



Penerapan ECS Stemmer untuk Modifikasi Nazief & Adriani Berbahasa Jawa

Nur Hidayatullah¹, Aji Prasetya Wibawa², Harits Ar Rosyid³

^{1 2 3}Teknik Informatika, Jurusan Teknik Elektro, Universitas Negeri Malang

¹hidayatnur067@gmail.com, ²aji.prasetya.ft@um.ac.id, ³haritz.ar.ft@um.ac.id

Abstract

Stemming Javanese affix words using Nazief & Adriani modifications still has problems that cannot be solved such as overstemming, understemming, and unchange. Then there needs to be improvements to improve the performance of Nazief & Adriani modifications. This study aims to improve the performance of Nazief & Adriani modifications using the Enhanced Confix Stripping (ECS) modification method. The results of this study indicate that Enhanced Confix Stripping can improve performance that previously had an accuracy of only 78.2% to 97.9% with an error rate of 2.1%. And fixing errors that originally numbered 98 to 9 errors. But Enhanced Confix Stripping still has problems with the words "ngetan, kumanggih, kumarut, kumasis, kumareg, kumadul, kumaras, katawakake, and pangenan". The next research is expected to be able to solve this problem.

Keywords: Javanese, *Stemming*, *Enhanced Confix Stripping*

Abstrak

Stemming kata imbuhan Bahasa Jawa dengan menggunakan modifikasi Nazief & Adriani masih memiliki permasalahan yang tidak dapat diselesaikan seperti *overstemming*, *understemming*, dan *unchange*. Maka perlu adanya perbaikan untuk meningkatkan performa dari modifikasi Nazief & Adriani. Pada penelitian ini bertujuan untuk meningkatkan performa modifikasi Nazief & Adriani menggunakan metode modifikasi *Enhanced Confix Stripping* (ECS). Hasil penelitian ini menunjukkan bahwa *Enhanced Confix Stripping* dapat meningkatkan performa yang sebelumnya memiliki *accuracy* hanya 78,2% menjadi 97,9% dengan *error rate* 2,1%. Dan memperbaiki kesalahan yang awalnya berjumlah 98 menjadi 9 kesalahan. Namun *Enhanced Confix Stripping* masih memiliki permasalahan pada kata "ngetan, kumanggih, kumarut, kumasis, kumareg, kumadul, kumaras, katawakake, dan pangenan". Pada penelitian selanjutnya diharapkan dapat menyelesaikan permasalahan tersebut.

Kata kunci: Bahasa Jawa, *Stemming*, *Enhanced Confix Stripping*

© 2019 Jurnal RESTI

1. Pendahuluan

Bahasa Jawa merupakan bahasa yang sering digunakan untuk orang pulau Jawa seperti Jawa Timur, Jawa Tengah, Yogyakarta dan lain-lain [1]. Bahasa Jawa sangat populer dengan cara berkomunikasi yang sopan dan lemah lembut dibandingkan dengan bahasa yang lain di Indonesia. Bahasa Jawa juga memiliki beberapa keunikan kata pada kata awalan, sisipan, akhiran dan kombinasi dari ketiganya. Namun, sebutan kata awalan, sisipan, dan akhiran berbeda dengan Bahasa Indonesia. Untuk Bahasa Jawa kata awalan disebut *ater-ater*, kata sisipan disebut *seselan* dan kata akhiran disebut *panembang*. Pada kata *ater-ater*

dikelompokkan menjadi tiga *ater-ater* yaitu : pertama, *ater-ater anuswara* yang terdiri dari "n-, ny-, m-, dan ng-". Kedua, *ater-ater tripurusa* yang terdiri dari "dak-, kok-, dan di-". Ketiga, *ater-ater liya* yang terdiri dari "a-, pa-, pan-, pang-, pi-, pra-, tak-, tar-, ka-, ke-, sa-, kuma-, dan kapi-". Pada kata *seselan* terdiri dari "-um-, -in-, -el-, dan -er-". Pada kata *panembang* terdiri dari "-a-, -i-, -e-, -an-, -en-, -ake-, -na-, -ne-, -ku-, dan -mu-". Karena keunikan Bahasa Jawa, maka perlu adanya proses *stemming*. Proses *stemming* yaitu teknik untuk mencari bentuk kata dasar dari suatu kata dengan menghilangkan kata awalan, sisipan, akhiran, dan gabungan kata tersebut [2][3][4][5]. Selain itu

Diterima Redaksi : 12-07-2019 | Selesai Revisi : 04-09-2019 | Diterbitkan Online : 01-12-2019

stemming banyak digunakan untuk *inormation retriveal* (IR) [6] [7]. Untuk menghasilkan proses pengembalian awalan, sisipan, akhiran dan kombinasinya perlu mempelajari morfologi dari suatu bahasa yang benar. Morfologi merupakan proses pembentukan kata dari satuan lain yang berupa kata, pokok kata, frase, kata dan kata, pokok kata dan kata, pokok kata dan pokok kata [8].

