JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Perbandingan Algoritma SVM, Logistic Regression, Naive Bayes, dan XGBoost Untuk Deteksi Fake News

Umar Farid Al Faqihi*, Afril Efan Pajri, Muhammad Jauhar Vikri

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
Email: [1*]umarfarid273@gmail.com, [2]afril@unugiri.ac.id,[3]vikri@unugiri.ac.id
Email Penulis Korespondensi: umarfarid273@gmail.com
Submitted **12-01-2026;** Accepted **28-02-2026;** Published **28-02-2026**

Abstrak

Kemajuan teknologi serta perkembangan internet memungkinkan masyarakat mendapatkan informasi dengan jauh lebih cepat dan luas. Hal ini juga diiringi dengan meningkatnya penyebaran berita palsu (fake news) yang dapat menyesatkan masyarakat. Berita hoax tidak hanya menyesatkan masyarakat, tetapi juga mengancam stabilitas sosial, politik, dan kesehatan masyarakat. Penelitian ini dilakukan untuk mengenali berita palsu dengan memanfaatkan teknik machine learning yang cocok digunakan pada text mining. Empat metode klasifikasi yang dipakai adalah Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), dan Extreme Gradient Boosting (XGBoost). Keempat algoritma ini digunakan untuk menemukan pola-pola dalam teks berita yang menunjukkan apakah berita tersebut asli atau palsu. Penelitian ini memiliki perbedaan dari studi-studi sebelumnya karena dataset yang digunakan diperoleh secara langsung melalui proses pengambilan data dari berbagai portal berita online berbahasa Indonesia. Jumlah total data dalam penelitian ini sebanyak 4909 data. Hasil yang didapat dalam penelitian berupa tingkat akurasi dari setiap model, akurasi yang dihasilkan sebesar 99.69% (SVM), 99.39% (LR), 99.29% (NB) dan 99.19% (XGBoost). Akurasi yang dihasilkan dari setiap model tidak jauh berbeda, meskipun hasil akurasi sudah tinggi, model juga akan melewati tahapan Cross validation untuk memaksimalkan akurasi yang didapat. Hasil Mean Cross validation untuk setiap model sebesar 99.57% (SVM), 99.52% (LR), 99.44% (NB) dan 99.49% (XGBoost). Dari perbandingan akurasi yang dihasilkan pada setiap model mencapai akurasi yang hampir maksimal, menunjukkan bahwa model-model tersebut relevan digunakan untuk klasifikasi berbasis teks mining.

Kata Kunci: Fake news; SVM; Logistic Regression; Naive Bayes; XGBoost

Abstract

The rapid growth of digital technology and internet access has completely changed how information is shared, enabling content to spread quickly across various online platforms. However, these advancements have also made it easier for misleading or entirely fabricated news to circulate, posing serious risks to social stability, political environments, and public health. This study tackles this problem by employing several machine learning-based classification methods for analyzing textual data. Four algorithms Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), and Extreme Gradient Boosting (XGBoost) were applied to detect linguistic patterns that differentiate genuine news from fake content. A major contribution of this research is the creation of a custom dataset gathered directly from Indonesian online news portals, comprising a total of 4,909 entries. The evaluation results demonstrate exceptionally high accuracy across the models: 99.69% for SVM, 99.39% for LR, 99.29% for NB, and 99.19% for XGBoost. To verify reliability, each model was further evaluated using cross-validation, yielding average accuracy scores of 99.57% (SVM), 99.52% (LR), 99.44% (NB), and 99.49% (XGBoost). These findings confirm that all four classifiers are highly effective and well-suited for identifying fake news in text-based data.

Keywords: Fake news; SVM; Logistic Regression; Naive Bayes; XGBoost

1. PENDAHULUAN

Di era digital saat ini, penyebaran informasi melalui *platform* online menjadi bagian penting dari kehidupan masyarakat. Kemajuan teknologi informasi ini juga membawa tantangan serius berupa penyebaran *fake news*, yang pada umumnya disebut sebagai hoaks. *Fake news* dapat diartikan sebagai informasi yang sengaja dibuat untuk menyesatkan, memanipulasi opini publik, atau memengaruhi keputusan sosial-politik [1]. Di Indonesia, fenomena ini semakin mengkhawatirkan, Kominfo mencatat terdapat 2.484 berita hoaks pada tahun 2022 dan sebagian besar tersebar melalui media sosial seperti Twitter, Facebook, dan TikTok [2]. Hingga saat ini tren terus meningkat tajam. Pada bulan Januari hingga Agustus 2025, Tim *Artificial Intelligence for Society* (AIS) Kementerian Kominfo telah mengidentifikasi 1.020 isu hoaks dan disinformasi. Pada bulan Januari terdapat 129 isu hoaks, Februari 124, Maret 104, April 103, Mei 124, Juni 133, Juli 160, dan Agustus 143 isu [3]. Fenomena tersebut dapat berdampak merusak kepercayaan masyarakat terhadap pemerintah, memperburuk perpecahan pendapat politik dan mengancam ketahanan nasional, sebagaimana dianalisis oleh penelitian [4] dalam konteks masyarakat digital Indonesia.

Fenomena ini menekankan pentingnya berpikir kritis dan literasi digital, karena rendahnya tingkat literasi membuat masyarakat rentan terhadap manipulasi. Sebagaimana dijelaskan dalam buku yang ditulis Pratama, hoaks berhasil menyebar luas karena kecenderungan bias berpikir manusia seperti bias konfirmasi (*confirmation bias*), pembuktian sosial (*social proof*), ketidak sesuaian cara berfikir, dan kesalahan dalam menilai, yang membuat orang cenderung menerima informasi tanpa verifikasi mendalam selama informasi itu sesuai dengan keyakinan atau kelompoknya [5].

Penyebaran *fake news* sering dimanfaatkan dalam periode krisis, seperti pemilu atau pandemi. Misalnya, pada Pemilu Presiden Indonesia 2024, hoaks dengan tema politik menjadi sarana untuk menggiring pemikiran dan opini publik, dengan 203 konten hoaks terkait pemilu yang ditangani oleh Kominfo pada awal tahun tersebut [6]. Hoaks politik sering kali berupa video atau teks yang dimanipulasi dan disebarkan untuk mendiskreditkan lawan politik. Selain itu, selama

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

pandemi Covid-19, penyebaran hoaks terkait virus corona berkontribusi terhadap keresahan masyarakat dan penurunan tingkat kepercayaan publik terhadap kebijakan pemerintah, sebagaimana dibahas oleh penelitian [7] dalam konteks dampak hukum dan sosial media. Fenomena ini menekankan pentingnya literasi digital, karena tingkat literasi yang rendah di masyarakat membuat penyelesaian semakin sulit [4][8].

Untuk mengatasi masalah tersebut, machine learning telah menjadi solusi potensial untuk deteksi fake news. Beberapa studi sebelumnya telah mengeksplorasi berbagai algoritma klasifikasi. Sebagaimana pada penelitian [9] membandingkan SVM dan *Naive Bayes* untuk mengklasifikasikan cyberbullying di Twitter, hasil yang didapat SVM mencapai akurasi 99,60% lalu dibandingkan akurasi yang didapat algoritma *Naive Bayes* sebesar 97,99%. Studi ini menunjukkan keunggulan SVM ketika bekerja dengan data yang berdimensi tinggi serta bersifat non-linear, meskipun *Naive Bayes* lebih sederhana dan efisien untuk dataset teks. Selain itu, penelitian [10] menerapkan *Naive Bayes* dan SVM untuk deteksi hoaks berita, dengan *Naive Bayes* mencapai akurasi 88% dan SVM 75,5%, menekankan peran fitur *term frequency* (TF) dalam meningkatkan performa model. *Naive Bayes* juga terbukti efektif dalam klasifikasi non-teks di konteks Indonesia, seperti identifikasi kualitas beras berdasarkan standar pangan nasional menggunakan electronic nose dan transformasi data *eksponensial*, dengan akurasi mencapai 97% [26]. Hal ini menunjukkan fleksibilitas *Naive Bayes* dalam berbagai domain klasifikasi, termasuk data sensorik yang memerlukan *feature extraction* statistik untuk meningkatkan akurasi[11].

