

PENERAPAN KAMUS DASAR PADA METODA *RULE BASE APPROACH* UNTUK MENGURANGI KESALAHAN *STEMMING* BAHASA BALI

Ni Wayan Wardani¹⁾ Putu Gede Surya Cipta Nugraha²⁾

Program Studi Teknik Informatika¹⁾²⁾

STMIK STIKOM Indonesia, Denpasar, Bali¹⁾²⁾

niwayan.wardani@stiki-indonesia.ac.id¹⁾ surya.cipta@stiki-indonesia.ac.id²⁾

ABSTRACT

The process of stemming Balinese text using the Rule Base Approach method produces a fairly good level of accuracy, reaching 77.82%, but if we refer to the results of the tests carried out with the wrong stemming test results, there are still many errors in the stemming process because of the Rule Base Approach method has a weakness if it is applied to words that have high additive complexity and the stemming application cannot, namely the rules, then an error will occur in the stemming process. In addition, the cause of errors also occurs due to overstemming. Overstemming is a problem in the stemming process where the root word from the stemming result lacks a syllable or letter which is considered an affix as a result of the stemming process. In this study, a dictionary was added to the Rule Base Approach method to reduce stemming errors and increase the accuracy of the results. The added dictionary will be used to check every word that has gone through the affix truncation process, then the cutting results are immediately matched with the dictionary containing the base word. The use of a basic dictionary in the Rule Base Approach method can increase the accuracy of stemming by 10.61%.

Keyword : Balinese language, Rule Base Approach, Stemming, Overstemming, Dictionary

ABSTRAK

Proses stemming teks Bahasa Bali dengan metoda *Rule Base Approach* menghasilkan tingkat akurasi yang cukup baik yaitu mencapai angka 77.82%, akan tetapi jika mengacu pada hasil pengujian tepatnya dengan hasil uji *stemming* salah, masih terdapat banyak kesalahan pada proses *stemming* hal ini di karenakan metoda *Rule Base Approach* memiliki kelemahan yaitu jika diterapkan pada kata yang memiliki kompleksitas imbuhan yang tinggi dan aplikasi *stemming* tidak dapat mengenali *rule* maka akan terjadi kesalahan pada proses *stemming*. Selain itu penyebab kesalahan juga terjadi karena *overstemming*. *Overstemming* adalah sebuah permasalahan pada proses *stemming* dimana kata dasar dari hasil *stemming* mengalami kekurangan suku kata atau huruf yang dianggap sebagai imbuhan akibat dari proses *stemming*. Pada penelitian ini diusulkan penambahan kamus pada metode *Rule Base Approach* untuk mengurangi kesalahan *stemming* dan meningkatkan ketepatan hasil. Kamus yang ditambahkan akan digunakan untuk pengecekan pada setiap kata yang telah melewati proses pemotongan imbuhan, kemudian hasil dari pemotongan langsung dicocokkan dengan kamus yang berisi kata dasar. Penggunaan kamus dasar pada metoda *Rule Base Approach* dapat meningkatkan ketepatan stemming sebesar 10.61%.

Kata Kunci : Bahasa Bali, Rule Base Approach, Stemming, Overstemming, Kamus Dasar

PENDAHULUAN

Bahasa Bali adalah bahasa yang mayoritas digunakan oleh penduduk pulau Bali, sebagian pulau Lombok dan di daerah Banyuwangi, Jawa Timur. Bahasa Bali memiliki tiga tingkatan sor-singgih yaitu Basa Kasar, Basa Madia dan Basa Alus dengan tata bahasa yang meliputi kata dasar, pangater (awalan), seselan (sisipan) dan pangiring (akhiran). Bahasa Bali masih

memegang peranan penting dalam penggunaannya seperti pada kegiatan keagamaan, adat istiadat maupun percakapan sehari – hari pada masyarakat Bali.

