

202010370311315  
Doni Ramadhani  
Prodi Informatika

# **Analisis Perbandingan Efektivitas Stemming Enhanced Confix Stripping dan IN-Idris dalam Klasifikasi Multi Kelas Naïve Bayes Terhadap Berita Online**

Diajukan Untuk Memenuhi  
Persyaratan Guna Meraih Gelar Sarjana  
Informatika Universitas Muhammadiyah Malang



**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH**  
**MALANG**  
**2025**

202010370311315  
Doni Ramadhani  
Prodi Informatika

## LEMBAR PERSETUJUAN

**Analisis Perbandingan Efektivitas Stemming Enhanced Confix Stripping dan IN-Idris  
dalam Klasifikasi Multi Kelas Naïve Bayes Terhadap Berita KompasTV**

**Doni Ramadhani**  
**202010370311315**

Telah Direkomendasikan Untuk Diajukan Sebagai  
Judul Tugas Akhir Di  
Program Studi Informatika Universitas Muhammadiyah Malang

Malang, 20 Mei 2025  
Menyetujui,

**Dosen 1**



**Christian Sri Kusuma Aditva, S.Kom., M.Kom**  
**NIDN:0727029101**



## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA : DONI RAMADHANI**

**NIM : 202010370311315**

**FAK./JUR. : Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**Analisis Perbandingan Efektivitas Stemming Enhanced Confix Stripping dan IN-Idris dalam Klasifikasi Multi Kelas Nave Bayes Terhadap Berita Online**” beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,  
Dosen Pembimbing

Christian Sri Kusuma Aditya S.Kom.,  
M.Kom

Malang, 15 Desember 2025  
Yang Membuat Pernyataan

DONI RAMADHANI



## ABSTRAK

Stemming merupakan tahapan penting dalam preprocessing teks bahasa Indonesia yang bertujuan mengubah kata berimbuhan menjadi bentuk dasarnya. Penelitian ini menganalisis perbandingan efektivitas dua algoritma stemming berbasis aturan, yaitu Enhanced Confix Stripping (ECS) dan IN-Idris, dalam klasifikasi multi-kelas berita menggunakan Naïve Bayes. Data yang digunakan bersumber dari dataset berita KompasTV yang terdiri dari 2.896 artikel dengan berbagai kategori. Metode penelitian meliputi tahapan preprocessing (case folding, cleansing, stopword removal, tokenizing, dan stemming), ekstraksi fitur menggunakan Bag of Words (BoW), serta klasifikasi dengan algoritma Naïve Bayes. Hasil penelitian menunjukkan bahwa ECS berhasil menghasilkan 14.588 kata dasar yang benar, sedikit lebih tinggi dibandingkan IN-Idris yang menghasilkan 14.579 kata. Namun dari segi akurasi klasifikasi, IN-Idris unggul dengan akurasi 65,11% dibandingkan ECS yang mencapai 63,73%. ECS lebih efektif dalam mengatasi overstemming melalui mekanisme pengembalian akhiran, sedangkan IN-Idris lebih baik dalam menangani imbuhan ganda melalui mekanisme pengurangan algoritma. Penelitian ini membuktikan bahwa kualitas stemming tidak hanya ditentukan oleh kuantitas kata yang benar, tetapi juga oleh jenis kesalahan stemming dan dampaknya terhadap representasi fitur dalam model klasifikasi.

**Kata Kunci:** Stemming, Enhanced Confix Stripping, IN-Idris, Klasifikasi Multi-Kelas, Naive Bayes Akurasi



## ABSTRACT

Stemming is a crucial stage in Indonesian text preprocessing, aiming to return inflected words to their basic word forms, which significantly impacts the quality of feature representation for classification. This study aims to analyze and compare the effectiveness of two rule-based stemming algorithms, namely Enhanced Confix Stripping (ECS) and IN-Idris, in improving the performance of multi-class classification on news articles. The methodology involved utilizing a dataset of 2,896 news articles from KompasTV across various categories. The steps included standard preprocessing (case folding, cleansing, stopword removal, and tokenizing), followed by the implementation of ECS and IN-Idris, feature extraction using the Bag of Words (BoW) model, and classification with the Naïve Bayes algorithm. Testing results indicate that ECS successfully generated slightly more correct basic words (14,588) compared to IN-Idris (14,579). However, from the perspective of model performance, IN-Idris demonstrated superiority with a classification accuracy of 65.11%, surpassing the ECS accuracy of 63.73%. The discussion highlights that ECS is more effective in addressing the overstemming problem through its suffix restoration mechanism, while IN-Idris excels in handling double affixes via its algorithm repetition. The conclusion of this study is that the quality of a stemming algorithm is not solely determined by the quantity of correct basic words, but primarily by its ability to minimize error types (overstemming and understemming) that directly impact the classification model's accuracy. Thus, the IN-Idris mechanism proved more effective in the context of this multi-class text classification.