Penelitian terdahulu juga menguji algoritma lain seperti *Logistic Regression* dan *XGBoost*. Penelitian [12] mengoptimalkan *Logistic Regression* dan *Random Forest* dengan TF-IDF dalam klasifikasi berita hoaks, dimana *Logistic Regression* mencapai akurasi 95,20% setelah tuning hiperparameter, unggul dalam keseimbangan presisi dan *recall*. Selain itu, penelitian [13] menggabungkan SVM, XGBoost, dan *ensemble stacking* dengan IndoBERT, pada model ensemble mencapai akurasi 82%, yang menunjukkan kekuatan kombinasi algoritma untuk menangani data tidak seimbang. Studi ini menekankan bahwa ensemble learning, seperti stacking, dapat menggabungkan kekuatan SVM untuk data berdimensi tinggi dan XGBoost untuk interaksi fitur kompleks, sehingga lebih kuat terhadap variasi teks berita. Namun, sebagian besar penelitian ini fokus pada perbandingan dua atau tiga algoritma saja, dan jarang membandingkan empat algoritma secara komprehensif pada dataset berita Indonesia yang beragam. Selain itu, aspek hukum dan pendidikan politik sering diabaikan, di mana kriminalisasi hoaks harus seimbang antara menjaga persatuan dan kebebasan berpendapat [14], serta pendidikan politik kritis diperlukan untuk mencegah hoaks pada pemilu [15]. Dampak hoaks terhadap preferensi sosial-politik remaja juga menunjukkan perlunya literasi yang lebih baik [16].

Penelitian sebelumnya dalam deteksi *fake news* berbahasa Indonesia umumnya mengandalkan dataset publik atau terbatas, seperti data hoaks Covid-19 [17][18], dataset *multilingual* umum [19], atau data spesifik dari *platform* media sosial[20], [21]. Dataset tersebut sering bersifat usang, terfokus pada topik tertentu, atau kurang mencerminkan pola hoaks terkini di portal berita nasional Indonesia. Gap utama terletak pada minimnya studi yang membandingkan secara bersamaan SVM, *Logistic Regression*, *Naive Bayes*, dan XGBoost pada dataset bersumber langsung dari portal berita terpercaya seperti CNN Indonesia, Detik.com, Kompas.com, Kompas Cek Fakta, dan Liputan6 Cek Fakta. Dataset ini mencakup berita valid-hoaks dengan variasi jurnalistik aktual serta memerlukan preprocessing khusus (stemming, stopword removal, TF-IDF) [2]. Meskipun penelitian [1] mengkaji dampak sosial *fake news*, integrasi dengan evaluasi *machine learning* masih terbatas. Penelitian ini mengisi gap tersebut melalui pengumpulan dataset orisinal via *web scraping*, serta perbandingan kinerja keempat algoritma menggunakan metrik akurasi, presisi, *recall*, f1-score, dan *cross-validation* untuk optimalisasi generalisasi.

Tujuan dari penelitian ini untuk mengembangkan pemodelan deteksi *fake news* yang akurat dan efisien untuk konteks bahasa Indonesia. Penelitian akan mengumpulkan dataset dari sumber berita online terverifikasi, menerapkan preprocessing, Ekstraksi fitur dengan TF-IDF, mengimplementasikan SVM dengan kernel linear, *Logistic Regression*, *Naive Bayes*, dan XGBoost serta membandingkan performa model untuk menentukan yang paling optimal. Kontribusi penelitian ini adalah menyediakan kerangka *machine learning* yang dapat diintegrasikan ke dalam platform cek fakta, seperti yang direkomendasikan oleh penelitian [4], untuk meningkatkan literasi digital dan ketahanan nasional terhadap disinformasi. Dengan ini, penelitian ini berkontribusi secara teknis dan juga sosial, sebagaimana diharapkan dalam studi [6] tentang komunikasi politik hoaks.

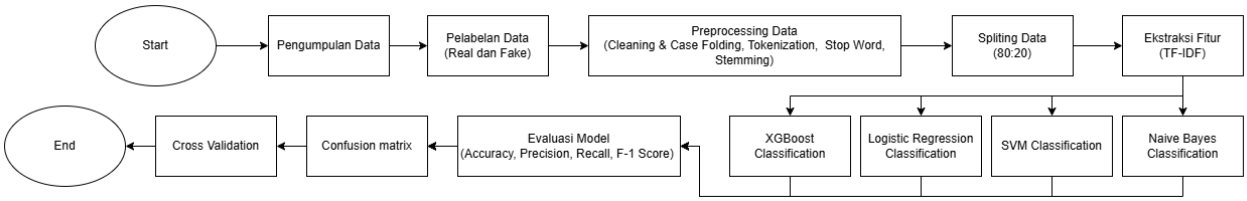
Melalui pendekatan ini, diharapkan dapat mengurangi dampak negatif *fake news*, seperti yang dialami selama pandemi [7] atau pemilu [6], dan mendukung pembangunan masyarakat yang lebih informasi. Penelitian selanjutnya dapat memperluas ke deep learning, seperti yang disarankan oleh [13], untuk menangani konteks multilingual.

2. METODOLOGI PENELITIAN

Penelitian ini memanfaatkan pendekatan text mining dengan metode *Knowledge Discovery in Database* (KDD) untuk menguji serta membandingkan kemampuan beberapa algoritma dalam mengenali berita hoax serta berita yang bukan hoax. Algoritma yang diuji meliputi SVM, *Logistic Regression*, *Naive Bayes*, dan XGBoost.Proses text mining dalam penelitian ini terdiri atas lima tahap utama sesuai metode KDD, yaitu pemilihan data, preprocessing, transformasi, penerapan data mining, dan tahap evaluasi. Pendekatan ini dipilih karena efektif dalam menangani data teks tidak terstruktur, sebagaimana diterapkan dalam studi deteksi hoaks oleh [17] dan [22]. Selain itu, *Naive Bayes* dapat dioptimalkan dengan PSO untuk sentiment analysis pada berita hoaks [21], yang mendukung integrasi aspek hukum dan

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

pendidikan politik dalam evaluasi [14][15]. Berikut merupakan alur proses yang ada pada penelitian ini sebagaimana dijelaskan pada Gambar 1.



Gambar 1. Flowchart Alur

Proses tahapan yang ada pada penelitian ini, dimana proses penelitian dimulai dengan pengumpulan data dari 6 sumber terverifikasi dengan menggunakan teknik *web scraping*, selanjutnya pelabelan data dengan 2 kategori label yaitu *real* dan *fake*, setelah pelabelan data akan melewati *preprocessing* data yang meliputi (*cleaning* dan *case folding*, *tokenization*, *stop word*, *stemming*), dilanjutkan dengan *splitting* data (pembagian data) dengan perbandingan 80:20 untuk data *training* dan data *testing*, kemudian proses ekstraksi fitur untuk mengukur bobot kata pada data teks, lalu pemodelan 4 algoritma dengan konfigurasi terbaik, dilanjutkan evaluasi model algortima yang meliputi (*accuracy*, *precision*, *recall*, dan *f1-score*), setelah itu, proses confusion matrix untuk memvisualisasikan kesalahan model, dan terakhir optimalisasi menggunakan *cross validation*.

