



## Stemming Teks Bahasa Bali dengan Algoritma Enhanced Confix Stripping

Ni Wayan Wardani<sup>1\*</sup>, Putu Gede Surya Cipta Nugraha<sup>2</sup> 

<sup>1,2</sup>STMIK STIKOM Indonesia, Denpasar, Indonesia

\*Corresponding author: [niwayan.wardani@stiki-indonesia.ac.id](mailto:niwayan.wardani@stiki-indonesia.ac.id)

### Abstract

Masih banyak yang mengalami permasalahan saat melakukan stemming dimana belum mampu melakukan stemming dengan tepat pada beberapa kata untuk aturan peluluhan prefix P3, P4, P5, P10, P11, dan P12. Tujuan penelitian ini adalah untuk mengkaji efektivitas algoritma Enhanced Confix Stripping Stemmer (ECS) terhadap stemming Bahasa Bali. Data yang digunakan dalam penelitian ini adalah 376 akar kata dalam bahasa Bali yang terdiri dari 240 kata yang mengandung prefiks, 17 akar kata yang mengandung infiks dan 199 akar kata yang mengandung sufiks. Hasil penelitian ini menunjukkan bahwa Enhanced Confix Stripping dapat meningkatkan performansi yang sebelumnya memiliki akurasi, dari hanya 77,82% menjadi 96,94% dengan tingkat kesalahan 3,06% dan memperbaiki kesalahan yang semula berjumlah 120 sampai 20 kesalahan. Berdasarkan hasil penelitian, dapat ditarik simpulan bahwa algoritma ECS Stemmer dapat memperbaiki kesalahan yang dilakukan oleh metoda Rule Based Approach

**Keywords:** Bahasa Bali; Stemming; Enhanced Confix Stripping Stemmer

### Abstract

There are still many who experience problems when doing stemming where they are not able to properly stemming a few words for the prefix kneading rules P3, P4, P5, P10, P11, and P12. The purpose of this study was to examine the effectiveness of the Enhanced Confix Stripping Stemmer (ECS) algorithm on Balinese stemming. The data used in this study were 376 root words in Balinese consisting of 240 words containing prefixes, 17 root words containing infixes and 199 root words containing suffixes. The results of this study indicate that Enhanced Confix Stripping can improve performance which previously had accuracy, from only 77.82% to 96.94% with an error rate of 3.06% and correcting the original 120 to 20 errors. Based on the research results, it can be concluded that the ECS Stemmer algorithm can correct errors made by the Rule Based Approach method.

**Keywords:** Bahasa Bali; Stemming; Enhanced Confix Stripping Stemmer

#### History:

Received: 24 August 2020

Revised: 1 September 2020

Accepted: 26 September 2020

Published: 7 October 2020

**Publisher:** Undiksha Press

**Licensed:** This work is licensed under a Creative Commons Attribution 3.0 License



## Introduction

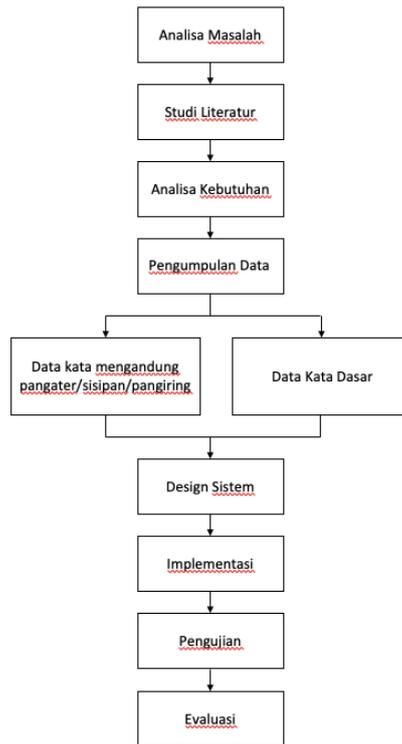
Bahasa daerah merupakan bagian dari kebudayaan yang hidup dan berkembang sesuai dengan kebutuhan masyarakat tersebut (Fallis, 2013). Bahasa daerah merupakan bahasa pendukung bahasa Indonesia yang keberadaannya diakui oleh negara (Rahmat & Mansyur, 2020; Riani, 2012). Salah satu Bahasa daerah yang ada di Indonesia adalah Bahasa Bali. Bahasa Bali merupakan bahasa ibu bagi masyarakat etnis Bali memiliki kedudukan dan fungsi yang sangat penting (Agus et al., 2019). Bahasa Bali merupakan bahasa Austronesia dari cabang sundik dan lebih spesifik dari anak cabang Bali-Sasak (Nata & Yudiastra, 2017).

