JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

Perbandingan Naïve Bayes dan Random Forest untuk Klasifikasi Sentimen Ulasan Produk Amazon Fire HD 7

Khabib Tri Anggara[1*], Rahmad Syukur Gea[2], Hendra Supendar[3], Riza Fahlapi[4]

[1,2,3,4]Machine Learning dan Kecerdasan Buatan, Universitas Bina Sarana Informatika, Jakarta, Indonesia
E-mail:[1]*khabibtri9@gmail.com, [2]rahmatsyukur782@gmail.com, [3]hendra.hds@bsi.ac.id, [4]riza.rzf@bsi.ac.id
(*Email Corresponding Author: valenkurnia12@gmail.com)
Received: 26 Juni 2026 | Revision: 30 Juni 2026 | Accepted: 3 Juli 2026

Abstrak

Di lingkungan e-commerce, penilaian yang ditulis pembeli pada kolom ulasan menyimpan opini yang berharga, baik untuk orang yang hendak berbelanja maupun untuk pihak yang menangani sebuah merek; sayangnya, banyaknya ulasan membuat penelaahan satu per satu menjadi tidak efisien. Riset ini menguji sejauh mana dua metode, yakni Naïve Bayes serta Random Forest, dapat memilah opini pada ulasan produk Amazon Fire HD 7 menjadi tiga kategori, yaitu positif, netral, dan negatif. Sejumlah 30.846 ulasan berbahasa Inggris terlebih dahulu dibersihkan, lalu diubah menjadi fitur bernilai numerik dengan pembobotan TF-IDF, dan dipisah secara terstratifikasi pada perbandingan 80:20; kualitas kedua model diperiksa lewat akurasi sekaligus precision, recall, dan F1-score yang dirata-ratakan secara makro. Hasilnya tidak seragam dan bergantung pada tolok ukur yang dipakai. Random Forest mencatat akurasi yang lebih besar (0,856 melawan 0,770), namun pada F1-score makro—ukuran yang lebih layak dipakai ketika data tidak berimbang—posisinya terbalik, dengan Naïve Bayes lebih tinggi (0,481 melawan 0,447). Kecenderungan Random Forest untuk menebak kategori yang paling banyak (positif) membuatnya lemah menangani kategori netral dan negatif, sementara Naïve Bayes membagi prediksinya secara lebih merata. Temuan tersebut mempertegas bahwa nilai akurasi tunggal mudah mengelabui pada data yang condong ke satu kelas, dan F1-score makro lebih mencerminkan mutu pemilahan opini bila kelasnya lebih dari dua.

Kata Kunci: analisis sentimen; Naïve Bayes; Random Forest; TF-IDF; ulasan produk

Abstract

On e-commerce sites, the opinions buyers leave in review sections are useful both for people about to make a purchase and for those responsible for a brand, yet examining such reviews one by one is impractical because there are so many of them. The present work investigates how well two methods—Naïve Bayes and Random Forest—separate the opinions expressed in Amazon Fire HD 7 reviews into three groups: positive, neutral, and negative. A collection of 30,846 English reviews was first cleaned and then turned into numeric features through TF-IDF weighting, after which the data were divided in a stratified 80:20 split; each model was judged using accuracy together with macro-averaged precision, recall, and F1-score. The outcome was not uniform and depended on the chosen measure. Random Forest reached the higher accuracy (0.856 against 0.770), but on the macro F1-score—the more suitable measure when the classes are unbalanced—the ranking reversed, with Naïve Bayes ahead (0.481 against 0.447). Because Random Forest leaned toward guessing the most frequent (positive) category, it handled the neutral and negative categories poorly, whereas Naïve Bayes spread its predictions more evenly. These results reaffirm that a single accuracy figure can be deceptive on skewed data and that the macro F1-score better reflects the quality of multiclass opinion classification.