2.1 Pengumpulan Data

Dataset yang digunakan dikumpulkan dengan teknik web scraping dari berbagai portal penyedia berita online yang terpercaya di Indonesia seperti CNN Indonesia, Detik News, Kompas, Kompas Cek Fakta, Liputan6 Cek Fakta, dan Turnbackhoax.id. Teknik web scraping dipilih karena memungkinkan pengumpulan data secara otomatis dan *real* time, mengatasi keterbatasan dataset publik yang sering kali tidak mencakup konteks lokal Indonesia [6]. Proses web scraping dilakukan menggunakan library Python seperti BeautifulSoup dan Selenium. Dataset disimpan dalam format CSV untuk kemudahan pemrosesan selanjutnya. Pendekatan ini mirip dengan studi [23] di mana data teks dikumpulkan dari sumber online untuk klasifikasi SVM, memastikan dataset asli dan relevan dengan konteks Indonesia.

2.2 Pelabelan Data

Proses pelabelan data menggunakan cara otomatis dengan memanfaatkan asal sumber berita. Jika sumber berita yang diperoleh dari website seperti CNN, Detik dan Kompas data akan otomatis diberikan label *real*. Kemudian, jika sumber data berasal dari website seperti Kompas cek fakta, Liputan 6 cek fakta dan Turnbackhoax data akan otomatis diberikan label *fake*. Metode berikut hampir mirip dengan penelitian [24].

2.3 *PreProcessing*

Tahap preprocessing terdiri beberapa langkah seperti *cleaning* dan *case folding* untuk menghilangkan kata-kata yang tidak dibutuhkan serta mengubah kapitalisasi huruf, dilanjutkan dengan *tokenization* untuk memisahkan teks menjadi kata-kata tersendiri. Selanjutnya dilakukan penghapusan *stopword* dengan bantuan library NLTK untuk membuang kata-kata umum seperti "dan" dan "yang", serta *stemming* menggunakan *library sastrawi* guna mengembalikan kata-kata ke bentuk dasar. Seluruh tahapan ini dilakukan agar representasi kata menjadi lebih baik, mengurangi noise dalam data, dan meningkatkan tingkat akurasi klasifikasi [2].

2.4 *Spliting* Data

Tahapan berikut merupakan tahapan pemisahan data menjadi 2 kategori (data latih dan data uji). Perbandingan menerapkan rasio 80:20 dikarenakan rasio 80:20 sudah dianggap ideal dan paling sering dipakai dalam berbagai penelitian serupa. Rasio 80:20 dipilih karena seimbang antara ketersediaan data latih yang cukup guna pembelajaran model dan data uji yang representatif guna pengujian model, sebagaimana diterapkan oleh penelitian [25] dalam deteksi *fake news* menggunakan SVM dan algoritma lain. Pendekatan ini juga membantu mencegah *overfitting*, terutama pada dataset teks yang cenderung tidak seimbang.

2.5 Ektraksi Fitur TF-IDF

Proses ekstraksi fitur dilakukan melalui pembobotan kata dan mengukur tingkat kepentingan kata didalam berita menggunakan teknik *Term Frequency Inverse Document Frequency* (TF-IDF). Dalam penerapannya, jumlah maksimum fitur ditetapkan sebanyak 5000 dengan penggunaan n-gram (1,2), karena kombinasi tersebut memberikan keseimbangan yang baik untuk dataset berukuran sedang. Pemilihan TF-IDF dilakukan karena metode ini terbukti efektif dalam mengolah data teks berbahasa Indonesia, sebagaimana dibuktikan oleh [12] dan [19] dalam deteksi *fake news* multilingual. Berikut merupakan rumus TF dan IDF:

$$TF : \frac{Jumlah\ kemunculan\ data}{Total\ kata\ dalam\ dokumen} \quad (1)$$

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$IDF : \log\left(\frac{Jumlah\ total\ dokumen}{Jumlah\ dokumen\ yang\ mengandung\ kata}\right) \quad (2)$$

2.6 Pemodelan Algoritma

Penelitian ini memilih empat algoritma, yaitu SVM, *Logistic Regression*, *Naive Bayes*, dan XGBoost, karena algoritma-algoritma tersebut relevan digunakan untuk teks mining. Empat algoritma klasifikasi diterapkan menggunakan *library Scikit-learn* di Python. SVM dikonfigurasi dengan *kernel linear*, *C = 1.0, max_iter* = 2000, dan *gamma = 'scale'* untuk menangani data teks secara efisien dan menghindari konvergensi prematur [20] [26] [27]. *Logistic Regression* dioptimalkan dengan regularisasi L2 untuk mencegah *overfitting* [12]. *Naive Bayes* menggunakan varian *Multinomial* untuk data teks [9]. XGBoost dikonfigurasi dengan *eval_metric='logloss'* untuk mengukur performa pada tugas *klasifikasi biner*, memastikan optimasi *loss function* yang tepat [13]. Optimatilasi dilakukan via *cross validation* dengan 10-fold.

2.7 Evaluasi Model

Untuk mengevaluasi model, tahapan ini menggunakan akurasi, presisi, recall, dan f1-score [25]. Model terbaik dipilih berdasarkan hasil akurasi tertinggi, mengingat data yang cukup seimbang. Berikut merupakan rumus-rumus yang digunakan untuk mengevaluasi model.

$$Accuracy = \frac{Jumlah\ prediksi\ benar}{Jumlah\ total\ data} \quad (3)$$

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (4)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (5)$$

$$F1 - Score = 2 * \frac{Presisi * Recall}{Presisi + Recall} \quad (6)$$

2.8 *Confusion Matrix*

Penggunaan *confusion matrix* pada penelitian yang saya kembangkan bertujuan untuk memvisualisasikan kemampuan model dalam mengklasifikasi berita ke dalam dua kategori, yaitu berita hoaks (*fake*) dan berita asli (*real*). Matriks ini berfungsi untuk memvisualisasikan jumlah data dengan prediksi benar dan salah di setiap label, sehingga memudahkan identifikasi pola kesalahan klasifikasi seperti *false positive* maupun *false negative* [12][25][13]. Didalam *confusion matrix* terdapat beberapa istilah seperti:

True Positive (TP), merupakan berita asli *(real)* yang berhasil diprediksi sebagai berita asli.

True Negative (TN), merupakan berita hoaks *(fake)* yang berhasil diprediksi sebagai berita hoax.

False Positive (FP), merupakan berita asli *(real)* yang salah diprediksi sebagai berita hoax.

False Negative (FN), merupakan berita hoaks *(fake)* yang salah diprediksi sebagai berita asli.

2.9 *Cross Validation*

Penelitian berikut menerapkan10 fold (k = 10) *cross validation* untuk mengoptimalisasi performa keempat algoritma secara adil dan kuat. Teknik ini dapat memberikan estimasi performa model yang lebih stabil dan dapat diandalkan dibandingkan hanya menggunakan satu kali pembagian data train-test tunggal, terutama pada dataset deteksi *fake news* yang seringkali memiliki distribusi kelas yang tidak seimbang dan variasi bahasa yang tinggi [12][13]. Metode ini diterapkan pada empat algoritma yang digunakan, yaitu SVM, *Logistic Regression*, *Naive Bayes* dan XGBoost. Setiap algoritma menghasilkan skor evaluasi pada masing-masing *fold*, kemudian nilai-nilai tersebut diambil nilai rata-rata untuk mendapatkan *mean score* yang menjadi hasil akhir. Pendekatan ini memastikan evaluasi model lebih stabil, tidak bergantung pada satu pembagian data, serta memungkinkan perbandingan performa antar algoritma secara lebih objektif dalam tugas deteksi *fake news*.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penggunaan dataset dalam penelitian yang saya kembangkan berjumlah 4909 berita dan terdapat dua kategori label, yaitu *real* (asli) dan *fake* (hoaks), dataset diambil menggunakan teknik *web scraping* yang mengambil data dari beberapa portal website terpercaya. Data yang diambil mempunyai beberapa variabel seperti Judul, Link, Isi berita dan sumber berita tersebut. Pendekatan ini juga mendukung literasi digital dengan fokus pada sumber kredibel, seperti yang direkomendasikan [8] untuk menanggulangi hoaks di media sosial. Tabel 1 menunjukkan contoh sampel berita dari 6 sumber data yang berbeda pada dataset penelitian dengan beberapa atribut meliputi Judul, Link, Isi Berita, dan Sumber Berita.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 1. Contoh Dataset