Bahasa Bali memiliki tiga tingkat penggunaannya yaitu basa bali alus, basa bali madya dan basa bali kasar. Tata Bahasa Bali sendiri meliputi kata dasar, pangater (awalan), seselan (sisipan) dan pangiring (akhiran) dalam sebuah kalimat (Gede et al., 2020; Wirayasa et al., 2019). Pangater yaitu imbuhan yang diletakkan di awal kata dasar, pangater terdiri dari : (a-, ma-, su-, ka-, pa-, pati-, pari-, maka-, saka-, kuma-, sa-, pa-, pi-, dur-, swa-) (Bawa & Jendra, 1981). Adapun contoh katanya yaitu: aukud, majalan, sudharma, kaejuk, pajalan, patigrepe, makadadua, sakabesik, kumalipan. Seselan yaitu penyisipan imbuhan di dalam kata dasar, dalam Bahasa Bali disebut dengan seselan terdiri dari : ( -in-, -um-, -el-, -er- ) (Bawa & Jendra, 1981). Adapun contoh katanya: sinurat, sumaur, telapak, gerigi. Pangiring yaitu imbuhan yang diletakkan di belakang kata dasar, dalam Bahasa Bali disebut dengan pangiring terdiri dari : ( -a, -e, -ne, -ang-,ing, -in, -an, -n, -wan, -nyane ) (Bawa & Jendra, 1981). Keunikan Bahasa Bali menjadikan perlu adanya proses *stemming*. *Stemming* adalah teknik untuk mencari bentuk kata dasar dari suatu kata dengan menghilangkan kata awalan, sisipan, akhiran, dan gabungan kata tersebut (Amin & Razaq, 2018; Guterres & Santoso, 2019; Maulidi, 2016; Simarangkir, 2017). Selain itu *stemming* banyak digunakan dalam bidang *information retrieval*. 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 (Hidayatullah et al., 2019).

Penelitian sebelumnya mengenai *stemming* Bahasa Bali menggunakan metoda *Rule Based Approach* menghasilkan rata-rata tingkat akurasi yang diperoleh aplikasi *stemming* Bahasa Bali sebesar 77.82%, dimana tingkat akurasi pangater 66,25% (uji coba 240 kata), tingkat akurasi seselan 100% (uji coba 17 kata) dan tingkat akurasi pangiring 67.22% ( uji coba 119 kata). Pada metoda *Rule Based Approach* masih banyak mengalami permasalahan saat melakukan *stemming* dimana belum mampu melakukan *stemming* dengan tepat pada beberapa kata untuk aturan peluluhan prefix P3, P4, P5, P10, P11, dan P12 seperti yang tercantum pada tabel 2 . Oleh karena itu perlu adanya perbaikan *stemming* dengan menggunakan algoritma *Enhanced Confix Stripping Stemmer* (ECS). Algoritma ECS Stemmer ini merupakan algoritma perbaikan dari algoritma *Confix Stripping* (CS) Stemmer. Perbaikan yang dilakukan oleh ECS Stemmer adalah perbaikan beberapa aturan pada tabel acuan pemenggalan imbuhan (Tahitoe & Purwitasari, 2010). ECS Stemmer adalah algoritma perbaikan dari *Confix Stripping* (CS) yang digunakan untuk metoda *stemming* Bahasa Indonesia (Fadziah & R, 2018). Selain diterapkan pada Bahasa Indonesia, Algoritma ECS Stemmer juga diterapkan pada *stemming* bahasa Jawa yaitu dimana algoritma ECS Stemmer dapat melakukan perbaikan dari algoritma Nazief & Adriani dengan tingkat akurasi 97,9% (Hidayatullah et al., 2019). Berdasarkan jabaran tersebut dapat dirumuskan tujuan penelitian yaitu mengkaji efektivitas algoritma *Enhanced Confix Stripping Stemmer* (ECS) terhadap *stemming* Bahasa Bali.