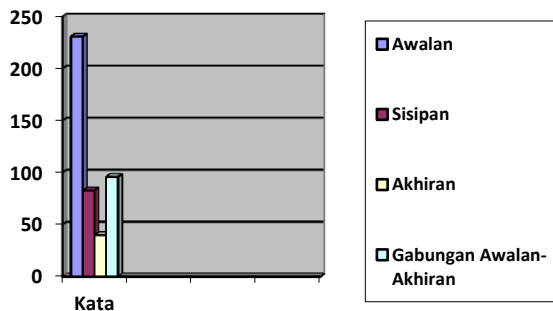
Algoritma *stemming* yang pernah digunakan sebelumnya adalah modifikasi Nazief & Adriani untuk Bahasa Jawa. Pada algoritma modifikasi Nazief & Adriani masih banyak mengalami permasalahan saat melakukan *stemming*. Maka perlu adanya perbaikan *stemming* dengan menggunakan algoritma *Enhanced Confix Stripping* (ECS). ECS adalah algoritma perbaikan dari *confix stripping* (CS) yang digunakan untuk metode *stemming* Bahasa Indonesia [9][10]. Namun ECS bukan metode untuk mengatasi masalah *overstemming* dan *understemming*. Contoh kata yang tidak bisa diselesaikan untuk diperbaiki menggunakan *Enhanced Confix Stripping* sebagai berikut : *kamitegan, kamigilan, pamulungan, pamriksa, koktekani, pamomong, pamudha, sadinane, kamban, gawana, segane, tekane, katawakake, pangenan*, dan *nithik*.

Artikel ini akan membahas tentang penerapan metode modifikasi *Enhanced Confix Stripping* (ECS) untuk *stemming* Bahasa Jawa. Dengan adanya ECS diharapkan akan memperbaiki pengembalian kata yang seharusnya tidak dihilangkan dan menambah beberapa *rule* perbaikan untuk Bahasa Jawa.

2. Metodologi Penelitian

2.1 Dataset

Data kata yang digunakan adalah data penelitian oleh *stemming* kata imbuhan Bahasa Jawa dengan menggunakan modifikasi algoritma Nazief & Adriani yang ditunjukkan pada Gambar 1.



Gambar 1. Jumlah Kata

Pada Gambar 1, kata berjumlah 449 kata, kata awalan sebanyak 231 kata, kata sisipan sebanyak 83 kata, kata akhiran sebanyak 40 kata, dan kata gabungan awalan-akhirannya sebanyak 96 kata.

2.2 Modifikasi Nazief & Adriani Bahasa Jawa

Pada bagian ini akan terdapat beberapa aturan pemenggalan awalan dari modifikasi Nazief & Adriani Bahasa Jawa yang ditunjukkan pada Tabel 1.

Tabel 1. Aturan Pemnggalan Awalan Modifikasi Nazief & Adriani

Aturan	Format Kata	Pemenggalan	Contoh
j1	ng{r l}V...	ng-{r l}V...	Ngrakit : Ng-Rakit
j2	m{1}V...	m-{1}V...	Mlembang : M-Lembang
j3	m{b u}A...	m-{b u}A...	Mbacok : M-Bacok
j4	n{j d}...	n-{j d}...	Njaba : N-Jaba; Ndableg : N-Dableg
j5	ng{g}...	ng-{g}...	Nggembol : Ng-Gembol
j6	kV...	k-V...	Kelingan : K-Eling-an
j7	kaA...	ka-A	Kapendhem : Ka-Pendhem
j8	keC...	ke-C...	Kegodhok : godhok
j9	ngA...	ng-A...	Ngababi : Ng-abab-i
j10	pV...	p-V...	Padusan : P-Adus-an
j11	nyA...	ny-cA... ny-sA...	Nyangkem : Ny-Cangkem
j12	mV...	m-pV...	Metani : M-Petan-i
j13	mV...	m-wV...	Minihi : M-Winihi
j14	nV... A!="j,d"	n-tV...	Napok : N-Tapok
j15	ngA...	ng-kA...	Nginthil : Ng-Kinthil
j16	ngV...	ng-wV...	Ngetan : Ng-Wetan

Berdasarkan Tabel 1, terdapat 16 aturan pemenggalan untuk Bahasa Jawa menggunakan modifikasi Nazief & Adriani.