Saat ini banyak dokumen berbahasa Bali berbentuk digital yang bertebaran di internet baik dalam bentuk artikel, jurnal, dll. Kebutuhan informasi berbahasa Bali yang relevan tentunya sangat diperlukan. Untuk mendapatkan informasi yang relevan

tersebut, dibutuhkan metoda yang tepat, salah satunya adalah metoda *Information Retrieval*. *Information Retrieval* merupakan metoda mencari material (dokumen), meliputi proses mencari dan mengambil informasi berbasis pengetahuan dari kumpulan dokumen [1].

Dalam *Information Retrieval*, sebelum pencarian informasi dokumen teks dilakukan prapengolahan terlebih dulu yang dikenal dengan *stemming*, yaitu proses untuk mendapatkan kata dasar dengan menghapus afiks dari kata yang dicari [2]. Oleh karena itu diperlukan metoda yang tepat untuk melakukan prapengolahan dengan *stemming*.

Penelitian mengenai *stemming* Bahasa Bali telah dilakukan sebelumnya yaitu dengan menggunakan metoda *Rule Base Approach* yang menghasilkan rata-rata tingkat akurasi 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) [3]. Pada metoda *Rule Based Approach* masih banyak mengalami permasalahan saat melakukan *stemming* dimana belum mampu melakukan *stemming* dengan tepat pada beberapa kata. Salah satu penelitian mengenai perbaikan ketepatan *stemming* adalah dengan menerapkan kamus kata dasar [4]. Penelitian tersebut menerapkan kamus kata dasar untuk memperbaiki proses *stemming* dari algoritma porter dimana dapat meningkatkan ketepatan *stemming* sebesar 13,33%.

Berdasarkan hal tersebut, untuk memperbaiki ketepatan proses *stemming* pada algoritma *Rule Base Approach*, peneliti mencoba menyempurnakan ketepatan *stemming* dengan menggunakan kamus kata dasar Bahasa Bali dengan harapan dapat meningkatkan ketepatan proses *stemming*.

TINJAUAN PUSTAKA

Text Mining

Text mining adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis [5]. Tahapan proses *text mining* dibagi menjadi 3 tahap utama yaitu *tokenizing*, *filtering*, dan *stemming* [6]. Tahapan-tahapan dalam *text mining* bertujuan untuk menemukan kata-kata yang mewakili isi dokumen.

1. *Tokenizing*

Tokenizing adalah proses untuk membagi teks yang dapat berupa kalimat paragraph atau dokumen bagian-bagian tertentu.

2. *Filtering*

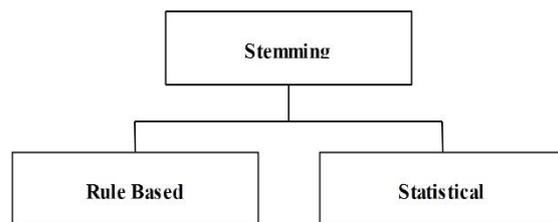
Filtering adalah mengambil kata-kata penting dari token dengan menggunakan *stopwords*.

3. *Stemming*

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya [7].

Rule Base Stemmer

Metode *Rule Base Approach* adalah metode yang membentuk akar kata dengan proses pengambilan awalan, sisipan dan akhiran. Metode *Rule Base* digunakan karena sifatnya yang fleksibel untuk digunakan sebagai *stemmer* berbagai macam bahasa dengan karakteristiknya yang lebih menekankan pada struktur morfologi suatu bahasa. Metode ini akan membuang (menghilangkan) *prefix* (awalan), *suffix* (akhiran) dan *infix* (sisipan) dari term bentuk menjadi suatu term kata dasar [8]. Algoritma *stemming* bisa diklasifikasikan menjadi 2 (dua) yaitu *Rule Based* dan *Statistical*. Menurut Sharma (2012) *Stemmer* berbasis aturan mengkodekan aturan khusus bahasa dimana *stemmer* statistik menggunakan informasi statistik dari yang besar korpus bahasa tertentu untuk mempelajari morfologi [9].



