PERBANDINGAN BERBAGAI TEKNIK FEATURE EXTRACTION PADA NAIVE BAYES UNTUK MENGOPTIMALKAN DETEKSI BERITA HOAX DI INDONESIA

LAPORAN MBKM PENELITIAN

CHRISTIAN IVAN WIBOWO
00000058450

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

PERBANDINGAN BERBAGAI TEKNIK FEATURE EXTRACTION PADA NAIVE BAYES UNTUK MENGOPTIMALKAN DETEKSI BERITA HOAX DI INDONESIA

LAPORAN MBKM PENELITIAN

CHRISTIAN IVAN WIBOWO
00000058450

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Christian Ivan Wibowo
Nomor Induk Mahasiswa	:	00000058450
Program Studi	:	Informatika

Skripsi dengan judul:

Perbandingan Berbagai Teknik Feature Extraction Pada Naive Bayes Untuk Mengoptimalkan Deteksi Berita Hoax Di Indonesia

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 3 Januari 2025

(Christian Ivan Wibowo)

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	CHRISTIAN IVAN WIBOWO
NIM	:	00000058450
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Perbandingan Berbagai Teknik Feature Extraction Pada Naive Bayes Untuk Mengoptimalkan Deteksi Berita Hoax Di Indonesia

Menyatakan dengan sesungguhnya bahwa saya bersedia:

☐ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☑ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

Tangerang, 3 Januari 2025

Yang menyatakan

CHRISTIAN IVAN WIBOWO

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

”Keberhasilan dimulai dengan keberanian untuk mencoba”

Walt Disney

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan laporan penelitian ini. Penelitian ini bertujuan untuk membandingkan teknik feature extraction pada Naive Bayes dalam mengoptimalkan deteksi berita hoax di Indonesia, yang diharapkan dapat memberikan kontribusi pada pengembangan teknologi dan cybersecurity. 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 Assoc. Prof. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak David Agustriawan, S.Kom., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Ibu Dr. Sy. Yuliani Yakub, S.Kom., M.T, selaku Kepala Tim Peneliti UNIIC UMN.
6. Rivo Juicer Wowor, Muhammad Alfarizky Ramadhani Oscandar, serta Fadhil Dzaky Muhammad selaku rekan-rekan tim Peneliti UNIIC UMN
7. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga penelitian ini bermanfaat untuk pengembangan ilmu pengetahuan dan teknologi.

Tangerang, 3 Januari 2025

CHRISTIAN IVAN WIBOWO

PERBANDINGAN BERBAGAI TEKNIK FEATURE EXTRACTION PADA NAIVE BAYES UNTUK MENGOPTIMALKAN DETEKSI BERITA HOAX DI INDONESIA

CHRISTIAN IVAN WIBOWO

ABSTRAK

Pada era digital saat ini, penyebaran berita hoax menjadi masalah yang semakin mendesak, yang memengaruhi stabilitas sosial dan politik, serta merusak kepercayaan publik. Deteksi berita hoax secara otomatis dapat membantu mengatasi masalah ini dengan menggunakan teknik klasifikasi teks berbasis machine learning. Penelitian ini menguji kinerja algoritma Naive Bayes dalam mendeteksi berita bohong (hoax), dengan menggunakan dua teknik ekstraksi fitur yaitu CountVectorizer dan TfidfVectorizer, dengan variasi n-gram (unigram, bigram, trigram). Data yang digunakan diperoleh dari sumber Mafindo API dan Kaggle, yang terdiri dari 29.552 entri berita yang terdiri dari berita hoax dan berita faktual. Hasil percobaan menunjukkan bahwa model Naive Bayes dengan unigram CountVectorizer (1,1) memberikan kinerja terbaik dengan akurasi sebesar 93% pada data uji, sedangkan unigram TfidfVectorizer (1,1) menghasilkan akurasi sebesar 90,6%. Selain itu, penggunaan n-gram yang lebih tinggi (bigram dan trigram) justru menurunkan kinerja model. Temuan ini menunjukkan bahwa dalam konteks pendeteksian hoax, penggunaan unigram CountVectorizer lebih efektif, sedangkan TfidfVectorizer lebih cocok untuk tugas yang memerlukan penekanan pada kata-kata yang jarang. Studi ini berkontribusi pada pengembangan sistem pendeteksian hoax yang lebih efisien dan akurat.

Kata kunci: CountVectorizer, Hoax, Klasifikasi Teks, Naive Bayes, TfidfVectorizer

COMPARISON OF VARIOUS FEATURE EXTRACTION TECHNIQUES IN NAIVE BAYES TO OPTIMIZE HOAX NEWS DETECTION IN INDONESIA

CHRISTIAN IVAN WIBOWO

ABSTRACT

In today's digital era, the spread of hoaxes has become an increasingly pressing problem, affecting social and political stability, and damaging public trust. Automatic detection of hoaxes can help overcome this problem by using machine learning-based text classification techniques. This study tests the performance of the Naive Bayes algorithm in detecting hoaxes, using two feature extraction techniques: CountVectorizer and TfidfVectorizer, with variations of n-grams (unigrams, bigrams, trigrams). The data used were obtained from Mafindo API and Kaggle sources, consisting of 29,552 news entries consisting of hoax and factual news. The experimental results show that the Naive Bayes model with CountVectorizer unigram (1,1) provides the best performance with an accuracy of 93% on the test data, while TfidfVectorizer unigram (1,1) produces an accuracy of 90.6%. In addition, the use of higher n-grams (bigrams and trigrams) actually decreases the model's performance. These findings suggest that in the context of hoax detection, the use of unigram CountVectorizer is more effective, while TfidfVectorizer is more suitable for tasks that require emphasis on rare words. This study contributes to the development of a more efficient and accurate hoax detection system.

Keywords: *CountVectorizer, Hoax, Naive Bayes, Text Classification, TfidfVectorizer*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN