

PERANCANGAN *MODEL MACHINE LEARNING* UNTUK MENDETEKSI
BERITA HOAKS MENGGUNAKAN ALGORITMA KLASIFIKASI



SKRIPSI

*Diajukan untuk memenuhi syarat dalam memperoleh gelar Sarjana Pendidikan di
Program Studi Pendidikan Sistem dan Teknologi Informasi*

Oleh:

Adrian Kusuma Widjaja Kardana
NIM. 2103456

PROGRAM STUDI S1
PENDIDIKAN SISTEM DAN TEKNOLOGI INFORMASI
KAMPUS UPI DI PURWAKARTA
UNIVERSITAS PENDIDIKAN INDONESIA

2025

LEMBAR HAK CIPTA

**PERANCANGAN *MODEL MACHINE LEARNING* UNTUK MENDETEKSI
BERITA HOAKS MENGGUNAKAN ALGORITMA KLASIFIKASI**

Oleh

Adrian Kusuma Widjaja Kardana

Sebuah skripsi yang diajukan untuk memenuhi syarat memperoleh gelar Sarjana
Pendidikan di Universitas Pendidikan Indonesia Kampus Daerah Purwakarta

© **Adrian Kusuma Widjaja Kardana** 2025

Universitas Pendidikan Indonesia

Agustus 2025

Hak Cipta dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difotokopi, atau cara lainnya tanpa izin dari penulis.

Adrian Kusuma Widjaja Kardana
PERANCANGAN MODEL MACHINE LEARNING
UNTUK MENDETEKSI BERITA HOAKS
MENGGUNAKAN ALGORITMA KLASIFIKASI

Disetujui dan disahkan oleh pembimbing:

Pembimbing



Rizki Hikmawan, S.Pd., M.Pd.

NIP.920171219880731101

Mengetahui:

Ketua Program Studi
Pendidikan Sistem dan Teknologi Informasi



Ir. Nuur Wachid Abdul Majid, S.Pd., M.Pd., IPM., ASEAN Eng.

NIPT.920171219910625101

PERANCANGAN *MODEL MACHINE LEARNING* UNTUK MENDETEKSI BERITA HOAKS MENGGUNAKAN ALGORITMA KLASIFIKASI

Adrian Kusuma Widjaja Kardana

2103456

ABSTRAK

Penelitian ini bertujuan mengembangkan aplikasi deteksi berita hoaks berbasis *Natural Language Processing* (NLP) dan algoritma *Machine Learning* dengan metode *Research and Development* (R&D) 4D, di mana tahap pengembangan aplikasi menggunakan model Waterfall. Dataset penelitian terdiri dari total 26.557 berita, yakni 14.665 berita valid dan 11.892 berita palsu. Empat model diuji: *Logistic Regression*, *Naive Bayes*, *Random Forest*, dan BERT. Berdasarkan evaluasi, *Logistic Regression* menghasilkan kinerja terbaik (*accuracy* 92,72%, *precision* 0,93, *recall* 0,93, *F1-score* 0,93) dengan performa stabil tanpa indikasi *overfitting*. Uji coba lapangan melibatkan 35 mahasiswa sebagai responden melalui pemberian materi, penggunaan aplikasi, dan post-test 10 soal yang telah divalidasi. Hasil uji *One-Sample Wilcoxon Signed Rank Test* menunjukkan $p < 0,001$, menandakan aplikasi berpengaruh signifikan terhadap peningkatan literasi digital peserta. Temuan ini menunjukkan bahwa sistem tidak hanya akurat mendeteksi berita hoaks, tetapi juga efektif sebagai sarana edukasi publik.