Gambar 1. *Types of Stemming Approach* [9]

Bahasa Bali

Beberapa imbuhan dalam Bahasa Bali, yaitu :

- Prefiks* yaitu imbuhan yang diletakkan di awal kata dasar, dalam Bahasa Bali disebut dengan pangater terdiri dari : (*a-*, *ma-*, *su-*, *ka-*, *pa-*, *pati-*, *pari-*, *maka-*, *saka-*, *kuma-*, *sa-*, *pa-*, *pi-*, *dur-*, *swa-*) [10]. Adapun contoh katanya yaitu : adiri, mayasa, sudharma, kaejuk, pajalan, patigrabe, paribasa, makaukud, sakabesik, kumajaum, sajjagat.
- Infiks* yaitu penyisipan imbuhan di dalam kata dasar, dalam Bahasa Bali

disebut dengan Seselan terdiri dari : (-in-, -um-, -el-, -er-). Adapun contoh katanya : sinurat, sumaur, telapak, gerigi [10].

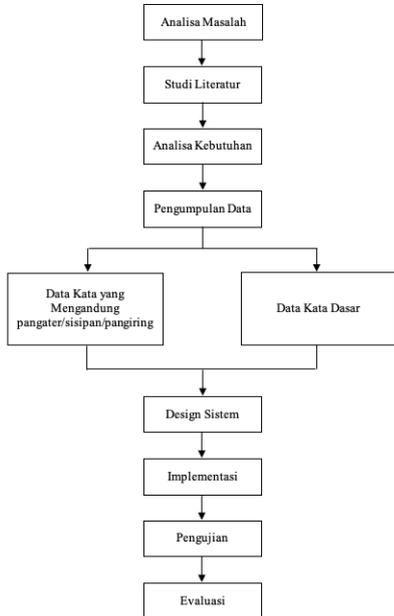
- c. *Sufiks* yaitu imbuhan yang diletakkan di belakang kata dasar, dalam Bahasa Bali disebut dengan pangiring terdiri dari : (-a, -e, -ne, -ang, -in, -an, -n, -wan, -nyane) [11]. Adapun contoh katanya : anggonna, batise, gigin, ngadepang, jagurin, menekan, bapan tiange, bukun, dharmawan.

Menghitung Akurasi Stemming

Pada pengujian proses *stemming* menggunakan perhitungan tingkat akurasi *stemming* seperti pada persamaan (1). Pada persamaan (1), terdapat simbol x yang merupakan tingkat akurasi *stemming*. Simbol T merupakan jumlah kata yang *di stemming* dengan hasil benar. Simbol N merupakan jumlah kata yang *di stemming* [12].

$$x = \frac{T}{N} \times 100\% \tag{1}$$

METODA PENELITIAN Rancangan Penelitian



Gambar 2. Rancangan Penelitian *Stemming* Bahasa Bali

Analisa Masalah

Proses *stemming* teks Bahasa Bali dengan metoda *Rule Base Approach* menghasilkan tingkat akurasi yang cukup baik yaitu mencapai angka 77.82%, akan tetapi jika mengacu pada hasil pengujian

tepatnya dengan hasil uji *stemming* salah, masih terdapat banyak kesalahan pada proses *stemming* hal ini di karenakan metoda *Rule Base Approach* memiliki kelemahan yaitu jika diterapkan pada kata yang memiliki kompleksitas imbuhan yang tinggi dan aplikasi *stemming* tidak dapat mengenali *rule* maka akan terjadi kesalahan pada proses *stemming*. Selain itu penyebab kesalahan juga terjadi karena *overstemming*. *Overstemming* adalah sebuah permasalahan pada proses *stemming* dimana kata dasar dari hasil *stemming* mengalami kekurangan suku kata atau huruf yang dianggap sebagai imbuhan akibat dari proses *stemming*. Pada penelitian ini diusulkan penambahan kamus pada metode *Rule Base Approach* untuk mengurangi kesalahan *stemming* dan meningkatkan ketepatan hasil. Kamus yang ditambahkan akan digunakan untuk pengecekan pada setiap kata yang telah melewati proses pemotongan imbuhan, kemudian hasil dari pemotongan langsung dicocokkan dengan kamus yang berisi kata dasar.

