

Vol. 2, No. 1, April 2013

# Jurnal Sarjana Teknik Informatika

**Sistem Informasi Dan  
Rekayasa Perangkat Lunak**

**Komputasi Cerdas**

**Sistem Terdistribusi**

**Digital Multimedia**

**Teknik Informatika  
Fakultas Teknik  
Universitas Trunojoyo**

ISSN 2338-5650



9 772338 565001



**Vol. 2, No 1, April 2013**

ISSN 2338-5650  
ISSN Online 2338-7114

Jurnal Sarjana Teknik Informatika, terbit dua kali setahun pada bulan Oktober dan April, memuat naskah hasil Tugas Akhir mahasiswa Program Studi Teknik Informatika Universitas Trunojoyo Madura

**Ketua Redaksi**

Yonathan Ferry Hendrawan

**Redaksi Pelaksana**

Dwi Kuswanto  
Arik Kurniawati  
Eza Rahmanita  
Sri Wahyuni

**Reviewer**

Rima Tri Wahyuningrum  
Fitri Damayanti  
Wahyudi Setiawan  
Bain Khusnul Khotimah  
M. Kautsar Sophan  
Andharini Dwi Cahyani  
Rika Yunitarini  
Ari Kusumaningsih  
Cucun Very Angkoso  
Iwan Santosa  
Fika Hastarita Rachman

**Tata Usaha**

Hery Roosmawan  
Husnul Hotimah

**Alamat Redaksi dan Tata Usaha**

Jurnal Sarjana Teknik Informatika, Program Studi Teknik Informatika  
Kampus Universitas Trunojoyo Madura  
Jl. Raya Telang PO BOX 2 Kamal, Bangkalan, 69192, Jawa Timur  
Telp. (031) 30 111 47, Fax (031) 30 115 06  
Email: [jstif@if.trunojoyo.ac.id](mailto:jstif@if.trunojoyo.ac.id)  
Website: <http://jstif.trunojoyo.ac.id/>



## DAFTAR ISI

Vol. 2, No. 1, April 2013

ISSN 2338-5650  
ISSN Online 2338-7114

### Bidang Minat Sistem Informasi dan Rekayasa Perangkat Lunak

<b>SISTEM INFORMASI GEOGRAFIS PEMETAAN WILAYAH BERDASARKAN TINGKAT KESEJAHTERAAN DI KABUPATEN BANGKALAN MENGGUNAKAN METODE FUZZY MAMDANI</b> Hendriatno, Firli Irhamni	<b>1</b>
<b>ANALISA DATA ANTARAN POS EXPRESS MENGGUNAKAN METODE <i>RADIAL BASIS FUNCTION NEURAL NETWORK</i> (RBF-NN) STUDY PT. POS INDONESIA SURABAYA</b> Siti Murtasyah Warda, Firli Irhamni	<b>13</b>
<b>RANCANG BANGUN SISTEM INFORMASI PEMETAAN TANAMAN PERTANIAN DI KABUPATEN BANGKALAN</b> Ridha Marista, Diana Rahmawati, Husni	<b>25</b>
<b>RANCANG BANGUN MANAJEMEN PEMBELAJARAN DAN TES TOEFL BERBASIS MOBILE</b> Susanto, Hermawan, Andharini Dwi Cahyani	<b>37</b>
<b>ANALISA VALIDITAS PENERIMA BEASISWA MENGGUNAKAN <i>ANALYTIC NETWORK PROCESS</i> (ANP) DAN TOPSIS</b> Hilmi Fairuz Abadi, Bain Khusnul Khotimah, Firli Irhamni	<b>45</b>
<b>PENGEMBANGAN MESIN PENCARIAN ANTI PLAGIASI PADA SIM JURNAL MAHASISWA MENGGUNAKAN ALGORITMA <i>WINNOWING FUZZY K-MEANS</i></b> Ilham Wibisono Aziz, Hermawan, Andharini Dwi Cahyani	<b>55</b>
<b>SISTEM PEMILIHAN SUPPLIER DI APOTEK MENGGUNAKAN METODE <i>FUZZY ANALYTICAL NETWORK PROCESS</i> (FANP)</b> Erwanto Indra Sasmita, Firli Irhamni, Diana Rahmawati	<b>65</b>
<b>ANALISIS DAN PENERAPAN ALGORITMA PARTICLE SWARM OPTIMIZATION PADA OPTIMASI PENJADWALAN MATA KULIAH DI JURUSAN TEKNIK INFORMATIKA UNIVERSITASS TRUNOJOYO MADURA</b> Qomarul Haryadi Irfan Rifa'i, Arik Kurniawati	<b>77</b>
<b>SISTEM PERAMALAN KEBUTUHAN OBAT MENGGUNAKAN METODE JARINGAN SYARAF TIRUAN ALGORITMA BACKPROPAGATION DENGAN FUNGSI LYAPUNOV</b> Rahmatina Hidayati, Bain Khusnul Khotimah, Haryanto	<b>89</b>
<b>SISTEM PERAMALAN PENJUALAN JANGKA PENDEK <i>SPARE PART</i> SEPEDA MOTOR MENGGUNAKAN <i>NEURAL NETWORK</i></b> Ana Qoimah, Mula'ab, Fika Hastarita Rachman	<b>101</b>