Pada Tabel 1, simbol KD digunakan untuk kata dasar, simbol V digunakan untuk huruf vokal yaitu (aiueo), simbol C digunakan untuk huruf konsonan selain huruf vokal, dan simbol A digunakan untuk semua huruf A sampai Z.

2.3 Kasus Morfologi Bahasa Jawa yang tidak bisa diselesaikan

Contoh kasus yang tidak bisa diselesaikan pada modifikasi algoritma Nazief & Adriani ditunjukkan pada Tabel 2.

Tabel 2. Morfologi yang tidak bisa diselesaikan

Aturan	Format Kata	Pemenggalan	Contoh
	kamiKD	kami-KD	kamigilan :kami-gila-n
	kamiKD	kami-KD	Kamitegan :kami-tega-n
	pamKD	pam-KD	Pamulungan:pam-wulung-an
	pamKD	pam-KD	Pamriksa:pam-priksa
	kokKD	kok-KD	Koktekani:kok-teka-ni
	paKD	pa-KD	Pamomong:pa-momong
	paKD	pa-KD	Pamudha:pa-mudha
	saKD	sa-KD	Sadinane:sa-dina-ne

j6	kV...	k-V...	Kamban:k-amba-n
j7	kaA...	ka-A...	Katawakake:ka-tawa-kake
j10	pV...	p-V...	Pangenan:p-angen-an
j14	nA... dimana A!="j,d"	n-tV...	Nithik:n-thithik
j16	ngV...	ng-wV...	Ngetan:ng-wetan
j19	krV...	kr-V...	Krogel:kr-ogel
j20	klV...	kl-V...	Kluwer : Kl-uwer
j21	{b c g k r s}um A...	{b c g k r s}A...	Kumropok:k-um-ropok
j22	k{um}A...	k{um}-pA...	Kuminter : k-um-pinter
j23	k{um}A...	k{um}-wA...	Kumasis : k-um-wasis
j24	g{um}A...	g{um}-bA	Gumagus : g-um-bagus
j25	C{in}V... dimana C!="n,m"	C{in}-CV	Tinulis : t-in-tulis
j26	{ing}V...	{ing}-V	Ingundur : Ing-undur
j27	{b c g j k p}{l}V...	{b c g j k p}-V...	Bleber : b-l-beber
j28	{b c g k p}{r}V...	{b c g k p}-V...	Grandhul : g-r-gandhul
j29	reA...	re-A...	Rerasan : Re-rasa-an

Pada Tabel 2, kasus morfologi bahasa Jawa yang tidak dapat diselesaikan oleh modifikasi Nazief & Adriani sebanyak 24 kesalahan. Pada kata awalan terdapat 8 kesalahan, 5 aturan untuk kata awalan, dan 11 aturan untuk kata sisipan.

2.4 Enhanced Confix Stripping (ECS)

Algoritma ECS merupakan algoritma perbaikan dari *Confix Stripping* (CS) untuk Bahasa Indonesia [11]. Algoritma ini dikembangkan oleh Arifin, Mahendra, dan Ciptaningtyas [12]. Algoritma ECS digunakan untuk melakukan perbaikan dan menambahkan terhadap beberapa aturan imbuhan pemenggalan. Selain itu, ECS digunakan untuk melakukan pengembalian kata akhiran yang seharusnya tidak dihilangkan [13][14][15].

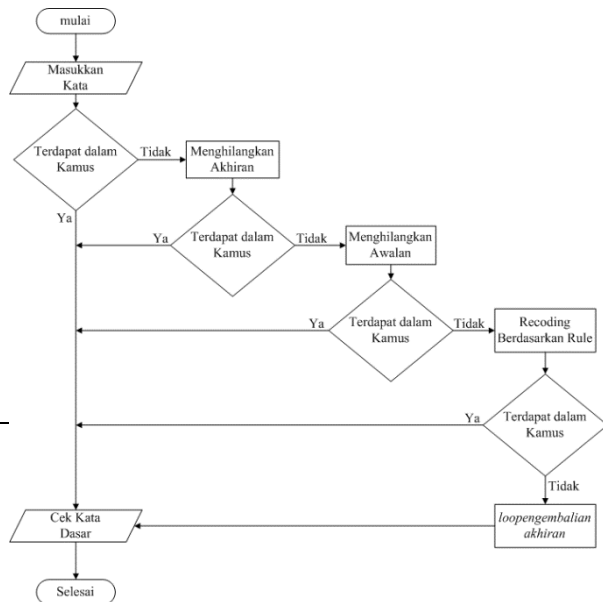
ECS akan dicoba dengan aturan baru untuk Bahasa Jawa dengan memperbaiki kesalahan modifikasi Nazief & Adriani untuk *stemming* kata imbuhan Bahasa Jawa.

