

STEMMING IN MADURESE LANGUAGE USING NAZIEF AND ADRIANI ALGORITHM

Moh Ashari^{*1}, Danang Arbian Sulisty², Fadhli Almu'iini Ahda^{*3}

^{1,2,3}Informatics Engineering, Faculty of Technology and Design, Asia Institut of Technology and Business
Malang, Indonesia

Email: ¹davinsiari@gmail.com, ²danangarbian@gamil.com, ³adhi32286@gmail.com

(Article received: April 26, 2024; Revision: June 14, 2024; published: August 31, 2024)

Abstract

Madurese is one of the regional languages in Indonesia, which dominates East Java and Madura Island in particular. However, the use of Madurese is declining compared to other regional languages. This is partly due to a sense of prestige and difficulty in learning it. As a result, the future of Madurese as one of the regional languages in Indonesia is increasingly threatened by the decline in its use. In addition, academic literature and scientific publications in Madurese are difficult to find in public and academic libraries, so previous research on Madurese stemming is still very little and needs to be developed further. Therefore, this research aims to find the base word of Madurese language using Nazief & Adriani algorithm based on Madurese language morphology. The Nazief & Adriani method in previous studies has good performance. Stemming can also be developed into a Madurese language translator application into other languages. This research uses 650 words in the form of datasets, consisting of 500 prefix words and 150 suffix words. The resulting accuracy for the whole is 96.61% with 628 correct words, the prefix has 95.6% accuracy, and the suffix has 100% accuracy. Overstemming was found in 22 prefix words and no words experienced Understemming.

Keywords: Bahasa Madura, Morfologi, Nazief & Adriani, NLP, Stemming.

STEMMING PADA BAHASA MADURA MENGGUNAKAN ALGORITMA NAZIEF DAN ADRIANI

Abstrak

Bahasa madura adalah salah satu Bahasa daerah di Indonesia, yang mendominasi Jawa Timur dan Pulau Madura pada khususnya. Namun, penggunaan bahasa Madura semakin menurun dibandingkan dengan Bahasa daerah lainnya. Hal ini Sebagian disebabkan oleh rasa gengsi dan kesulitan dalam mempelajarinya. Akibatnya, masa depan bahasa Madura sebagai salah satu bahasa daerah di indonesia semakin terancam oleh penurunan penggunaannya. Selain itu, literatur akademis dan publikasi ilmiah berbahasa Madura sulit ditemukan di perpustakaan umum dan akademis, sehingga penelitian sebelumnya mengenai stemming bahasa Madura masih sangat sedikit dan perlu dikembangkan lebih lanjut. Oleh karena itu, penelitian ini bertujuan untuk menemukan kata dasar bahasa Madura menggunakan algoritma Nazief & Adriani berdasarkan morfologi bahasa Madura. Metode Nazief & Adriani pada penelitian-penelitian sebelumnya memiliki performa yang baik. Stemming juga dapat dikembangkan menjadi aplikasi penerjemah bahasa Madura ke dalam bahasa lainnya.. Penelitian ini menggunakan 650 kata berbentuk dataset, yang terdiri dari 500 kata awalan dan 150 kata akhiran. Akurasi yang dihasilkan untuk keseluruhan adalah 96,61% dengan 628 kata yang benar, awalan memiliki akurasi 95,6%, dan akhiran memiliki akurasi 100%. Overstemming ditemukan pada 22 kata awalan dan tidak ada kata yang mengalami Understemming.

Kata kunci: Bahasa Madura, Morfologi, Nazief & Adriani, NLP, Stemming.

1. PENDAHULUAN

Indonesia adalah negara terpadat keempat di dunia dan memiliki keberagaman bahasa yang luar biasa. Hampir 800 bahasa daerah yang tercatat di Indonesia, namun penelitian tentang pemrosesan

bahasa dan pengolahan bahasa alami masih perlu ditingkatkan[1].

Salah satu bahasa daerah di indonesia adalah bahasa Madura, yang digunakan sebagai bahasa sehari-hari di pulau Madura dan juga beberapa daerah

lain seperti Probolinggo, Bondowoso, Pasuruan, Jember dan Banyuwangi. Sebagian besar penduduk pulau Madura menggunakan bahasa Madura untuk berkomunikasi baik di dalam maupun di luar pulau Madura.