Tabel 1. Hasil Pengujian dengan *Rule Base Approach*

Imbuhan	Jumlah Kata Dasar	Uji Stemming (Benar)	Uji Stemming (Salah)
Pangate	240	159	81
r Seselan	17	17	0
Pangirin	119	80	39

Studi Literatur

Tahapan studi literatur adalah melakukan pengumpulan referensi baik dari buku, artikel, jurnal dan sebagainya yang diperoleh dari perpustakaan, internet, dll serta mempelajari referensi tersebut.

Analisa Kebutuhan

Analisa kebutuhan pada penelitian ini adalah menganalisa proses *stemming* pada penelitian sebelumnya terutama pada kebutuhan data hasil proses *stemming* pada kata dasar Bahasa Bali yang diujikan.

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.

Design Sistem

Arsitektur umum pada sistem pembentukan akar kata dalam dokumen teks Bahasa Bali memiliki beberapa tahapan yaitu sebagai berikut :

1. Input

Pada tahap pertama sistem adalah input 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 kategori.

3. Proses

Pada tahap proses adalah proses *stemming* dengan metoda *Rule Base Approach* dan penerapan kamus kata dasar Bahasa Bali.

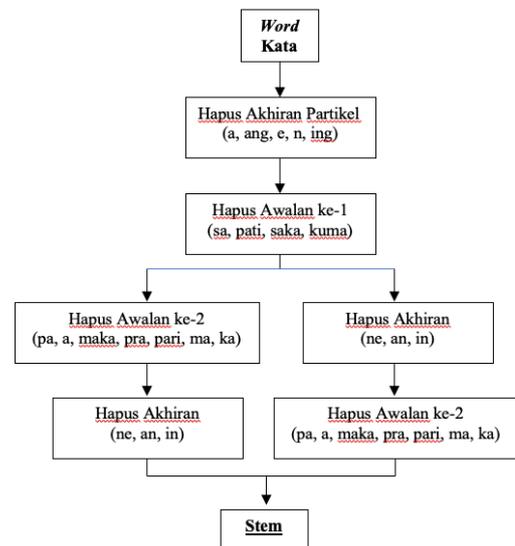
4. Output

Menampilkan hasil proses *stemming*, yang selanjutnya akan dievaluasi ketepatan *stemming* dan melakukan perhitungan tingkat akurasi keberhasilan *stemming*.

Implementasi

Proses *stemming* dengan metoda *Rule Base Approach* adalah (Gambar 2) :

- 1) Menghapus akhiran partikel
- 2) Menghapus awalan pertama, jika tidak ada maka lanjut untuk menghapus awalan ke-2, dan jika tidak ditemukan maka menghapus akhiran.
- 3) Menghapus awalan ke-2 dan menghapus akhiran dan kata akhir diartikan sebagai kata dasar.
- 4) Menghapus akhiran, jika tidak ada maka kata tersebut diartikan sebagai kata dasar, namun jika ditemukan maka dilanjutkan dengan menghapus awalan ke-2 dan kata akhir diartikan sebagai kata dasar.



Gambar 3. Desain *Stemming Rule Base Approach* Untuk Bahasa Bali

Aturan Imbuhan (Tiron)

Berdasarkan aturan morfologi dari Gambar 3, terdapat 4 aturan affix yang telah dibuat untuk *stemming* Bahasa Bali dan akan

diterapkan dalam sistem aplikasi ini. 4 (empat) dari aturan tersebut akan didefinisikan dalam beberapa tabel. Yang pertama adalah aturan penghapusan partikel yang berupa akhiran (pangater), pada Tabel 2 menjelaskan aturan penghapusan partikel akhiran.

Tabel 2. Kelompok Rule Pertama: Akhiran Partikel

Akhiran	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
a	NULL	2	NULL	Daara - daar
ang	NULL	2	NULL	Gelahang - gelah
e	NULL	2	NULL	Payuke - payuk
n	NULL	2	NULL	Bukun - buku
ing	NULL	2	NULL	Rikalaning - rikal

Tabel 2 menjelaskan akhiran kata yang harus dihilangkan ada pada aturan tersebut adalah a, ang, e, n, ing.

