil dapat dikatakan memiliki keakuratan yang tinggi jika nilai akurasi semakin besar[14]. Pengujian performa dapat dilakukan dengan rumus persamaan 1[15]:

$$\text{Akurasi} = \frac{(\text{Jumlah root words} - \text{Jumlah yang diklasifikasi benar}) \times 100\%}{\text{Total sampel yang di uji}}$$

Agar mendapatkan hasil pengujian yang baik antara kedua alur *stemming* yang ada, maka dilakukan pengujian terhadap kedua alur tersebut menggunakan *case* dari data yang sudah

diperoleh dari *twitter*. Hasil pengujian menunjukkan bahwa nilai akurasi *stemming* porter 80% dan nilai akurasi *stemming* optimasi 84%.

Selain melakukan pengujian akurasi seperti persamaan rumus 1, berikutnya dilakukan pengujian menggunakan *confusion matrix* untuk mengetahui kinerja dari metode *stemming* yang diusulkan [6]. Pada pengujian *confusion matrix*, eksperimen yang dilakukan menggunakan alur *stemming* optimasi menghasilkan nilai *True Positive* (TP) yang tinggi karena hasil pengujian sesuai dengan hipotesis awal pada percobaan (pengembangan alur *stemming* akan meningkatkan akurasi). Nilai *True Positive* (TP) yang diperoleh yaitu 96, *False Positive* (FP) sebanyak 4. *False Negative* (FN) yang diperoleh dari penelitian ini yaitu 19. Lalu untuk nilai *True Negative* (TN) yaitu 26. Perhitungan *confusion matrix* dapat dilihat pada Tabel 5.

Tabel 5. Dataset *Twitter*

	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	96	4
<i>Negative</i>	19	26

Hasil perhitungan nilai akurasi dari *stemming* porter dan *stemming* optimasi yang diajukan, hasil nilai akurasi pada *stemming* optimasi memiliki nilai yang lebih tinggi yaitu lebih unggul 4%. Dari hasil percobaan yang dilakukan peneliti, pengembangan *stemming* belum cukup signifikan karena terbatasnya dataset dan *rule* yang dipakai peneliti. Pada penelitian selanjutnya, diharapkan dapat menambah *rule* untuk *stemming* dan menambah jumlah dataset.

Selain itu untuk menguji performa dari pengembangan *stemming*, penelitian selanjutnya diharapkan dapat melakukan pembobotan atau perhitungan untuk menimbang kedekatan kata sebelum *stemming* dan sesudah *stemming*. Serta melakukan penelitian untuk menemukan pola kalimat yang berubah maknanya setelah melalui proses *stemming*. Penelitian berkaitan kalimat yang tidak perlu melalui proses *stemming* diharap kedepannya mampu membentuk *library* baru untuk meningkatkan nilai akurasi pada analisis sentimen.

KESIMPULAN

Berdasarkan hasil percobaan yang sudah dilakukan untuk mengatasi kegagalan *stemming* kata pada penelitian sebelumnya, pengembangan alur *stemming* yang dilakukan berhasil melakukan proses *stemming* terhadap kata yang belum berhasil ter-*stemming* sebelumnya. Kegagalan alur porter pada penelitian [7] terdapat pada bagian penghapusan *prefix* bertingkat, sehingga pada alur pengembangan *stemming* optimasi ini menghapus tahapan *possessive pronoun* sehingga kata bertingkat dapat ter-*stemming*. Hal ini didukung juga atas hasil *stemming* terhadap *case* baru dari dataset *twitter* yang digunakan untuk melakukan *stemming* kata dan mampu menghasilkan performa akurasi sebesar 84%. Sehingga ke depannya alur pengembangan yang dapat digunakan sebagai alternatif pada proses *stemming*.

Perlu dilakukan tindak lanjut terhadap pengaruh *profiling* dataset yang diambil berdasarkan lokasinya. Khususnya dengan dataset dari seluruh Indonesia yang memiliki keragaman gaya bahasa untuk menentukan apakah suatu kalimat masuk kedalam sentiment positif ataupun negatif. Penelitian selanjutnya diharap mampu mengembangkan *rule* dan menambah dataset sehingga kata yang di *stemming* akan lebih akurat. Selain itu untuk menguji performa akurasi dari pengembangan *stemming*, penelitian selanjutnya diharapkan dapat melakukan pembobotan kedekatan kata sebelum *stemming* dan sesudah *stemming*.