Kata kunci: Deteksi Hoaks, *Natural Language Processing*, *Machine Learning*

DEVELOPMENT OF A MACHINE LEARNING MODEL FOR HOAX NEWS DETECTION USING A CLASSIFICATION ALGORITHM

Adrian Kusuma Widjaja Kardana

2103456

ABSTRACT

This study aims to develop a hoax news detection application based on Natural Language Processing (NLP) and Machine Learning algorithms using the Research and Development (R&D) 4D method, with the development stage implemented through the Waterfall model. The dataset comprises 26,557 news articles, consisting of 14,665 valid and 11,892 hoax articles. Four models were tested: Logistic Regression, Naive Bayes, Random Forest, and BERT. Logistic Regression achieved the best performance (accuracy 92.72%, precision 0.93, recall 0.93, F1-score 0.93) with stable results and no significant overfitting. A field experiment involved 35 university students as respondents, including material delivery, application usage, and a 10-item validated post-test. The One-Sample Wilcoxon Signed Rank Test yielded $p < 0.001$, indicating a significant effect of the application on improving participants' digital literacy. These findings suggest that the system is not only accurate in detecting hoax news but also effective as a public educational tool.

Keywords: Hoax Detection, Natural Language Processing, Machine Learning

DAFTAR ISI

LEMBAR HAK CIPTA	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN BEBAS PLAGIARISME	iv
KATA PENGANTAR	v
UCAPAN SYUKUR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.4.1 Manfaat Teoritis	5
1.4.2 Manfaat Praktis	5
1.5 Ruang Lingkup Penelitian	5
BAB II TINJAUAN PUSTAKA	7
2.1 <i>Artificial Intelligence (AI)</i>	7
2.2 <i>Machine Learning</i>	8
2.3 <i>Natural Language Processing</i>	10
2.3.1 <i>Logistic Regression</i>	12
2.3.2 <i>Naive Bayes</i>	12
2.3.3 <i>Random Forest</i>	13
2.3.4 BERT	13
2.4 Hoaks	14
2.5 Penelitian Terdahulu	16
2.5.1 Penelitian yang Dilakukan oleh Mitranikasih Laia et al. (2025)	16
2.5.2 Penelitian yang Dilakukan oleh Thota A, Tilak P, Ahluwalia S, Lohia	17

2.5.3 Penelitian yang Dilakukan oleh Gururaj H, Lakshmi H, Soundarya B, Flammini F, Janhavi V (2022)	18
2.5.4 Penelitian yang Dilakukan oleh William & Handhayani (2025).....	18
BAB III METODE PENELITIAN	20
3.1 Jenis dan Desain Penelitian	20
3.2 Prosedur Penelitian	22
3.3 Objek Penelitian	26
3.4 Teknik Pengumpulan Data	26
3.5 Teknik Pengujian Data	28
3.5.1 Instrumen Penelitian Aplikasi	28
3.5.2 Instrumen penelitian Lapangan	29
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Hasil	34
4.1.1 <i>Define</i> (Identifikasi).....	34
4.1.2 <i>Design</i> (Perancangan Aplikasi)	36
4.1.3 <i>Develop</i> (Pengembangan Aplikasi)	39
4.1.4 Uji Model Klasifikasi	61
4.1.5 Eksperimen Lapangan	66
4.2 Pembahasan	74
4.2.1 Perancangan Sistem Aplikasi	74
4.2.2 Komparasi dengan Penelitian Terdahulu.....	75
4.2.3 Pengaruh terhadap Skill Literasi Digital	77
BAB V KESIMPULAN DAN SARAN	78
5.1 Kesimpulan.....	78
5.2 Saran.....	79
DAFTAR PUSTAKA.....	80
LAMPIRAN	83
RIWAYAT PENULIS.....	172

DAFTAR TABEL

Tabel 3. 1 Rubrik Penilaian	28
Tabel 3. 2 Tabel Nilai Kritis Wilcoxon	32
Tabel 4. 1 Hasil Pengujian Model untuk Berita Asli	59
Tabel 4. 2 Hasil Pengujian Model untuk Berita Palsu.....	60
Tabel 4. 3 Komparasi Pengujian dengan Berbagai Algoritma	61
Tabel 4. 4 Daftar Pertanyaan yang Digunakan dalam Eksperimen	69
Tabel 4. 5 Hasil Validasi oleh Para Panelis.....	70
Tabel 4. 6 Hasil Uji Deskriptif	71
Tabel 4. 7 Hasil Uji Normalitas Shapiro-Wilk.....	72
Tabel 4. 8 Hasil Uji Hipotesis One Sample Wilcoxon	73