Kemudian aturan ke 2 dalam stemming dengan *Rule Base Approach* Bahasa Bali ini adalah penghilangan awalan pada kata. Pada Tabel 3 akan menjelaskan beberapa awalan yang ada pada aturan.

Tabel 3. Kelompok Rule Ketiga: Awalan Pertama

Awalan	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
sa	NULL	2	NULL	Sajagat - jagat
pati	NULL	2	NULL	Patigrabe - grape
saka	NULL	2	NULL	Sakabesik - besik
kuma	NULL	2	NULL	Kumalipan - lipan

Tabel 3 merupakan awalan-awalan yang harus dihilangkan pada kata yang akan dilakukan *stemming*. Kemudian aturan yang keempat adalah penghapusan awalan yang kedua dijelaskan pada Tabel 4.

Tabel 4. Kelompok Rule Keempat yang Mencakup Awalan yang Kedua

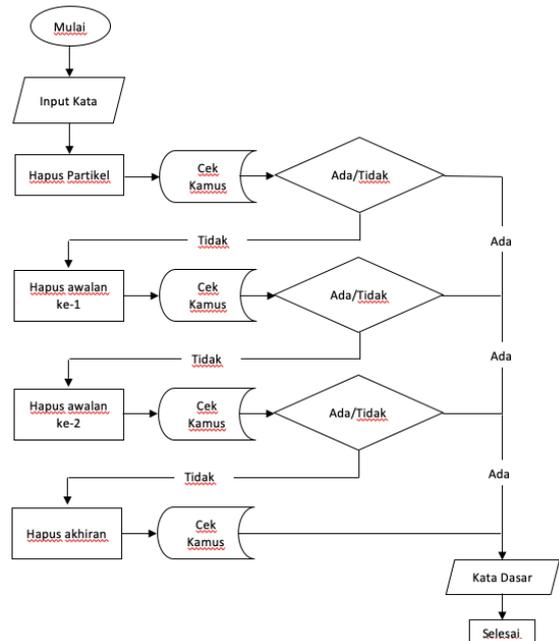
Awalan	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
ma	NULL	2	NULL	Majalan - jalan
pa	NULL	2	NULL	Paiieh - pileh
ka	NULL	2	NULL	Kaicen - icen
a	NULL	2	NULL	Aukud - ukud
pra	NULL	2	NULL	Prajani - jani
pari	NULL	2	NULL	Paribasa - basa
maka	NULL	2	NULL	Makaukud - ukud

Aturan yang keempat adalah aturan untuk akhiran, dapat dilihat pada Tabel 5.

Tabel 5. Aturan Untuk Akhiran

Akhiran	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
an	NULL	2	NULL	Cenikan - cenik
in	NULL	2	NULL	Jagurin - jagur
ne	NULL	2	NULL	Gigine - gigi

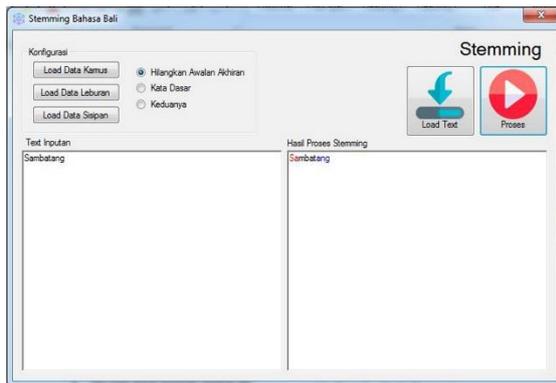
Pada gambar 3 menampilkan diagram proses *stemming* tanpa menggunakan kamus. Pada gambar 4 menampilkan diagram proses *stemming* dengan menggunakan kamus kata dasar. Setelah input kata, akan dilakukan penghapusan partikel, dilanjutkan dengan mengecek kamus. Jika terdapat pada kamus, maka kata dasar ditemukan, jika tidak maka dilanjutkan dengan menghapus awalan ke-1. Setelah menghapus awalan ke-1, dilanjutkan dengan mengecek kamus. Jika terdapat pada kamus, maka kata dasar ditemukan, jika tidak maka dilanjutkan dengan menghapus awalan ke-2. Setelah menghapus awalan ke-2, dilanjutkan dengan mengecek kamus. Jika terdapat pada kamus, maka kata dasar ditemukan, jika tidak maka dilanjutkan dengan menghapus akhiran jika terdapat akhiran, bila tidak maka kata tersebut adalah kata dasar dan proses *stemming* dengan kamus selesai.