Judul	Link	Isi Berita	Sumber
Olah TKP, Polisi Temukan 6 Barang Bukti Kasus Pembunuhan Pegawai Minimarket	https://bandung.kompas.com/read/2025/10/10/134004778/olah-tkp-polisi-temukan-6-barang-bukti-kasus-pembunuhan-pegawai-minimarket	KARAWANG, KOMPAS.com- Kepolisian Resor (Polres) Purwakarta telah menemukan enam barang bukti dalam kasus pembunuhan dan pemerkosaan DO (21), pegawai minimarket, yang jasadnya ditemukan mengambang di Sungai Citarum, Karawang, Jawa Barat	Kompas
Anggota DPR Gerindra Minta Prabowo Evaluasi Penyegelan Wisata PuncakNasional â€¢ 24 menit yang lalu	https://www.cnnindonesia.com/nasional/20251010194606-32-1283327/anggota-dpr-gerindra-minta-prabowo-evaluasi-penyegelan-wisata-puncak	Anggota DPR dari Fraksi Gerindra MulyadiÂ meminta Presiden PrabowoÂ Subianto untuk mengevaluasi kebijakan penyegelan tempat wisata di kawasan Puncak, Bogor, oleh Kementerian Lingkungan Hidup (LH	CNN Indonesia
Penumpang Tak Terangkut, Trans Banten Akan Ganti Armada Sebesar Transjakarta	https://news.detik.com/berita/d-8155346/penumpang-tak-terangkut-trans-banten-akan-ganti-armada-sebesar-transjakarta	Dinas Perhubungan (Dishub) Provinsi Banten mengatakanTrans Bantendiminati hingga ada penumpang yang tak terangkut. Karena itu, mereka akan mengganti model bus menjadi mirip yang dipakaiTransjakarta.	Detik News
[HOAKS] Menteri Yusril Jenguk Ammar Zoni di Nusakambangan	https://www.kompas.com/cekfakta/read/2025/10/24/185800882/-hoaks-menteri-yusril-jenguk-ammar-zoni-di-nusakambangan	KOMPAS.com- Menteri Koordinator Bidang Hukum, HAM, Imigrasi, dan Pemasyarakatan, Yusril Ihza Mahendra diklaim menjenguk artis, Ammar Zoni yang kini ditahan di Lapas Nusakambangan karena kasus peredaran narkoba di Rutan Salemba.	Kompas Cek Fakta
Cek Fakta: Hoaks Dedi Mulyadi Bagikan Rp 50 Juta untuk Warga dari Sabang-Merauke	https://www.liputan6.com/cek-fakta/read/6193961/cek-fakta-hoaks-dedi-mulyadi-bagikan-rp-50-juta-untuk-warga-dari-sabang-merauke	Liputan6.com, Jakarta -Beredar postingan di media sosial klaim Gubernur Jawa BaratDedi Mulyadimembagikan hadiah sebesar Rp 50 juta kepada masyarakat dari Sabang sampai Merauke. Informasi tersebut diunggah salah satu akun Facebook pada 22 Oktober 2025	Liputan6 Cek Fakta
[SALAH] Elon Musk Luncurkan Tesla Pi Phone, Gratis Internet Seumur Hidup	https://turnbackhoax.id/2025/10/10/salah-elon-musk-luncurkan-tesla-pi-phone-gratis-internet-seumur-hidup/	Akun Facebook “Warta Ekonomi” pada Minggu, (05/10/25) mengunggahfoto[arsip] yang menginformasikan bahwa Tesla resmi meluncurkan Pi Phone seharga USD 789 (sekitar Rp12 juta) yang diklaim mengubah peta persaingan smartphone dunia.	TurnBackHoax

3.2 Pelabelan

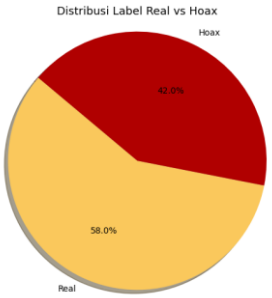
Pemberian label pada data yang dilakukan secara otomatis dengan memanfaatkan asal sumber pengambilan data. Berita yang bersumber dari portal website CNN, Detik dan Kompas akan diberikan label *real* dan data yang diambil dari portal website Kompas cek fakta, Liputan 6 cek fakta dan Turnbackhoax akan diberikan label *fake*. Seperti yang dijelaskan pada Tabel 2 dibawah.

Tabel 2. Pelabelan Data Berdasarkan Sumber

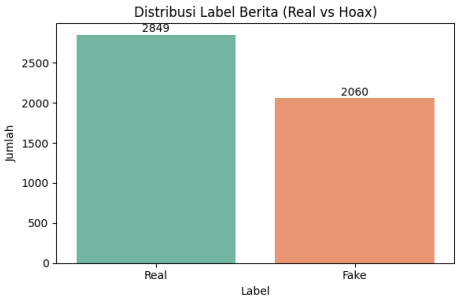
Sumber	Label
CNN Indonesia	*Real*
Kompas	*Real*
Detik News	*Real*
Liputan6 Cek Fakta	*Fake*
Kompas Cek Fakta	*Fake*
Turnback Hoax	*Fake*

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Distribusi kedua label data (*real* dan *fake*) cukup seimbang dengan jumlah data *real* 2849 dan data *fake* 2060 data seperti yang dijelaskan pada Gambar 3. Kemudian, persentase dari kedua label berjumlah 58.0% data *real* dan 42.0% data *fake,* seperti yang di gambarkan pada Gambar 2.

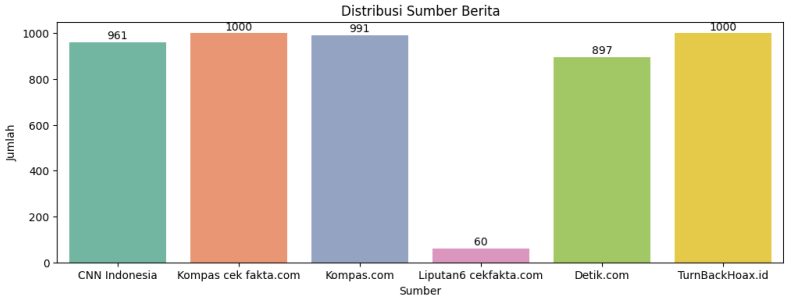


Gambar 2. Persentase Label



Gambar 3. Jumlah Label *Real* dan *Fake*

Data bersumber dari 6 platform terverifikasi meliputi CNN Indonesia, Kompas, Detik, Kompas Cek Fakta, Liputan6 Cek Fakta, dan Turnbackhoax. Distribusi data yang diambil dari sumber-sumber tersebut berjumlah 961 data berita dari CNN Indonesia, 1000 data berita dari Kompas Cek Fakta, 991 data berita dari Kompas.com, 60 data berita dari Liputan6 Cek Fakta, 897 data berita dari Detik.com, 1000 data berita dari Turnbackhoax.id. seperti yang dijelaskan pada Gambar 4.



Gambar 4. Jumlah Data dari Setiap Sumber

3.3 *Pre Processing*

3.3.1 *Cleaning dan Case folding*

Dataset mentah pertama-tama akan melewati proses *cleaning* dan *case folding*. Tahapan ini merupakan tahapan awal pada pre processing data, tahapan ini merupakan normalisasi kolom seperti menghapus kata-kata yang ikut ter scaping secara tidak sengaja yang dapat membocorkan data (advertisement, scroll, continue dan lain-lain), menghapus angka dan simbol, menghapus data duplikat, mengubah kapitalisasi data serta menghapus kolom yang kosong pada kolom Isi Berita.