<b>SISTEM PENDUKUNG KEPUTUSAN PENENTUAN PEMBERIAN BANTUAN RASKIN MENGGUNAKAN METODE <i>SMARTER</i></b> Ayu Winda Istara, Andharini Dwi Cahyani, Fika Hastarita Rachman	<b>113</b>
<b>KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN <i>KOHONEN SELF ORGANIZING MAP</i></b> Uhty Zunairoh, Bain Khusnul Khotimah, Aeri Rachmad	<b>121</b>
<b>PENERAPAN METODE <i>RADIAL BASIS FUNCTION NETWORK</i> DENGAN <i>K-MEANS CLUSTER</i> UNTUK PERAMALAN KEBUTUHAN <i>STRAW</i></b> Dwi Nuzulul Heriyanto, Fika Hastarita Rachman, Budi Dwi Satoto	<b>133</b>
<b>APLIKASI DAN DESAIN MODEL DECISION AID PELANGGAN PADA E-COMMERCE TOKO BATIK TULIS MADURA</b> Ardiyanto Setiawan, Mochammad Kautsar Sophan, Yeni Kustiyah Ningsih	<b>143</b>
<b>SISTEM PARAMALAN TINGKAT PENGANGGURAN WILAYAH BANGKALAN MENGGUNAKAN ALGORITMA BACKPROPAGATION</b> Ali Imron Zamzami, Bain Khusnul Khotimah, Firli Irhamni	<b>155</b>
<b>PENJADWALAN PRODUKSI ROKOK MENGGUNAKAN METODE SHORTEST PROCESSING TIME UNTUK Mendukung PENERAPAN SUPPLY CHAIN MANAGEMENT</b> Erwin Andriyansyah, Yeni Kustiyah Ningsih	<b>163</b>
<b>SISTEM PERAMALAN DISTRIBUSI BERAS MENGGUNAKAN METODE EXPONENTIAL SMOOTHING</b> Hendra Pramana Putra, Fika Hastarita Rachman	<b>173</b>
<b>IMPLEMENTASI PERFORMANCE PRISM UNTUK SISTEM PENGUKURAN KINERJA PERUSAHAAN</b> Luhur Priyantoko Aji, Bain Khusnul Khotimah, Khamdi Mubarak	<b>181</b>
<b>PERANCANGAN SISTEM INFORMASI PERENCANAAN PRODUKSI DAN PERSEDIAAN BAHAN BAKU TINTA MENGGUNAKAN METODE EOQ</b> Muhammad Arif Rahman, Yeni Kustiyahningsih, Heri Awalul Ilhamsyah	<b>193</b>
<b>ANALISIS PERAMALAN PENJUALAN PRODUK ASESORIS KOMPUTER MENGGUNAKAN METODE <i>EXPONENTIAL SMOOTHING (ES)</i> DAN <i>ORDINARY LEAST SQUARE (OLS)</i></b> Mohammad Hasbi Ash Shiddieqy, Mula'ab, Yeni Kustiyahningsih	<b>201</b>
<b>SISTEM PENENTU CALON PENERIMA BEASISWA IKSASS MENGGUNAKAN METODE <i>SMARTER &amp; FORWARD CHAINING</i></b> Moh Rudi, Achmad Ubaidillah	<b>209</b>
<b>PENERAPAN METODE <i>ANALYTICAL HIERARCHY PROCESS</i> DAN <i>FUZZY MULTI-OBJECTIVE PROGRAMMING</i> UNTUK PEMILIHAN <i>SUPPLIER</i></b> Heliyatin, Bain Khusnul Khotimah, Andharini Dwi Cahyani	<b>219</b>
<b>IMPLEMENTASI SISTEM PAKAR PADA PENGAMBILAN KEPUTUSAN PENENTUAN TINDAKAN MEDIS DI POLI MATA BERBASIS WEB</b> Rizky Maulidya, Firli Irhamni	<b>229</b>

<b>MODEL PENJUALAN PRODUK UNGGULAN BATIK KHAS PAMEKASAN BERBASIS <i>ELECTRONIC COMMERCE</i></b> Murni Rotifah, Moch. Kautsar Sophan	<b>239</b>
<b>PENGELOMPOKAN KUALITAS KELAS PADA SISWA MENGGUNAKAN <i>INDEKS DIVIES-BOULDIN SOM</i></b> Hamayah, Bain Khusnul Khotimah, Firli Irhamni	<b>251</b>
<b>RANCANG BANGUN CRM BERBASIS SMS <i>GATEWAY</i> DISERTAI PERAMALAN PERGUDANGAN DENGAN METODE <i>FUZZY TIME SERIES</i></b> Akhmad Tajuddin Tholaby MS., Fika Hastarita Rachman, Firli Irhamni	<b>263</b>
<b>SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN KARYAWAN BERPRESTASI DENGAN INTEGRASI FAHP DAN ELECTRE II</b> Catur Ngesti Waluyo, Mula'ab, Riza alfita	<b>275</b>
<b>APLIKASI PENILAIAN KINERJA DENGAN METODOLOGI <i>CUSTOMER RELATIONSHIP MANAGEMENT (CRM)</i></b> Luthfi Zahro, Mochammad Kautsar Sophan, Rachmad Hidayat	<b>285</b>
 <b><u>Bidang Minat Sistem Terdistribusi</u></b>	
<b>APLIKASI PENERJEMAH BAHASA MADURA-INDONESIA DAN INDONESIA-MADURA MENGGUNAKAN <i>FREE CONTEX PARSING ALGORITHM</i></b> Rahilah, Firdaus Solihin, Fika Hastarita Rachman	<b>295</b>
<b>PENERAPAN MODIFIKASI METODE <i>ENHANCED CONFIX STRIPPING STEMMER</i> PADA TEKS BERBAHASA MADURA</b> Andi Sholihin, Firdaus Solihin, Fika Hastarita Rachman	<b>305</b>
<b>RANCANG BANGUN APLIKASI SEARCH ENGINE DAN SISTEM PENDETEKSI PLAGIARISME MENGGUNAKAN METODE LSA-SOM</b> Aris Prama Aulia, Hermawan, Indah Agustin Siradjuddin	<b>315</b>
 <b><u>Bidang Minat Komputasi Dan Sistem Cerdas</u></b>	
<b>SISTEM PEROLEHAN CITRA BERBASIS ISI MENGGUNAKAN WAVELET HAAR DAN EUCLIDIEN DISTANCE</b> Vembi Pratama Pristyono	<b>327</b>
<b>PENGENALAN TELAPAK TANGAN MENGGUNAKAN METODE <i>NAÏVE BAYES</i> BERBASIS REDUKSI DIMENSI <i>PRINCIPAL COMPONENT ANALYSIS (PC A)</i></b> Daril Ulumiyah	<b>339</b>
<b>PENGENALAN TULISAN TANGAN CARAKAN JAWA SECARA <i>REAL TIME</i> DENGAN MENGGUNAKAN <i>DOMINANT POINT</i></b> Nining Latifah	<b>349</b>
<b>APLIKASI CITRA <i>MOSAIC PANORAMIC</i> MENGGUNAKAN METODE SIFT (<i>SCALE INVARIANT FEATURE TRANSFORM</i>)</b> Gendra Budiarti	<b>359</b>

