

**ANALISIS PERBANDINGAN MODEL NAÏVE BAYES DAN
LOGISTIC REGRESSION BERBASIS TF-IDF DALAM
MENDETEKSI BERITA HOAKS BERBAHASA INDONESIA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

ZIDNA ROHMATAL IKHSAN

22.12.2308

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PERBANDINGAN MODEL NAÏVE BAYES DAN LOGISTIC
REGRESSION BERBASIS TF-IDF DALAM MENDETEKSI BERITA
HOAKS BERBAHASA INDONESIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

ZIDNA ROHMATAL IKHSAN

22.12.2308

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

HALAMAN PERSETUJUAN

SKRIPSI

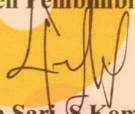
**ANALISIS PERBANDINGAN MODEL NAIVE BAYES DAN
LOGISTIC REGRESSION BERBASIS TF-IDF DALAM
MENDETEKSI BERITA HOAKS BERBAHASA INDONESIA**

yang disusun dan diajukan oleh

ZIDNA ROHMATAL IKHSAN
22.12.2308

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 – 01 – 2026

Dosen Pembimbing,


Bety Wulan Sari, S.Kom., M.Kom
NIK. 190302254

HALAMAN PENGESAHAN
SKRIPSI

ANALISIS PERBANDINGAN MODEL NAIVE BAYES DAN
LOGISTIC REGRESSION BERBASIS TF-IDF DALAM
MENDETEKSI BERITA HOAKS BERBAHASA INDONESIA

yang disusun dan diajukan oleh

ZIDNA ROHMATAL IKHSAN

22.12.2308

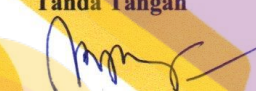
Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 – 01 – 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Krisnawati, S.Si., M.T.
NIK. 190302038



Hendra Kurniawan, S.Kom., M.Kom
NIK. 190302244



Bety Wulan Sari, S.Kom., M.Kom.
NIK. 190302254



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 – 01 – 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : ZIDNA ROHMATAL IKHSAN
NIM : 22.12.2308

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS PERBANDINGAN MODEL NAÏVE BAYES DAN LOGISTIC REGRESSION BERBASIS TF-IDF DALAM MENDETEKSI BERITA HOAKS BERBAHASA INDONESIA

Dosen Pembimbing : Bety Wulan Sari, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Januari 2026

Yang Menyatakan,



Zidna Rohmatal Ikhsan

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil ‘alamin, dengan penuh rasa syukur ke hadirat Allah SWT, skripsi ini saya persembahkan sebagai bentuk rasa terima kasih atas segala rahmat, karunia, dan kemudahan yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

Skripsi ini saya persembahkan kepada kedua orang tua tercinta yang senantiasa memberikan doa, dukungan, kasih sayang, serta motivasi tanpa henti dalam setiap langkah perjalanan hidup penulis. Tanpa doa dan pengorbanan mereka, pencapaian ini tidak akan terwujud. Penulis juga mempersembahkan karya ini kepada keluarga besar yang selalu memberikan semangat dan dukungan moral, serta kepada dosen pembimbing dan seluruh dosen yang telah dengan sabar memberikan bimbingan, ilmu, dan arahan selama proses perkuliahan dan penyusunan skripsi ini.

Tidak lupa, skripsi ini dipersembahkan kepada sahabat dan rekan seperjuangan yang telah memberikan dukungan, kebersamaan, dan semangat selama proses penyusunan skripsi, serta kepada almamater tercinta yang telah menjadi tempat penulis menimba ilmu dan pengalaman berharga.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Perbandingan Model Naive Bayes dan Logistic Regression dalam Mendeteksi Berita Hoaks” dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada program studi yang penulis tempuh.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa keberhasilan penyelesaian skripsi tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan arahan serta masukan yang sangat berharga selama proses penyusunan skripsi. Ucapan terima kasih juga penulis sampaikan kepada seluruh dosen dan staf akademik yang telah memberikan ilmu pengetahuan dan pelayanan selama masa perkuliahan.