2.5 Flowchart

Flowchart dari *Enhanced Confix Stripping* (ECS) ditunjukkan pada Gambar 2.

Pada Gambar 2 menunjukkan Proses *stemming* untuk ECS. Proses awal memasukkan kata. Selanjutnya, dilakukan proses kedua kamus akan mengecek kata tersebut berada pada kamus atau tidak jika terdapat pada kamus maka dianggap sebagai kata dasar. Jika tidak terdapat dalam kamus, maka dilakukan proses ketiga menghilangkan kata akhiran. Jika kata akhiran ditemukan dalam kamus, maka dianggap sebagai kata dasar. Tetapi, jika kata akhiran tidak ditemukan, maka

dilakukan proses ketiga menghilangkan kata awalan. Jika menghilangkan kata awalan terdapat dalam kamus, maka dianggap sebagai kata dasar. Namun, jika kata tidak terdapat dalam kamus, maka dilakukan proses *recoding*. Apabila proses *recoding* kata terdapat dalam kamus, maka proses dihentikan. Jika masih gagal dilanjutkan dengan proses pengembalian akhiran atau *looppengembalianakhir*. Jika masih tetap gagal kata akan dikembalikan kebentuk semula dan dianggap sebagai kata dasar.



Gambar 2. Flowchart Enhanced Confix Stripping

Contoh proses *stemming* pada kata “*sisihe*”. Kata “*sisihe*” terlebih dahulu dicek dalam kamus. Namun, kata “*sisihe*” tidak terdapat dalam kamus, maka dilakukan proses menghilangkan kata. Proses pertama, menghilangkan kata akhiran “*an*” sehingga kata menjadi “*sisih*”. Kata “*sisih*” dicek kembali dalam kamus. Karena di dalam kamus terdapat kata “*sisih*” sehingga proses selanjutnya dihentikan dan dianggap sebagai kata dasar.

2.6 Pengujian

Pada bagian pengujian ini digunakan *tools* NLTK dengan *library sastrawi* python yang dimodifikasi untuk *stemming* Bahasa Jawa. Pengujian menggunakan semua dataset yang ada. Kemudian, dilakukan perbaikan pemenggalan kata yang salah dalam algoritma sebelumnya.

2.7 Evaluasi

Pada bagian ini cara perhitungannya secara manual untuk mengetahui akurasi dari ECS dengan menggunakan rumus sebagai berikut :

$$Accuracy = \frac{Jumlah\ Kata\ benar}{Jumlah\ Kata\ seluruhnya} \times 100\% \quad (1)$$

Selain itu juga dihitung *error rate* data menggunakan rumus sebagai berikut :

$Error\ Rate = 100\% - Accuracy$

(2) “*kum*” pada aturan j22 pada Tabel 2. Sehingga proses *stemming* gagal menjadi kata “*panggih*”.

3. Hasil dan Pembahasan

Berdasarkan kesalahan modifikasi Nazief & Adriani untuk *stemming* kata imbuhan Bahasa Jawa, maka perbaikan yang dilakukan *Enhanced Confix Stripping* yang ditunjukkan pada Tabel 2.

Tabel 3. Jumlah Perbaikan ECS

Imbuhan	Jumlah Kata Benar	Jumlah Kata Salah	Akurasi(%)
Awalan	230	1	99,6
Sisipan	77	6	92,8
Akhiran	40	0	100
Awalan & Akhiran	93	2	97,9
Jumlah	440	9	97,9

Berdasarkan Tabel 3, jumlah kata yang salah adalah 9 kata. Kata yang salah atau tidak dapat melakukan proses *stemming* seperti kata “*ngetan*, *kumanggih*, *kumarut*, *kumasis*, *kumareg*, *kumadul*, *kumaras*, *katawakake*, dan *pangenan*”. Berikut ulasan dari kesalahan-kesalahan kata sebagai berikut :

