

**IMPLEMENTASI LOGISTIC REGRESSION DAN RIDGE CLASSIFIER
UNTUK DETEKSI BERITA HOAKS BAHASA INDONESIA**



LAPORAN MBKM PENELITIAN

**MOHAMMAD ALFARIZKY RAMADHANI OSCANDAR
00000058407**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**IMPLEMENTASI LOGISTIC REGRESSION DAN RIDGE CLASSIFIER
UNTUK DETEKSI BERITA HOAKS BAHASA INDONESIA**



UMMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Mohammad Alfarizky Ramadhani Oscandar
NIM : 00000058407
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan MBKM Penelitian saya yang berjudul:

Implementasi Logistic Regression dan Ridge Classifier untuk Deteksi Berita Hoaks Bahasa Indonesia

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 03 Januari 2025



(Mohammad Alfarizky Ramadhani Oscandar)

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

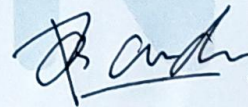
Nama : Mohammad Alfarizky Ramadhani
Oscandar
NIM : 00000058407
Program Studi : Informatika
Jenjang : S1
Jenis Karya : MBKM Penelitian

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 03 Januari 2025

Yang menyatakan



Mohammad Alfarizky Ramadhani Oscandar

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN

Halaman Persembahan / Motto

”Dan bahwasanya seorang manusia tiada memperoleh selain apa yang telah diusahakannya,”

(QS. An-Najm: 39)



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

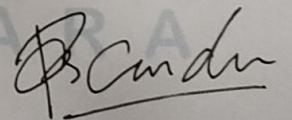
KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan MBKM Penelitian ini dengan judul: Implementasi Logistic Regression dan Ridge Classifier untuk Deteksi Berita Hoaks Bahasa Indonesia dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan laporan magang ini, sangatlah sulit bagi saya untuk menyelesaikan laporan magang ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, CEH, CEI, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Dr. Maria Irmina Prasetyowati, S.Kom., M.T., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya karya ilmiah ini.
5. Ibu saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan karya ilmiah ini.

Penulis berharap agar karya ilmiah ini dapat memberikan manfaat bagi pembaca dan dapat dijadikan referensi bagi peneliti yang ingin melakukan penelitian lebih lanjut.

Tangerang, 03 Januari 2025



Mohammad Alfarizky Ramadhani Oscandar

IMPLEMENTASI LOGISTIC REGRESSION DAN RIDGE CLASSIFIER UNTUK DETEKSI BERITA HOAKS BAHASA INDONESIA

Mohammad Alfarizky Ramadhani Oscandar

ABSTRAK

Penelitian ini mengimplementasikan dan membandingkan model *Logistic Regression* dan *Ridge Classifier* untuk mendeteksi berita hoaks berbahasa Indonesia. Dataset yang digunakan terdiri dari 29.482 artikel berita yang dikumpulkan dari portal berita Kompas, Tempo, dan database Mafindo. Proses *preprocessing* data meliputi pembersihan teks, tokenisasi, dan *stemming* menggunakan *library* Sastrawi. Fitur teks diekstraksi menggunakan *Count Vectorizer* dengan total 5.000 fitur yang terdiri dari 2.500 fitur judul dan 2.500 fitur konten berita. Evaluasi performa model menunjukkan bahwa *Logistic Regression* mencapai akurasi 98% sementara *Ridge Classifier* mencapai 97%. Meskipun terdapat perbedaan tipis pada metrik evaluasi, kedua model menunjukkan performa yang sangat baik dalam mengklasifikasikan berita hoaks dengan nilai *precision*, *recall*, dan *F1-score* yang tinggi. Hasil penelitian juga mengungkapkan bahwa fitur dari isi berita memiliki pengaruh yang lebih signifikan dibandingkan fitur judul dalam menentukan klasifikasi berita.

Kata kunci: *Logistic Regression*, *Ridge Classifier*, deteksi hoaks, pemrosesan bahasa alami, klasifikasi teks



**IMPLEMENTING LOGISTIC REGRESSION AND RIDGE CLASSIFIER
FOR INDONESIA HOAX NEWS DETECTION**

Mohammad Alfarizky Ramadhani Oscandar

ABSTRACT

This research implements and compares Logistic Regression and Ridge Classifier models for detecting Indonesian hoax news. The dataset consists of 29,482 news articles collected from Kompas, Tempo news portals, and the Mafindo database. The data preprocessing involves text cleaning, tokenization, and stemming using the Sastrawi library. Text features were extracted using Count Vectorizer with a total of 5,000 features comprising 2,500 title features and 2,500 news content features. Performance evaluation using cross-validation with 5-fold and StratifiedKfold showed that Logistic Regression achieved 98% accuracy while Ridge Classifier achieved 97%. Despite slight differences in evaluation metrics, both models demonstrated excellent performance in classifying hoax news with high precision, recall, and F1-score values. The findings also revealed that features from news content had a more significant influence compared to title features in determining news classification. This research provides a strong foundation for developing more sophisticated hoax detection systems in the future.