<b>SISTEM PEROLEHAN CITRA BERBASIS ISI MENGGUNAKAN GRLM BERDASARKAN CIRI TEKSTUR PADA POLA BATIK</b> Priagung Safara Dila	<b>371</b>
<b>Pengenalan Tipe Huruf Aksara di Indonesia Menggunakan Edge Direction Matrix</b> Sulayhah, Bain Khusnul Khotimah, Rima Tri Wahyuningrum	<b>381</b>
<b>Pengenalan Wajah Menggunakan <i>KERNEL PRINCIPAL COMPONENT ANALYSIS</i> (KCPA)</b> Yulia Fatmawati, Rima Tri Wahyuningrum, Fitri Damayanti	<b>389</b>
<b>Segmentasi Parasit Malaria dalam Sel Darah Menggunakan Operasi Morfologi dan Transformasi <i>WATERSHED</i></b> Yenni Rahmawati, Rima Tri Wahyuningrum, Fitri Damayanti	<b>401</b>
 <b><u>Bidang Minat Multimedia</u></b>	
<b>AUTO LEVELING BERBASIS <i>FINITE STATE MACHINE</i> (FSM) PADA GAME PEMBELAJARAN BAHASA INGGRIS</b> Syafrial Nur Maulana, Arik Kurniawati, Ari Kusumaningsih	<b>409</b>
<b>APLIKASI CITRA MOSAIK PANORAMIK MENGGUNAKAN METODE <i>WATERSHED</i></b> Sri Rahma Puspa Sari	<b>421</b>
<b>IMPLEMENTASI METODE N-GRAM PADA PEMBUATAN GAME <i>FIGHTING MADURA</i> BERBASIS ANDROID <i>OS MOBILE</i></b> Farizal Arifin, Arik Kurniawati, Cucun Very Angkoso	<b>429</b>
<b>APLIKASI PENENTUAN STATUS GIZI BALITA MENGGUNAKAN NAIVE BAYES</b> Maimuna Mahdalena, Diana Rahmawati, Riza Alfita	<b>439</b>



# **PENERAPAN MODIFIKASI METODE *ENHANCED CONFIX STRIPPING STEMMER* PADA TEKS BERBAHASA MADURA**

**Andi Sholihin, Firdaus Solihin, Fika Hastarita Rachman**

Program Studi Teknik Informatika, Universitas Trunojoyo

Jl. Raya Telang, PO BOX 2, Kamal, Bangkalan - 69162

*E-mail:* dieka.91@gmail.com, fsolihin@gmail.com, hastarita.fika@gmail.com

## **ABSTRAK**

*Stemming* adalah teknik untuk menemukan kata dasar dari sebuah kata yang telah mengalami imbuhan dengan cara menghilangkan atau menghapus imbuhan-imbuhan tersebut. Algoritma *Enhanced Confix Stripping Stemmer* adalah algoritma *stemming* yang akurat untuk mencari bentuk dasar dari suatu kata dalam Bahasa Indonesia. Bahasa Madura memiliki struktur morfologi yang mirip dengan Bahasa Indonesia. Bahasa Madura memiliki kata yang dapat dibentuk dari awalan (*ter-ater*), akhiran (*panoteng*) dan sisipan (*sesselan*) dan Kata Ulang (*Rangkebbhan*). Modifikasi algoritma *Enhanced Confix Stripping Stemmer* dilakukan pada tabel aturan pemenggalan, langkah-langkah algoritma dan penambahan langkah algoritma sesuai dengan morfologi Bahasa Madura. Dengan demikian dapat dihasilkan sebuah algoritma baru yang dikembangkan dari algoritma *Enhanced Confix Stripping Stemmer*. Setelah melakukan beberapa analisa dan pengujian dengan menggunakan sistem yang telah dibangun maka diperoleh hasil dari modifikasi algoritma *Enhanced Confix Stripping Stemmer* pada teks berbahasa Madura. Persentase kata Bahasa Madura yang berhasil *distemming* adalah 95,75 % dari 400 kata yang diuji. Dengan demikian modifikasi *Enhanced Confix Stripping Stemmer* pada teks berbahasa Madura dinilai efektif untuk diterapkan.

Kata kunci : *stemming*, *Enhanced Confix Stripping Stemmer*, Madura.

## **ABSTRACT**

*Stemming* is a technique to find the root of a word that has been experienced by removing affixes. *Enhanced Algorithm Confix Stripping Stemmer* is accurate stemming algorithm to find the basic form of a word in Indonesian. Madurese language morphological structure similar to Indonesian. Madurese language has a word that can be formed from the prefix (*ter-ater*), suffix (*panoteng*) and insertions (*sesselan*) and Reduplication (*Rangkebbhan*). Modifications *Enhanced Confix Stripping Stemmer* algorithm performed on the table decapitation rules, measures and additional step algorithm according to the morphology of the Madurese Language. Thus it can be produced by a newly developed algorithm of algorithms *Enhanced Confix Stripping Stemmer*. After doing some analysis and testing using a system that has built the obtained results of the modified algorithm *Confix Enhanced Stripping Stemmer* in Madurese language text. The percentage of successful Madurese language word stemming is 95.75% of the 400 words that were tested. Thus the modification of *Enhanced Confix Stripping Stemmer* on Madura language text considered effective to implement.