Bahasa Madura memiliki banyak keunikan dalam struktur morfologi. Salah satu bidang ilmu pengetahuan yang dikenal sebagai morfologi mempelajari satuan bahasa sebagai satuan gramatikal. Proses morfologi adalah pembentukan kata dari bentuk dasar melalui pembubuhan afiks, pengulangan, dan pemajemukan[2]. Keunikan lain dari bahasa Madura dari unsur fonologi khususnya dari variasi fonem, yakni pada huruf vocal dan konsonan. bahasa Madura memiliki enam vocal yakni /a/, /i/, /u/, /e/, /ə/ dan /o/. keunikan fonem dalam bahasa Madura berpengaruh dalam pelafalan atau artikulasi dan penulisan kata, kesalahan dalam penulisan kata khususnya penggunaan fonem yang kurang tepat akan berpengaruh pada pelafalan dan arti kata[3].

Di bidang kecerdasan buatan dan linguistik komputasi, penelitian tentang pemrosesan bahasa alami (NLP) telah menjadi topik yang menarik, dimana tujuan utama NLP adalah melatih mesin untuk memahami bahasa manusia dan memberikan umpan balik yang sesuai[4]. Stemming merupakan proses perubahan kata berimbuhan ke bentuk kata dasarnya (root word) dengan menggunakan aturan tertentu[5]. Algoritma stemming akan berbeda untuk setiap bahasa yang digunakan [6]. Sebagai contoh, kata bersama, kebersamaan, menyamai, akan distem ke root word nya yaitu “sama”. Proses ini memungkinkan penggabungan kata-kata dengan akar kata yang sama. Overstemming dan understemming adalah masalah yang sering terjadi dalam proses stemming. Overstemming menghilangkan imbuhan yang berlebih pada sebuah kata, menghasilkan kata dengan arti yang berbeda [7]. Sebaliknya, understemming tidak menghilangkan imbuhan, sehingga kata yang dihasilkan sama dengan kata aslinya [8]. Proses stemming adalah menentukan kata dasar dengan menghilangkan imbuhan.

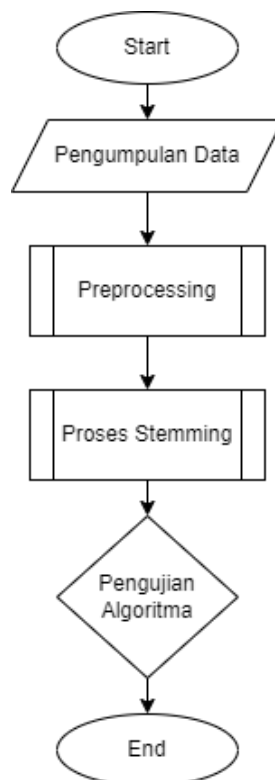
Beberapa penelitian telah menemukan bahwa tren penggunaan bahasa Madura telah bergeser ke arah bahasa Indonesia, yang dianggap sebagai bahasa resmi [9]. Hal ini disebabkan oleh rasa gengsi dan kesulitan mempelajari bahasa Madura, yang memiliki berbagai dialek dan tingkat bahasa. Akibatnya, masa depan bahasa Madura sebagai salah satu bahasa daerah di Indonesia semakin terancam oleh penurunan penggunaannya [10]. Hal ini dapat menyebabkan kehancuran bahasa daerah jika dibiarkan. Menurut berita harian, lebih dari 50 bahasa daerah terancam punah karena jumlah penutur yang telah lanjut usia dan sangat sedikit anak muda yang menuturkannya. Bahkan sembilan bahasa Papua telah punah karena tidak ada yang menuturkannya. Berita ini juga didukung oleh dikti yang memasang berita tentang bahasa daerah yang terancam punah [11].

Selain itu, literatur akademis dan publikasi ilmiah berbahasa Madura sulit ditemukan di perpustakaan umum dan akademis, sehingga penelitian sebelumnya mengenai stemming bahasa Madura masih sangat sedikit dan perlu dikembangkan lebih lanjut. Algoritma stemming digunakan untuk bahasa Madura [12]. Hal inilah yang melatarbelakangi untuk membuat penelitian mengenai “Stemming Pada Bahasa Madura Menggunakan Algoritma Nazief & Andriani”.

Algoritma Nazief & Andriani digunakan untuk mencari kata dasar dalam bahasa Indonesia dengan menggunakan kaidah morfologi bahasa Indonesia, membuang awalan dan akhiran kata seperti yang ditemukan dalam kamus bahasa Indonesia [13]. Algoritma ini juga sering digunakan untuk stemming bahasa lain [14]-[15]. Understemming, yaitu pengurangan kata yang terlalu sedikit dari yang seharusnya, understemming dan overstemming yaitu pemenggalan kata yang lebih banyak dari yang seharusnya sehingga memiliki arti yang berbeda [16]. Oleh karena itu, Tujuan dari penelitian ini untuk menemukan kata dasar bahasa Madura menggunakan algoritma Nazief & Andriani berdasarkan morfologi bahasa Madura.