3.3.2 *Tokenization*

Setelah melewati *cleaning* dan *case folding* tahap selanjutnya merupakan *tokenization*. Proses tokenisasi dalam penelitian deteksi berita hoax ini bertujuan mengubah teks mentah yang masih tidak beraturan menjadi bentuk yang terdiri atas kata-kata terpisah, sehingga lebih mudah dipahami dan diproses oleh algoritma machine learning. Dengan melakukan tokenisasi, setiap kata dapat dikenali sebagai satuan analisis yang terpisah, sehingga langkah-langkah berikutnya seperti *stopword removal*, *stemming*, dan *vectorization* dapat dilakukan dengan lebih efektif.

3.3.3 *Stopword*

Kemudian, penerapan tahapan *stopword* pada preprocessing berfungsi untuk menghapus kata yang umum dan tidak bermakna seperti di, yang, ke, dan lain-lain menggunakan *library* python *sastrawi*.

3.3.4 *Stemming*

Tahapan ini menggunakan *library* python *sastrawi* yang berfungsi mengembalikan kata ke bentuk dasar seperti “berita-berita” menjadi “berita”. *Stemming* merupakan tahapan terakhir dalam preprocessing, dengan demikian *stemming* juga sangat berpengaruh terhadap representasi data. Tabel 3 berikut menunjukkan contoh penerapan hasil *preprocessing* data berita pada satu sampel berita yang diambil dari dataset penelitian dengan 4 tahapan yang meliputi *Cleaning* dan *Case Folding*, *Tokenization*, *Stopword*, dan *Stemming*.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 3. Tahapan Preprocessing Data

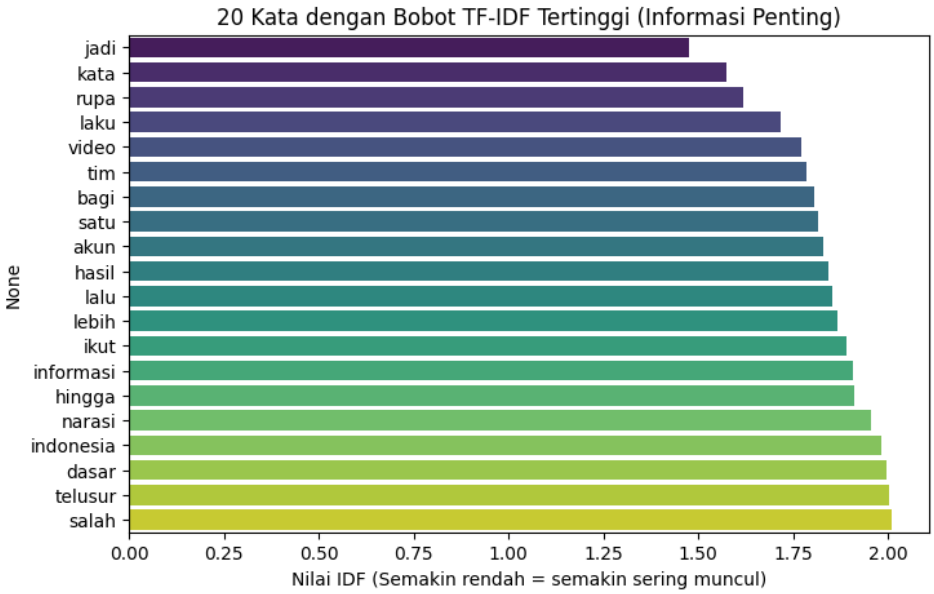
Isi Berita	*Cleaning* dan *Case folding*	*Tokenization*	*Stopword*	*Stemming*
Timnas Indonesia U-23 tertinggal 1-2 dari India dalam babak pertama uji coba internasional di Stadion Madya, Senayan, Jakarta, Jumat (10/10).	timnas indonesia u tertinggal dari india dalam babak pertama uji coba internasional di stadion madya senayan jakarta jumat	['timnas', 'indonesia', 'u', 'tertinggal', 'dari', 'india', 'dalam', 'babak', 'pertama', 'uji', 'coba', 'internasional', 'di', 'stadion', 'madya', 'senayan', 'jakarta', 'jumat', …]	['timnas', 'indonesia', 'u', 'tertinggal', 'india', 'babak', 'pertama', 'uji', 'coba', 'internasional', 'stadion','senayan', 'jakarta', 'jumat', ...]	['timnas', 'indonesia', 'u', 'tinggal', 'india', 'babak', 'pertama', 'uji', 'coba', 'internasional', 'stadion', 'madya', 'senayan', 'jakarta', 'jumat', ...]

3.4 *Spliting* Data

Tahapan ini merupakan proses pemisahan dataset dengan 2 kategori (data latih dan data uji). Kedua data ini mempunyai fungsi tersendiri, perbandingan kedua data tersebut juga berpengaruh terhadap penelitian. Perbandingan *spliting* data pada penelitian ini berjumlah 80:20. Pada data latih jumlah label dengan label *real* yaitu 2279 dan label *fake* 1648 data untuk melatih pembelajaran mesin. Sementara itu, tahap pengujian dilakukan menggunakan data uji yang terdiri atas 570 sampel berkategori *real* dan 412 sampel berkategori *fake*. Data tersebut tergolong masih seimbang oleh sebab itu, tahapan penyeimbang tidak digunakan dalam penelitian berikut.

3.5 Ekstraksi Fitur TF-IDF

Tahapan berikut merupakan tahapan setelah *spliting*, penelitian ini mendahulukan *spliting* dari pada ekstraksi fitur, ekstraksi fitur hanya dilakukan pada data latih dikarenakan untuk mencegah adanya data *leakage* didalam evaluasi model nantinya. Beberapa penelitian pada umumnya melakukan ekstraksi fitur sebelum *spliting* data. Mengenai hal tersebut, kemungkinan besar akan menyebabkan kebocoran data dikarenakan data test nantinya sudah terkontaminasi sebelum *spliting* data. Pada tahap ekstraksi fitur, metode TF-IDF diterapkan dengan batas maksimum fitur sebanyak 5000 serta penggunaan n-gram (1,2) pada data latih dan data uji. Dimensi matrik yang dihasilkan data latih adalah (3927, 5000) dan (982, 5000) untuk data uji. Gambar 5 dibawah ini menunjukkan hasil dari penerapan ekstraksi fitur menggunakan TF-IDF yang menghasilkan visualisasi 20 kata dengan bobot TF-IDF tertinggi.



Gambar 5. 20 Kata Dengan Bobot Tertinggi

3.6 Pemodelan Algoritma

Model yang digunakan dalam perbandingan penelitian deteksi *fake news* ini yaitu SVM, *Naive Bayes*, *Logistic Regression*, dan XGBoost. Pengambilan model tersebut dikarenakan model-model tersebut relevan digunakan untuk penelitian berbasis text mining. Model-model ini nantinya akan di *comparasion* berdasarkan hasil akurasi dari evaluasi model. Setiap model akan dikonfigurasi dengan peforma terbaik yang disesuaikan dengan jumlah dataset yang digunakan.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3.7 Evaluasi Model

Tabel 4 menunjukkan hasil evaluasi peforma model SVM dalam mendeteksi *fake news* dengan menggunakan 4 metrik evaluasi standar.