Keywords: *sentiment analysis; Naïve Bayes; Random Forest; TF-IDF; product reviews*

1. PENDAHULUAN

Dalam perdagangan elektronik, ulasan produk kini termasuk rujukan yang paling banyak diandalkan. Calon pembeli umumnya menengok dulu cerita orang lain yang sudah memakai produk sebelum memutuskan bertransaksi; karena itu ulasan ikut menentukan keputusan beli sekaligus menggambarkan bagaimana publik memandang suatu merek. Amazon, yang tergolong pemain e-commerce terbesar secara global, menghimpun ulasan dalam jumlah yang sangat banyak pada tiap produknya, dan kumpulan ulasan itu menyimpan keterangan yang bernilai untuk menakar tingkat kepuasan pelanggan [1].

Nilai ulasan tidak berhenti pada keputusan pembelian perorangan. Bagi produsen dan pengelola merek, agregasi opini pelanggan berfungsi sebagai umpan balik langsung atas kualitas produk, fitur yang diapresiasi, maupun keluhan yang berulang, sehingga dapat menjadi dasar perbaikan produk dan penyusunan strategi pemasaran. Meski demikian, nilai tersebut baru dapat dimanfaatkan apabila opini yang tersebar pada ribuan hingga puluhan ribu ulasan diringkas secara cepat dan konsisten, bukan ditelaah satu per satu.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

Skala perdagangan elektronik yang terus tumbuh memperbesar persoalan ini. Setiap hari bermunculan ulasan baru pada beragam produk, dan sebuah produk populer saja dapat mengumpulkan puluhan ribu ulasan dalam kurun waktu singkat. Volume sebesar itu mustahil ditelaah secara manual dengan cermat, sementara opini yang terkandung di dalamnya justru semakin berharga karena mencerminkan pengalaman banyak pengguna. Ketersediaan data dalam jumlah besar inilah yang sekaligus menjadi tantangan dan peluang bagi penerapan metode otomatis.

Masalahnya, jumlah ulasan yang membludak dan tak tertata rapi membuat penilaian manual memakan waktu sekaligus mudah terpengaruh penafsiran pribadi [1]. Penjual maupun pengelola merek kesulitan mengikuti arah opini pelanggan secara utuh, sedangkan pembeli sendiri kerepotan menarik kesimpulan dari ribuan komentar. Keadaan ini menuntut adanya cara kerja otomatis yang sanggup menggolongkan ulasan ke dalam sentimen positif, netral, maupun negatif secara taat asas.

Penggolongan tiga kelas dipilih karena lebih informatif daripada pembedaan biner positif-negatif semata. Kelas netral menampung ulasan yang tidak condong ke salah satu kutub, misalnya penilaian bintang tiga yang memuat sisi positif sekaligus negatif, sehingga keberadaannya penting agar opini yang bercampur tidak dipaksakan masuk ke kategori positif atau negatif. Bagi pengelola merek, membedakan ulasan netral dari yang benar-benar negatif membantu memilah keluhan serius dari umpan balik yang masih moderat. Namun, penambahan kelas peralihan ini juga mempersulit klasifikasi karena batas antarkelasnya menjadi kurang tegas.

Kebutuhan tersebut dijawab oleh analisis sentimen, yaitu salah satu bentuk pemanfaatan Natural Language Processing (NLP) yang secara otomatis menentukan arah opini pada sebuah teks [2]. Untuk keperluan ini, sudah cukup banyak algoritma machine learning yang dicoba. Naïve Bayes sering menjadi pilihan sebab cara kerjanya ringkas dan hemat dalam memproses teks [2], termasuk ketika diterapkan pada ulasan di lokapasar maupun pada penilaian opini terhadap sebuah merek [3]. Sementara itu, Random Forest yang merupakan teknik ensemble dengan banyak pohon keputusan dipandang kuat menghadapi gangguan data dan fitur berdimensi besar, dan sudah pernah dipakai untuk membedah sentimen pada ulasan layanan digital [4], [5].