DAFTAR PUSTAKA

- [1] R. J. Kasim and E. Utami, “Penerapan Algoritma Boyer Moore yang di Modifikasi untuk stemmer Bahasa Indonesia,” *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1657–1667, Aug. 2024, doi: 10.29100/jipi.v9i3.5449.
- [2] M. U. Albab, Y. Karuniawati P, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic,” vol. 20, no. 2, pp. 1–10, 2023, doi: 10.26623/transformatika.v20i2.5374.
- [3] A. Siswandi and N. Surojudin, “ANALISIS DAN PERBANDINGAN STEMMING ALGORITMA PORTER DENGAN ALGORITMA AHMAD YUSOFF SEMBOK DALAM DOKUMEN TEKS BAHASA INDONESIA,” *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK)*, vol. 4, no. 1, 2020.
- [4] A. Arif Siswandi, Y. Permana, and A. Emarilis, “Stemming Analysis Indonesian Language News Text with Porter Algorithm,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1845/1/012019.
- [5] 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.
- [6] Z. Abidin, A. Junaidi, and Wamiliana, “Text Stemming and Lemmatization of Regional Languages in Indonesia: A Systematic Literature Review,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 2, pp. 217–231, Jun. 2024, doi: 10.20473/jisebi.10.2.217-231.
- [7] F. Riza, S. Rifai, A. Dirgantara, Sfenrianto, Rasenda, and S. Herdyansyah, “Information Retrieval Technique for Indonesian PDF Document with Modified Stemming Porter Method Using PHP,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, 2020. doi: 10.1088/1742-6596/1477/3/032016.
- [8] K. Ezra Rofran and F. Joanda Kaunang, “ANALISIS SENTIMEN PENGGUNA INSTAGRAM TERHADAP KEBIJAKAN NADIEM MAKARIM YANG MEMPERBOLEHKAN MAHASISWA LULUS TANPA SKRIPSI MENGGUNAKAN METODE ANALISIS VADER DAN METODE KLASIFIKASI NAÏVE BAYES ANALYSIS OF INSTAGRAM USER SENTIMENTS TOWARDS NADIEM MAKARIEM’S POLICY OF ALLOWING STUDENTS TO GRADUATE WITHOUT A THESIS USING THE VADER ANALYSIS METHOD AND NAÏVE BAYES CLASSIFICATION METHOD,” *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 7, no. 2, 2024.
- [9] A. Yudi and D. M. Makmun, “Optimasi Stemming Porter KBBI dan Cross Validation NaïveNaïve Bayes untuk Klasifikasi Topik Soal UN (Ujian Nasional) Bahasa Indonesia.”
- [10] S. Djusar, S. Firhan Abdillah, P. Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Lancang Kuning, and J. K. Yos Sudarso, “Perbandingan Analisis Sentimen Masyarakat Terhadap Calon Presiden Negara Indonesia 2024 Berdasarkan Opini Di Twitter Menggunakan Metode Rule Based Perbandingan Analisis Sentimen Masyarakat Terhadap Calon Presiden Negara Indonesia 2024 Berdasarkan Opini Di Twitter Menggunakan Metode Rule Based”.

- [11] M. Fikri and Z. Abidin, “Analysis Of The Use Of Nazief-Adriani Stemming And Porter Stemming In Covid-19 Twitter Sentiment Analysis With Term Frequency-Inverse Document Frequency Weighting Based On K-Nearest Neighbor Algorithm,” *Recursive Journal of Informatics*, vol. 2, no. 2, pp. 80–87, Sep. 2024, doi: 10.15294/rji.v2i2.74267.
- [12] G. Mandar, A. H. Muhamamd, and S. Sudin, “Klasifikasi Berita Indonesia Menggunakan Naïve Bayes dengan Porter Stemmer,” *Jurnal Teknik Informatika (J-Tifa)*, vol. 3, no. 2, pp. 17–22, Sep. 2020, doi: 10.52046/j-tifa.v3i2.1121.
- [13] R. K. and A. Sunyoto, “Skin Cancer Classification Using EfficientNetV2 and ViT B16,” in *2023 6th International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia: IEEE, 2023, pp. 395–400. doi: 10.1109/ICOIACT59844.2023.10455921.
- [14] R. Kusumastuti, E. Utami, and A. Yaqin, “Detection of Sarcasm Sentences in Indonesian Tweets using SentiStrength,” in *Proceeding - 6th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Sciences and Artificial Intelligence Technologies for Environmental Sustainability, ICITISEE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 93–98. doi: 10.1109/ICITISEE57756.2022.10057904.
- [15] F. Ramadhan Wiratama and S. Astuti, “EKSPLORA INFORMATIKA 127 Implementasi Algoritma C4.5 untuk Analisa Performa Pelayanan Bank Terhadap Nasabah Implementation of C4.5 Algorithm for Bank Service Performance Analysis To Customer.”