## Materials and Methods

Penelitian ini diselesaikan dengan melalui tahapan-tahapan yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram ikhtisar penelitian

**a. Analisa Masalah**

Analisa masalah dari penelitian ini adalah berdasarkan pada penelitian *stemming* bahasa Bali sebelumnya yang menggunakan metoda *Rule Based Approach* yang masih banyak mengalami permasalahan dan bagaimana pada penelitian ini dilakukan pengembangan atau memodifikasi tabel aturan pemenggalan dari algoritma sebelumnya yaitu *Rule Based Approach* dengan algoritma *ECS Stemmer*. *Rule Based Approach* adalah metoda yang meluluhkan prefiks, infiks dan sufiks dan sangat bergantung pada penalaran manusia sebagai *expert* dalam memecahkan masalah (Agus et al., 2019). Hasil pengujian dengan metoda *Rule Based Approach* ditunjukkan pada tabel 1.

Tabel 1. Hasil pengujian dengan metoda *Rule Based Approach*

Imbuhan	Jumlah Kata Dasar	Uji Stemming (Benar)	Uji Stemming (Salah)	Akurasi
Pengater	240	159	81	66,25%
Seselan	17	17	0	100%
Pangiring	119	80	39	67,22%

**b. Studi Literatur**

Pada tahapan ini peneliti melakukan tinjauan pustaka yaitu mengumpulkan dan mempelajari literatur, buku, artikel, jurnal dan sebagainya yang diperoleh dari berbagai sumber seperti perpustakaan, internet, dll.

**c. Analisa Kebutuhan**

Pada proses analisa kebutuhan dilakukan pengumpulan kebutuhan yang diperlukan untuk memperkuat data penelitian yang selanjutnya dilakukan pemenuhan semua kebutuhan sebagai dasar untuk proses pengumpulan data. Pada analisa kebutuhan ini dilakukan

pengumpulan kebutuhan mengenai aturan pemenggalan prefiks pada Bahasa Bali yaitu pada Tabel 2.

**Tabel 2.** Aturan Peluluhan Prefiks Bahasa Bali

Aturan	Format	Peluluhan	Contoh
P1	ngV...	ng-V...	Ngidih : ng-Idih
P2	ngC...	ngC...	Ngwangun : ng-Wangun
P3	nV...	n-tV...	Negul : n-Tegul Nundun : n-Dundun
P4	nyA...	ny-CA... ny-JA... ny-SA...	Nyacad : ny-Cacad Nyaring : ny-Jaring Nyampat : ny-Sampat
P5	ngV...	ng-V...	Ngutang : ng-Kutang Ngambar : ng-Gambar
P6	ngaC...	nga-CV...	Ngamaling : nga-Maling Nganengneng : nga-Nengneng
P7	maC...	paC... baC...	Mapag : Papag Matek : Batek
P8	maC...	ma-CV...	Makesiab : ma-Kesiab
P9	maA...	ma-AV...	Mayasa : ma-Yasa
P10	mV...	ma-VA...	Mikut : ma-Ikut Mubad : ma-Ubad
P11	pV...	pa-V...	Pileh : pa-Ileh
P12	kV...	ka-V...	Kicen : ka-Icen
P13	saC...	sa-CV...	Sajagat : sa-Jagat
P14	saV...	sa-V...	Sausan : sa-Usan
P15	aC...	a-C...	Adiri : a-Diri
P16	aV...	a-V...	Aukud : a-Ukud
P17	praC...	pra-C...	Prajani : pra-Jani
P18	pariC...	pari-C...	Paribasa : pari-Basa
P19	pariV...	pari-V...	Pariindik : pari-Indik
P20	patiC...	pati-C...	Patigrabe : pati-Grape
P21	makaC...	maka-C...	Makasami : maka-Sami
P22	makaV...	maka-V...	Makaukud : maka-ukud
P23	sakaC...	saka-C...	Sakabesik : saka-Besik
P24	sakaV...	saka-V...	Sakaukud : saka-Ukud
P25	kumaC...	kuma-C...	Kumajaum : kuma-Jaum
P26	suC...	su-C...	Sudharma : su-Dharma
P27	swaC...	swa-C...	Swadharma : swa-Dharma
P28	durC...	dur-C...	Durlaba : dur-Laba

Contoh kasus yang tidak dapat diselesaikan dengan metoda *Rule Based Approach* ditunjukkan pada Tabel 3.