Penulis juga mengucapkan terima kasih kepada kedua orang tua dan keluarga yang senantiasa memberikan doa, dukungan moral, dan motivasi sehingga penulis mampu menyelesaikan skripsi ini. Selain itu, terima kasih penulis sampaikan kepada sahabat dan rekan-rekan seperjuangan yang telah memberikan semangat, bantuan, serta kebersamaan selama proses penyusunan skripsi. Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan skripsi ini di masa mendatang.

Yogyakarta, 20 Januari 2026

Zidna Rohmatal Ikhsan

DAFTAR ISI

HALAMN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Konsep Dasar Permasalahan	10
2.3 Pre-Processing.....	11
2.4 Term Frequency-inverse Document Frequency (TF-IDF).....	12

2.5	Machine Learning	13
2.6	Matrix Evaluasi	18
BAB III METODE PENELITIAN		20
3.1	Objek Penelitian	20
3.2	Alur Penelitian.....	21
BAB IV HASIL DAN PEMBAHASAN		25
4.1	Pengumpulan Data	25
4.2	Preprocessing	29
4.3	Ekstraksi Fitur TF-IDF.....	34
4.4	Training dan Testing	36
4.5	Naive Bayes.....	36
4.6	Logistic Regression	38
4.7	Evaluasi Model.....	41
4.8	Analisis Perbandingan Model	42
BAB V PENUTUP		44
5.1	Kesimpulan.....	44
5.2	Saran.....	44
REFERENSI		46
LAMPIRAN.....		48

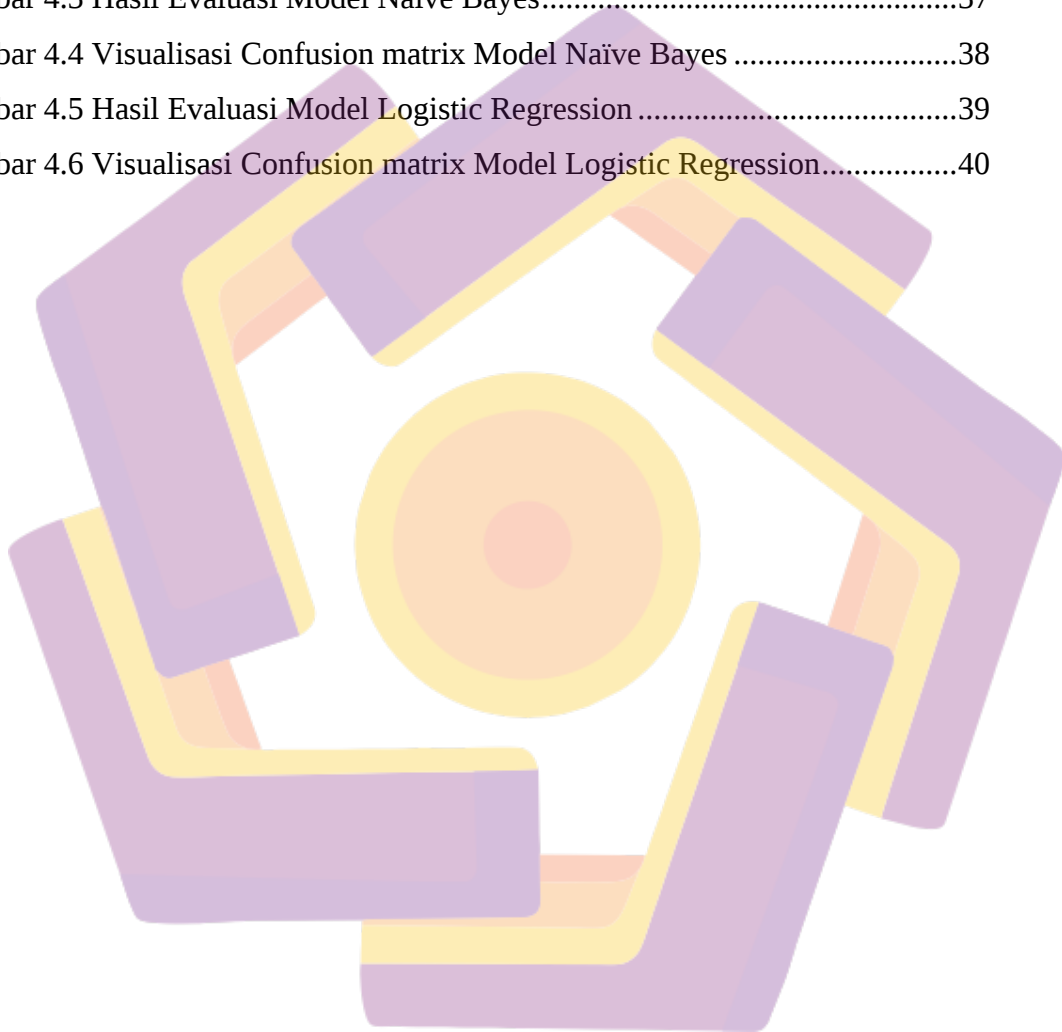
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 4.1 Contoh Data Mentah Dataset Berita Hoaks	25
Tabel 4.2 Data Setelah Tahap Cleaning & Case Folding	29
Tabel 4.3 Data Setelah Tahap Tokenizing	31
Tabel 4.4 Data Setelah Tahap Stemming	33
Tabel 4.5 Hasil Perhitungan TF-IDF	35
Tabel 4.6 Hasil Evaluasi Model Naïve Bayes & Logistic Regression	41