Analisis sentimen terhadap ulasan produk maupun opini publik telah banyak dikaji dengan beragam algoritma pembelajaran mesin. Naïve Bayes kerap dijadikan titik tolak karena proses pelatihannya ringan dan kebutuhan datanya kecil. Ramadhan dkk. [2] menerapkan Naïve Bayes untuk menggolongkan ulasan pada aplikasi e-commerce dan menunjukkan bahwa pendekatan probabilistik ini mampu memberikan hasil yang memadai pada data teks. Pada ranah opini merek, Putri dkk. [3] memanfaatkan Naïve Bayes classifier guna menakar sentimen terhadap brand skincare lokal. Kedua kajian memperlihatkan bahwa Naïve Bayes tetap relevan sebagai pembanding dasar, terutama ketika fitur teks direpresentasikan melalui pembobotan seperti TF-IDF.

Random Forest, sebagai metode ensemble berbasis banyak pohon keputusan, banyak dipilih ketika dimensi fitur teks membesar dan data mengandung derau. Larasati dkk. [4] menggunakan Random Forest untuk menganalisis sentimen ulasan aplikasi Dana, sementara Indrayanto dkk. [5] menerapkannya pada ulasan pengguna aplikasi MyPertamina di Google Play Store. Keduanya memanfaatkan kemampuan Random Forest dalam menangani interaksi antarfitur yang kompleks tanpa banyak penyetelan, sehingga metode ini kerap tampil kompetitif pada klasifikasi teks.

Sejumlah penelitian membandingkan langsung kedua algoritma, tetapi simpulannya tidak seragam dan sangat dipengaruhi karakteristik data. Pada klasifikasi ulasan aplikasi Kredivo, Random Forest unggul dengan akurasi 91% berbanding 82% milik Naïve Bayes [6], dan pola serupa muncul pada ulasan aplikasi DeepSeek dengan Random Forest mencapai 96,38% [7]. Sebaliknya, pada penggolongan sentimen unggahan media sosial, Naïve Bayes justru jauh lebih akurat (90,41%) dibandingkan Random Forest (39,74%) [8]. Miftahusalam dkk. [9] yang membandingkan Random Forest, Naïve Bayes, dan Support Vector Machine pada sentimen Twitter turut menegaskan bahwa peringkat antaralgoritma dapat berpindah bergantung pada jenis teks dan sebaran kelasnya.

Dari tinjauan tersebut, dua hal menjadi catatan. Pertama, mayoritas studi berfokus pada ulasan aplikasi di Google Play Store atau cuitan berbahasa Indonesia, sedangkan perbandingan langsung Naïve Bayes dan Random Forest untuk klasifikasi tiga kelas pada ulasan produk berbahasa Inggris di lokapasar—khususnya produk Amazon—masih jarang dijumpai. Kedua, sebagian besar kajian menilai kinerja terutama melalui akurasi, padahal pada data yang sangat timpang akurasi dapat menyesatkan. Penelitian ini mengisi celah tersebut dengan menyandingkan kedua algoritma pada dataset ulasan Amazon Fire HD 7 serta menempatkan F1-score makro sebagai tolok ukur utama di samping akurasi.