Gambar 4. Diagram Proses *Stemming Rule Base Approach* dengan Kamus

HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan eksperimen dengan menggunakan komputer untuk melakukan proses *stemming*. Spesifikasi perangkat keras dan sistem operasi yang digunakan pada penelitian ini adalah menggunakan sebuah laptop dengan prosesor Intel® core™ i3-3217U CPU @ 1.80GHz (4 CPUs), ~1.7GHz, memori (RAM) 2 GB, dan menggunakan sistem operasi Windows 8 Pro 64-bit (6.2, build 9200). Sedangkan perangkat lunak yang digunakan untuk menyimpan database kamus adalah MySQL.



Gambar 5. Tampilan Sistem Pencarian Kata Dasar

Pengujian Prefix (Pengater)

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

Tabel 6. 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 penghapusan 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 7. 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 penghapusan 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 8. Pengujian Seselan

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

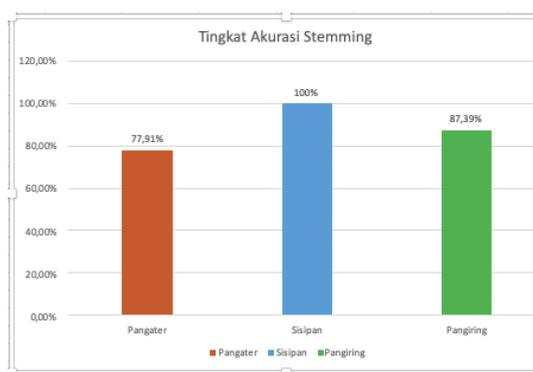
Hasil Pengujian

Setelah dilakukan pengujian *stemming* teks kata dasar pada pengater, seselan dan pangiring menggunakan aplikasi *stemming* kata dasar Bahasa Bali, dilihat hasil pengujian menghasilkan uji *stemming* benar pada pengater sebanyak 91 kata dan uji *stemming* salah pada pengater sebanyak 53 kata. Uji *stemming* benar pada seselan sebanyak 17 kata dan uji *stemming* salah pada seselan sebanyak 0 kata. Uji *stemming* benar pada pangiring sebanyak 104 kata dan uji *stemming* salah pada pangiring sebanyak 15 kata.

Tabel 9. Hasil Pengujian

Imbuhan	Jumlah Kata Dasar	Uji Stemming (Benar)	Uji Stemming (Salah)
Pengater	240	187	53
Seselan	17	17	0
Pangiring	119	104	15

Tingkat Akurasi



Gambar 6. Tingkat Akurasi Stemming

Pada gambar 6 dapat dilihat tingkat akurasi aplikasi *stemming* kata dasar Bahasa Bali dengan metode *Rule Base Approach* dan penerapan kamus kata dasar, menghasilkan tingkat akurasi pengater 77.91%, tingkat akurasi seselan 100% dan tingkat akurasi pangiring 87.39%. Hasil rata-rata tingkat akurasi yang diperoleh adalah sebesar 88.43%.

Tabel 10. Perbandingan Tingkat Akurasi

Imbuhan	Jumlah Kata Dasar	Tanpa Kamus	Dengan Kamus
Pengater	240	66.25%	77.91%
Seselan	17	100%	100%
Pangiring	119	67.22%	87.39%
Rata-rata	376	77.82%	88.34%

Dari tabel 10, perbandingan tingkat akurasi dari hasil proses *stemming* tanpa kamus dan dengan kamus, terdapat perbaikan hasil *stemming* sebesar 10.61% setelah menerapkan kamus kata dasar dalam proses *stemming* dengan metoda *Rule Base Approach*.