DAFTAR GAMBAR

Gambar 1. 1 Hasil Temuan Isu Hoaks Komdigi pada Tahun 2024.....	2
Gambar 1. 2 Laporan Indeks Literasi Digital Komdigi dari tahun 2020-2022	3
Gambar 1. 3 Hasil Survei IMDI Perilaku Berpikir Kritis dalam Merespon Informasi di Media Digital 2024	3
Gambar 2. 1 10 Cabang AI.....	7
Gambar 2. 2 Gambar Ilustrasi <i>Supervised Learning vs Unsupervised Learning</i>	9
Gambar 3. 1 Alur <i>Research and Development (R&D)</i> 4D Model yang diadaptasi	21
Gambar 3. 2 Desain Penelitian <i>One-Shot Case Study</i>	22
Gambar 3. 3 Alur Konsep Aplikasi Bekerja.....	23
Gambar 3. 4 Visualisasi Alur Eksperimen	25
Gambar 3. 5 Perhitungan Uji Model dengan <i>Precision, Recall, Accuracy</i>	27
Gambar 3. 6 Perhitungan Uji Model dengan <i>F1-Score</i>	27
Gambar 4. 1 Alur kerja aplikasi	37
Gambar 4. 2 Antarmuka Pada Aplikasi.....	38
Gambar 4. 3 Alur Pengembangan Aplikasi.....	39
Gambar 4. 4 Desain Arsitektur Aplikasi	40
Gambar 4. 5 Struktur Dataset Berita Palsu.....	42
Gambar 4. 6 Struktur Dataset Berita Asli.....	42
Gambar 4. 7 Instalasi dan Impor Pustaka.....	43
Gambar 4. 8 Pembacaan File Dataset dengan Kode Python	43
Gambar 4. 9 Tampilan Hasil Penggabungan Kumpulan File Menjadi Satu	44
Gambar 4. 10 Pengaturan Ulang Index Pada Data dengan Kode Python.....	44
Gambar 4. 11 Pembagian Dataset Menjadi Training dan <i>Testing</i>	45
Gambar 4. 12 Hasil pembagian dataset	45
Gambar 4. 13 Proporsi Data Training dan <i>Testing</i>	46
Gambar 4. 14 Kode Python untuk pelatihan dan evaluasi <i>Logistic Regression</i> dengan <i>TF-IDF Vectorizer</i> pada data teks.....	47
Gambar 4. 15 Hasil Visualisasi Confusion Matrix.....	48
Gambar 4. 16 Visualisasi <i>Confusion Matrix</i> untuk data pelatihan dan pengujian menggunakan <i>heatmap</i>	49
Gambar 4. 17 Visualisasi <i>Confusion Matrix</i> untuk <i>Data Training</i>	49
Gambar 4. 18 Visualisasi <i>Confusion Matrix</i> untuk <i>Data Testing</i>	50
Gambar 4. 19 Penyimpanan model <i>Logistic Regression</i> dan <i>TF-IDF Vectorizer</i> menggunakan <i>joblib</i>	50
Gambar 4. 20 Instalasi dan Impor Pustaka untuk <i>User Interface</i>	51
Gambar 4. 21 Dekorator <code>@st.cache_resource</code> untuk menyimpan sumber daya agar tidak dimuat ulang di Streamlit	52
Gambar 4. 22 Fungsi <code>load_models()</code> untuk memuat <i>TF-IDF Vectorizer</i> dan model <i>Logistic Regression</i> dari file menggunakan <i>joblib</i>	52
Gambar 4. 23 Fungsi <code>configure_genai()</code> untuk mengatur API key Gemini dan menginisialisasi model <i>gemini-2.5-flash-lite</i>	53
Gambar 4. 24 Fungsi <code>get_explanation()</code> untuk menghasilkan penjelasan klasifikasi berita menggunakan model Gemini	54
Gambar 4. 25 Kode antarmuka aplikasi Streamlit untuk deteksi berita palsu dengan konfigurasi API Gemini dan deskripsi aplikasi	55