Kesalahan pertama, pada kata “*ngetan*” dicek terlebih dahulu di dalam kamus. Namun, kata “*ngetan*” tidak terdapat dalam kamus, maka dilakukan proses menghilangkan akhiran “-an” sehingga kata menjadi kata “*nget*”. Kata “*nget*” dicek kembali dalam kamus. Tetapi, kata “*nget*” tidak terdapat dalam kamus. Maka kata akhiran “-an” dikembalikan lagi menjadi kata “*ngetan*” dan dilanjutkan dengan proses menghilangkan kata awalan “ng-” berdasarkan aturan j9 pada Tabel 1. Sehingga kata menjadi “*wetan*”. Kata dicek kembali dalam kamus. Kata “*wetan*” terdapat dalam kamus. Namun proses tidak dihentikan, dilanjutkan dengan algoritma “*looppengembalianakhir*”. Sehingga kata yang berakhiran “-an” akan dikembalikan menjadi akhiran “a”. Karena kata akhir “*wetan*” adalah “an”, maka kata menjadi “*weta*”. Akan tetapi, kata “*weta*” tidak terdapat dalam kamus, maka proses *stemming* gagal dan kata dikembalikan menjadi “*ngetan*”.

Kesalahan kedua, pada kata “*kumanggih*” dicek terlebih dahulu di dalam kamus. Akan tetapi, kata “*kumanggih*” tidak terdapat dalam kamus. Maka dilakukan proses menghilangkan kata. Proses pertama, menghilangkan akhiran. Tapi kata “*kumanggih*” tidak terdapat kata akhiran sehingga dilanjutkan proses selanjutnya. Proses kedua menghilangkan kata awalan “*kuma-*”. Sehingga kata menjadi “*nggah*”. Kata “*nggah*” dicek kembali dalam kamus. Namun, kata “*nggah*” tidak terdapat dalam kamus, maka kata awalan “*kuma-*” dikembalikan lagi menjadi kata “*kumanggih*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*kumanggih*” dilakukan *recoding* menggunakan aturan j22 pada Tabel 2. Namun kata tetap menjadi kata “*kumanggih*”. Karena terdapat kata awalan “*kuma-*” yang diproses terlebih dahulu, sebelum kata sisipan

Kesalahan ketiga, pada kata “*kumarut*” dicek terlebih dahulu di dalam kamus. Tetapi, kata “*kumarut*” tidak terdapat dalam kamus, maka dilakukan proses selanjutnya menghilangkan kata. Proses pertama, menghilangkan kata akhiran. Akan tetapi, kata “*kumarut*” tidak memiliki kata akhiran sehingga dilakukan proses selanjutnya. proses kedua, menghilangkan kata awalan “*kuma-*” sehingga kata menjadi “*rut*”. Kata “*rut*” dicek kembali dalam kamus. Namun, kata “*rut*” tidak terdapat dalam kamus, maka kata dikembalikan lagi menjadi kata “*kumarut*” dan dilanjutkan dengan proses ketiga, yaitu proses *recoding* berdasarkan aturan j22 pada Tabel 2. Kata “*kumarut*” tetap menjadi kata “*kumarut*”. Karena terdapat kata awalan “*kuma-*” yang diproses terlebih dahulu, sebelum kata sisipan “*kum*” pada aturan j22. Sehingga kata tidak dapat dilakukan proses *stemming* menjadi kata “*parut*”.

Kesalahan keempat, pada kata “*kumareg*” dicek terlebih dahulu di dalam kamus. Namun, kata “*kumareg*” tidak terdapat dalam kamus. Sehingga dilakukan proses selanjutnya menghilangkan kata. Proses pertama, menghilangkan kata akhiran. Akan tetapi, kata “*kumareg*” tidak memiliki kata akhiran, maka dilakukan proses selanjutnya. Proses kedua, menghilangkan kata awalan “*kuma-*” sehingga kata menjadi “*reg*”. Kata “*reg*” dicek kembali dalam kamus. Tetapi, kata “*reg*” tidak terdapat dalam kamus, maka kata dikembalikan lagi menjadi kata “*kumareg*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*kumareg*” di-*recoding* berdasarkan aturan j22 pada Tabel 2. Namun kata tetap menjadi kata “*kumareg*”. Karena terdapat kata awalan “*kuma-*” yang diproses terlebih dahulu, sebelum kata sisipan “*kum*”. Sehingga kata tidak dapat dilakukan proses *stemming* menjadi kata “*wareg*”.

