

**PENERAPAN METODE *MACHINE LEARNING*  
DALAM MENGIDENTIFIKASI BERITA HOAKS**

**SKRIPSI**



Oleh  
**Vijay Andika Saputra**  
201510033

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2024**

**PENERAPAN METODE *MACHINE LEARNING*  
DALAM MENGIDENTIFIKASI BERITA HOAKS**

**SKRIPSI**

**Untuk Memenuhi salah satu syarat  
memperoleh gelar sarjana**



**Oleh:  
Vijay Andika Saputra  
201510033**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2024**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Vijay Andika Saputra

NPM : 201510033

Fakultas : Teknik dan Komputer

Program Studi : Sistem Informasi

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

### **PENERAPAN METODE *MACHINE LEARNING* DALAM MENGIDENTIFIKASI BERITA HOAKS**

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 15 Januari 2024



**Vijay Andika Saputra**

201510033

**PENERAPAN METODE *MACHINE LEARNING*  
DALAM MENGIDENTIFIKASI BERITA HOAKS**

**SKRIPSI**

**Untuk Memenuhi salah satu syarat  
memperoleh gelar sarjana**

**Oleh  
Vijay Andika Saputra  
201510033**

**Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera di bawah ini**

**Batam, 15 Januari 2024**



**Sasa Ani Arnomo, S.Kom., M.SI.  
Pembimbing**

## ABSTRAK

Dalam era informasi digital yang semakin berkembang, berita hoaks atau palsu menjadi tantangan serius yang dapat mempengaruhi persepsi dan keputusan masyarakat. Penelitian ini bertujuan untuk mengembangkan metode deteksi berita hoaks menggunakan pendekatan *machine learning* pada data berita. Fokus penelitian ini adalah mengidentifikasi berita hoaks yang terkait. Penelitian ini melibatkan pengumpulan data berita dari sumber berita daring terkemuka dan penerapan algoritma *machine learning*, termasuk *Logistic Regression*, *Decision Tree*, *Gradient Boosting*, dan *Random Forest*, untuk melakukan klasifikasi berita hoaks. Performa model diukur menggunakan metrik akurasi, presisi, *recall*, dan *f1-score*, dengan hasil pengujian dibandingkan dengan evaluasi manusia yang telah terlatih dalam mengenali berita hoaks. Hasil penelitian ini menunjukkan bahwa pendekatan *machine learning* dapat berhasil digunakan untuk mengidentifikasi berita hoaks dengan tingkat akurasi yang tinggi. Namun, penelitian ini juga mengidentifikasi batasan dalam pendekatan ini, seperti keterbatasan data pelatihan dan kompleksitas bahasa. Meskipun demikian, penelitian ini memberikan kontribusi penting dalam upaya untuk mengatasi masalah berita hoaks dan memberikan landasan untuk pengembangan lebih lanjut dalam deteksi berita hoaks menggunakan teknologi *machine learning*.

**Kata Kunci:** Analisis Berita, Berita Hoaks, Klasifikasi Teks, *Machine Learning*

## **ABSTRACT**

*In the era of increasingly developing digital information, hoax or fake news is a serious challenge that can influence people's perceptions and decisions. This research aims to develop a hoax news detection method using a machine learning approach on news data. The focus of this research is to identify related hoax news. This research involves collecting news data from leading online news sources and applying machine learning algorithms, including Logistic Regression, Decision Tree, Gradient Boosting, and Random Forest, to classify hoax news. Model performance is measured using accuracy, precision, recall and f1-score metrics, with test results compared with human evaluations that have been trained to recognize hoax news. The results of this research show that the machine learning approach can be successfully used to identify hoax news with a high level of accuracy. However, this study also identified limitations in this approach, such as limited training data and language complexity. Nevertheless, this research makes an important contribution in efforts to overcome the problem of hoax news and provides a basis for further development in hoax news detection using machine learning technology.*

**Keywords:** *Hoax News, Machine Learning, News Analysis, Text Classification*

## KATA PENGANTAR

Puji syukur penulis ucapkan kepada Tuhan Yang Maha Esa, yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. selaku rektor Universitas Putera Batam;
2. Bapak Welly Sugianto, S.T., M.M. selaku dekan Fakultas Teknik dan Komputer Universitas Putera Batam;
3. Bapak Muhammad Rasid Ridho, S.Kom., M.SI. selaku ketua Program Studi Sistem Informasi Universitas Putera Batam;
4. Sasa Ani Arnomo, S.Kom., M.SI. selaku pembimbing Skripsi pada Program Studi Universitas Putera Batam;
5. Mesri Silalahi S.KOM., M.SI. selaku pembimbing Akademik pada Program Studi Universitas Putera Batam;
6. Dosen dan Staff Universitas Putera Batam;
7. Orang tua, keluarga dan teman penulis yang telah memberikan dukungan dan motivasi dalam menyelesaikan skripsi ini.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 15 Januari 2024