**Keywords:** Stemming, Enhanced Confix Stripping, IN-Idris, Multi-Class Classification, Naive Bayes Accuracy



## DAFTAR ISI

<b>LEMBAR PERSETUJUAN</b> .....	2
<b>LEMBAR PENGESAHAN</b> .....	3
<b>LEMBAR PERNYATAAN</b> .....	4
<b>ABSTRAK</b> .....	5
<b>ABSTRACT</b> .....	6
<b>DAFTAR ISI</b> .....	7
<b>BAB I PENDAHULUAN</b> .....	9
<b>1.1. Latar Belakang</b> .....	9
<b>1.2. Rumusan Masalah</b> .....	12
<b>1.3. Tujuan Penelitian</b> .....	12
<b>1.4. Batasan Masalah</b> .....	12
<b>1.5. Sistematika Penulisan</b> .....	13
<b>BAB II TINJAUAN PUSTAKA</b> .....	15
<b>2.1. Preprocessing dalam Pemrosesan Teks</b> .....	15
<b>2.2. Stemming dalam Bahasa Indonesia</b> .....	15
<b>2.2.1. Nazief &amp; Adriani</b> .....	15
<b>2.2.2. Confix Stripping</b> .....	16
<b>2.2.3. Arifin Setiono</b> .....	16
<b>2.2.4. Idris Streamer</b> .....	16
<b>2.2.5. Enhanced Confix Stripping</b> .....	17
<b>2.2.6. IN-Idris</b> .....	19
<b>2.3. Overstemming dan Understemming</b> .....	20
<b>2.4. Ekstraksi Fitur Bag of Words(BoW)</b> .....	21
<b>2.5. Klasifikasi Naïve Bayes</b> .....	21
<b>2.6. Penelitian Sebelumnya</b> .....	22
<b>BAB III METODE PENELITIAN</b> .....	23
<b>3.1. Pendekatan Penelitian</b> .....	23
<b>3.2. Pengumpulan Data</b> .....	23
<b>3.2.1. Sumber dan Akses Data</b> .....	24
<b>3.2.2. Deskripsi Dataset</b> .....	24
<b>3.2.3. Metode Pengumpulan Data</b> .....	24
<b>3.3. Preprocessing</b> .....	24

3.3.1. Case Folding .....	25
3.3.2. Cleansing .....	25
3.3.3. Stopword Removal .....	25
3.3.4. Tokenizing .....	26
3.3.5. Stemming .....	26
3.4. Ekstrasi Fitur .....	26
3.5. Klasifikasi .....	27
3.6. Evaluasi .....	27
<b>BAB IV PEMBAHASAN DAN HASIL .....</b>	<b>28</b>
4.1. Hasil Setelah Preprocessing(tanpa stemming) .....	28
4.2. Hasil Setelah Stemming .....	28
4.2.1. Hasil ECS(Enhanced Confix Stripping) .....	28
4.2.2. Hasil IN-Idris .....	28
4.3. Analisa Hasil Kata yang Telah di Stemming .....	29
4.3.1. Overstemming dan Understemming Enhanced Confix Stripping (ECS) .....	30
4.3.2. Overstemming dan Understemming Enhanced IN-Idris .....	30
4.4. Hasil Klasifikasi .....	30
4.5. Pembahasan Hasil .....	32
<b>BAB V PENUTUP .....</b>	<b>34</b>
5.1. Kesimpulan .....	34
5.2. Saran Penelitian Selanjutnya .....	34
<b>DAFTAR PUSTAKA .....</b>	<b>36</b>