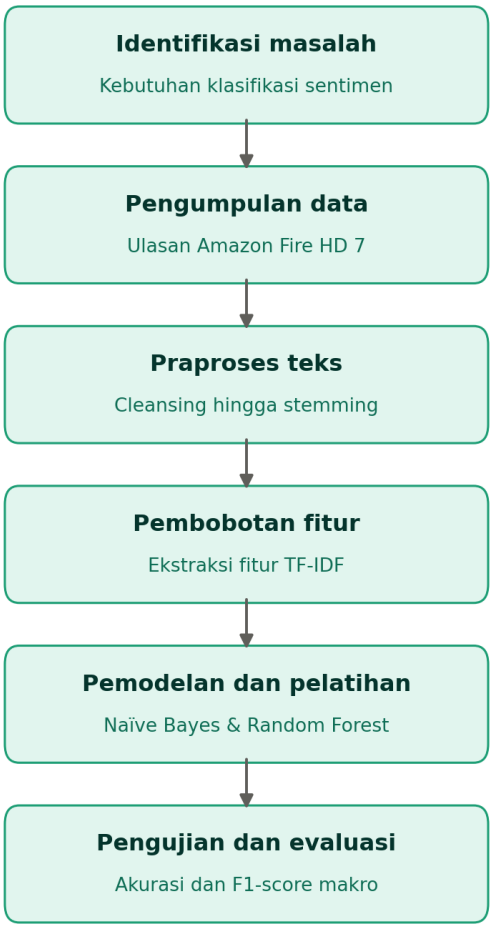
Kontribusi penelitian ini mencakup tiga hal. Pertama, tersedianya bukti empiris perbandingan Naïve Bayes dan Random Forest pada klasifikasi sentimen tiga kelas untuk ulasan produk berbahasa Inggris di Amazon, ranah yang belum banyak disentuh kajian sejenis. Kedua, penekanan pada F1-score makro memperlihatkan bagaimana kesimpulan mengenai keunggulan algoritma dapat berubah bergantung pada metrik ketika data sangat timpang. Ketiga, hasilnya memberikan pertimbangan praktis bagi pengelola merek dalam memilih algoritma untuk memantau opini pelanggan secara otomatis.

Secara ringkas, permasalahan yang diangkat dapat dirumuskan sebagai berikut: bagaimana perbandingan kinerja Naïve Bayes dan Random Forest dalam mengklasifikasikan sentimen ulasan produk berbahasa Inggris ke dalam tiga kelas, dan bagaimana pengaruh ketidakseimbangan kelas terhadap pemilihan metrik yang tepat untuk menilai kinerja tersebut. Selebihnya, artikel ini disusun sebagai berikut: bagian kedua memaparkan data dan tahapan metode yang ditempuh, bagian ketiga menyajikan hasil pengujian beserta pembahasannya, dan bagian keempat memuat simpulan.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

2. METODE PENELITIAN

Kajian ini bersifat kuantitatif dan disusun sebagai eksperimen yang mempertandingkan dua algoritma penggolong. Keseluruhan prosesnya dijalankan di atas aplikasi Orange Data Mining. Alur kerjanya mencakup enam tahap yang ditempuh berurutan, mulai dari perumusan masalah, penghimpunan data, pembersihan teks, pembobotan fitur, pembentukan serta pelatihan model, hingga pengujian dan evaluasi, seperti dirangkum dalam Gambar 1.



Gambar 1. Diagram alur tahapan penelitian.

Alur pada Gambar 1 ditempuh secara berurutan. Tahap identifikasi masalah menetapkan kebutuhan klasifikasi sentimen atas ulasan produk. Tahap pengumpulan data menghimpun ulasan Amazon Fire HD 7 sebagai bahan analisis. Data mentah kemudian melewati tahap praproses teks untuk membersihkan dan menyeragamkan kata, dilanjutkan tahap pembobotan fitur yang mengubah teks menjadi representasi numerik TF-IDF. Representasi tersebut dipakai pada tahap pemodelan dan pelatihan untuk membangun model Naïve Bayes dan Random Forest, lalu diakhiri tahap pengujian dan evaluasi yang mengukur kinerja masing-masing model melalui akurasi dan F1-score makro. Rincian tiap tahap diuraikan pada subbab berikut.

Secara teknis, keseluruhan alur pada Gambar 1 diwujudkan sebagai rangkaian widget yang saling terhubung di Orange. Korpus ulasan dimuat melalui widget Corpus, lalu dialirkan ke widget Preprocess Text untuk pembersihan, tokenisasi, penyaringan, dan stemming. Keluarannya masuk ke widget Bag of Words yang menghitung bobot TF-IDF, kemudian ke widget Data Sampler yang membagi data secara berstrata pada perbandingan 80:20. Data latih dihubungkan ke widget Naïve Bayes dan Random Forest, sedangkan kinerja masing-masing model dinilai melalui widget yang menghitung skor pengujian sekaligus confusion matrix pada data uji. Perancangan visual semacam ini memudahkan penelusuran ketika terjadi kekeliruan konfigurasi pada salah satu tahap.