**Tabel 3.** Morfologi yang tidak dapat diselesaikan

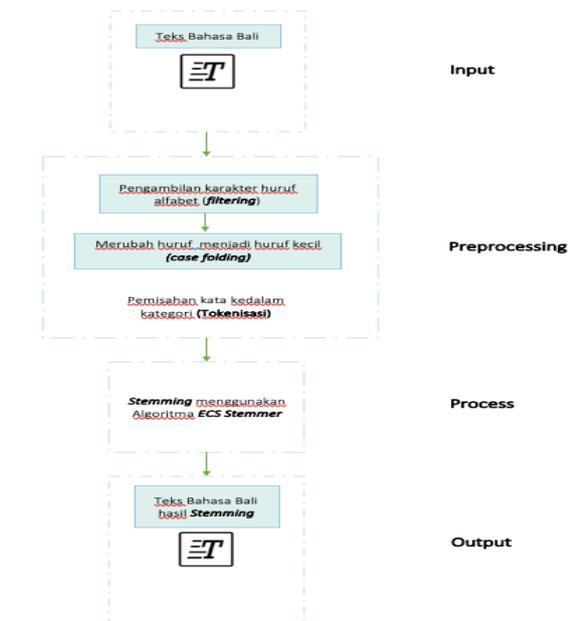
Aturan	Format	Peluluhan	Contoh
P3	nV...	n-tV...	Negul : n-Tegul Nundun : n-Dundun
P4	nyA...	ny-CA...	Nyacad : ny-Cacad

Aturan	Format	Peluluhan	Contoh
P5	ngV...	ny-JA...	Nyaring : ny-Jaring
		ny-SA...	Nyambat : ny-Sambat
		ng-V...	Ngutang : ng-Kutang
P10	mV...	ma-VA...	Ngambar : ng-Gambar
			Mikut : ma-Ikut
P11	pV...	pa-V...	Mubad : ma-Ubad
P12	kV...	ka-V...	Pileh : pa-Ileh
			Kicen : ka-Icen

#### d. Pengumpulan Data

Data kata yang digunakan pada penelitian ini adalah data kata yang digunakan pada penelitian sebelumnya mengenai *stemming* Bahasa Bali menggunakan metoda *Rule Based Approach* yaitu data kata Bahasa Bali berjumlah 376 kata yang terdiri dari 240 kata dasar yang mengandung pangater, 17 kata dasar yang mengandung seselan dan 119 kata dasar yang mengandung pangiring. Selain itu dataset yang akan digunakan dalam proses *stemming* adalah data kata dasar Bahasa Bali berjumlah 6.124 kata dasar. Pada penelitian sebelumnya, dilakukan pengujian *stemming* teks kata dasar pada pangater, seselan dan pangiring menggunakan aplikasi *stemming* kata dasar Bahasa Bali, dilihat hasil pengujian menghasilkan uji *stemming* benar pada pangater sebanyak 159 kata dan uji *stemming* salah pada pangater sebanyak 81 kata. Uji *stemming* benar pada seselan sebanyak 17 kata dan uji *stemming* salah pada seselan sebanyak 0 kata. Uji *stemming* benar pada pangiring sebanyak 80 kata dan uji *stemming* salah pada pangiring sebanyak 39 kata.

#### e. Design Sistem



Gambar 2. Design Sistem

Arsitektur umum pada sistem pembentukan akar kata dalam dokumen teks Bahasa Bali memiliki beberapa tahapan. Adapun tahapan yang ada dalam sistem sebagai berikut :

1. *Input*

Pada tahap pertama sistem adalah input dokumen teks Bahasa Bali.

## 2. *Preprocessing*

Pada tahap *preprocessing* dilakukan proses perubahan bentuk teks yang belum terstruktur menjadi teks yang terstruktur, yang mana teks tersebut disesuaikan dengan kebutuhan aplikasi. Tahap yang dilakukan, yaitu :

### a. *Filtering*

Pada proses *filtering* dilakukan pengambilan karakter alfabet dari dokumen teks Bahasa Bali, selanjutnya pada proses ini, dilakukan penghapusan karakter, tanda baca dan angka.

### b. *Case Folding*

Pada proses *case folding* dilakukan perubahan untuk semua huruf dalam dokumen teks Bahasa Bali menjadi huruf kecil. Huruf yang dapat diterima yaitu huruf 'a' sampai dengan huruf 'z'.