*Keywords:* *stemming*, *Enhanced Confix Stripping Stemmer*, *madurese*.

## PENDAHULUAN

Penelitian dan pengembangan Sistem Temu Kembali Informasi semakin menunjukkan peningkatan yang signifikan baik kualitas maupun kuantitasnya. Hal tersebut ditandai dengan semakin banyaknya penelitian tentang *stemming*, *search engine* dan lain sebagainya yang memusatkan perhatian pada Bahasa Indonesia. Selain itu perbaikan demi perbaikan terhadap metode-metode *Information Retrieval* untuk teks dapat memperkecil tingkat kesalahan yang terjadi.

Penelitian tentang *stemming* antara lain adalah Algoritma *Confix Stripping Stemmer* untuk mencari bentuk kata dasar dari kata berimbuhan pada Bahasa Indonesia sebagai perbaikan terhadap algoritma *stemming* pendahulunya yaitu algoritma Nazief Andriani [1]. *Confix* dapat diartikan sebagai *combination of prefix and suffix* (kombinasi imbuhan awalan dan akhiran). Algoritma *Confix Stripping Stemmer* terbukti dapat melakukan *stemming* dengan hasil yang lebih baik daripada algoritma Nazief Andriani [1].

Algoritma *Confix Stripping Stemmer* memiliki beberapa kelemahan sehingga dilakukan perbaikan terhadap algoritma tersebut. Hasil perbaikan dari algoritma *Confix Stripping Stemmer* diberi nama *Enhanced Confix Stripping Stemmer*. Algoritma *Enhanced Confix Stripping Stemmer* dapat menangani kesalahan yang tidak dapat diatasi dengan algoritma *Confix Stripping Stemmer* [2].

Tujuan penelitian ini adalah untuk menerapkan modifikasi algoritma *ECS Stemmer* pada teks berbahasa Madura. Batasan masalah dalam penelitian ini adalah metode yang akan digunakan dalam penelitian ini hanya metode *Enhanced Confix Stripping Stemmer* yang akan dimodifikasi pada tabel aturan pemenggalan imbuhan untuk disesuaikan dengan Bahasa Madura, kata berbahasa Madura yang akan digunakan dalam penelitian adalah kata yang umum digunakan di kabupaten Bangkalan karena kamus yang digunakan umumnya memuat kata dalam dialek Bangkalan, serta pengujian modifikasi *ECS Stemmer* untuk teks berbahasa Madura ini menggunakan aplikasi yang sederhana dan tidak menggunakan koleksi dokumen (*corpus*) berbahasa Madura.

## METODE

### Algoritma *Enhanced Confix Stripping Stemmer*

Algoritma *Enhanced Confix Stripping Stemmer* dapat menangani kesalahan yang tidak dapat diatasi dengan algoritma *Confix Stripping Stemmer*. Algoritma ini dinilai lebih akurat dibandingkan algoritma pendahulunya [2].

Perbaikan yang dilakukan oleh algoritma *Enhanced Confix Stripping Stemmer* dapat dijelaskan sebagai berikut:

1. Merevisi aturan 19 pada Tabel 1 agar *stemming* berhasil pada kata-kata dengan format “mem+p...”.

2. Merevisi aturan 14 pada Tabel 1 agar *stemming* berhasil pada kata-kata dengan format “men+s...”.
3. Merivisi aturan 17 pada Tabel 1 agar *stemming* berhasil pada kata-kata dengan format “menge+...”.
4. Merivisi aturan 30 pada Tabel 1 agar *stemming* berhasil pada kata-kata dengan format “penge+...”. Merevisi aturan 29 pada Tabel 1 agar *stemming* berhasil pada kata-kata dengan format “peng+k...”.
5. Menambahkan suatu algoritma tambahan untuk mengatasi kesalahan pemenggalan akhiran yang seharusnya tidak dilakukan. Algoritma ini disebut *loop Pengembalian Akhiran*, dan dilakukan apabila proses *recoding* gagal. Algoritma *loop Pengembalian Akhiran* dideskripsikan sebagai berikut:
  - a. Kembalikan seluruh awalan yang telah dihilangkan sebelumnya, sehingga menghasilkan model kata seperti berikut:  
**[DP+[DP+[DP]]] + Kata Dasar.**
  - b. Pemenggalan awalan dilanjutkan dengan proses pencarian di kamus kemudian dilakukan pada kata yang telah dikembalikan menjadi model tersebut.
  - c. Kembalikan akhiran sesuai dengan urutan model. Ini berarti bahwa pengembalian dimulai dari DS (“-i”, “-kan”, “-an”), lalu PP (“-ku”, “-mu”, “-nya”), dan terakhir adalah P (“-lah”, “-kah”, “-tah”, “-pun”). Untuk setiap pengembalian, lakukan langkah 3) hingga 5) berikut. Khusus untuk akhiran “-kan”, pengembalian pertama dimulai dengan “k”, baru kemudian dilanjutkan dengan “an”.
  - d. Lakukan pengecekan di kamus. Apabila ditemukan, proses dihentikan. Apabila gagal, maka lakukan proses pemenggalan awalan berdasarkan aturan pada Tabel 1 (dengan revisi Tabel 2).
  - e. Lakukan *recoding* apabila diperlukan.

Apabila pengecekan di kamus tetap gagal setelah *recoding*, maka awalan-awalan yang telah dihilangkan dikembalikan lagi.

### **Morfologi Bahasa Madura**

Bahasa Madura merupakan salah satu bahasa daerah yang terdapat di Indonesia. Bahasa Madura digunakan secara umum di pulau Madura yaitu di Kabupaten Bangkalan, Sampang, Pamekasan, dan Sumenep. Selain itu, Bahasa Madura juga digunakan di daerah lain di luar pulau Madura baik di dalam negeri maupun di luar negeri.