Gambar 4. 26 Kode utama aplikasi Streamlit untuk mendeteksi berita palsu, menampilkan hasil klasifikasi, penjelasan, dan tingkat kepercayaan model	56
Gambar 4. 27 Tampilan antarmuka aplikasi Streamlit Detektor Berita Palsu dengan fitur input teks berita dan konfigurasi API Gemini	57
Gambar 4. 28 Hasil Pengujian Model	58
Gambar 4. 29 Performa model pada data latih dan uji dengan <i>Logistic Regression</i>	63
Gambar 4. 30 Performa model pada data latih dan uji dengan Naive Baiyes	64
Gambar 4. 31 Performa model pada data latih dan uji dengan <i>Random Forest</i>	65
Gambar 4. 32 Pelaksanaan Eksperimen Lapangan	66
Gambar 4. 33 Alur Aktivitas Eksperimen	67
Gambar 4. 34 Contoh Soal yang Digunakan dalam Eksperimen	68
Gambar 4. 35 Hasil Uji Hipotesis One Sample Wilcoxon	73
Gambar 4. 36 Flowchart Penggunaan Sistem Deteksi yang Dibangun.....	75

DAFTAR LAMPIRAN

Lampiran 1 Dataset Berita.....	83
Lampiran 2 Code app.py	84
Lampiran 3 <i>Machine Learning</i> Model	86
Lampiran 4 Uji <i>Testing</i> Aplikasi	133
Lampiran 5 Format Lembar Validasi	144
Lampiran 6 Bukti Panelis Validasi.....	149
Lampiran 7 <i>Content Validity Ratio</i> (CVR).....	152
Lampiran 8 Materi Ajar Berita Hoaks Eksperimen.....	153
Lampiran 9 Format Post-Test.....	158
Lampiran 10 Hasil Responden Post-Test	159
Lampiran 11 Uji Normalitas.....	160
Lampiran 12 Uji Hipotesis	164
Lampiran 13 Dokumentasi Lainnya	166
Lampiran 14 SK Judul dan Bimbingan Skripsi.....	168
Lampiran 15 Lembar Bimbingan Skripsi.....	171

DAFTAR PUSTAKA

- Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *SECURITY AND PRIVACY*, 1(1). <https://doi.org/10.1002/spy2.9>
- Arikunto Suharsimi. (2010). Prosedur penelitian suatu pendekatan praktik. In *Jakarta: Rineka Cipta*.
- Bauskar, S., Badole, V., Jain, P., & Chawla, M. (2019). Natural Language Processing based Hybrid Model for Detecting Fake News Using Content-Based Features and Social Features. *International Journal of Information Engineering and Electronic Business*, 11(4), 1–10. <https://doi.org/10.5815/ijieeb.2019.04.01>
- Bennett, N., Borg, W. R., & Gall, M. D. (1984). Educational Research: An Introduction. *British Journal of Educational Studies*, 32(3). <https://doi.org/10.2307/3121583>
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of Natural Language Processing research. In *IEEE Computational Intelligence Magazine* (Vol. 9, Issue 2, pp. 48–57). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/MCI.2014.2307227>
- Chowdhury, G. G. (2003). *Natural Language Processing. Annual Review of Information Science and Technology*, 37(1), 51–84. <https://doi.org/https://doi.org/10.1002/ariss.1440370103>
- Clara, S., Laksmi Prianto, D., Al Habsi, R., Friscila Lumbantobing, E., & Chamidah, N. (2021). Implementasi Seleksi Fitur Pada Algoritma Klasifikasi *Machine Learning* Untuk Prediksi Penghasilan Pada Adult Income Dataset. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, 2(1).
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://github.com/tensorflow/tensor2tensor>
- Ertel, W. (2024). *Undergraduate Topics in Computer Science Introduction to Artificial Intelligence*. <http://www.springer.com/series/7592>
- European Commission. (2018). *A DEFINITION OF AI: MAIN CAPABILITIES AND SCIENTIFIC DISCIPLINES*. <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>
- GeeksForGeeks. (2025, July 23). *Top 10 branches of Artificial Intelligence*.
- Gururaj, H. L., Lakshmi, H., Soundarya, B. C., Flammini, F., & Janhavi, V. (2022). Machine Learning-Based Approach for Fake news detection. *Journal of ICT Standardization*, 10(4), 509–530. <https://doi.org/10.13052/jicts2245-800X.1042>
- Handhayani, T. (2025). Perbandingan Kinerja Naïve Bayes dan *Random Forest* dalam Mendeteksi Berita Palsu. In *Jurnal Informatika Sunan Kalijaga* (Vol. 10, Issue 2). MEI. <https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset?select=True.csv>.

- Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models Third Edition draft Summary of Contents*.
- Lawhon, D. (1974). Instructional development for training teachers of exceptional children: A sourcebook. *Journal of School Psychology*, 14(1). [https://doi.org/10.1016/0022-4405\(76\)90066-2](https://doi.org/10.1016/0022-4405(76)90066-2)
- M. I. Jordan, & T. M. Mitchell. (2015). *Machine Learning: Trends, perspectives, and prospects*. *Science*, 349(6245), 253–255. <https://doi.org/10.1126/science.aac4520>
- Media rankings: Indonesia*. (2025, August 8). Scimago Media Ranking.
- Meskó, B., & Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. In *npj Digital Medicine* (Vol. 3, Issue 1). Nature Research. <https://doi.org/10.1038/s41746-020-00333-z>
- Mitranikasih Laia, Ayuliana, Wasiran, Muhammad Lukman Hakim, & Dikky Suryadi. (2025). Analisis Big Data untuk Deteksi Hoaks dan Disinformasi di Platform Berita Online. *Jurnal Teknologi Informasi Dan Komunikasi*, 776–782.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning: Vol. MIT Press*.
- Norman, D. A. . (2013). *The Design of Everyday Things : Revised and Expanded Edition*. Basic Books.
- Oshikawa, R., Qian, J., & Wang, W. Y. (2020). *A Survey on Natural Language Processing for Fake news detection*. <https://www.snopes.com/fact-check/>
- Pardha, R. H., M<, P., & Kaufmann, M. (2019). *The UX Book Agile UX Design for a Quality User Experience Second Edition*.
- Prachi, N. N., Habibullah, M., Rafi, M. E. H., Alam, E., & Khan, R. (2022). Detection of Fake News Using *Machine Learning and Natural Language Processing Algorithms*. *Journal of Advances in Information Technology*, 13(6), 652–661. <https://doi.org/10.12720/jait.13.6.652-661>
- Rismanta Silalahi, R., & Sevilla, V. (2020). *REKONSTRUKSI MAKNA HOAKS DI TENGAH ARUS INFORMASI DIGITAL*.
- Royce, W. W. (1970). Managing the Development of Large Software Systems (1970). In *Ideas That Created the Future*. <https://doi.org/10.7551/mitpress/12274.003.0035>
- Siti Badriyah. (2023). *Pengertian Hoaks: Sejarah, Jenis, Contoh, Penyebab dan Cara Menghindarinya*. Gramedia.
- Sommerville, Ian. (2016). *Software engineering*. Pearson.
- Steven Bird, Ewan Klein, & and Edward Loper. (2009). *Natural Language Processing with Python*.

- Sugiyono. (2017). Metode penelitian kuantitatif kualitatif dan R&D. In *Penerbit Alfabeta*. <https://doi.org/WWW.cvalfabeta.com>
- Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). *Fake news detection: A Deep learning Approach*. *SMU Data Science Review*, 1(3), 10. <https://scholar.smu.edu/datasciencereview> Available at: <https://scholar.smu.edu/datasciencereview/vol1/iss3/10http://digitalrepository.smu.edu>.
- Ulfa Batoebara, M., & Hasugian, B. S. (2023). *ISU HOAKS MENINGKAT MENJADI POTENSI KEKACAUAN INFORMASI*. 4(2). <https://www.kominfo.go.id/content/detail/51727/siaran-pers-no>
- Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37. <https://doi.org/10.2478/jagi-2019-0002>
- Yau, K. L. A., Lee, H. J., Chong, Y. W., Ling, M. H., Syed, A. R., Wu, C., & Goh, H. G. (2021). Augmented Intelligence: Surveys of Literature and Expert Opinion to Understand Relations between Human Intelligence and Artificial Intelligence. *IEEE Access*, 9, 136744–136761. <https://doi.org/10.1109/ACCESS.2021.3115494>
- Zhou, Z. H. (2021). *Machine Learning*. In *Machine Learning*. Springer Nature. <https://doi.org/10.1007/978-981-15-1967-3>