2. METODE PENELITIAN

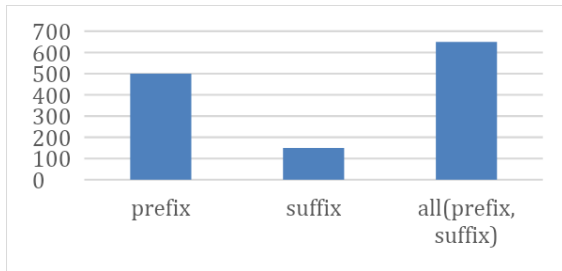
Penelitian ini akan berkonsentrasi pada penggunaan algoritma Nazief & Andriani untuk melakukan proses stemming pada bahasa Madura. Berikut adalah tahapan penelitian yang dapat dilihat pada gambar 1.



Gambar 1. Flowchart tahapan penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah kata berimbuhan bahasa Madura berjumlah 650 kata yang terdiri dari 500 imbuhan awalan dan 150 imbuhan akhiran, beserta 650 kata dasar yang diambil dari kumpulan buku bahasa Madura dan dataset yang diperoleh dari Mendeley Data, seperti yang terlihat pada gambar 2.



Gambar 2. Grafik jumlah kata

Data tersebut dipilih dan disatukan dalam bentuk dataset secara manual dengan cara menyetik ulang isi kata dalam buku dan dataset yang diperoleh.

2.2. Preprocessing Data

Pada tahap *Preprocessing*, bentuk teks yang tidak terstruktur diubah menjadi teks yang terstruktur sehingga data siap untuk diproses [17]. Prosesnya terdiri dari:

1. Filtering

Pada proses *filtering* mengubah alfabet beraksen menjadi huruf biasa pada alfabet dan kemudian menghapus tanda baca, karakter, dan angka.

2. Case Folding

Proses ini dilakukan pada semua huruf dalam kata menjadi huruf kecil semua [14].

2.3. Proses Stemming

Algoritma Nazief & Adriani pertama kali dikembangkan oleh Bobby Nazief dan mirna Adriani. Algoritma ini berdasarkan aturan morfologi pada bahasa Indonesia yang membagi imbuhan menjadi kategori yang diperbolehkan dan tidak diperbolehkan. Imbuhan di depan (awalan), imbuhan di belakang (akhiran), imbuhan di tengah (sisipan), dan kombinasi imbuhan awalan dan akhiran (konfiks) termasuk dalam kategori ini.

Algoritma Nazief & Adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahapan sebagai berikut [18] :

1. Cari kata yang akan distemming dalam kamus. Jika ditemukan maka diasumsikan kata adalah *root word*. Maka algoritma berhenti.
2. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah”, atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “-nya”), jika ada.

3. Hapus Derivation Suffixes (“-i”, “-an”, atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka kembali ke langkah 3a.
 - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an”, atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus Derivation Prefix. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
 - a. Periksa tabel kombinasi awalan-akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
 - b. For $i = 1$ to 3, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan : jika awalan kedua sama dengan awalan pertama algoritma berhenti.
5. melakukan *Recording*.
6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

Algoritma Nazief & Adriani mengusung tiga konsep dasar dalam penghilangan imbuhan pada kata-kata dalam bahasa Indonesia [19].

1. Menghapus imbuhan (afiksasi) sesuai dengan kaidah morfologi bahasa Indonesia
 - a. Inflection suffixes (IS)
 - Particle (P) {“-lah”, “-pun”, “-tah”, “-kah”}.
 - Possesive pronoun (PP) {“-ku”, “-mu”, “-nya”}.
 - b. Derivation suffixes (DS) {“-i”, “-kan”, “-an”}.
 - c. Derivation prefixes (DP)
 - morphological DP {“me-”, “be-”, “te-”, “pe-”}.
 - Plain DP {“di-”, “ke-”, “se-”}.
2. Algoritma sangat bergantung pada kamus kata dasar.
3. Algoritma ini mendukung recording, yang berarti menyusun kembali kata-kata yang telah mengalami stemming berlebihan.

Penulis menggunakan konsep yang sama dengan algoritma Nazief & Adriani, metode stemmer untuk bahasa Madura adalah sebagai berikut :

Pertama, metode ini mengimplementasikan penghapusan imbuhan (afiksasi) sesuai dengan aturan morfologi bahasa Madura. Selanjutnya, terdapat pemrosesan akhiran infleksi, termasuk kata ganti posesif (PP) seperti “-na”, derivation suffixes (DS) seperti “-a”, “-e”, “-na”, “-an”, “-aghi”, dan derivation prefixes (DP) yang mencakup morfologi seperti {“N”} dan awalan derivasi yang umum seperti “-a-“, “-e-“, “-ta-“, “-ma-“, “-ka-“, “-epa-“, “-sa-“, “-pa-“, “-pe-“, “-pra-“, “-nga-“, “-eka-“.