Kesalahan kelima, pada kata “*kumasis*” dicek terlebih dahulu di dalam kamus. Tetapi, kata “*kumasis*” tidak terdapat dalam kamus, maka dilakukan proses menghilangkan kata. Proses pertama, menghilangkan kata akhiran. Karena kata “*kumasis*” tidak memiliki akhiran sehingga dilakukan proses selanjutnya. Pada proses kedua, menghilangkan kata awalan “*kuma-*” sehingga kata menjadi “*sis*”. Kata “*sis*” dicek kembali dalam kamus. Akan tetapi, kata “*sis*” tidak terdapat dalam kamus, maka kata dikembalikan lagi menjadi kata “*kumasis*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*kumasis*” di-*recoding* berdasarkan aturan j23 pada Tabel 2. Namun kata tetap menjadi kata “*kumasis*”. Karena terdapat kata awalan “*kuma-*” yang diproses terlebih dahulu, sebelum kata sisipan “*kum*”. Sehingga Kata tidak dapat dilakukan proses *stemming* menjadi kata “*wasis*”.

Kesalahan keenam, pada kata “*kumadul*” dicek terlebih dahulu di dalam kamus. Tetapi, kata “*kumadul*” tidak

terdapat dalam kamus, maka dilakukan proses menghilangkan kata. Proses pertama, menghilangkan kata akhiran. Karena kata “*kumadul*”, tidak memiliki akhiran sehingga dilakukan proses selanjutnya. Pada proses kedua, menghilangkan kata awalan “*kuma-*” sehingga kata menjadi “*dul*”. Kata “*dul*” dicek kembali dalam kamus. Namun, kata “*dul*” tidak terdapat dalam kamus sehingga kata dikembalikan lagi menjadi kata “*kumadul*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*kumadul*” di-*recoding* berdasarkan pemenggalan aturan j23 pada Tabel 2. Namun kata tetap menjadi kata “*kumadul*”. Karena terdapat kata awalan “*kuma-*” yang diproses terlebih dahulu, kata sisipan “*kum*”. Sehingga kata tidak dapat dilakukan proses *stemming* menjadi kata “*wadul*”.

Kesalahan ketujuh, pada kata “*kumaras*” dicek terlebih dahulu di dalam kamus. Namun, kata “*kumaras*” tidak terdapat dalam kamus, maka dilakukan proses selanjutnya yaitu menghilangkan kata. Proses pertama, menghilangkan kata akhiran. Akan tetapi, kata “*kumaras*” tidak memiliki kata akhiran sehingga dilakukan proses selanjutnya. Pada proses kedua, menghilangkan kata awalan “*kuma-*” sehingga kata menjadi “*ras*”. Kata “*ras*” dicek kembali dalam kamus. Tetapi, kata “*ras*” tidak terdapat dalam kamus sehingga kata dikembalikan lagi menjadi kata “*kumaras*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*kumaras*” di-*recoding* berdasarkan aturan pemenggalan j23 pada Tabel 2. Kata tetap menjadi kata “*kumaras*”. Karena terdapat kata awalan “*kuma-*” yang diproses terlebih dahulu, sebelum kata sisipan “*kum*”. Sehingga kata tidak dapat dilakukan proses *stemming* menjadi kata “*waras*”.

Kesalahan kedelapan, pada kata “*katawakake*” dicek terlebih dahulu di dalam kamus. Akan tetapi, kata “*katawakake*” tidak terdapat dalam kamus, maka dilakukan proses menghilangkan kata. Proses pertama, menghilangkan kata akhiran “*-ake*” sehingga kata menjadi “*katawak*”. Kata “*katawak*” dicek kembali dalam kamus. Namun, kata “*katawak*” tidak terdapat dalam kamus, maka dilakukan proses selanjutnya. Pada proses kedua, menghilangkan kata awalan “*ka-*” sehingga kata menjadi “*tawak*”. Kata “*tawak*” dicek kembali dalam kamus. Tetapi, kata “*tawak*” tidak terdapat dalam kamus sehingga kata dikembalikan lagi menjadi kata “*katawakake*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*katawakake*” di-*recoding* berdasarkan aturan pemenggalan j7 pada Tabel 2. Setelah, di-*recoding* berdasarkan aturan pemenggalan j7 kata “*katawakake*” tetap tidak dapat dilakukan proses *stemming*. Karena tidak terdapat kata akhiran “*kake-*”.