## DAFTAR PUSTAKA

- [1] M. F. Wajdi and M. I. Hajiri, "Tantangan Adaptasi Kecerdasan Buatan dalam Layanan Perpustakaan Perguruan Tinggi," *Jurnal Ilmu Perpustakaan dan Informasi*, Vol. 8, No. 2, 2024, doi: 10.29240/tik.v8i2.10901.
- [2] E. M. Pusunga and I. N. Dewia, "Optimasi RoBERTa dengan Hyperparameter Tuning untuk Deteksi Emosi berbasis Teks," *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol. 10, no. 3, pp. 240–248, 2024, doi: 10.25077/TEKNOSI.v10i3.2024.240-248.
- [3] N. T. Hariyanti, T. Rahmawati, and A. Wirapraja, "Perancangan Aplikasi Mobile Menggunakan Machine Learning Untuk Menentukan Klasifikasi Kategori Berita," *Teknika*, Vol. 13, no. 3, doi: 10.34148/teknika.v13i3.1093.
- [4] Y. Widhiyasana, T. Semiawan, I. G. A. Mudzakir, and M. R. Noor, "Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, Vol. 10, no. 4, pp. 354–361, doi: 10.22146/jnteti.v10i4.2438.
- [5] E. E. Pratama and B. R. Trilaksono, "Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM)," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, Vol. 1, no. 2, doi: 10.26418/jp.v1i2.11023.
- [6] E. Y. Hidayat and M. A. Rizqi, "Klasifikasi Dokumen Berita Menggunakan Algoritma Enhanced Confix Stripping Stemmer dan Naïve Bayes Classifier," *Jurnal Teknologi dan Sistem Informasi*, Vol. 6, no. 2, pp. 90–99, doi: 10.25077/TEKNOSI.v6i2.2020.90-99.
- [7] Y. D. Pramudita, S. S. Putro, and N. Makhmud, "KLASIFIKASI BERITA OLAHRAGA MENGGUNAKAN METODE NAÏVE BAYES DENGAN ENHANCED CONFIX STRIPPING STEMMER," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, Vol. 5, No. 3, pp. 269–276, 2018, doi: 10.25126/jtiik.201853810.
- [8] R. P. N. Budianti, H. C. Marvyna, and S. Sukaridhoto, "Analyzing Twitter Users' Sentiments on the Surge of Fuel Oil Prices in Indonesia using the K-Nearest Neighbor Algorithm," *E3S Web of Conferences*, Vol. 482, p. 02004, doi: 10.1051/e3sconf/202448202004.
- [9] N. Pamungkas et al., "Comparison of Stemming Test Results of Tala Algorithms with Nazief Adriani in Abstract Documents and National News," *Jurnal Ilmiah Bidang Teknologi dan Komunikasi*, Vol. 8, no. 1, 2023, doi: 10.25139/inform.v8i1.5569.
- [10] F. W. Suci, N. Hayatin, and Y. Munarko, "IN-IDRIS: MODIFICATION OF IDRIS STEMMING ALGORITHM FOR INDONESIAN TEXT," *IIUM Engineering Journal*, Vol. 23, no. 1, 2022, doi: 10.31436/iiumej.v23i1.1783.
- [11] N. W. Wardani and P. G. S. C. Nugraha, "Stemming Teks Bahasa Bali dengan Algoritma Enhanced Confix Stripping," *International Journal of Natural Science and Engineering*, Vol. 4, no. 3, pp. 103–113, 2020, doi: 10.23887/ijnse.v4i3.30309.