Bahasa Madura memiliki karakter khusus dengan tujuan supaya mudah dibaca baik oleh orang madura ataupun bukan orang madura. Karakter tersebut adalah sebagai berikut [5].

- a. a = dibaca a biasa seperti pada kata awas
- b. â = dibaca e seperti pada kata belajar

- c. e = dibaca e biasa seperti pada kata kertas
- d. è = dibaca e seperti pada kata bebas
- e. bh, dh, gh dan jh = dibaca tebal, contoh : bhâjâ, dhâbu, ghâbây dan bâjhâ.
- f. tanda petik (‘) = dibaca seperti pada kata tidak.

Tabel 1 Aturan pemenggalan awalan

Aturan	Format Kata	Pemenggalan	Aturan	Format Kata	Pemenggalan
1	berV...	ber-V...   be-rV...	9	teC <sub>1</sub> erC <sub>2</sub> ...	te-C <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> !=’r’
2	berCAP...	ber-CAP... dimana C!=’r’ & P!=’er’	10	me{l r w y}V...	me- {l r w y} V...
3	berCAerV...	ber-CaerV... dimana C!=’r’	11	mem{b f v}...	mem- {b f v}...
4	belajar	bel-ajar	12	mempe...	mem-pe...
5	beC <sub>1</sub> erC <sub>2</sub> ...	be-C <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> != {’r’ ’l’}	13	mem{rV V}...	me-m {rV V}...   mem {rV V}...
6	terV...	ter-V...   te-rV...	14	men{c d j z}...	men- {c d j z}...
7	TerCerV...	ter-CerV... dimana C!=’r’	15	menV...	me-nV...   me-tV...
8	terCP...	ter-CP... dimana C!=’r’ dan P!=’er’	16	meng{g h q k}...	meng- {g h q k}...
19	mempV...	mem-pV dengan V!=’e’	17	meng-V...	meng-V...   meng-kV...
20	pe{w y}V...	pe- {w y} V...	18	menyV...	meny-sV...
21	perV...	per-V...   pe-rV...	30	pengV...	peng-V...   peng-kV...
23	perCAP	per-CAP... dimana C!=’r’ dan P!=’er’	31	penyV...	peny-sV...
24	perCAerV...	per-CaerV... dimana C!=’r’	32	peIV	pe-IV... kecuali “pelajar” yang menghasilkan “ajar”
25	pem{b f V}...	pem- {b f V}...	33	peCerV...	per-erV... dimana C!= {r w y l m n}
26	pem{rV V}...	pe-m {rV V}...   pe-p {rV V}...	34	peCP...	pe-CP... dimana C!= {r w y l m n} dan P!=’er’
27	pen{c d j z}...	pen- {c d j z}...	35	terC <sub>1</sub> erC <sub>2</sub> ...	ter-C <sub>1</sub> erC <sub>2</sub> ... dimana C!=’r’
28	penV...	pe-nV...   pe-tV...	36	peC <sub>1</sub> erC <sub>2</sub> ...	pe-C <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> != {r w y l m n}
29	peng{g h q}...	peng- {g h q}...			

Tabel 2. Revisi untuk Tabel 1

Aturan	Format Kata	Pemenggalan
14	men{c d j s z}...	men- {c d j s z}...
17	mengV...	meng-V...   meng-kV... (mengV-... jika V=’e’)
19	mempA...	mem-pA... dengan A!=’e’
29	pengC...	peng-C...
30	pengV...	peng-V...   peng-kV... (pengV-... jika V=’e’)

Keterangan simbol huruf :

- C : huruf konsonan
- V : huruf vokal
- A : huruf vokal atau konsonan
- P : partikel atau fragmen dari suatu kata, misalnya “er”

Imbuhan dalam Bahasa Madura adalah sebagai berikut:

1. Ter-ater (awalan) yaitu : e-, a-, ta-, ma-, ka-, sa-, pa-, koma-, kame-, kape-, pe-, an-, any-, am-, ang-, pam-, pang-, pan-.
2. Sessellan (sisipan) yaitu: -ar-, -al-, -om-/ -um-, -en-.
3. Panoteng (akhiran) macamnya yaitu : -e, -a, -an, -en, -na, -ana, -aghi.
4. Rangkebban atau kata dalam bahasa Madura ada tiga macam, yaitu: kata ulang dwilingga, dwipurwa, dan kata ulang belakang [3].

### **Keterkaitan Bahasa Madura dengan Bahasa Indonesia**

Bahasa Madura dan Bahasa Indonesia memiliki keterkaitan bahasa baik dari sisi morfologi, fonologi maupun sintaknya [4]. Pola kata dalam Bahasa Madura banyak memiliki kemiripan dengan Bahasa Indonesia. Pola kata berimbuhan dalam Bahasa Madura juga memiliki pola yang mirip dengan Bahasa Indonesia.

Dalam pengamatan yang dilakukan penulis, Bahasa Madura memiliki model kata yang mirip dengan Bahasa Indonesia. Dalam Bahasa Madura dikenal adanya imbuhan kata berupa ter-ater (*Prefiks*), Panoteng (*suffiks*) dan Sesselan (sisipan) [3]. Jika menurut Jelita Asian dalam penelitiannya tentang algoritma *Confix Stripping Stemmer* memodelkan kata dalam Bahasa Indonesia sebagai berikut :

**[DP+[DP + [DP+]]] Kata Dasar [[+DS][+PP][+P]]**

Maka dalam Bahasa Madura, kata dasar yang mengalami afiksasi (imbuhan) dapat dimodelkan sebagai berikut [3]:

**[DP+[DP + [DP+]]] Kata Dasar [[+DS][+PP]]**

Perlu diketahui bahwa dalam Bahasa Madura tidak dikenal penggunaan partikel (P) yang umum digunakan dalam Bahasa Indonesia [3].