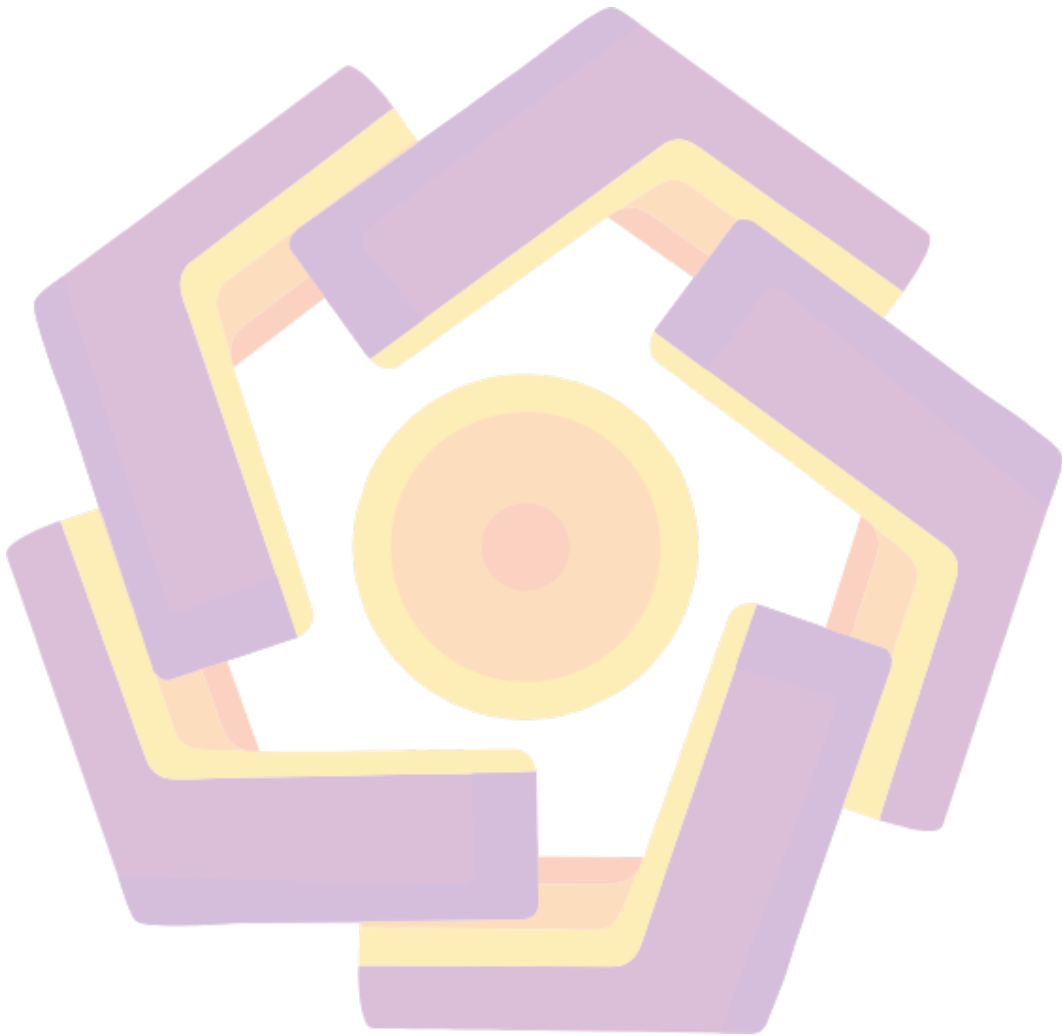


DAFTAR GAMBAR

Gambar 2.1 Konsep Dasar Machine Learning.....	13
Gambar 2.2 Tahapan Supervised Learning.....	14
Gambar 3.1 Alur Penelitian Deteksi Hoaks	21
Gambar 4.1 Word Cloud Narasi Hoaks	28
Gambar 4.2 Word Cloud Narasi Non-Hoaks	28
Gambar 4.3 Hasil Evaluasi Model Naïve Bayes.....	37
Gambar 4.4 Visualisasi Confusion matrix Model Naïve Bayes	38
Gambar 4.5 Hasil Evaluasi Model Logistic Regression	39
Gambar 4.6 Visualisasi Confusion matrix Model Logistic Regression.....	40