Tabel 4. Evaluasi Model SVM

	Precision	*Recall*	F1-score	Support
Real	0.9948	1.0000	0.9974	570
Fake	1.0000	0.9927	0.9963	412
Accuracy			0.9969	982
Macro Avg	0.9974	0.9964	0.9969	982
Whited Avg	0.9970	0.9969	0.9969	982

Algoritma SVM dengan konfigurasi *kernel linear*, C= 1.0, *max_iter*= 2000, dan *gamma = scale* menunjukkan performa paling unggul di antara keempat model. SVM mencapai akurasi 99.69% pada 982 data uji, dengan presisi sempurna 100% pada kelas *Fake* dan *recall* sempurna 100% pada kelas *Real*. Hal ini menandakan bahwa SVM sangat mampu mengenali semua berita asli tanpa kesalahan. Model ini hampir tidak pernah salah mengklasifikasikan berita palsu sebagai berita asli. Kesalahan yang terjadi sangat sedikit, hanya 3 kasus dari total 982 sampel.

Tabel 5 menunjukkan evaluasi peforma algoritma Logistic Regression dalam mendeteksi fake news dengan metrik evaluasi standar.

Tabel 5. Evaluasi Model *Logistic Regression*

	Precision	*Recall*	F1-score	Support
Real	0.9930	0.9965	0.9947	570
Fake	0.9951	0.9903	0.9927	412
Accuracy			0.9939	982
Macro Avg	0.9941	0.9934	0.9937	982
Whited Avg	0.9939	0.9939	0.9939	982

Algoritma *Logistic Regression* dengan konfigurasi *solver* 'lbfgs' dan *regularisasi* L2 memperoleh akurasi 99.39% dengan performa yang sangat seimbang pada kedua kelas. Model ini hanya salah mengklasifikasikan 6 sampel dan menunjukkan peforma yang sangat baik dalam memisahkan antara berita asli dan palsu.

Tabel 6 menunjukkan evaluasi peforma algoritma Naive Bayes dalam mendeteksi fake news dengan metrik evaluasi standar.

Tabel 6. Evaluasi Model *Naive Bayes*

	Precision	*Recall*	F1-score	Support
Real	0.9930	0.9947	0.9939	570
Fake	0.9927	0.9903	0.9915	412
Accuracy			0.9929	982
Macro Avg	0.9928	0.9925	0.9927	982
Whited Avg	0.9929	0.9929	0.9929	982

Algortima *Naive Bayes* dengan konfigurasi varian *Multinomial* menghasilkan akurasi 99.29% dengan distribusi kesalahan yang merata. Meskipun merupakan algoritma paling sederhana dan tercepat dalam pelatihan, *Naive Bayes* mampu bersaing dengan model-model yang lebih kompleks dengan kualitas representasi TF-IDF dan preprocessing yang optimal.

Tabel 7 menunjukkan evaluasi peforma algoritma XGBoost dalam mendeteksi fake news dengan metrik evaluasi standar

Tabel 7. Evaluasi Model XGBoost

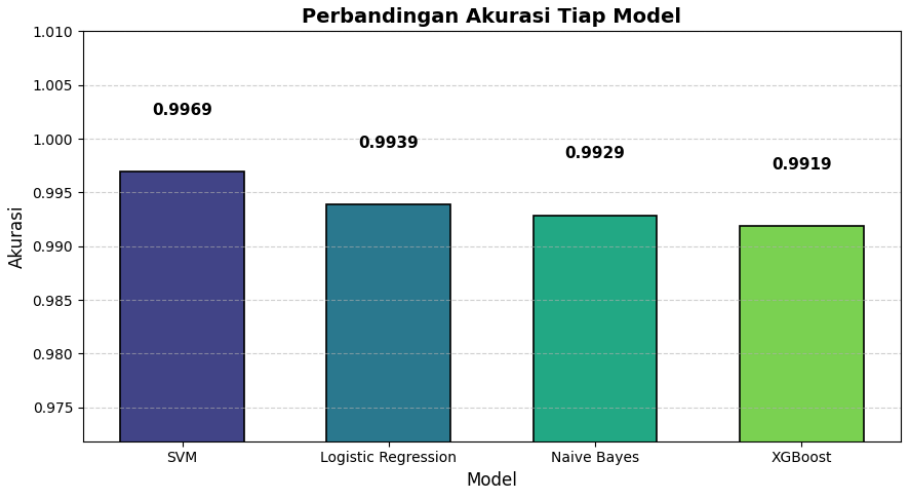
	Precision	*Recall*	F1-score	Support
Real	0.9878	0.9982	0.9930	570
Fake	0.9975	0.9830	0.9902	412
Accuracy			0.9919	982
Macro Avg	0.9927	0.9906	0.9916	982
Whited Avg	0.9919	0.9919	0.9918	982

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Dalam evaluasi model XGBoost yang dikenal sangat kuat pada dataset tidak seimbang, pada penelitian ini dengan konfigurasi *parameter n_estimators* 200, *learning_rate* 0.1, dan *max_depth* 6 menempati posisi terendah dengan akurasi 99.19%. Hal ini menunjukkan bahwa pada dataset berbahasa Indonesia yang telah dibersihkan dan di-balance dengan baik, algoritma klasik seperti SVM dan *Logistic Regression* justru lebih optimal dan stabil dibandingkan *ensemble tree-based* yang cenderung *overfit* pada pola-pola kecil.

Secara keseluruhan, keempat algoritma memperoleh akurasi di atas 99.19%, dengan SVM menjadi yang terbaik (99.69%). Tingginya nilai presisi, *recall*, dan F1-score di semua model membuktikan bahwa pendekatan berbasis TF-IDF yang dikombinasikan dengan preprocessing bahasa Indonesia yang tepat (*stemming sastrawi* dan *stopword* lokal) mampu menghasilkan representasi teks yang sangat diskriminatif. Hanya terdapat total kesalahan klasifikasi sebanyak 3 (SVM), 6 (LR), 7 (NB), dan 8 (XGBoost) dari 982 data uji, menegaskan bahwa model-model ini sudah sangat siap untuk diimplementasikan dalam sistem deteksi *fake news real time* di Indonesia.

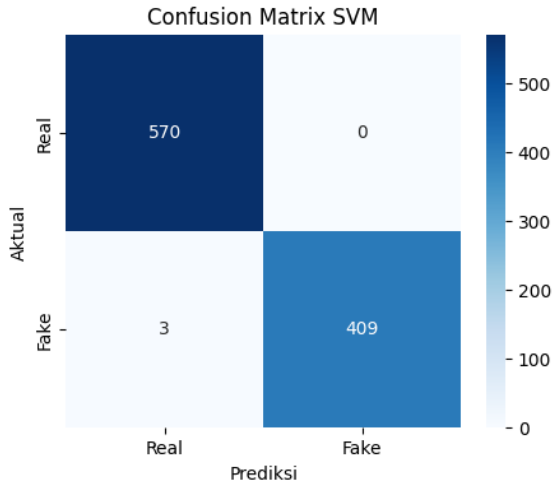
Gambar 6 dibawah menunjukkan hasil perbandingan akurasi yang dihasilkan pada keempat algoritma dalam mendeteksi *fake news* dalam bentuk bar plot.



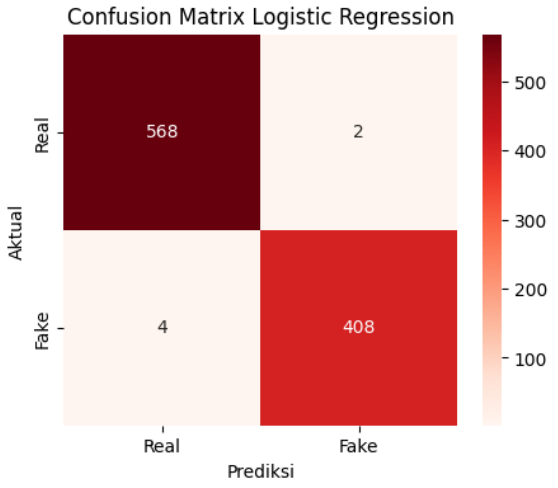
Gambar 6. Perbandingan Akurasi Setiap Model

3.8 Confusion matrix

Proses *confusion matrix* pada penelitian ini bertujuan untuk memvisualisasikan evaluasi kesalahan yang dilakukan oleh ke-empat algortima. Pada Gambar 7, Gambar 8, Gambar 9, dan Gambar 10 dibawah ini merupakan hasil evaluasi dari keempat algortima yang meliputi SVM, Logistic Regression, Naive Bayes, dan XGBoost.