Penelitian ini disesuaikan untuk bahasa Madura berdasarkan prinsip dasar Nazief & Adriani. gambar 3 menjelaskan langkah pada proses stemming.



Gambar 3. Proses stemmer bahasa madura

Pada gambar 3 menggambarkan alur proses stemming bahasa Madura dalam penelitian ini adalah stemming tanpa menggunakan kamus, namun memanfaatkan dataset bahasa Madura. pendekatan ini dipilih karena sebagian kata yang ada dalam kamus bahasa Madura memiliki kekeliruan atau tidak lengkap, sehingga penggunaan kamus menjadi kurang efektif.

2.4. Pengujian Algoritma

Pada tahap ini dilakukan pengujian terhadap algoritma Nazief & Adriani dalam mendapatkan kata dasar dengan menghitung tingkat akurasi yakni persentase dari total kata yang benar dibandingkan jumlah total kata yang diuji dikali 100% [20]. Persamaan yang digunakan untuk menghitung akurasi adalah sebagai berikut:

$$Accuracy = \frac{\text{jumlah kata yang distem benar}}{\text{total jumlah kata yang diuji}} \times 100\% \quad (1)$$

Akurasi menunjukkan seberapa besar persentase kata-kata yang berhasil distem dengan

benar oleh sistem. Semakin tinggi nilai akurasi, semakin baik kinerja sistem melakukan stemming.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pada penelitian ini bahasa pemrograman python digunakan untuk pengembangan model dan google collab sebagai tools. Data yang digunakan dalam penelitian ini tidak berasal dari kamus. Namun, dikumpulkan secara manual dari kata-kata yang umum digunakan dalam bahasa Madura. Proses pengumpulan dilakukan dengan cermat untuk memastikan bahwa data yang digunakan relevan dan representatif terhadap bahasa Madura. Dataset yang digunakan berjumlah 650 kata, yang terdiri dari 500 kata dengan awalan dan 150 kata dengan akhiran. Tabel 1 menggambarkan data dan artinya.

Tabel 1. Data kata

no	kata	indonesia
1	abhâjâng	melaksanakan sholat
2	alako	bekerja
3	aberri'	memberi
4	asarong	memakai sarung
5	acopa	meludah
6	atanè	bertani
7	anapso	bernafsu
8	abhânto	membantu
9	ètegghu'	dipegang
...
650	parlona	butuhnya

Sebagian dari dataset yang digunakan dalam penelitian ini digambarkan dalam Tabel 1, yang mencakup kata-kata dalam bahasa Madura dan terjemahannya. Tabel ini menunjukkan bagaimana setiap kata dalam bahasa Madura dihubungkan dengan arti atau terjemahan dalam bahasa Madura.

penelitian ini dapat memastikan bahwa kata-kata yang dimasukkan adalah representatif dan relevan dengan tujuan analisis, karena kamus bahasa Madura yang tersedia sangat terbatas dan sebagian besar memiliki kekeliruan atau kurang lengkap dalam mencakup variasi kata yang digunakan dalam komunikasi sehari-hari. oleh karena itu, dibuat keputusan untuk menyusun dataset secara manual.

3.2. Preprocessing Data

Pada tahap preprocessing dilakukan proses pembersihan data sebelum proses stemming dilakukan. Mempersiapkan teks yang tidak terstruktur menjadi data yang baik dan siap diolah, sehingga hasil stemming menjadi lebih akurat dan relevan. Pada penelitian ini proses yang digunakan adalah Filtering dan Case Folding. Hasil preprocessing dapat dilihat pada tabel 2.

Tabel 2. Hasil preprocessing

no	kata	indonesia
1	abhâjâng	abhâjâng
2	alako	alako
3	aberri'	aberri'
4	asarong	asarong

5	acopa	acopa
6	atanè	atanè
7	anapso	anapso
8	Tabhendem	tabhendem
9	Taangka'	taangka'
...
650	Matappor	matappor

Tabel 2 menunjukkan hasil preprocessing dari dataset yang digunakan dalam penelitian ini, yang memperlihatkan perubahan kata-kata asli dalam bahasa Madura setelah di Filtering dan Case folding. Proses ini memastikan bahwa teks bebas dari gangguan dan siap untuk tahap pemrosesan berikutnya, meningkatkan akurasi dan relevansi hasil stemming. Sebagai contoh, kata "Tabhendem" diubah menjadi "tabhendem" dan kata "Taangka'" diubah menjadi "taangka'". Proses ini penting karena pencocokan kata yang peka terhadap huruf besar dan kecil dapat menghasilkan duplikasi dan inkonsistensi dalam analisis teks. Preprocessing merupakan langkah krusial untuk mengoptimalkan proses stemming dan memastikan bahwa data yang digunakan bersih dan konsisten.