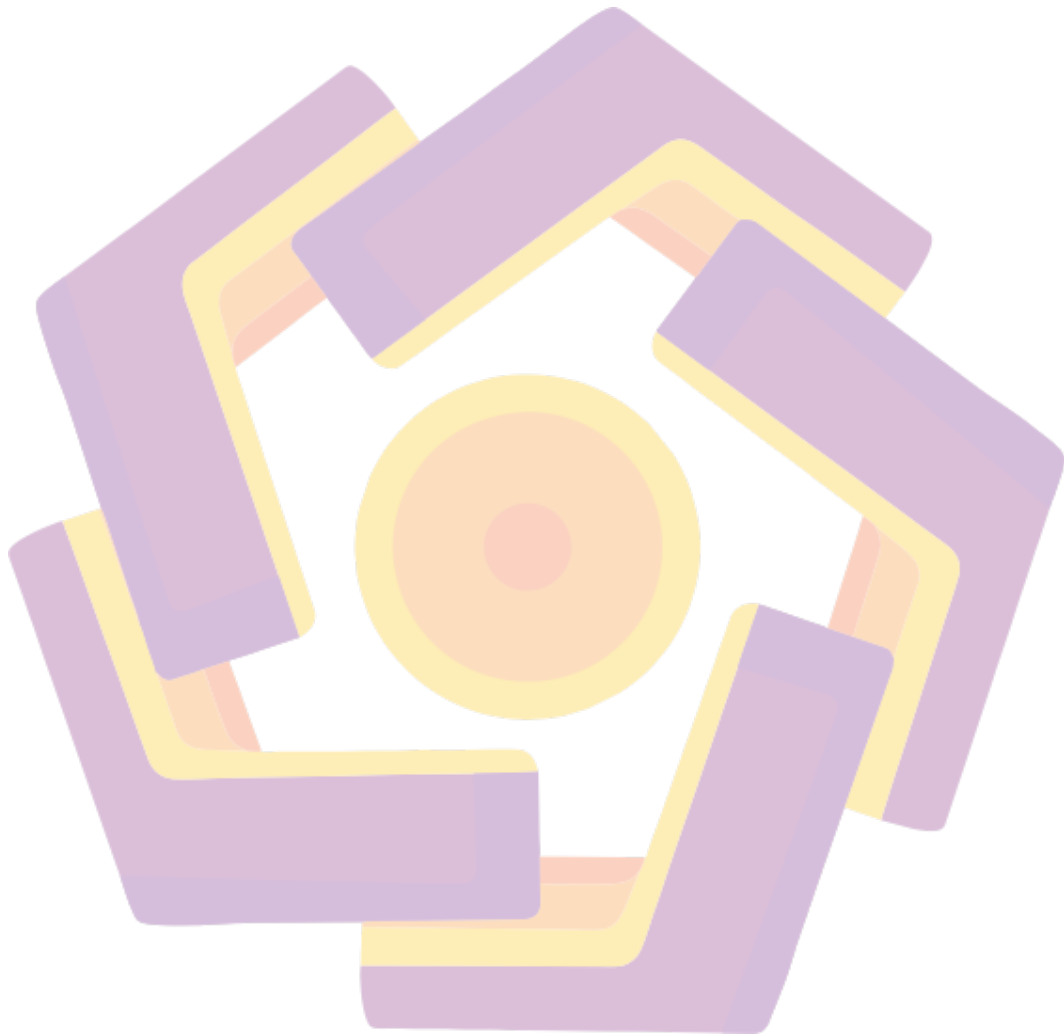


DAFTAR LAMPIRAN

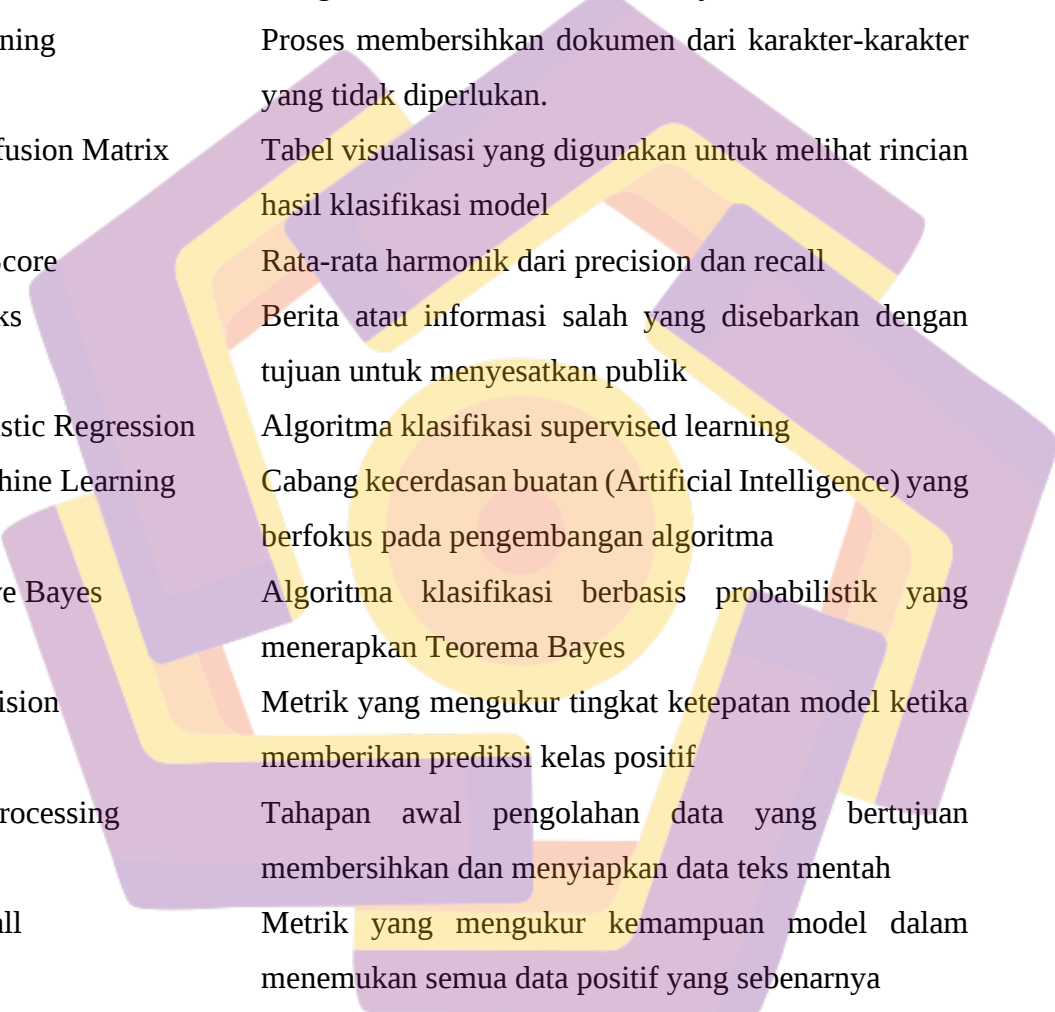


DAFTAR LAMBANG DAN SINGKATAN

Ω	Tahanan Listrik
μ	Konstanta gesekan
ANFIS	Adaptive Network Fuzzy Inference System
SVM	Support Vector Machines