Keywords: *Logistic Regression, Ridge Classifier, hoax detection, natural language processing, text classification*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERSEMBAHAN/MOTO	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR KODE	xii
DAFTAR LAMPIRAN	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	3
1.5 Urgensi Penelitian	3
1.6 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Berita Hoaks Berbahasa Indonesia	5
2.1.1 Sumber Dataset	5
2.2 Metode Preprocessing Teks	6
2.2.1 Pembersihan Teks	7
2.2.2 Tokenisasi	7
2.2.3 Stemming	8
2.2.4 Feature Extraction	9
2.3 Logistic Regression	9
2.4 Ridge Classifier	10
BAB 3 METODOLOGI PENELITIAN	12
3.1 Knowledge Discovery in Databases Framework	12
3.2 Pengumpulan Data	13
3.3 Pembersihan dan Pengolahan Data	13
3.4 Transformasi Data	14
3.4.1 Sastrawi Stemmer	14
3.4.2 Count Vectorizer	15
3.5 Pembuatan Model	15
3.6 Evaluasi Model	16
3.7 Pembuatan Laporan dan Dokumentasi	16
BAB 4 HASIL DAN PEMBAHASAN	17
4.1 Logistic Regression	17
4.1.1 Jumlah Fitur	17
4.1.2 Dataset	17
4.1.3 Keseimbangan Kelas	17
4.1.4 Koefisien Fitur	18
4.1.5 Evaluasi Performa	18
4.1.6 Confusion Matrix	18
4.1.7 Feature Importance	19
4.1.8 Model Prediksi Berita	20
4.2 Ridge Classifier	21

4.2.1	Jumlah Fitur	21
4.2.2	Dataset	22
4.2.3	Keseimbangan Kelas	22
4.2.4	Koefisien Fitur	22
4.2.5	Evaluasi Performa	22
4.2.6	Confusion Matrix	23
4.2.7	Feature Importance	23
4.2.8	Model Prediksi Berita	24
4.3	Visualisasi perbandingan	26
4.3.1	Perbandingan Akurasi	27
4.3.2	Perbandingan <i>Precision</i> dan <i>Recall</i>	27
4.3.3	Perbandingan <i>F1-Score</i>	28
4.3.4	Perbandingan ROC - AUC	29
4.3.5	Perbandingan Waktu Training	30
BAB 5	SIMPULAN SARAN	31
5.1	Simpulan	31
5.2	Saran	31
	DAFTAR PUSTAKA	33