3.3. Hasil Stemming

Dalam tahap ini, kata-kata yang telah di preprocessing kemudian di stemming, proses stemming yang dilakukan dalam penelitian ini dibagi menjadi dua langkah, yaitu stemming awalan dan stemming akhiran. Langkah pertama, stemming awalan, dilakukan untuk menghilangkan awalan pada kata bahasa Madura, sedangkan langkah kedua, stemming akhiran, dilakukan untuk menghilangkan akhiran pada kata bahasa Madura.

Dengan memisahkan proses stemming menjadi dua langkah, dapat membantu meningkatkan akurasi proses stemming. Sehingga kata-kata yang dihasilkan lebih sesuai dengan kata dasar yang dimaksudkan dalam bahasa Madura. Selain itu, juga membantu dalam mengatasi kompleksitas morfologi bahasa Madura. Bahasa Madura memiliki struktur morfologi yang kaya, dengan banyak variasi awalan dan akhiran yang digunakan untuk membuat kata-kata. Dengan memisahkan proses stemming, dapat memperhitungkan berbagai variasi ini secara lebih efektif, sehingga hasil stemming menjadi lebih akurat dan relevan.

1. Stemming awalan

Proses stemming dilakukan terhadap 500 kata dari awalan bahasa Madura. Salah satu contohnya adalah kata "alako", yang hasil stemmingnya adalah "lako", yang berasal dari awalan "a" dan kata "lako". Hasil stemming awalan dapat dilihat pada tabel 3.

Tabel 3. Hasil stemming awalan

no	kata	stemming	status
1	abhâjâng	bhâjâng	Success
2	alako	lako	Success
3	aberri'	berri'	Success
4	asarong	sarong	Success
5	acopa	copa	Success
6	atanè	tanè	Success

7	anapso	napso	Success
8	abhânto	bhânto	Success
9	aghellâ	dipegang	Success
10	atellor	tellor	Success
11	abhudu'	bhudu'	Success
12	akole'	kole'	Success
13	arasan	ras	Overstemming
...
650	takakan	kak	Overstemming

Tabel 3 menunjukkan hasil stemming awalan kata-kata dalam bahasa Madura. Contohnya adalah kata "abhâjâng" distemming menjadi "bhâjâng", dan kata "alako" distemming menjadi "lako". Proses ini berhasil dilakukan pada sebagian besar kata dengan status "Success".

Namun, terdapat beberapa kata yang mengalami overstemming, yaitu kondisi ketika terlalu banyak karakter dihilangkan sehingga mengubah makna dasar kata secara signifikan. Salah satu contohnya adalah kata "arasan" yang diubah menjadi "ras".

2. Stemming akhiran

Proses stemming akhiran dilakukan terhadap 150 kata akhiran bahasa Madura. Salah satu contohnya adalah kata "bâccoaghi", yang hasil stemmingnya adalah "bâcco", yang berasal dari kata "bâcco" dan akhiran "aghi". tabel 4 menunjukkan hasil stemming akhiran.

Tabel 4. Hasil stemming akhiran

no	kata	indonesia	status
1	mellea	melle	Success
2	molea	mole	Success
3	bâkkèlaghi	bâkkèl	Success
4	kabâlâaghi	kabâlâ	Success
5	afalan	afal	Success
6	pèkkèrè	pèkkèr	Success
7	ontongnga	ontong	Success
8	robbhui	robbhu	Success
...
650	robâna	robâ	Success

Tabel 4 menampilkan hasil stemming akhiran dari kata-kata dalam bahasa Madura. Contohnya adalah kata "bâkkèlaghi" setelah melalui proses stemming menjadi "bâkkèl", dimana akhiran "aghi" dihapus untuk mendapatkan kata dasar "bâkkèl", dan kata "mellea" distemming menjadi kata "melle".

Setiap kata pada tabel 4 menampilkan status "Success", menunjukkan bahwa proses stemming berhasil dan kata dasar yang dihasilkan sesuai dengan makna aslinya. Dalam proses ini, tidak ada kasus overstemming, yaitu penghapusan berlebihan yang mengubah makna dasar kata secara signifikan, maupun understemming, dimana penghapusan tidak cukup sehingga masih menyisakan elemen yang tidak diperlukan. Ini menunjukkan bahwa algoritma stemming yang digunakan telah disesuaikan dengan baik untuk menangani akhiran dalam bahasa Madura.