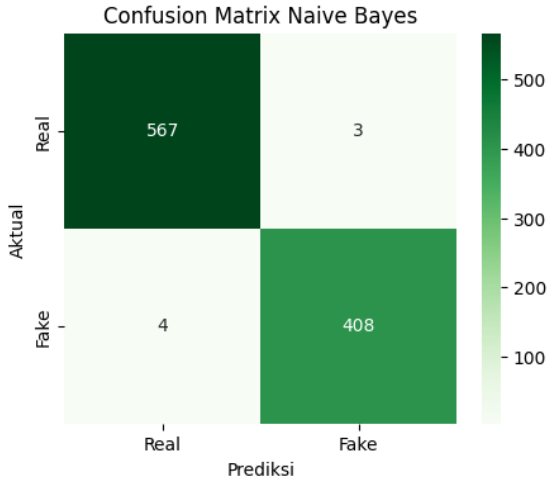


Gambar 7. *Confusion Matrix* SVM

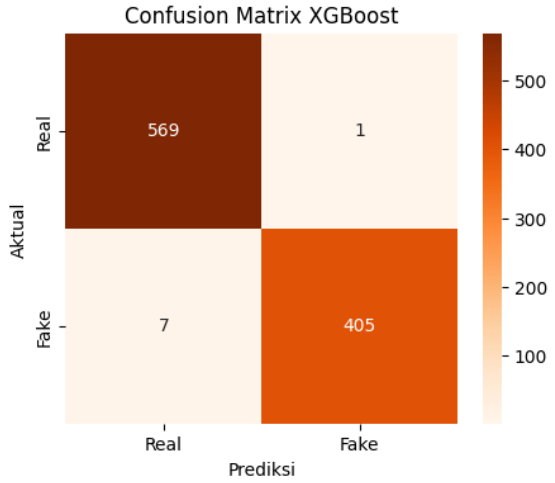


Gambar 8. *Confusion Matrix Logistic Regression*

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9492
Hal 462-473
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 9. *Confusion Matrix Naive Bayes*



Gambar 10. *Confusion Matrix* XGBoost

Dari hasil visualisasi diatas dapat disimpulkan bahwasannya *false negative* (FP) pada algoritma SVM pada gambar 7 berjumlah 3, FP tersebut merupakan data dengan kategori aktualnya *fake news* diprediksi sebagai *real* news sedangkan *false negative* (FN) merupakan data *fake news* yang diprediksi sebagai *real news*. Dilain sisi, algoritma *Logistic Regression* mempunyai *confusion matrix* dengan FN berjumlah 2 dan FP berjumlah 4 pada gambar 8. Kemudian, pada gambar 9 algoritma naive bayes mempunyai *confusion matrix* dengan FN berjumlah 3 dan FP yang sama dengan logistic berjumlah 4 dan. Lalu, dapat dilihat pada gambar 10 algoritma XGBoost mempunyai *confusion matrix* dengan FN 1 data dan FP tertinggi berjumlah 7 data. Analisis *confusion matrix* memungkinkan penilaian yang lebih mendalam dibandingkan hanya melihat nilai akurasi, karena dapat mengidentifikasi kecenderungan model dalam melakukan kesalahan tertentu, misalnya terlalu sering memprediksi berita sebagai *fake* atau *real*. Dengan demikian, *confusion matrix* membantu memahami kekuatan dan kelemahan model serta memberikan wawasan penting untuk perbaikan performa dalam mendeteksi *fake news*.

3.9 Cross Validation

Hasil *cross validation* semakin menguatkan performa luar biasa keempat model. SVM tetap unggul secara jelas dengan rata-rata akurasi *cross-validation* 99.57% dan variansi yang sangat rendah (std 0.20%), menunjukkan bahwa model ini sangat stabil dan hampir tidak terpengaruh oleh perbedaan komposisi fold. Selisih antara akurasi uji (99.69%) dan CV Mean (99.57%) hanya 0.12%, membuktikan tidak adanya overfitting serta kemampuan generalisasi yang sangat tinggi. *Logistic Regression*, Naïve Bayes, dan XGBoost juga menunjukkan konsistensi yang sangat baik dengan CV Mean masing-masing 99.52%, 99.44%, dan 99.49%, serta standar deviasi yang masih berada di bawah 0.25%. Ketidakstabilan performa antar fold sangat kecil, sehingga keempat model terbukti kuat dan dapat diandalkan pada data berita berbahasa Indonesia yang asli. Dengan nilai CV Std yang rendah dan CV Mean yang tetap tinggi di atas 99.44%, hasil ini memperkuat kesimpulan bahwa SVM merupakan model paling optimal dan paling stabil untuk tugas deteksi *fake news* dalam penelitian ini. Keunggulan SVM tidak hanya terlihat pada pengujian tunggal, tetapi juga terkonfirmasi secara kuat melalui evaluasi *cross-validation* yang lebih ketat dan representatif. Berdasarkan hasil yang diperoleh, model SVM (*kernel linear)* ditetapkan sebagai model terbaik yang direkomendasikan untuk diimplementasikan pada sistem deteksi *fake news* berbasis portal berita Indonesia. Tabel 8 dibawah ini menjelaskan penerapan *mean cross validation* dengan 10 fold yang dibandingan dengan akurasi yang dihasilkan algortima sebelum melalui proses *cross validation*.

Tabel 8. Perbandingan Akurasi dengan Mean *Cross validation*

Model	Akurasi	CV Mean	CV std
SVM	99.69%	99.57%	00.20
Logistic Regression	99.39%	99.52%	00.24
Naïve Bayes	99.29%	99.44%	00.22
XGBoost	99.19%	99.49%	00.20

4. KESIMPULAN

Penerapan dan perbandingan peforma empat algoritma (SVM, *Logistic Regression*, *Naive Bayes* dan XGBoost) untuk deteksi fake news pada dataset berbahasa Indonesia berhasil dilakukan. Dataset tersebut dikumpulkan langsung melalui web scraping dari platform terpercaya seperti CNN Indonesia, Detik News, Kompas, Kompas Cek Fakta, Liputan6 Cek Fakta, dan Turnbackhoax.id. Dataset berjumlah 4.909 entri ini memiliki distribusi seimbang, dengan kebaruan utama pada sumber data autentik dan segar tanpa bergantung pada dataset publik yang usang. Hasil evaluasi menunjukkan akurasi tinggi SVM 99.69%, *Logistic Regression* 99.39%, *Naive Bayes* 99.29%, dan XGBoost 99.19%. SVM dengan *kernel linear*, C=1.0, *max_iter* = 2000, *gamma* = *scale* menjadi model terbaik dengan presisi 100% pada kelas *fake*, *recall* 100% pada kelas *real*, serta hanya 3 kesalahan dari 982 sampel uji. *Cross-validation* (10-fold) mengonfirmasi stabilitas dengan *mean score* SVM 99.57% (std 0.20%), *Logistic Regression* 99.52% (std 0.24%), *Naive Bayes* 99.44% (std 0.22%), dan XGBoost 99.49% (std 0.20%). Dibandingkan penelitian terdahulu, akurasi ini jauh lebih unggul, Ropikoh et al hanya 97,06% dengan SVM pada hoaks Covid-19, Putri & Athoillah sekitar 78,96-79%, Hamdikatama 82% meski menggunakan stacking dan IndoBERT, serta Wahid et al. (2024) 95,20% dengan *Logistic Regression* dan Random Forest. Peningkatan hingga di atas 99% disebabkan kualitas dataset yang lebih segar serta tuning hiperparameter optimal, sehingga berhasil menutup gap akurasi pada kajian sebelumnya.

REFERENCES

[1] F. Olan, U. Jayawickrama, E. O. Arakpogun, J. Suklan, and S. Liu, "Fake news on Social Media: the Impact on Society," *Inf. Syst. Front.*, vol. 26, no. 2, pp. 443–458, 2024, doi: 10.1007/s10796-022-10242-z.
[2] O. N. Cahyani and F. Budiman, "Performa *Logistic Regression* dan *Naive Bayes* dalam Klasifikasi Berita Hoax di Indonesia," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 60–68, 2025, doi: 10.29408/edumatic.v9i1.28987.
[3] Admidppid, "Rekap Isu Hoaks & Disinformasi Tahun 2025," Ppid.Diskominfo.Jatengprov.Go.Id. Accessed: Nov. 18, 2025. [Online]. Available: https://ppid.diskominfo.jatengprov.go.id/rekap-isu-hoaks-disinformasi-tahun-2025/
[4] A. Sarjito, "Hoaks, Disinformasi, dan Ketahanan Nasional: Ancaman Teknologi Informasi dalam Masyarakat Digital Indonesia," *J. Gov. Local Polit.*, vol. 5, no. 2, pp. 175–186, 2021.
[5] H. S. Pratama, "Menghadapi Berita Palsu," p. 24, 2019.
[6] C. Juditha and J. J. Darmawan, "Komunikasi Politik Terkait Hoaks Pada Pemilu Presiden Indonesia 2024," *J. Stud. Komun. dan Media Komdigi*, vol. 28, p. 178, 2024, doi: 10.17933/jskm.2024.5682.
[7] A. Anshari, "Penyuluhan Hukum Dampak Beredarnya Berita Hoax dan Pencegahannya di Situasi Covid-19 Melalui Sosial Media," *J. Bul. Al-Ribaath*, vol. 20, no. 1, p. 30, 2023, doi: 10.29406/br.v20i1.5802.
[8] N. Amaly and A. Armiah, "Peran Kompetensi Literasi Digital Terhadap Konten Hoaks dalam Media Sosial [The Role of Digital Literacy Competence in Hoax Content on Social Media]," *Alhadharah J. Ilmu Dakwah*, vol. 20, no. 2, pp. 43–52, 2021.
[9] N. Chamidah and R. Sahawaly, "Comparison Support Vector Machine and *Naive Bayes* Methods for Classifying Cyberbullying in Twitter," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 7, no. 2, p. 338, 2021, doi: 10.26555/jiteki.v7i2.21175.
[10] N. E. Febriyanty, M. A. Hariyadi, and C. Crysdian, "Hoax Detection News Using Naïve Bayes and Support Vector Machine Algorithm," *Int. J. Adv. Data Inf. Syst.*, vol. 4, no. 2, pp. 191–200, 2023, doi: 10.25008/ijadis.v4i2.1306.
[11] M. Jauhar Vikri, I. Wisma Dwi Prastya, U. Pradema Sanjaya, and M. Agung Barata, "Rice Quality Identification Built on Indonesian Food Standards Based on Electronic Nose using Naïve Bayes Algorithm," *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 1, pp. 49–60, 2025, doi: 10.35314/0y0xct32.
[12] A. M. Wahid, Turino, K. A. Nugroho, D. Titi Safitri4, and F. S. Utomo, "Optimasi *Logistic Regression* dan Random Forest untuk Deteksi Berita Hoax Optimasi *Logistic Regression* dan Random Forest untuk Deteksi Berita Hoax Berbasis Hyperparameter Optimization of *Logistic Regression* and Random Forest for Hoax News Detection Using T," *J. Pendidik. dan Teknol. Indones.*, no. January, 2025, doi: 10.52436/1.jpti.602.
[13] B. Hamdikatama, "Beyond Algorithms: an Integrated Approach To Fake News Detection Using Machine Learning Techniques," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 3, pp. 609–622, 2025, doi: 10.33480/jitk.v10i3.6061.
[14] F. Hukum, C. B. Devina, D. C. Iswari, G. Christian, B. Goni, and D. Kimberly, "Kosmik Hukum," vol. 21, no. 1, pp. 44–58, 2021.
[15] Febriansyah Putra and H. Patra, "Analisis Hoax pada Pemilu: Tinjauan dari Perspektif Pendidikan Politik," *Naradidik J. Educ. Pedagog.*, vol. 2, no. 1, pp. 95–102, 2023, doi: 10.24036/nara.v2i1.119.
[16] Dinda Marta Almas Zakirah, "Pengaruh Hoax Di Media Sosial Terhadap Preferensi Sosial Politik Remaja Di Surabaya," *Mediakita*, vol. 4, no. 1, pp. 37–36, 2020, doi: 10.30762/mediakita.v4i1.2446.
[17] I. A. Ropikoh, R. Abdulhakim, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 64–73, 2021, doi: 10.30871/jaic.v5i1.3167.
[18] R. K. Putri and M. Athoillah, "Support Vector Machine Untuk Identifikasi Berita Hoax Terkait Virus Corona (Covid-19)," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 162–167, 2021, doi: 10.30591/jpit.v6i3.2489.
[19] S. Murugesan and K. P. Kaliyamurthie, "A Machine Learning Framework for Automatic Fake News Detection in Indian Tamil News Channels," *Ing. des Syst. d'Information*, vol. 28, no. 1, pp. 205–209, 2023, doi: 10.18280/isi.280123.
[20] Y. Sibaroni and S. S. Prasetiyowati, "Buzzer Detection on Indonesian Twitter using SVM and Account Property Feature Extension," *J. RESTI*, vol. 6, no. 4, pp. 663–669, 2022, doi: 10.29207/resti.v6i4.4338.
[21] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using *Naive Bayes* optimization," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 2, pp. 799–806, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14744.
[22] N. W. S. Saraswati, I. P. K. S. Putra, I. D. M. K. Muku, and G. D. Pramitha, "Support Vector Machine For Hoax Detection," *SINTECH (Science Inf. Technol. J.*, vol. 6, no. 2, pp. 107–117, Aug. 2023, doi: 10.31598/sintechjournal.v6i2.1366.
[23] N. H. Ovirianti, M. Zarlis, and H. Mawengkang, "Support Vector Machine Using A Classification Algorithm," *SinkrOn*, vol. 7,

no. 3, pp. 2103–2107, 2022, doi: 10.33395/sinkron.v7i3.11597.
[24] D. K. Sharma and S. Garg, "IFND: a benchmark dataset for fake news detection," *Complex Intell. Syst.*, vol. 9, no. 3, pp. 2843–2863, 2023, doi: 10.1007/s40747-021-00552-1.
[25] M. Sudhakar and K. P. Kaliyamurthie, "Detection of fake news from social media using support vector machine learning algorithms," *Meas. Sensors*, vol. 32, no. January, p. 101028, 2024, doi: 10.1016/j.measen.2024.101028.
[26] A. Frenica, L. Lindawati, L. Lindawati, S. Soim, and S. Soim, "Implementasi Algoritma Support Vector Machine (SVM) untuk Deteksi Banjir," *INOVTEK Polbeng - Seri Inform.*, vol. 8, no. 2, p. 291, 2023, doi: 10.35314/isi.v8i2.3443.
[27] K. K. Ray *et al.*, "Guava leaf disease detection using support vector machine (SVM)," *Smart Agric. Technol.*, vol. 12, Dec. 2025, doi: 10.1016/j.atech.2025.101190.