2.1 Pengumpulan Data

Bahan kajian berjumlah 30.846 ulasan untuk produk Amazon Fire HD 7 pada kategori PC di marketplace Amerika Serikat; karena berasal dari pasar tersebut, keseluruhan ulasan tertulis dalam bahasa Inggris dan terkumpul sepanjang periode Oktober 2014 sampai Agustus 2015. Himpunan data ini bersumber dari koleksi ulasan pelanggan Amazon yang dipublikasikan secara terbuka [11]. Tiap entri menyimpan beberapa kolom, antara lain review_body (isi ulasan), review_headline (judul ulasan), star_rating (skor 1–5 bintang), dan verified_purchase. Dalam studi ini, kolom yang

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

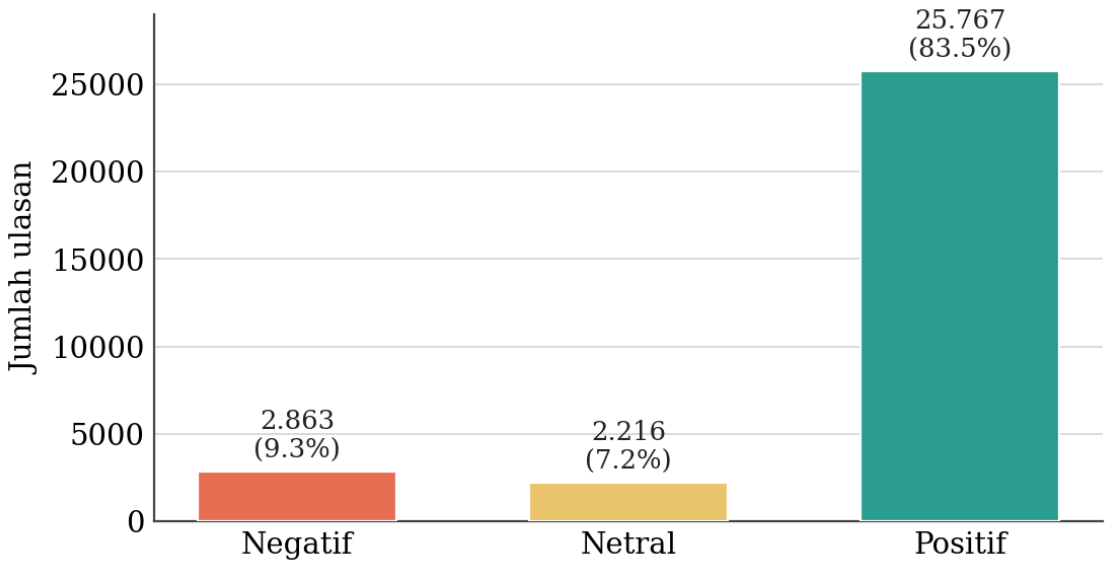
dipakai sebagai masukan teks hanyalah review_body, sementara kolom selebihnya diperlakukan sebagai metadata dan tidak difungsikan sebagai fitur prediktor.

Pemusatan pada satu produk dengan jumlah ulasan besar dipilih agar perbandingan kedua algoritma dilakukan pada kondisi data yang seragam, sehingga perbedaan kinerja lebih dapat dikaitkan dengan algoritma alih-alih keberagaman produk. Rentang waktu pengumpulan yang mencakup hampir satu tahun turut menghadirkan variasi ulasan dari waktu ke waktu. Adapun pembatasan masukan pada review_body didasari pertimbangan bahwa isi ulasanlah yang paling langsung memuat opini, sementara judul ulasan cenderung ringkas dan metadata seperti status pembelian tidak menyampaikan muatan sentimen.