### c. Tokenisasi

Proses ini bertujuan untuk memisahkan setiap kata yang menyusun dokumen menjadi kata per kata dengan menandai karakter spasi. Tiap kata hasil tokenisasi akan disimpan ke dalam katagori.

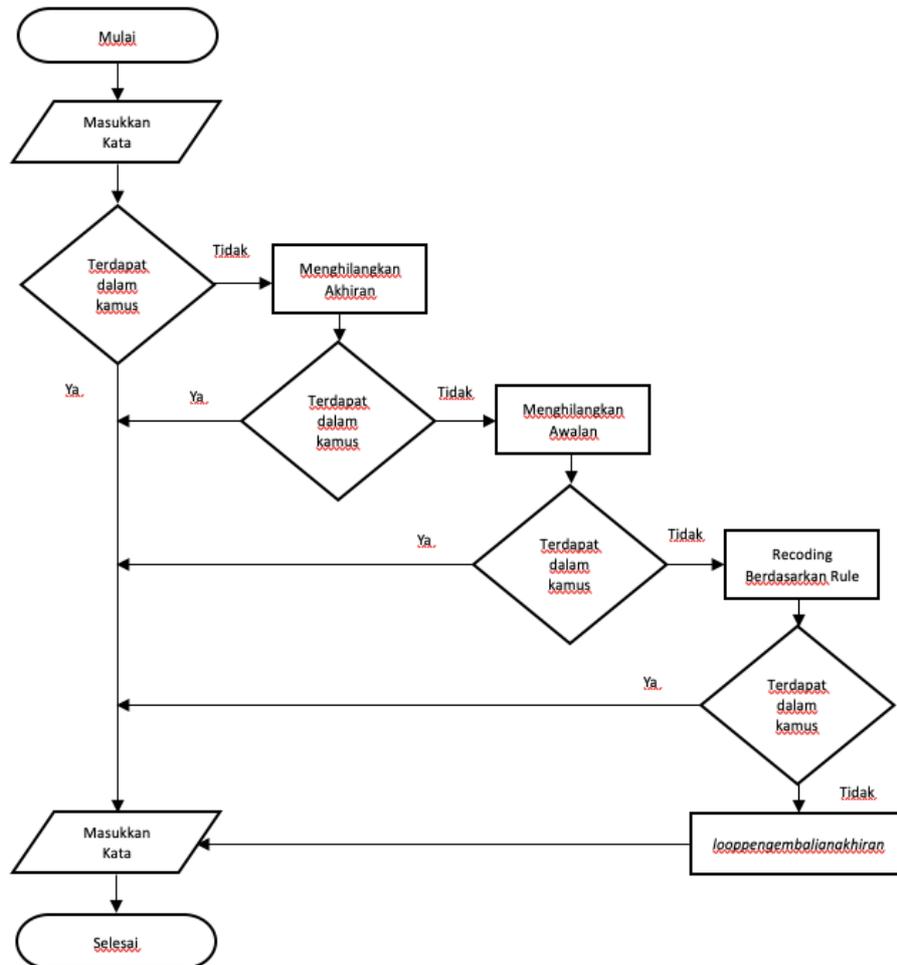
## 3. *Process*

Algoritma ECS merupakan algoritma yang dikembangkan oleh (Arifin & Ciptaningtyas, 2009) dalam penelitian yang berjudul "*Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying New Document in Indonesian Language*", dimana merupakan algoritma perbaikan dari *Confix Stripping* (CS) untuk Bahasa Indonesia. 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. ECS akan dicoba dengan aturan baru untuk Bahasa Bali dengan memperbaiki kesalahan metoda *Rule Based Approach* untuk *stemming* kata imbuhan Bahasa Bali.

Proses *stemming* dari *Enhanced Confix Stripping* (ECS) ditunjukkan pada Gambar 2, yaitu :