- [12] Y. T. Arifin, “KOMPARASI FITUR SELEKSI PADA ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN REVIEW,” *Jurnal Informatika*, Vol. 3, no. 2, pp. 191–199, 2016, doi: 10.31294/eco.v3i2.2016.191-199.
- [13] S. A. Azzahra and A. Wibowo, “ANALISIS SENTIMEN MULTI-ASPEK BERBASIS KONVERSI IKON EMOSI DENGAN ALGORITME NAÏVE BAYES UNTUK ULASAN WISATA KULINER PADA WEB TRIPADVISOR,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, Vol. 7, no. 4, pp. 737–744, 2020, doi: 10.25126/jtiik.202071907.
- [14] A. Z. Arifin, P. A. K. Mahendra, and H. T. Ciptaningtyas, “ENHANCED CONFIX STRIPPING STEMMER AND ANTS ALGORITHM FOR CLASSIFYING NEWS DOCUMENT IN INDONESIAN LANGUAGE,” Vol. 149, pp. 2085–1944.
- [15] J. Asian, H. Williams, and S. M. M. Tahaghoghi, “Stemming Indonesian,” *ACM Transactions on Asian Language Information Processing*, Vol. 38, no. 4, pp. 307–314, doi: 10.1145/1316457.1316459.
- [16] M. Adriani, J. Asian, B. Nazief, H. Williams, and S. M. M. Tahaghoghi, “Stemming Indonesian: A confix-stripping approach,” *ACM Transactions on Asian Language Information Processing*, Vol. 6, no. 4, doi: 10.1145/1316457.1316459.
- [17] J. Pardede and D. Darmawan, “PERBANDINGAN ALGORITMA STEMMING PORTER, SASTRAWI, IDRIS, DAN ARIFIN & SETIONO PADA DOKUMEN TEKS BAHASA INDONESIA,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, Vol. 12, No. 1, pp. 69–76, 2025, doi: 10.25126/jtiik.2025128860.
- [18] S. Devi, C. Dhaval, L. Moharkar, and S. Khanvilkar, “Impact of Online Education and Sentiment Analysis from Twitter Data using Topic Modeling Algorithms,” *Jurnal Teknologi dan Sistem Informasi*, Vol. 4, Issue 1, pp. 21–34, doi: 10.24071/ijasst.v4i1.4637.
- [19] N. Sari et al., “Implementation of Naive Bayes and Support Vector Machine Classification Algorithms for Sentiment Analysis of Bilingual Cyberbullying on X Application,” *Jurnal Sistem Informasi*, Vol. 14, no. 1, pp. 211–224, doi: 10.32520/stmsi.v14i1.4799.
- [20] A. Prasadhatama and K. M. Suryaningrum, “PERBANDINGAN ALGORITMA NAZIEF & ADRIANI DENGAN ALGORITMA IDRIS UNTUK PENCARIAN KATA DASAR,” *Jurnal Teknologi dan Manajemen Informatika*, Vol. 4, no. 1, doi: 10.26905/jtmi.v4i1.1773.
- [21] I. O. Suzanti et al., “COMPARISON OF STEMMING AND SIMILARITY ALGORITHMS IN INDONESIAN TRANSLATED AL-QUR'AN TEXT SEARCH,” *Jurnal Ilmiah KURSOR*, Vol. 11, No. 2, doi: 10.21107/kursor.v11i2.280.



UNIVERSITAS  
MUHAMMADIYAH  
MALANG



## FAKULTAS TEKNIK

### INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Doni Ramadhani  
NIM : 202010370311315  
Judul TA : Analisis Perbandingan Efektivitas Stemming Enhanced Confix Stripping dan IN-Idris dalam Klasifikasi Multi Kelas Naïve Bayes Terhadap Berita Online

#### Hasil Cek Plagiarisme dengan Turnitin

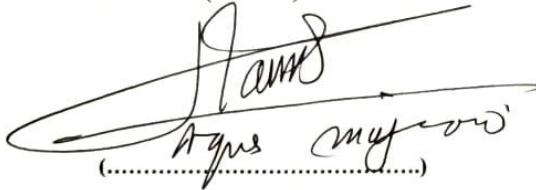
No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	10 % ✓
2.	Bab 2 – Daftar Pustaka	25 %	11 % ✓
3.	Bab 3 – Analisis dan Perancangan	25 %	5 % ✓
4.	Bab 4 – Implementasi dan Pengujian	15 %	0 % ✓
5.	Bab 5 – Kesimpulan dan Saran	5 %	41 % ✓
6.	Makalah Tugas Akhir	20%	4 % ✓

\*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)

  
(.....)



Kampus I  
Jl. Bandung 1 Malang, Jawa Timur  
P. +62 341 851 253 (Hunrei)  
F. +62 341 480 438

Kampus II  
Jl. Bendungan Sutarni No 188 Malang, Jawa Timur  
P. +62 341 551 149 (Hunrei)  
F. +62 341 582 060

Kampus III  
Jl. Raya Tlogomas No 248 Malang, Jawa Timur  
P. +62 341 404 319 (Hunrei)  
F. +62 341 480 435  
E. webmaster@umm.ac.id