2.2 Pelabelan

Label sentimen tiga kelas diturunkan secara otomatis dari star_rating dengan ketentuan: peringkat 1–2 bintang dikategorikan negatif, 3 bintang netral, dan 4–5 bintang positif. Skema ini menghasilkan distribusi kelas sebesar 25.767 ulasan positif, 2.863 negatif, dan 2.216 netral. Distribusi tersebut sangat timpang (kelas positif mendominasi sekitar 83,5%), sehingga pemilihan metrik yang sesuai diperlukan agar hasil tidak bias terhadap kelas mayoritas. Sebanyak empat ulasan dengan review_body kosong dihapus sebelum pemrosesan.

Ketimpangan sebaran kelas tersebut ditampilkan pada Gambar 2. Kelas positif mendominasi hampir seluruh korpus, sedangkan kelas netral dan negatif hanya menempati porsi kecil, sehingga penilaian kinerja tidak dapat bertumpu pada akurasi semata.



Gambar 2. Distribusi tiga kelas sentimen pada dataset.

2.3 Praproses Teks

Mengingat ulasan berbahasa Inggris, tahap praproses ditangani lewat widget Preprocess Text di Orange yang dijalankan secara bertahap: (1) transformation, untuk menurunkan seluruh huruf menjadi lowercase, merapikan markah HTML (parse html) semisal
, sekaligus melenyapkan URL; (2) tokenization, untuk mengurai kalimat menjadi token kata; (3) filtering, untuk menyaring stopword bahasa Inggris berikut token yang berupa angka; dan (4) normalization, untuk mengembalikan setiap token ke bentuk dasarnya dengan algoritma Porter Stemmer.

Setiap tahap praproses memikul peran tertentu. Penurunan huruf dan pembersihan markah HTML menyeragamkan bentuk kata sekaligus membuang elemen non-tekstual yang tidak membawa makna sentimen. Penyaringan stopword menghilangkan kata fungsi berfrekuensi tinggi yang cenderung menambah derau tanpa nilai pembeda, sedangkan penghapusan token berupa angka menekan fitur yang tidak relevan. Adapun stemming menyatukan berbagai bentuk infleksi sebuah kata ke satu akar, misalnya bentuk jamak dan kata berimbuhan, sehingga dimensi fitur menyusut dan kemunculan kata yang bermakna sama tidak terpecah menjadi banyak fitur berbeda.

2.4 Pembobotan Fitur TF-IDF

Token yang dihasilkan praproses selanjutnya dipetakan ke bentuk numerik via widget Bag of Words memakai skema pembobotan TF-IDF, dengan term frequency dihitung berbasis kemunculan (count), document frequency dipakai untuk IDF, serta penormalan L2. Pada skema ini, bobot suatu term t di dokumen d dirumuskan $tfidf(t,d) = tf(t,d) \times idf(t)$, sedangkan $idf(t) = \log(N / df(t))$, dengan N menyatakan total dokumen dan df(t) menyatakan banyaknya dokumen yang mengandung term t.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

Pemilihan TF-IDF di atas pencacahan kata mentah didasari kemampuannya menimbang pentingnya sebuah kata secara lebih proporsional. Kata yang muncul di hampir semua ulasan, meskipun frekuensinya tinggi, memperoleh bobot rendah karena daya pembedanya kecil; sebaliknya, kata yang khas pada sebagian ulasan memperoleh bobot lebih besar. Penormalan panjang dokumen juga menjaga agar ulasan yang panjang tidak serta-merta mendominasi hanya karena memuat lebih banyak kata. Dengan cara ini, representasi fitur menjadi lebih peka terhadap istilah yang benar-benar menandai perbedaan sentimen.

2.5 Algoritma Klasifikasi

Kedua algoritma dijajarkan dan dibangun lewat widget pemodelan di Orange. Naïve Bayes adalah pengklasifikasi probabilistik yang berpijak pada teorema Bayes, $P(c|d) = [P(c) \times \Pi\, P(t_i|c)] / P(d)$, yang mengandaikan tiap fitur bersifat independen bila kelasnya telah diketahui, dan dipanggil melalui widget Naive Bayes pada konfigurasi default. Adapun Random Forest tergolong metode ensemble yang menumbuhkan sejumlah besar pohon keputusan via bagging serta penarikan subset fitur secara acak, kemudian merangkum keluaran tiap pohon lewat majority vote [10]; pada penelitian ini Random Forest dijalankan melalui widget terkait dengan banyak pohon ditetapkan 100.