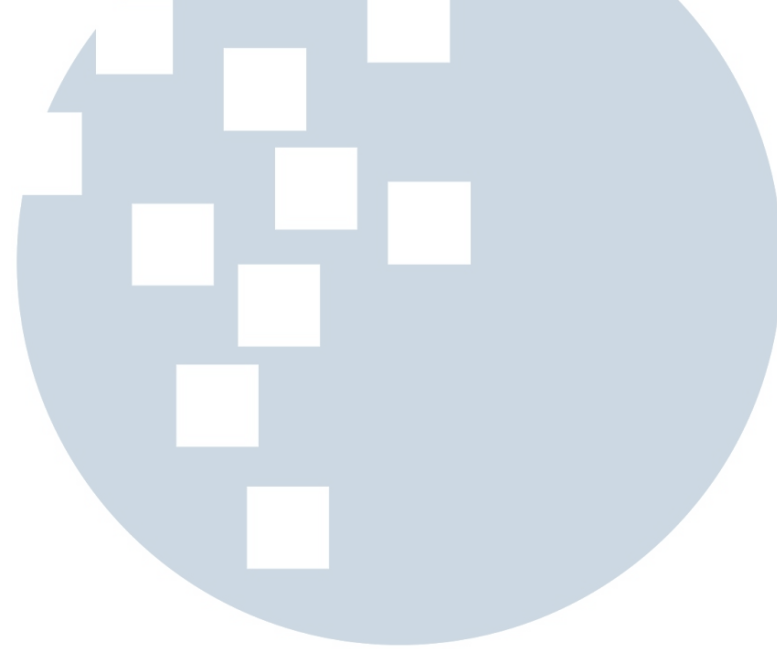
DAFTAR GAMBAR

Gambar 3.1	Research Framework based on Knowledge Discovery in Databases (KDD) Methodology	12
Gambar 4.1	Confusion Matrix <i>Logistic Regression</i>	19
Gambar 4.2	Feature Importance <i>Logistic Regression</i>	19
Gambar 4.3	Proses Prediksi Berita Menggunakan <i>Logistic Regression</i>	21
Gambar 4.4	Confusion Matrix <i>Ridge Classifier</i>	23
Gambar 4.5	Feature Importance <i>Ridge Classifier</i>	24
Gambar 4.6	Proses Prediksi Berita Menggunakan <i>Ridge Classifier</i>	26
Gambar 4.7	Perbandingan Akurasi <i>Logistic Regression</i> dan <i>Ridge Classifier</i>	27
Gambar 4.8	Perbandingan <i>Precision</i> dan <i>Recall Logistic Regression</i> dan <i>Ridge Classifier</i>	28
Gambar 4.9	Perbandingan <i>F1-Score Logistic Regression</i> dan <i>Ridge Classifier</i>	29
Gambar 4.10	Perbandingan <i>ROC - AUC Logistic Regression</i> dan <i>Ridge Classifier</i>	30
Gambar 4.11	Perbandingan Waktu <i>Training Logistic Regression</i> dan <i>Ridge Classifier</i>	30



DAFTAR TABEL

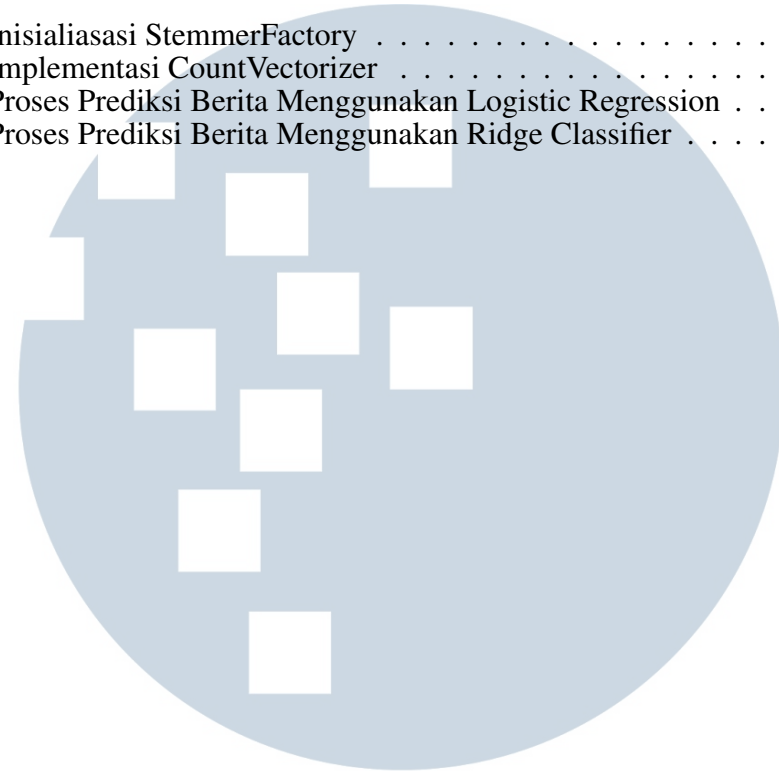
Tabel 2.1	Matriks frekuensi kata dari dua dokumen menggunakan <i>Count Vectorizer</i>	9
Tabel 3.1	Distribusi Dataset Berdasarkan Jenis Berita dan Sumber	13



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR KODE

3.1	Inisialiasasi StemmerFactory	14
3.2	Implementasi CountVectorizer	15
4.1	Proses Prediksi Berita Menggunakan Logistic Regression	20
4.2	Proses Prediksi Berita Menggunakan Ridge Classifier	25



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	MBKM-01 Cover Letter MBKM Research	34
Lampiran 2	MBKM-02 MBKM Research Card	35
Lampiran 3	MBKM-03 Daily Task MBKM Research	36
Lampiran 4	MBKM-04 Verification Form of Report MBKM Research	37
Lampiran 5	Form Bimbingan	38
Lampiran 6	Kontrak Kerjasama Penelitian	39
Lampiran 7	LOA Mahasiswa	40
Lampiran 8	Turnitin Report	41
Lampiran 9	Draft Paper Penelitian	46