Dengan adanya pola tersebut maka dapat dimungkinkan penggunaan algoritma *Enhanced Confix Stripping Stemmer* untuk Bahasa Madura dengan syarat harus ada modifikasi baik dari segi tabel aturan pemenggalan imbuhan maupun dari segi langkah-langkah jalannya algoritma tersebut sehingga dapat sesuai dengan Bahasa Madura. Dengan demikian dapat ditemukan algoritma stemming yang baik untuk mencari bentuk dasar dari kata berimbuhan dalam Bahasa Madura.

### **Modifikasi *Enhanced Confix Stripping Stemmer* untuk Bahasa Madura**

Algoritma *Enhanced Confix Stripping Stemmer* adalah algoritma *stemming* yang akurat untuk mencari bentuk dasar dari suatu kata dalam Bahasa Indonesia [2]. Hal ini karena algoritma ECS menggunakan kamus sebagai acuan untuk menemukan kata dasar. Selain itu, tabel aturan pemenggalan yang dimiliki algoritma ECS dapat menyelesaikan permasalahan pemotongan pada kata yang bermorfologi.

Penerapan Algoritma ECS dalam Bahasa Madura dinilai efektif karena Bahasa Madura memiliki struktur kata yang mirip dengan Madura. Tentunya algoritma tersebut tidak dapat langsung diterapkan karena sejatinya penggunaan algoritma ECSS hanya terbatas untuk Bahasa Indonesia. Perlu adanya modifikasi pada algoritma ECSS sehingga dapat diterapkan untuk mencari bentuk dasar dari kata berbahasa Madura.

Modifikasi Algoritma *Enhanced Confix Stripping Stemmer* harus dilakukan agar dapat berfungsi dengan baik dan dapat diterapkan untuk Bahasa Madura. Modifikasi harus dilakukan pada tabel aturan pemenggalan, pada langkah-langkah algoritma dan penambahan langkah

algoritma. Dengan demikian dapat dihasilkan sebuah algoritma baru yang dikembangkan dari algoritma *Enhanced Confix Stripping Stemmer*. *Flowchart* algoritma dapat dilihat pada Gambar 1.

Langkah-langkah untuk mencari bentuk kata dasar dari teks berbahasa madura adalah sebagai berikut.

1. kata yang di *stemming* dicari di kamus, jika ada maka algoritma berhenti, jika tidak maka lakukan langkah 2.
2. Cek *rule precedence*, apabila kata tersebut memiliki kombinasi awalan dan akhiran yang tidak diperbolehkan (Tabel 3) maka urutan langkah berikutnya adalah langkah 5, 6, 3, 4, 7, 8, 9, 10. Apabila kata tersebut tidak memiliki pasangan awalan dan akhiran yang dilarang maka algoritma berjalan normal sesuai langkah 3, 4, 5, 6, 7, 8, 9, 10.
3. Hilangkan kata ganti kepemilikan (*possessive pronoun*) (“-na”, ”-èpon”).
4. Hilangkan *derivation suffixes* (akhiran) (“-e”, ”-a”, “-an”, “-en”, “-ana”, “-aghi”).
5. Hilangkan *derivational prefixes* (awalan) dengan iterasi maksimum sampai 3 kali:
  - a. Langkah 5 berhenti jika:
    1. Terjadi imbuhan terlarang (*rule precedence*).
    2. Awalan yang terdeteksi saat ini sama dengan yang dihilangkan sebelumnya.
    3. Tiga awalan telah dihilangkan.
  - b. Identifikasikan tipe awalan dan hilangkan, awalan ada 2 tipe:
    1. Standar : a-, è-, dll dapat langsung dihilangkan.
    2. Kompleks : am-, an-, any-, ang adalah tipe-tipe awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya. Gunakan tabel aturan pemenggalan awalan (Tabel 4).
    3. Cari kata yang telah dihilangkan awalannya ini dalam kamus. Jika ada maka keseluruhan proses dihentikan. Jika tidak ada maka ulangi langkah 5 ini.
6. Jika setelah langkah 5 kata dasar masih belum ditemukan maka proses *recoding* dilakukan dengan mengacu pada tabel aturan pemenggalan awalan pada Tabel 4.
7. Hilangkan sisipan sesuai dengan tabel aturan pemenggalan pada Tabel 5.
8. Hilangkan imbuhan pada kata ulang dwipurwa jika terdeteksi dengan kata tersebut memiliki pola  $C+V+C$  dan ternyata  $C=C$  maka kemungkinan kata tersebut adalah kata ulang dwipurwa (perulangan depan) sehingga hapus  $C+V$ . dengan  $C$  adalah konsonan dan  $V$  adalah huruf vokal
9. Apabila kata tidak ditemukan di kamus, maka lakukan *LoopPengembalianAkhiran* dengan algoritma sebagai berikut:
  - a. Kembalikan seluruh awalan yang telah dihilangkan sebelumnya sehingga menghasilkan model kata seperti berikut:  
$$[DP+[DP+[DP]]]+kata\ dasar$$
  - b. Pemenggalan awalan dilanjutkan dengan proses pencarian di kamus kemudian dilakukan dengan proses pencarian di kamus kemudian dilakukan pada kata yang telah dikembalikan menjadi model tersebut.

- c. Kembalikan akhiran sesuai dengan urutan model

**[DP+[DP+[DP]]]+katadasar+[DS+[PP]]**

untuk setiap pengembalian, lakukan langkah 3 hingga 5 berikut.

- d. Lakukan pengecekan di kamus, apabila ditemukan maka proses dihentikan, apabila tidak ada, maka lakukan pemenggalan berdasarkan Tabel 4.
- e. Lakukan *recoding* bila dibutuhkan.
- f. Apabila pengecekan di kamus tetap gagal maka awalan-awalan dikembalikan lagi.

10. Jika sampai langkah ke-9 kata dasar belum ditemukan dalam kamus maka kata awal yang dimasukkan dalam proses *stemming* dianggap sebagai kata dasar.