3.4. Pengujian Algoritma

Proses pengujian menggunakan persamaan untuk menghitung akurasi (Accuracy). Pengujian

terhadap hasil stemming dilakukan secara detail dari awal sampai terakhir.

Dari hasil stemming yang dilakukan pada kata bahasa Madura, algoritma yang digunakan dapat menemukan kata dasar yang memiliki awalan dan akhiran. Namun, ada beberapa kata yang mengalami overstemming, dan tidak ada satupun kata yang mengalami understemming. Hasil uji coba stemming disajikan pada tabel 5.

Tabel 5. Hasil uji coba stemming

no	Imbuan	Benar	Overstemming	Understemming	akurasi
1	Awalan	478	22	0	95,6%
2	Akhiran	150	0	0	100%
3	All(awalan,akhir)	628	22	0	96,61%

Pada tabel 5 menunjukkan hasil dari pengujian algoritma Nazief & Adriani yang telah disesuaikan dengan aturan morfologi bahasa Madura. Sebanyak 500 kata diuji dalam pengujian kata dengan awalan. Algoritma berhasil menemukan kata dasar yang benar untuk 478 kata, dengan akurasi 95,6%. Namun, terdapat 22 kata dimana algoritma menghapus terlalu banyak karakter sehingga mengubah makna dasar kata. Misalnya, stemming kata "rasan" seharusnya menghasilkan kata "ras", tetapi akhirnya menghasilkan kata "ras", yang menghilangkan sebagian dari makna awalnya. Dalam pengujian awalan, tidak ada kasus understemming, ini menunjukkan bahwa algoritma cukup baik dalam mengidentifikasi dan menghapus awalan.

Pengujian kata dengan akhiran yang dilakukan pada 150 kata menunjukkan bahwa algoritma mampu menemukan kata dasar yang benar untuk semua kata yang diuji dengan tingkat akurasi 100%. pengujian ini tidak menemukan overstemming atau understemming. Contohnya, kata "bakkèlaghi" distemming menjadi "bakkèl" dengan tepat, menghapus akhiran "aghi" tanpa mengubah makna dasar. Ini menunjukkan bahwa aturan untuk menghilangkan akhiran telah disesuaikan dengan sangat baik dengan struktur morfologi bahasa Madura.

4. DISKUSI

Dalam penelitian-penelitian sebelumnya mengenai stemming menggunakan algoritma Nazief & Adriani, telah terbukti bahwa penggunaan algoritma Nazief & Adriani terhadap bahasa Madura dapat memberikan hasil akurasi yang baik. Pada tahun 2023, penelitian oleh Enni Lindrawati yang berjudul "Comparison of Modified Nazief & Adriani and Modified Enhanced Confix Stripping algorithms for Madurese Language Stemming" menunjukkan bahwa Nazief & Adriani memiliki tingkat akurasi yang baik dibanding Enhanced Confix Stripping (ESC) untuk stemming pada bahasa madura, akurasi yang diperoleh untuk Modifikasi Nazief & Adriani adalah 88,80% dan untuk ESC adalah 74,00% [21]. Pada tahun 2020, Aji Prasetya melakukan penelitian

yang berjudul "Stemming Javanese affix word using Nazief and Adriani modifications", dari 366 kata yang diujikan dihasilkan 351 kata yang benar dan 15 kata mengalami kesalahan. Hasil penelitian ini memperoleh akurasi sebesar 95,9% [22].

Dalam keseluruhan penelitian-penelitian tersebut, penggunaan algoritma Nazief & Adriani telah terbukti efektif dalam penggunaan stemming terhadap bahasa Madura. Dengan demikian, dalam konteks penelitian ini, algoritma Nazief & Adriani dapat dipilih sebagai algoritma terbaik untuk penggunaan stemming terhadap bahasa yang diteliti.

Berdasarkan dari proses pengujian pada penelitian ini, algoritma yang dikembangkan dapat menemukan kata dasar dengan awalan dan akhiran. Namun, ada beberapa kata yang mengalami overstemming tanpa ada kata yang mengalami understemming.

Tabel 6. Kata overstemming

No	kata	indonesia
1	atari	tar
2	arasan	ras
3	èkakan	kak
4	èangko'	suh
5	èasuh	sang
6	èpasang	sang
7	èpakala	pakala
8	èpaalos	paalos
9	apagherra	pagherra
10	ngellu'	ellu'
11	ngaca	aca
12	ngakan	kan
13	ngabin	bin
...
22	tabukka'	bukk

Tabel 6, menunjukkan kata-kata yang mengalami overstemming yang hanya terjadi pada awalan, kesalahan tersebut seperti pada kata "atari", hasil stemmingnya adalah "tar". Pemotongan kata yang benar adalah "a-tari". Hal ini terjadi karena huruf awal "a" dan huruf "akhir" "i" dibaca sebagai imbuan.