Kesalahan kesembilan, pada kata “*pangenan*” kata dicek terlebih dahulu di dalam kamus. Tetapi, kata “*pangenan*” tidak terdapat dalam kamus, maka dilakukan proses menghilangkan kata. Proses pertama, menghilangkan kata akhiran “*-an*” sehingga kata

menjadi “*pangen*”. Kata “*pangen*” dicek kembali dalam kamus. Namun, kata “*pangen*” tidak terdapat dalam kamus sehingga dilakukan proses selanjutnya. Pada proses kedua, menghilangkan kata awalan “*pang-*” sehingga kata menjadi “*en*”. Kata “*en*” dicek kembali dalam kamus. Namun, kata “*en*” tidak terdapat dalam kamus sehingga kata dikembalikan lagi menjadi kata “*pangenan*” dan dilakukan proses selanjutnya. Proses ketiga, kata “*pangenan*” di-*recoding* berdasarkan aturan pemenggalan j10 yang ditunjukkan pada Tabel 2. Namun kata tetap tidak dapat dilakukan proses *stemming* menjadi kata “*angen*”. Karena terdapat kata awalan “*pang-*” dan kata awalan “*pa-*” yang diproses terlebih dahulu, sebelum dilakukan proses *recoding* berdasarkan aturan j10.

Berdasarkan kesalahan satu sampai sembilan, ada beberapa kesalahan seperti : kata yang seharusnya tidak dihilangkan menjadi hilang, kata sisipan yang terdapat dalam aturan masih harus diperbaiki karena terdapat kata awalan yang sama dengan aturan kata sisipan.

Berdasarkan Tabel 3, ECS dapat memperbaiki kesalahan yang awalnya 98 kata dalam penelitian sebelumnya sehingga menjadi 9 kata. Perbaikan yang dilakukan oleh ECS sebanyak 89 kata pada kata awalan, sisipan, akhiran dan gabungan awalan-akhiran. Sehingga akurasi yang diperoleh sebagai berikut :

$$Accuracy = \frac{440}{449} \times 100\% = 97,9\%$$

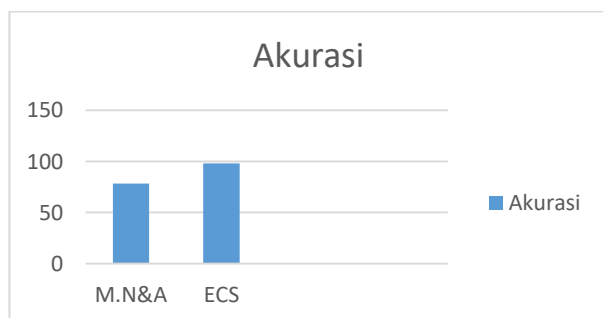
Akurasi yang diperoleh oleh ECS sebanyak 97,9%. Maka ECS dapat meningkatkan performa modifikasi Nazief & Adriani yang hanya memiliki akurasi 78,2%. Peningkatan yang dilakukan oleh ECS sebanyak 19,7% terhadap modifikasi Nazief & Adriani untuk *stemming* kata imbuhan Bahasa Jawa.

Berdasarkan akurasi yang didapatkan dari ECS, maka untuk menghitung *error rate* sebagai berikut :

$$Error Rate = 100\% - 97,9\% = 2,1\%$$

ECS juga dapat memperkecil kesalahan yang awalnya 21,8% menjadi 2,1%.

Pada Gambar 3. Menunjukkan grafik peningkatan ECS dalam modifikasi Nazief & Adriani untuk *stemming* Bahasa Jawa.



Gambar 3. Grafik Akurasi

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan [5] bahwa ECS dapat memperbaiki kesalahan yang dilakukan oleh modifikasi Nazief & Adriani yang awalnya 98 kata salah dapat diperbaiki hingga menjadi 9 kata salah. ECS juga dapat meningkatkan performa dari modifikasi Nazief & Adriani yang awalnya memiliki akurasi 78,2% menjadi 97,9%. Peningkatan performa yang dilakukan oleh ECS sebanyak 19,7% dari modifikasi Nazief & Adriani. ECS juga dapat memperkecil kesalahan yang awalnya 21,8% menjadi 2,1%. Namun ECS masih memiliki beberapa permasalahan pada kata “ngetan, kumanggih, kumarut, kumasis, kumareg, kumadul, kumaras, katawakake, dan pangenan”.