Tabel 3. Tabel Kombinasi Awalan dan Akhiran yang Tidak Diperbolehkan

No	Awalan	Akhiran	No	Awalan	Akhiran
1	è	èpon	15	pe	wen
2	a	èpon	16	pan	è
3	ta	è	17	pan	wi
4	ta	wi	18	pan	ne
5	ta	en	19	pan	en
6	ta	ana	20	pam	è
7	ta	aghi	21	pam	wi
8	sa	è	22	pam	ne
9	sa	wi	23	pam	en
10	sa	ne	24	pang	è
11	pè	è	25	pang	wi
12	pè	wi	26	pang	ne
13	pe	ne	27	pang	en
14	pe	en			

Tabel 4. Tabel Aturan Pemenggalan Awalan

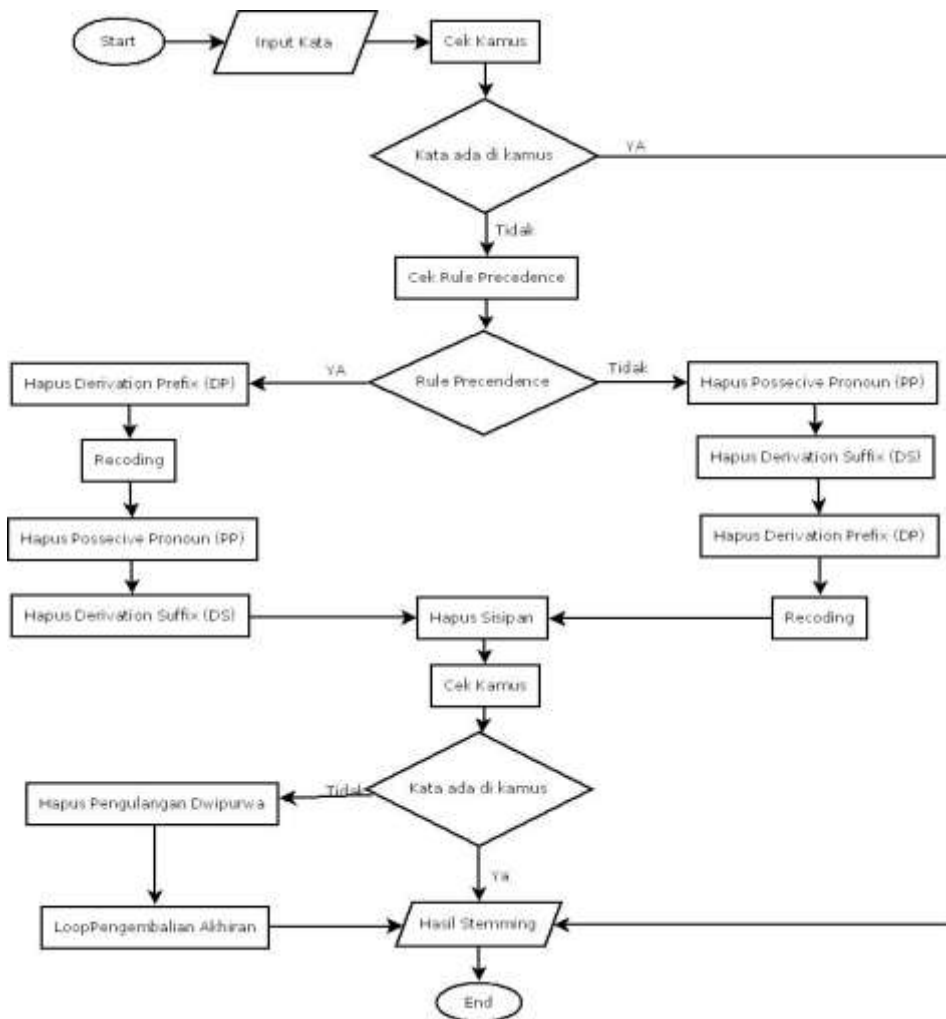
No	Format Kata	Pemenggalan	Contoh Kata
1	eA	e-A	eatorè, ecokor, etapok
2	aA	a-A	aghâbây, alonca', akemmor
3	taA	ta-A	tabâca, talèbât, tapalè'
4	maA	ma-A	mataber, maalos, mapotè
5	kaA	ka-A	kaator, katello, kasebbhut
6	saA	sa-A	saalam, saarè, sakampong
7	paA	pa-A   pa-nV	paloros, pakonèng, patao
8	komaC	koma-C	komalancang
9	kameC	kame-C	kamèporon
10	kapeC	kape-C	kapèderreng
11	peA	pe-A	pètodhu, pètolong
12	an → nV	n-tV   n-dV	namen, noghel, notop
13	any → nyV	ny-sV   ny-cV   ny-jV	nyate, nyoro, nyotok
14	am → mV	m-pV   m-bV	mancèng, maghâr, maca, mabâ
15	ang → ngV	ng-V   ng-kV   ng-gV	ngoca', ngobâ, ngakan, ngamar
16	panC	pan-C	panjâi'
17	pamC	pam-C	pambâjâr
18	pangC	Pang-C	panglèpor

Tabel 5. Tabel aturan pemenggalan sisipan

No	Format Kata	Pemenggalan	Contoh Kata
1	CalV	C-al-V → CV	gâlâggâ'
2	CarV	C-ar-V → CV	garibbung
3	CenV	C-en-V → CV	tenembang
4	CinV	C-in-V → CV	binâreng
5	ComV	C-om-V → CV	tomeba
6	CumV	C-um-V → CV	gumate
7	CamV	C-am-V → CV	pamenta