- a. Memasukkan kata sebagai proses awal.
- b. Kamus memeriksa kata masukan. Jika kata masukan terdapat di dalam kamus maka dianggap sebagai kata dasar. Jika tidak, maka dilakukan proses menghilangkan kata akhiran
- c. Proses ketiga adalah menghilangkan kata akhiran pada kata masukan. Jika hasil proses ketiga menghasilkan kata yang terdapat di dalam kamus, maka dianggap sebagai kata dasar. Jika tidak, maka dilakukan proses menghilangkan kata awalan.
- d. Proses keempat adalah menghilangkan kata awalan. Jika hasil proses keempat menghasilkan kata yang terdapat di dalam kamus, maka dianggap sebagai kata dasar. Namun, jika kata tidak terdapat dalam kamus, maka dilakukan proses *recoding*.
- e. Apabila proses *recoding* kata terdapat dalam kamus, maka proses dihentikan. Proses *recoding* merupakan proses pemenggalan kata atau penyusunan kembali kata-kata yang mengalami proses *stemming* yang berlebih (Pramudita et al., 2018).
- f. Jika masih gagal dilanjutkan dengan proses pengembalian akhiran atau loop pengembalian akhiran.
- g. Jika masih tetap gagal kata akan dikembalikan ke bentuk semula dan dianggap sebagai kata dasar.

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



Gambar 3. Enhanced Confix Stripping (ECS) Flowchart

4. Output

Pada bagian pengujian, akan menggunakan semua dataset yang ada dan dilakukan perbaikan pemenggalan kata yang salah dalam algoritma sebelumnya. Untuk evaluasi, menggunakan akurasi dari ECS *Stemmer* dengan menggunakan rumus sebagai berikut :

$$Accuracy = \frac{Jumlah\ kata\ benar}{Jumlah\ kata\ seluruhnya} \times 100\%$$

Selain itu, evaluasi juga diukur dari *error rate* dengan rumus sebagai berikut :

$$Error\ Rate = 100\% - Accuracy$$

## f. Pengujian

### Pengujian Prefix (Pangater)

Pada tahap ini dilakukan pengujian prefix (pangater) dari teks Bahasa Bali yang telah di masukan. Sebagai contoh “mabaju” bentuk dasarnya “baju”. Pengujian dilakukan dengan menggunakan teks kata dasar pangater sebanyak 240 kata dasar.

**Tabel 4.** Pengujian Pengater

Pengater	Teks	Hasil Stemming
a	abungkul	bungkul
ma	mabaju	baju
ka	kajagur	jagur
su	subhakti	bhakti
pa	pangayah	ngayah
pati	patikaplug	kaplug
pari	parisolah	solah
maka	makatelu	telu
saka	sakabesik	besik
kuma	kumajaum	jaum
sa	sadurung	durung
pi	pitutur	tutur
dur	durlaba	laba
swa	swabawa	bawa
ng	ngalap	alap

### Pengujian Sufiks (Pangiring)

Pada tahap ini dilakukan pengujian sufiks (pangiring) dari teks Bahasa Bali yang telah di inputkan. Sebagai contoh “jemakin” bentuk dasarnya “jemak”. Pengujian dilakukan dengan menggunakan teks kata dasar pangiring sebanyak 119 kata dasar.

**Tabel 5.** Pengujian Pangiring

Pangiring	Teks	Hasil Stemming
- a	jemaka	jemak
- e	carike	carik
- ne	bapanne	bapan
- ang	dingehang	dingeh
- in	negakin	negak
- an	cenikan	cenik
- n	mejan	meja

### Pengujian Infiks (Seselan)

Pada tahap ini dilakukan pengujian infiks (seselan) dari teks Bahasa Bali yang telah di inputkan. Sebagai contoh “sinurat” bentuk dasarnya “surat”. Pengujian dilakukan dengan menggunakan teks kata dasar seselan sebanyak 17 kata dasar.

**Tabel 6.** Pengujian Seselan

Sisipan	Teks	Hasil Stemming
-in-	sinurat	surat
-um-	rumaksa	maksa
-el-	telapak	tapak
-er-	keresek	kesek

### Results and Discussion

Berdasarkan kelemahan metoda *Rule Based Approach* untuk stemming kata imbuhan Bahasa Bali, maka perbaikan yang dilakukan algoritma ECS *Stemmer* ditunjukkan pada Tabel 7 berikut.