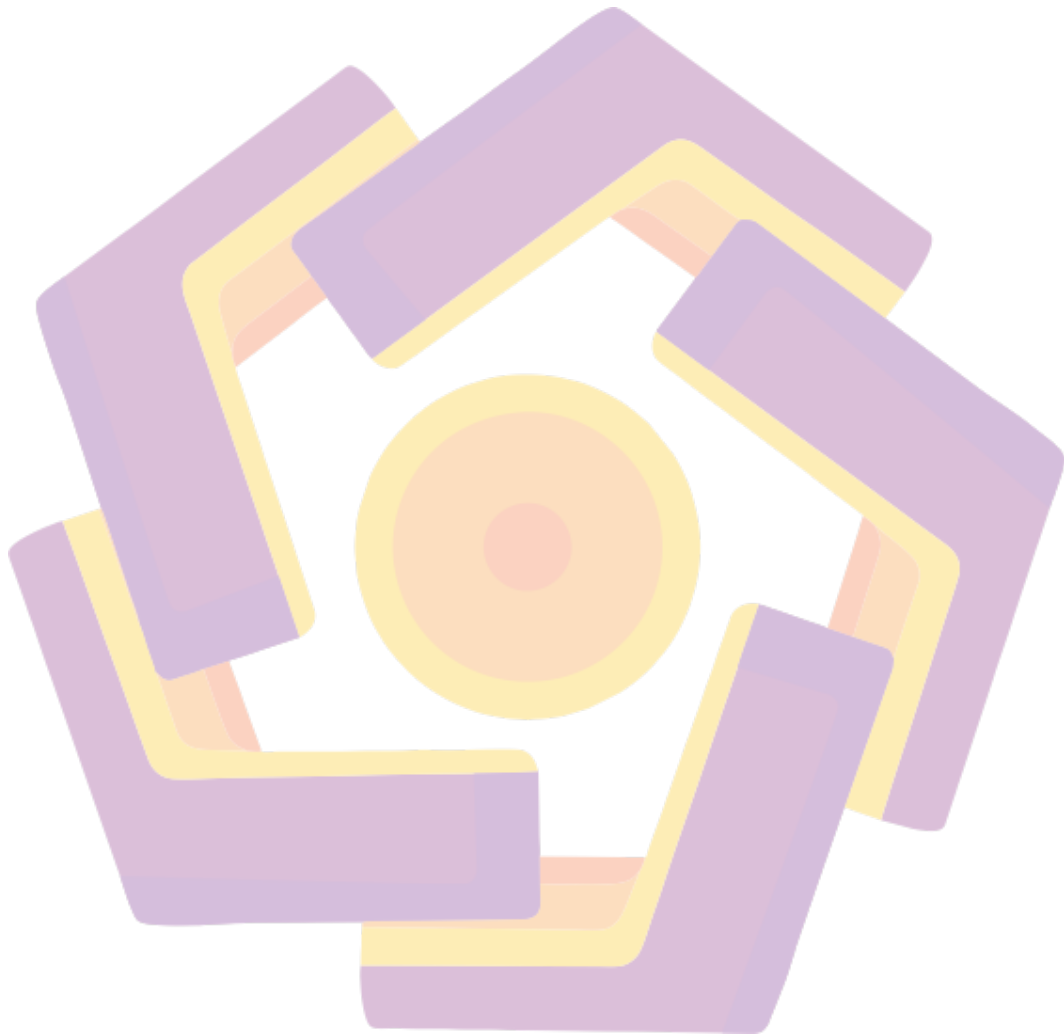
DAFTAR ISTILAH



Accuracy	Metrik evaluasi yang mengukur proporsi prediksi model yang benar (baik hoaks maupun non-hoaks) terhadap keseluruhan data uji.
Case Folding	Salah satu tahapan preprocessing yang bertujuan untuk menyeragamkan bentuk huruf dalam dokumen dengan mengubah semua huruf besar menjadi huruf kecil
Cleaning	Proses membersihkan dokumen dari karakter-karakter yang tidak diperlukan.
Confusion Matrix	Tabel visualisasi yang digunakan untuk melihat rincian hasil klasifikasi model
F1-Score	Rata-rata harmonik dari precision dan recall
Hoaks	Berita atau informasi salah yang disebarakan dengan tujuan untuk menyesatkan publik
Logistic Regression	Algoritma klasifikasi supervised learning
Machine Learning	Cabang kecerdasan buatan (Artificial Intelligence) yang berfokus pada pengembangan algoritma
Naive Bayes	Algoritma klasifikasi berbasis probabilistik yang menerapkan Teorema Bayes
Precision	Metrik yang mengukur tingkat ketepatan model ketika memberikan prediksi kelas positif
Preprocessing	Tahapan awal pengolahan data yang bertujuan membersihkan dan menyiapkan data teks mentah
Recall	Metrik yang mengukur kemampuan model dalam menemukan semua data positif yang sebenarnya
Stemming	Proses mengubah kata menjadi bentuk dasarnya (<i>root word</i>) dengan cara menghilangkan imbuhan awal dan akhir
TF-IDF	Metode pembobotan kata untuk mengubah teks menjadi representasi numerik
Tokenizing	Proses pemecahan kalimat atau teks menjadi unit-unit kata atau token terpisah

Word Cloud

Visualisasi frekuensi kata kunci yang menampilkan kata-kata yang paling sering muncul dalam ukuran yang lebih besar



INTISARI

Penyebaran berita hoaks di Indonesia semakin meningkat seiring dengan berkembangnya media digital dan media sosial. Hal ini menimbulkan kebutuhan terhadap sistem pendeteksi otomatis yang mampu mengidentifikasi berita hoaks secara cepat dan akurat. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja dua algoritma klasifikasi teks, yaitu Naive Bayes dan Logistic Regression, dengan menggunakan representasi fitur TF-IDF (*Term Frequency–Inverse Document Frequency*). Dataset yang digunakan terdiri dari sekitar 2.600 data berita hoaks dan non-hoaks berbahasa Indonesia. Tahapan penelitian meliputi pengumpulan data, preprocessing teks, ekstraksi fitur TF-IDF, pelatihan model, dan evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa kedua model mampu melakukan klasifikasi berita hoaks dengan baik, namun Logistic Regression memberikan performa yang lebih stabil dan optimal dibandingkan Naïve Bayes berdasarkan nilai evaluasi yang diperoleh. Dengan demikian, Logistic Regression berbasis TF-IDF direkomendasikan sebagai model yang lebih efektif untuk mendeteksi berita hoaks berbahasa Indonesia.

Kata kunci: Naive Bayes, Logistic Regression, TF-IDF, Deteksi Hoaks, Klasifikasi Teks.

ABSTRACT

The spread of hoax news in Indonesia has increased significantly with the rise of digital and social media platforms. This situation highlights the need for an automated system capable of detecting hoax news quickly and accurately. This study aims to analyze and compare the performance of two text classification algorithms, Naive Bayes and Logistic Regression, using the TF-IDF (Term Frequency–Inverse Document Frequency) feature representation. The dataset consists of approximately 2000 Indonesian-language news articles, both hoax and non-hoax. The research stages include data collection, text preprocessing, TF-IDF feature extraction, model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The results indicate that both models perform well in hoax news classification, however Logistic Regression demonstrates more stable and optimal performance compared to Naïve Bayes based on the evaluation metrics. Therefore, TF-IDF-based Logistic Regression is recommended as a more effective model for detecting Indonesian-language hoax news.

Keyword: Naive Bayes, Logistic Regression, TF-IDF, Hoax Detection, Text Classification.