Dalam penelitian sebelumnya terdapat permasalahan *overstemming*. Salah satu contoh permasalahan *overstemming* seperti

proses pengiring “-ang” pada kata “sambatang” menjadi “mbat” padahal yang benar adalah “sambat”. Dari hasil tersebut dapat kita lihat ada huruf atau suku kata yang menjadi bagian kata dasar yang terhapus atau hilang. Yang menjadi penyebab kegagalan pada proses *stemming* adalah karena ada bagian huruf yang dianggap sebagai pengater oleh aplikasi yaitu “sa”, sehingga huruf atau suku kata tersebut dihapus. Dengan menerapkan kamus kata dasar, permasalahan *overstemming* dapat diperbaiki sebesar 10.61%.

SIMPULAN

Berdasarkan proses penelitian yang sudah dilakukan maka peneliti dapat mengambil kesimpulan bahwa penerapan kamus dasar pada metode *Rule Base Approach* dapat meningkatkan ketepatan *stemming* sebesar 10.61%, dimana tingkat akurasi tanpa menggunakan kamus sebesar 77.82% dan tingkat akurasi dengan menggunakan kamus sebesar 88.34%.

DAFTAR PUSTAKA

- [1] A. Roshdi and A. Roohparvar, “Review: Information Retrieval Techniques and Applications,” *Int. J. Comput. Networks Commun. Secur.*, vol. 3, no. 9, pp. 373–377, 2015.
- [2] R. Setiawan, A. Kurniawan, W. Budiharto, I. H. Kartowisastro, and H. Prabowo, “Flexible affix classification for stemming Indonesian Language,” *2016 13th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. ECTI-CON 2016*, 2016.
- [3] P. Gede, S. Cipta, N. W. Wardani, P. T. Informatika, and R. B. Approach, “Stemming Dokumen Teks Bahasa Bali Dengan Metode Rule Base Approach,” vol. 7, no. 3, pp. 510–521, 2020.
- [4] N. J. M. Verdaningroem and A. Saifudin, “Penerapan Kamus Dasar pada Algoritma Porter untuk Mengurangi Kesalahan Stemming Bahasa Indonesia,” *J. Teknol.*, vol. 10, no. 2, pp. 103–112, 2018.
- [5] R. Feldman and J. Sanger, *The Text Mining Handbook*. Cambridge University Press, 2007.
- [6] N. Luh, W. Sri, R. Ginantra, and N. W. Wardani, “Implementasi Metoda

- Naïve Bayes Dan Vector Space Model Dalam Deteksi Kesamaan Artikel Jurnal Berbahasa,” *J. Infomedia*, vol. 4, no. 2, pp. 94–100, 2019.
- [7] N. Luh, W. Sri, R. Ginantra, and N. W. Wardani, “Measurement of the Similarity of Indonesian Papers on One Journal Topic with the Naive Bayes Algorithm and Vector Space Model,” no. September, pp. 20–26, 2019.
- [8] F. Amin and J. A. Razaq, “Implementasi Stemmer Bahasa Jawa dengan Metode Rule Base Approach pada Sistem Temu Kembali Informasi Dokumen Teks Berhasa Jawa,” pp. 978–979, 2018.
- [9] D. Sharma, “Stemming Algorithms: A Comparative Study and their Analysis,” *Int. J. Appl. Inf. Syst. (IJ AIS)*. *www.ijais.org.*, vol. 4, no. 3, 2012.
- [10] N. Wardani and P. Nugraha, “Stemming Teks Bahasa Bali dengan Algoritma Enhanced Confix Stripping,” *Int. J. Nat. Sci. Eng.*, vol. 4, no. 3, 2020.
- [11] M. Agus, P. Subali, and C. Fatichah, “Kombinasi Metode Rule-Based Dan N-Gram Stemming Untuk a Combination of Methods Rule-Based and N-Gram Stemming To Recognize Balinese Language Stemmer,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, 2019.
- [12] M. Shahid Husain, “An Unsupervised Approach to Develop Stemmer,” *Int. J. Nat. Lang. Comput.*, vol. 1, no. 2, pp. 15–23, 2012.