Sebagai pengklasifikasi probabilistik, Naïve Bayes menaksir peluang posterior setiap kelas dari kemunculan kata pada dokumen, kemudian memilih kelas dengan peluang tertinggi. Untuk mencegah peluang bernilai nol pada kata yang tidak pernah muncul bersama suatu kelas, diterapkan penghalusan (smoothing) sederhana, dan perkalian antarpeluang lazim dihitung dalam ruang logaritma agar stabil secara numerik. Asumsi independensi bersyarat antarkata memang jarang terpenuhi pada bahasa alami, tetapi dalam praktik membuat model ringkas, cepat dilatih, serta tahan terhadap fitur berdimensi tinggi seperti TF-IDF; karena itulah Naïve Bayes lazim ditempatkan sebagai garis dasar pada klasifikasi teks.

Aturan keputusan Naïve Bayes dapat dinyatakan sebagai pemilihan kelas dengan peluang posterior tertinggi seperti pada persamaan (1); dalam praktik, perkalian peluang dihitung pada ruang logaritma sebagaimana persamaan (2) agar tetap stabil secara numerik, dengan C menyatakan himpunan kelas dan t_i menyatakan token ke-i pada dokumen.

$$\text{prediksi} = \operatorname{argmax} P(c) \times \Pi\, P(t_i \mid c),\ \ c \in C \qquad (1)$$

$$\text{prediksi} = \operatorname{argmax} [\ \log P(c) + \Sigma \log P(t_i \mid c)\ ],\ \ c \in C \qquad (2)$$

Random Forest menekan kelemahan pohon keputusan tunggal yang mudah mengalami overfitting dengan menumbuhkan banyak pohon pada sampel bootstrap yang berlainan serta memilih subset fitur secara acak pada tiap pemecahan simpul. Pengacakan ganda ini membuat pohon-pohon saling tidak berkorelasi sehingga penggabungan keputusannya melalui suara terbanyak menghasilkan prediksi yang lebih stabil. Sebagian data yang tidak terpilih pada tiap bootstrap (out-of-bag) dapat dipakai untuk menaksir kesalahan tanpa memerlukan data uji terpisah, sedangkan kontribusi tiap kata terhadap keputusan dapat diperkirakan melalui ukuran kepentingan fitur. Kemampuan inilah yang membuat Random Forest sanggup menangkap pola taklinear pada data teks berdimensi besar.

2.6 Skema Pengujian dan Evaluasi

Pemisahan data latih dan data uji ditetapkan pada perbandingan 80:20 melalui widget Data Sampler yang opsi stratifikasinya diaktifkan supaya komposisi ketiga kelas tetap proporsional. Persoalan kelas yang tidak seimbang diredam baik oleh pembagian berstrata itu maupun—dan ini yang terpenting—oleh penetapan F1-score makro sebagai tolok ukur utama, sebab metrik tersebut memperlakukan setiap kelas dengan bobot yang sama. Pengukuran kinerja berpijak pada confusion matrix yang disertai akurasi serta precision, recall, dan F1-score; lantaran kelasnya berjumlah tiga dan timpang, ketiga metrik terakhir lebih dahulu dihitung pada tiap kelas, lalu dirata-ratakan secara makro (macro-average).

2.7 Metrik Evaluasi

Penilaian kinerja bertumpu pada confusion matrix, yaitu tabel yang memetakan kelas sebenarnya terhadap kelas hasil prediksi. Dari tabel tersebut diperoleh empat besaran dasar bagi tiap kelas: true positive (TP) dan true negative (TN) sebagai prediksi yang benar, serta false positive (FP) dan false negative (FN) sebagai prediksi yang keliru. Berdasarkan keempat besaran itu dihitung akurasi, precision, recall, dan F1-score sebagaimana dinyatakan pada persamaan (3)–(6).