Proses pengujian yang dilakukan terhadap 650 kata yang diujikan dihasilkan 628 kata yang benar dan 22 kata yang mengalami kesalahan. Dengan akurasi keseluruhan 96,61%. nilai akurasi ini menunjukkan bahwa Nazief & Adriani dapat digunakan untuk stemming bahasa Madura dengan melakukan beberapa penyesuaian aturan berdasarkan morfologi bahasa Madura. Tingkat keberhasilan tertinggi terdapat pada prefix, dan suffix terendah. Hal ini salah satunya disebabkan oleh tidak adanya aturan morfonemik pada penelitian ini, seperti pelepasan konsonan dari "ng" menjadi "k". Sehingga kata "ngakan" menjadi "kakan". Penyesuaian algoritma juga belum dapat menyelesaikan konflik, infiks, dan pengulangan. Sehingga perlu ditambahkan aturan yang sesuai dengan morfologi bahasa Madura. Penambahan aturan ini diharapkan dapat meningkatkan akurasi stemming.

5. KESIMPULAN

Berdasarkan dari hasil penelitian stemming pada bahasa Madura menggunakan algoritma Nazief & Adriani, diperoleh beberapa kesimpulan. Akurasi stemming dengan Nazief & Adriani didapatkan akurasi sebesar 96,61%, dimana akurasi untuk awalan (prefix) mencapai 95,6%, dan untuk akhiran (suffix) mencapai 100%. dalam analisis kesalahan, ditemukan bahwa kesalahan pada hasil stemming yang dilakukan cenderung terjadi pada overstemming tanpa adanya understemming. Hal ini juga menunjukkan bahwa algoritma Nazief & Adriani dapat disesuaikan dengan baik untuk digunakan dalam stemming bahasa Madura. Penyesuaian aturan pembentukan kata dalam bahasa Madura dapat membantu meningkatkan akurasi stemming dan kebergunaan algoritma ini.

DAFTAR PUSTAKA

- [1] F. A. Ahda, A. P. Wibawa, D. Dwi Prasetya, and D. Arbian Sulisty, "Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation," *JOIV Int. J. Inform. Vis.*, vol. 8, no. 1, p. 231, Mar. 2024, doi: 10.62527/joiv.8.1.1818.
- [2] I. Irwiandi and M. Norman, "Proses Morfologis pada Bahasa Madura: Studi pada Mahasiswa Madura di Universitas Trunojoyo," *AJER Algazali Int. J. Educ. Res.*, vol. 5, no. 1, pp. 68–75, Oct. 2022, doi: 10.59638/ajer.v5i1.329.
- [3] R. Maulidi, "STEMMER UNTUK BAHASA MADURA DENGAN MODIFIKASI METODE ENHANCED CONFIX STRIPPING STEMMER," *ISSN.*, 2016.
- [4] Indri Tri Julianto, D. Kurniadi, and B. B. Balilo Jr, "ENHANCING SENTIMENT ANALYSIS WITH CHATBOTS: A COMPARATIVE STUDY OF TEXT PRE-PROCESSING," *J. Tek. Inform. Jutif*, vol. 4, no. 6, pp. 1419–1430, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1448.
- [5] V. Amrizal, A. Munandar, and A.- Arini, "IDENTIFIKASI MATAN HADITS MENGGUNAKAN NATURAL LANGUAGE PROCESSING DAN ALGORITMA KNUTH MORRIS PRATT BERBASIS WEB," *J. CoreIT J. Has. Penelit. Ilmu Komput. Dan Teknol. Inf.*, vol. 5, no. 2, p. 56, Dec. 2019, doi: 10.24014/coreit.v5i2.8477.
- [6] S. Tuhpatussania, E. Utami, and A. D. Hartanto, "COMPARISON OF PORTERS STEMMING ALGORITHM AND NAZIEF & ADRIANI'S STEMMING ALGORITHM IN DETERMINING INDONESIAN LANGUAGE LEARNING MODULES," *J. Pilar Nusa Mandiri*, vol. 18, no. 2, pp. 203–210, Sep. 2022, doi: 10.33480/pilar.v18i2.3940.
- [7] S. B. Rossi Hersianie, "ANALISA MODIFIKASI ALGORITMA STEMMING UNTUK KASUS OVERSTEMMING," *TEKNOKOM*, vol. 3, no. 2, pp. 23–28, Dec. 2020, doi: 10.31943/teknokom.v3i2.51.
- [8] I. M. A. Agastya, "PENGARUH STEMMER BAHASA INDONESIA TERHADAP PEFORMA ANALISIS SENTIMEN TERJEMAHAN ULASAN FILM," *J. Tekno Kompak*, vol. 12, no. 1, p. 18, Feb. 2018, doi: 10.33365/jtk.v12i1.70.
- [9] A. F. Aji *et al.*, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 7226–7249. doi: 10.18653/v1/2022.acl-long.500.
- [10] D. A. Sulisty, "LSTM-Based Machine Translation for Madurese-Indonesian," *J. Appl. Data Sci.*, vol. 4, no. 3, pp. 189–199, Sep. 2023, doi: 10.47738/jads.v4i3.113.
- [11] I Putu Satwika, S. Kom., M. Kom. and Helmy Syahk Alam, "ALGORITMA STEMMING DALAM BAHASA BALI MENGGUNAKAN PENDEKATAN N-GRAM," *Smart Techno Smart Technol. Inform. Technopreneurship*, vol. 2, no. 2, pp. 55–63, Sep. 2020, doi: 10.59356/smart-techno.v2i2.22.
- [12] F. H. Rachman, N. Ifada, S. Wahyuni, G. D. Ramadani, and A. Pawitra, "ModifiedECS (mECS) Algorithm for Madurese-Indonesian Rule-Based Machine Translation," in *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, Denpasar, Bali, Indonesia: IEEE, Nov. 2022, pp. 51–56. doi: 10.1109/ICSINTESA56431.2022.10041470.
- [13] W. Hidayat, E. Utami, and A. D. Hartanto, "Effect of Stemming Nazief & Adriani on the Ratcliff/Obershelp algorithm in identifying level of similarity between slang and formal words," in *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia: IEEE, Nov. 2020, pp. 22–27. doi: 10.1109/ICOIACT50329.2020.9331973.
- [14] S. Firman Sodiq, W. Desena, and A. Wibowo, "Penerapan Algoritma Stemming