**Tabel 7.** Hasil pengujian dengan algoritma ECS *Stemmer*

Imbuhan	Jumlah Kata Dasar	Uji Stemming (Benar)	Uji Stemming (Salah)	Akurasi
Pengater	240	229	11	95,41%
Seselan	17	17	0	100%
Pangiring	119	110	9	92,43%

Berdasarkan Tabel 7, jumlah kata yang salah adalah 20 kata. Dari pengujian, didapatkan hasil bahwa Algoritma ECS *Stemmer* dapat melakukan *stemming* pada kata yang memiliki aturan peluluhan P3, P4, P5, P10, P11 dan P12 walaupun tidak semua kata dengan aturan peluluhan tersebut dapat di *stemming* dengan tepat. Contoh kata yang tidak dapat di *stemming* dengan tepat adalah kata "kajanang". Pada kata "kajanang" dicek terlebih dahulu di dalam kamus. Namun, kata "kajanang" tidak terdapat dalam kamus, maka dilakukan proses menghilangkan akhiran "-ang" sehingga kata menjadi kata "kajan". Kata "kajan" dicek kembali dalam kamus. Tetapi, kata "kajan" tidak terdapat dalam kamus. Proses kedua menghilangkan awalan "ka-", sehingga kata menjadi "janang". Kata "janang" dicek kembali dalam kamus, namun kata "janang" tidak terdapat dalam kamus, maka kata awalan "ka" dikembalikan lagi menjadi kata "kajanang" dan dilakukan proses selanjutnya. Proses ketiga, kata "kajanang" dilakukan recoding menggunakan aturan kaja-n-ang. Namun kata tetap menjadi "kajanang", sehingga proses *stemming* gagal menjadi kata "kaja". Contoh berikut adalah kata "nundun". Pada kata "nundun" dicek terlebih dahulu di dalam kamus. Namun, kata "nundun" tidak terdapat dalam kamus, maka dilakukan proses menghilangkan akhiran, tetapi kata "nundun" tidak terdapat kata akhiran, maka dilanjutkan dengan proses menghilangkan awalan. Kata "nundun" tidak terdapat awalan, maka dilakukan proses recoding n-Dundun. Namun kata tetap menjadi "nundun".

Jadi, algoritma ECS *Stemmer* dapat memperbaiki kesalahan yang dilakukan oleh metoda *Rule Based Approach* yang awalnya 120 kata salah dapat diperbaiki hingga menjadi 20 kata salah. ECS *Stemmer* juga dapat meningkatkan performa dari metoda *Rule Based Approach* yang awalnya memiliki akurasi 77,82% menjadi 96,94%. Peningkatan performa yang dilakukan oleh ECS *Stemmer* sebanyak 19,12% dari metoda *Rule Based Approach*. ECS *Stemmer* juga dapat memperkecil kesalahan yang awalnya 22,18% menjadi 3,06%. Namun ECS *Stemmer* belum mampu secara sempurna melakukan *stemming* pada beberapa kata yang memiliki aturan peluluhan prefix P3, P4, P5, P10, P11 dan P12. Hasil penelitian ini didukung oleh penelitian yang dilakukan oleh Rizki et al., (2019) menunjukkan bahwa menunjukkan bahwa algoritma Enhanced confix stripping (ECS) adalah yang terbaik untuk keperluan

pemrosesan teks bahasa Indonesia dengan nilai skor terbobot 0,648. Penelitian yang dilakukan Tahitoe & Purwitasari, (2010) menunjukkan uji coba yang telah dilakukan dan setelah menganalisis hasil pengujian terhadap implementasi metode corpus based stemming untuk memperbaiki kesalahan stemming dari algoritma ECS Stemmer ini dapat diambil beberapa kesimpulan antara lain : a) Perbaikan yang dilakukan dapat memperbaiki seluruh kesalahan stemming yang dilakukan oleh algoritma ECS Stemmer. b) Untuk kesalahan overstemming dan understemming, di mana hasil stemming dari suatu term dapat berjumlah lebih dari satu, penggunaan metode corpus based stemming dapat digunakan untuk memilih hasil stemming yang tepat berdasarkan koleksi dokumen yang digunakan.

## Conclusion

Berdasarkan hasil penelitian yang telah dilakukan bahwa algoritma ECS Stemmer dapat memperbaiki kesalahan yang dilakukan oleh metoda *Rule Based Approach* yang awalnya 120 kata salah dapat diperbaiki hingga menjadi 20 kata salah. ECS Stemmer juga dapat meningkatkan performa dari metoda *Rule Based Approach* yang awalnya memiliki akurasi 77,82% menjadi 96,94%. Peningkatan performa yang dilakukan oleh ECS Stemmer sebanyak 19,12% dari metoda *Rule Based Approach*. ECS Stemmer juga dapat memperkecil kesalahan yang awalnya 22,18% menjadi 3,06%. Namun ECS Stemmer belum mampu secara sempurna melakukan *stemming* pada beberapa kata yang memiliki aturan peluluhan prefix P3, P4, P5, P10, P11 dan P12.