Tabel 6. Tabel Aturan Pemenggalan Akhiran

No	Format Kata	Pemenggalan	Contoh Kata	No	Format Kata	Pemenggalan	Contoh Kata
1	Aè	A-è	toraè, kacaè, pèntaè	11	CCa	C-Ca	kaloarra, kasokanna
2	Vwi	V-wi	toghuwi, tabbhuwi	12	Vana	V-ana	mola'âna, mossa'ana
3	Vne	V-ne	lakonè, nemmonè, astanè	13	CCana	C-Cana	majârana, nolèsana, malessana
4	Aa	A-a	èntara, ajârâ, toju'â	14	Vwana	V-wana	noguwana, ngataowana
5	Aan	A-an	pèkkèran, kakanan	15	Vyana	V-yana	mèlèyana
6	Vyan	V-yan	belliyân, gengsèyan	16	Vaghi	V-aghi	sala'aghi, baca'aghi
7	Vwan	V-wan	kalakowan, robbuwan	17	Vwaghi	V-waghi	kokowaghi, nopowaghi
8	Aen	A-en	pèttengèn, kacèlleben	18	Vyaghi	V-yaghi	gâjijaghi, pèlèyaghi
9	Vwen	V-wen	kabiruwen,	19	Aepon	A-epon	rabuepon, soroepon, toguepon
10	Vna	V-na	robbuna, macana	20	CCepon	C-Cepon	pèccottepon,



Gambar 1. Flowchart Algoritma ECS Madura.



## HASIL DAN PEMBAHASAN

### Pengujian Sistem Stemming Bahasa Madura

Proses pengujian menggunakan koleksi kata dalam bahasa madura sejumlah 400 kata yang terdiri dari kombinasi awalan, akhiran, sisipan dan kata ulang. Berikut ini adalah 10 contoh hasil pengujian sistem.

Tabel 7. Tabel pengujian kombinasi imbuhan kata Bahasa Madura

No	Kata	Hasil	Status Hasil
1	èpakalowarraghina	kalowar	Sukses
2	malakohaghina	lakoh	Sukses
3	kaatorra	ator	Sukses
4	kabâlâaghi	bâlâ	Sukses
5	èkakannaghina	kakan	Sukses
6	akancaan	kanca	Sukses
7	atokaran	tokar	Sukses
8	ka-bhungkaan	bhungka	Sukses
9	kadimmaa	kadimmaa	Gagal
10	asareng	sareng	Sukses

*Enhanced Confix Stripping Stemmer* adalah algoritma *stemming* yang sangat bergantung pada kelengkapan kamus. Begitu juga dengan hasil modifikasi dari algoritma ECS sangat bergantung dengan kelengkapan kamus. Walaupun hasil stemming telah benar tapi setelah dicek di kamus tidak terdapat kata yang dimaksud maka hasil stemming dinyatakan gagal.

Pada tabel hasil percobaan, umumnya kata yang dinyatakan gagal dalam hasil *stemming* adalah kata yang tidak terdapat di kamus. Sementara beberapa kata yang dinyatakan gagal karena terjadi kesalahan pemotongan baik *overstemming* (pemotongan berlebihan) atau *understemming* (pemotongan yang kurang maksimal). Hal tersebut terjadi karena faktor kemiripan suatu kata Bahasa Madura terhadap pola-pola kata tertentu sehingga sistem gagal melakukan *stemming* pada kata tersebut.

## SIMPULAN DAN SARAN

Pada tahap perancangan sistem diperoleh algoritma baru hasil modifikasi dari metode *Enhanced Confix Stripping Stemmer* sehingga dapat melakukan pemotongan (*stripping*) pada kata berimbuhan dalam Bahasa Madura. Tahap penyesuaian algoritma stemming tersebut Bahasa Madura meliputi modifikasi langkah-langkah algoritma dengan menghilangkan langkah pemotongan *inflection particle*, menambah langkah pemotongan sisipan (*infiks*), menambah langkah pemotongan kata ulang dwipurwa dan mengubah aturan pemotongan imbuhan.

Setelah melakukan beberapa analisa dan pengujian dengan menggunakan sistem yang telah dibangun maka diperoleh hasil dari modifikasi algoritma *Enhanced Confix Stripping Stemmer* pada teks berbahasa Madura. Hasil pengujian menunjukkan bahwa dari 400 kata yang diuji, 383 kata dinyatakan sukses di stemming dan 17 dinyatakan gagal. Dengan demikian presentase kata Bahasa Madura yang berhasil di stemming adalah 95,75 %. Waktu yang dibutuhkan untuk melakukan

proses *stemming* pada 400 kata tersebut adalah 25,2 detik dengan menggunakan spesifikasi komputer uji coba yang telah disebutkan. Dengan demikian modifikasi *Enhanced Confix Stripping Stemmer* pada teks berbahasa Madura dinilai efektif untuk diterapkan. Saran yang dapat disampaikan terkait dengan penelitian ini antara lain perlu adanya penelitian lebih lanjut tentang *stemming* Bahasa Madura untuk lebih meningkatkan presentasi keberhasilan *stemming* pada teks berbahasa Madura, dan perlu adanya penelitian tentang topik Bahasa Madura guna melestarikan Bahasa Madura yang merupakan kekayaan bangsa.

## **DAFTAR PUSTAKA**

- [1] Asian J., *Effective Techniques for Indonesian Text Retrieval*. PhD thesis School of Computer Science and Information Technology RMIT University Australia. 2007.
- [2] Mahendra, I. P. A. K. *Penggunaan Algoritma Semut Dan Confix Stripping Stemmer Untuk Klasifikasi Dokumen Berita Berbahasa Indonesia*. Tugas Akhir Jurusan Teknik Informatika Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember (ITS). 2008.
- [3] Tim Nabara Sumenep. *Bahasa dan Sastra Madura*. Sumenep: Panitia Pelatihan Bahasa dan Sastra Madura. 2001.
- [4] Effendy, Hafid.M., *Pramasastra Bhâsâ Madhurâ*. Pamekasan : Fakultas Keguruan dan Ilmu Pendidikan Jurusan Bahasa dan Sastra Indonesia Universitas Madura. 2006