- Nazief & Adriani Pada Proses Klasterisasi Berita Berdasarkan Tematik Pada Laman (Web) Direktorat Jenderal HAM Menggunakan Rapidminer,” *Syntax J. Inform.*, vol. 11, no. 02, pp. 10–21, Nov. 2022, doi: 10.35706/syji.v11i02.7192.
- [15] I. P. M. Wirayasa, I. M. A. Wirawan, and I. M. A. Pradnyana, “ALGORITMA BASTAL: ADAPTASI ALGORITMA NAZIEF & ADRIANI UNTUK STEMMING TEKS BAHASA BALI,” *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 8, no. 1, p. 60, Jun. 2019, doi: 10.23887/janapati.v8i1.13500.
- [16] M. Fauziah, “JURUSAN TEKNIK INFORMATIKA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG 2019”.
- [17] M. Anjani and H. Nurramdhani, “COMPARISON PERFORMANCE OF WORD2VEC, GLOVE, FASTTEXT USING SUPPORT VECTOR MACHINE METHOD FOR SENTIMENT ANALYSIS”.
- [18] A. Prasadhatama and K. M. Suryaningrum, “PERBANDINGAN ALGORITMA NAZIEF & ADRIANI DENGAN ALGORITMA IDRIS UNTUK PENCARIAN KATA DASAR,” *J. Teknol. Dan Manaj. Inform.*, vol. 4, no. 1, Jan. 2018, doi: 10.26905/jtmi.v4i1.1773.
- [19] M. A. Nq, L. P. Manik, and D. Widiatmoko, “Stemming Javanese: Another Adaptation of the Nazief-Adriani Algorithm,” in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia: IEEE, Dec. 2020, pp. 627–631. doi: 10.1109/ISRITI51436.2020.9315420.
- [20] G. Septian, A. Susanto, and G. F. Shidik, “Indonesian news classification based on NaBaNA,” in *2017 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang: IEEE, Oct. 2017, pp. 175–180. doi: 10.1109/ISEMANTIC.2017.8251865.
- [21] E. Lindrawati, E. Utami, and A. Yaqin, “Comparison of Modified Nazief&Adriani and Modified Enhanced Confix Stripping algorithms for Madurese Language Stemming,” *INTENSIF J. Ilm. Penelit. Dan Penerapan Teknol. Sist. Inf.*, vol. 7, no. 2, pp. 276–289, Aug. 2023, doi: 10.29407/intensif.v7i2.20103.
- [22] A. P. Wibawa, F. A. Dwiyanto, I. A. E. Zaeni, R. K. Nurrohman, and A. Afandi, “Stemming javanese affix words using nazief and adriani modifications,” *J. Inform.*, vol. 14, no. 1, p. 36, Jan. 2020, doi: 10.26555/jifo.v14i1.a17106.