Saran untuk penelitian selanjutnya diharapkan dapat memperbaiki kesalahan-kesalahan tersebut dengan metode lain atau memodifikasi aturan-aturan yang sebelumnya agar kata-kata tersebut dapat dilakukan proses *stemming*.

Ucapan Terimakasih

Terimakasih kepada Rizal Kholif Nurrohman yang telah memberikan referensi artikel terkait yang berjudul *Stemming* Kata Imbuhan Bahasa Jawa dengan Menggunakan Modifikasi Nazief & Adriani. Namun artikel yang terkait masih dalam proses review dalam jurnal lain.

Daftar Rujukan

- [1] I. B. Arif Antono, Ida Zulaeha, “Pemertahanan Fonologis dan Leksikal Bahasa Jawa di Kabupaten Wonogiri,” *J. Sastra Indones.*, vol. 8, no. 1, pp. 47–56, 2019.
- [2] P. F. Ariyani, A. Rahmala, and N. Juliasari, “Implementasi Metode Stemming Tala dan Fungsi Jaccard pada Aplikasi Katalog Perpustakaan,” *Semin. Nas. Inov. dan Apl. Teknol. di Ind. 2019*, vol. 5, pp. 128–133, 2019.
- [3] M. N. Kassim, S. Hisham, M. Jali, and M. A. Maarof, “Towards Stemming Error Reduction for Malay Texts,” pp. 13–23, 2019.
- [4] D. Sa, W. B. Zulfikar, C. Slamet, M. A. Ramdhani, and Y. A. Gerhana, “An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media,” no. Citsm, 2018.
- [5] D. F. H. P. Amalia Sahira Ragma, Vit Zuraida, “Penggunaan Dictionary-Based dan Corpus-Based Thesaurus untuk Pembobotan Term pada Pengelompokan Dokumen Berita Berbahasa Indonesia,” vol. 2, no. 1, 2017.
- [6] D. Farrar and J. H. Hayes, “A Comparison of Stemming Techniques in Tracing,” pp. 1–8, 2019.
- [7] A. S. Rizki, A. Tjahyanto, and R. Trialih, “Comparison of Stemming Algorithms on Indonesian Text Processing,” vol. 17, no. 1, pp. 95–103, 2019.
- [8] N. Husni Mubarak, “Analisis Morfologi pada Bahasa Mandar dalam Ruang Lingkup Keluarga di Desa Tanjung Lalak Kecamatan Pulau Laut Kepulauan Kabupaten Kotabaru,” vol. 6, no. 2, pp. 63–79, 2018.
- [9] T. Winarti and S. Arief, “Determining Term on Text Document Clustering using Algorithm of Enhanced Confix Stripping Stemming,” vol. 157, no. 9, pp. 8–13, 2017.
- [10] Y. D. Pramudita, S. S. Putro, N. Makhmud, B. Olahraga, E. Confix, and S. Stemmer, “Klasifikasi Berita Olahraga Menggunakan Metode Naive Bayes dengan Enhanced Confix Stripping Stemmer,” vol. 5, no. 3, 2018.
- [11] M. N. Khidfi and J. Y. Sari, “Rancang Bangun Aplikasi Pendeteksian Kesamaan pada Dokumen Teks Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma Winnowing,” no. September, 2018.
- [12] K. N. Sistem, A. Ridok, R. Latifah, F. Unibraw, and N. Neighbor, “Klasifikasi Teks Bahasa Indonesia pada Corpus Tak Seimbang Menggunakan NWKNN,” pp. 9–10, 2015.
- [13] T. Yusnitasari *et al.*, “Uji Coba Stemming ECS (Enhance Confix Stripping) Ayat- Ayat Al Qur ’ an Dan Hadist Terjemahan Bahasa Indonesia,” pp. 24–26, 2018.
- [14] T. Yusnitasari, L. Wulandari, D. Iksari, I. Humaini, K. K. Informasi, and I. Retrieval, “Perancangan Smart Digital Al Quran dan Hadist Bukhori Muslim untuk Platform Mobile Application,” no. September, 2018.
- [15] Y. N. Fadziyah and E. F. R., “Penerapan Algoritma Enhanced Confix Stripping dalam Pengukuran Keterbacaan Teks Menggunakan Gunning Fog Index,” vol. 1, no. 1, pp. 15–24, 2018.
- [16] P. D. & T. D. Andita, “Implementasi Modifikasi Enhanced Confix Stripping Stemmer untuk Bahasa Indonesia dengan Metode Corpus Based Stemming,” pp. 1–15, 2012.