## References

- Agus, M., Subali, P., & Fatichah, C. (2019). Kombinasi Metode Rule-Based Dan N-Gram Stemming Untuk a Combination of Methods Rule-Based and N-Gram Stemming To Recognize Balinese Language Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 6(2). <https://doi.org/10.25126/jtik.201961105>
- Amin, F., & Razaq, J. A. (2018). *Implementasi Stemmer Bahasa Jawa dengan Metode Rule Base Approach pada Sistem Temu Kembali Informasi Dokumen Teks Berbahasa Jawa*. 978–979.
- Arifin, A. Z., & Ciptaningtyas, H. T. (2009). *Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language*.
- Bawa, I. W., & Jendra, I. W. (1981). *Struktur Bahasa Bali*. Pusat Pembinaan dan Pengembangan Bahasa.
- Fadziah, Y. N., & R, E. F. (2018). *Penerapan Algoritma Enhanced Confix Stripping dalam Pengukuran Keterbacaan Teks Menggunakan Gunning Fog Index*. 1(1), 15–24.
- Fallis, A. . (2013). Peran Bahasa Indonesia dan Bahasa Daerah Dalam Pendidikan Karakter. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Gede, P., Cipta, S., & Wardani, N. W. (2020). Stemming Dokumen Teks Bahasa Bali Dengan Metode Rule Base Approach. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(3), 510–521. <https://doi.org/https://doi.org/10.35957/jatisi.v7i3.538>
- Guterres, A., & Santoso, J. (2019). *Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based*. 8(November), 142–147. <https://doi.org/10.34148/teknika.v8i2.224>
- Hidayatullah, N., Aji Prasetya Wibawa, & Harits Ar Rosyid. (2019). Penerapan ECS Stemmer untuk Modifikasi Nazief & Adriani Berbahasa Jawa. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 343–348. <https://doi.org/10.29207/resti.v3i3.994>

- Maulidi, R. (2016). *Stemmer untuk Bahasa Madura dengan Modifikasi Metode Enhanced Confix Stripping Stemmer*. December.
- Nata, G. N. M., & Yudiastra, P. P. (2017). Stemming teks sor-singgih Bahasa Bali. *E-Proceedings KNS&I STIKOM Bali*, 608–612. <http://knsi.stikom-bali.ac.id/index.php/e proceedings/article/view/111>
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 269. <https://doi.org/10.25126/jtiik.201853810>
- Rahmat, & Mansyur, U. (2020). Pengaruh Bahasa Daerah Terhadap Pola Komunikasi Mahasiswa Fakultas Sastra Universitas Muslim Indonesia. *Jurnal Pembelajaran Bahasa Dan Sastra Indonesia Berada*, 1(3), 156–160. <https://ojs.unm.ac.id/Indonesia/article/view/15189/pdf>
- Riani. (2012). Permasalahan dalam Perencanaan Bahasa pada Masyarakat Multikultur. *Madah*, 3(2), 133–140. <https://doi.org/https://doi.org/10.31503/madah.v32>.
- Rizki, A. S., Tjahyanto, A., & Trialih, R. (2019). Comparison of stemming algorithms on Indonesian text processing. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17(1), 95–102. <https://doi.org/10.12928/TELKOMNIKA.v17i1.10183>
- Simarangkir, M. S. H. (2017). Studi Perbandingan Algoritma - Algoritma Stemming Untuk Dokumen Teks Berbahasa Indonesia. *Jurnal Inkofar*, 1(1), 41–47.
- Tahitoe, A. D., & Purwitasari, D. (2010). Implementasi Modifikasi Enhanced Confix Stripping Stemmer Untuk Bahasa Indonesia Dengan Metode Corpus Based Stemming. *Jurnal Ilmiah*, 1–15. <http://digilib.its.ac.id/public/TTS-Undergraduate-14255-paperpdf.pdf>
- Wirayasa, I. P. M., Wirawan, I. M. A., & Pradnyana, I. M. A. (2019). Algoritma Bastal: Adaptasi Algoritma Nazief & Adriani untuk Stemming Teks Bahasa Bali. *JANAPATI*, 8, 60–69.