Penulis  
Vijay Andika Saputra

## DAFTAR ISI

<b>HALAMAN SAMPUL DEPAN</b> .....	<b>I</b>
<b>HALAMAN JUDUL</b> .....	<b>II</b>
<b>SURAT PERNYATAAN ORISINALITAS</b> .....	<b>III</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>IV</b>
<b>ABSTRAK</b> .....	<b>V</b>
<b>ABSTRACT</b> .....	<b>VI</b>
<b>KATA PENGANTAR</b> .....	<b>VII</b>
<b>DAFTAR ISI</b> .....	<b>VIII</b>
<b>DAFTAR GAMBAR</b> .....	<b>X</b>
<b>DAFTAR TABEL</b> .....	<b>XI</b>
<b>DAFTAR RUMUS</b> .....	<b>XII</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1    LATAR BELAKANG .....	1
1.2    IDENTIFIKASI MASALAH.....	3
1.3    BATASAN MASALAH .....	4
1.4    RUMUSAN MASALAH .....	5
1.5    TUJUAN PENELITIAN .....	6
1.6    MANFAAT PENELITIAN.....	7
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>8</b>
2.1    TINJAUAN TEORI UMUM .....	8
2.1.1    Berita .....	8
2.1.2    Berita <i>Hoax</i> .....	8
2.1.3    Identifikasi.....	9
2.1.4 <i>Web Scraping</i> .....	10
2.1.5 <i>Dataset</i> Berita.....	12
2.1.6 <i>Python</i> .....	12
2.2    TINJAUAN TEORI KHUSUS .....	13
2.2.1 <i>Machine learning</i> .....	13
2.2.2    Pustaka Visualisasi Data .....	15
2.2.3    Paket <i>Machine learning</i> .....	16
2.2.4 <i>Kaggle</i> .....	17
2.2.5    Algoritma Klasifikasi .....	18
2.2.6    Platform Cloud .....	19
2.3    KERANGKA BERFIKIR.....	20
2.4    PENELITIAN TERDAHULU .....	23



<b>BAB III METODE PENELITIAN .....</b>	<b>31</b>
3.1 DESAIN PENELITIAN.....	31
3.2 OBJECT PENELITIAN.....	34
3.3 POPULASI DAN SAMPEL.....	35
3.3.1 Populasi.....	35
3.3.2 Sampel.....	35
3.4 TEKNIK PENGUMPULAN DATA .....	36
3.5 <i>EKSTRASI</i> FITUR .....	36
3.6 MODEL KLASIFIKASI.....	38
3.7 PELATIHAN DAN EVALUASI.....	40
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>44</b>
4.1 HASIL PENELITIAN.....	44
4.2 PEMBAHASAN.....	50
4.2.1 Pengumpulan Data dari Web .....	51
4.2.2 Penggunaan <i>Dataset</i> .....	54
4.2.3 Proses Pelabelan Data .....	56
4.2.4 <i>Preprocessing</i> Teks.....	58
4.2.5 Analisis Data <i>Eksploratif</i> .....	61
4.2.6 Ekstraksi Fitur dengan <i>TF-IDF</i> .....	64
<b>BAB V SIMPULAN DAN SARAN.....</b>	<b>67</b>
5.1 SIMPULAN .....	67
5.2 SARAN.....	68
<b>DAFTAR PUSTAKA.....</b>	<b>70</b>
<b>LAMPIRAN.....</b>	<b>73</b>
LAMPIRAN 1. DATA PENELITIAN .....	73
LAMPIRAN 2. DAFTAR RIWAYAT HIDUP .....	81
LAMPIRAN 3. SURAT KETERANGAN PENELITIAN.....	82

## DAFTAR GAMBAR

<b>Gambar 2 1</b> Kerangka Berfikir .....	21
<b>Gambar 3 1</b> Deain Penelitian.....	31
<b>Gambar 3 2</b> Sampel .....	35
<b>Gambar 4 1</b> Perbandingan Rata-Rata Algoritma Klasifikasi.....	46
<b>Gambar 4 2</b> Model Testing Data Valid .....	49
<b>Gambar 4 3</b> Model Testing Data <i>Hoax</i> .....	50
<b>Gambar 4 4</b> Dataset Valid .....	55
<b>Gambar 4 5</b> Dataset <i>Hoax</i> .....	56
<b>Gambar 4 6</b> Jumlah Setiap <i>Category</i> .....	62
<b>Gambar 4 7</b> Jumlah Data <i>Hoax</i> dan Valid.....	63
<b>Gambar 4 8 (a)</b> <i>Wordcloud</i> Berita Valid (b) <i>Wordcloud</i> Berita <i>Hoax</i> .....	64
<b>Gambar 4 9</b> Hasil pembobotan dengan <i>TF-IDF</i> .....	66

## DAFTAR TABEL

<b>Tabel 2 1</b> Penelitian Terdahulu.....	23
<b>Tabel 4 1</b> Evaluasi Model.....	44
<b>Tabel 4 2</b> Perhitungan <i>Logistic Regression</i> .....	47

## DAFTAR RUMUS

<b>Rumus 4 1</b> Perhitungan Logistic Regression .....	48
--	----