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN) \qquad (3)$$

$$\text{Precision} = TP / (TP + FP) \qquad (4)$$

$$\text{Recall} = TP / (TP + FN) \qquad (5)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (6)$$

Karena persoalan yang dihadapi bersifat multikelas dan sebarannya timpang, precision, recall, dan F1-score lebih dahulu dihitung pada masing-masing kelas, lalu dirata-ratakan secara makro (macro-average) seperti pada persamaan (7), dengan k menyatakan banyaknya kelas dan metrik_i menyatakan nilai metrik pada kelas ke-i. Rata-rata makro memberi bobot yang sama kepada setiap kelas sehingga kegagalan pada kelas minoritas tidak tertutup oleh keberhasilan pada kelas mayoritas. Sebaliknya, akurasi memperlakukan seluruh ulasan secara setara sehingga cenderung mengikuti kelas yang paling banyak; hal inilah yang menjadikan F1-score makro lebih layak dijadikan tolok ukur utama pada data seperti ini.

$$\text{Makro} = (1 / k) \times \Sigma\, \text{metrik\_i},\ \ i = 1, 2, \ldots, k \qquad (7)$$

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

3. HASIL DAN PEMBAHASAN

Pada bagian ini diuraikan hasil penyandingan kinerja Naïve Bayes dan Random Forest ketika memilah sentimen ulasan menjadi tiga kelas, yang dilanjutkan dengan telaah per kelas serta pembahasan atas temuan yang diperoleh. Seluruh pengujian dijalankan pada himpunan data uji sebanyak 6.169 ulasan.

Tabel 1. Ringkasan capaian Naïve Bayes dan Random Forest pada data uji.

Algoritma	Akurasi	Precision (makro)	Recall (makro)	F1-score (makro)
Naïve Bayes	0,770	0,479	0,492	0,481
Random Forest	0,856	0,706	0,422	0,447

Berdasarkan Tabel 1, kedua algoritma memberikan gambaran yang berbeda bergantung pada metrik yang digunakan. Random Forest memperoleh akurasi lebih tinggi (0,856 berbanding 0,770) dan precision makro lebih tinggi (0,706 berbanding 0,479), sedangkan Naïve Bayes unggul pada recall makro (0,492 berbanding 0,422) dan, yang terpenting, pada F1-score makro (0,481 berbanding 0,447). Karena distribusi kelas sangat timpang, F1-score makro dijadikan acuan utama; berdasarkan metrik tersebut, Naïve Bayes sedikit lebih unggul dibandingkan Random Forest.

Tabel 2. Rincian capaian tiap kelas bagi masing-masing model.

Algoritma	Kelas	Precision	Recall	F1-score
Naïve Bayes	Negatif	0,371	0,347	0,359
Naïve Bayes	Netral	0,165	0,266	0,204
Naïve Bayes	Positif	0,902	0,861	0,881
Random Forest	Negatif	0,732	0,248	0,370
Random Forest	Netral	0,524	0,025	0,047
Random Forest	Positif	0,861	0,995	0,923

Tabel 2 memerinci capaian precision, recall, dan F1-score untuk tiap kelas pada kedua model. Keduanya sama-sama kuat pada kelas positif, tetapi berbeda tajam pada kelas minoritas: recall Random Forest pada kelas netral hanya 0,025, jauh di bawah Naïve Bayes, sehingga F1-score kelas netralnya pun anjlok menjadi 0,047.

Tabel 3. Matriks kekeliruan Naïve Bayes (baris menyatakan kelas sebenarnya, kolom menyatakan kelas hasil prediksi).

Aktual \ Prediksi	Negatif	Netral	Positif
Negatif	199	120	254
Netral	95	118	230
Positif	242	475	4436

Tabel 3 menyajikan sebaran prediksi Naïve Bayes terhadap kelas sebenarnya. Model ini mengenali 199 ulasan negatif dan 118 ulasan netral secara benar; meskipun sebagian ulasan minoritas masih terklasifikasi ke kelas positif, sebaran kekeliruannya relatif merata di ketiga kelas.

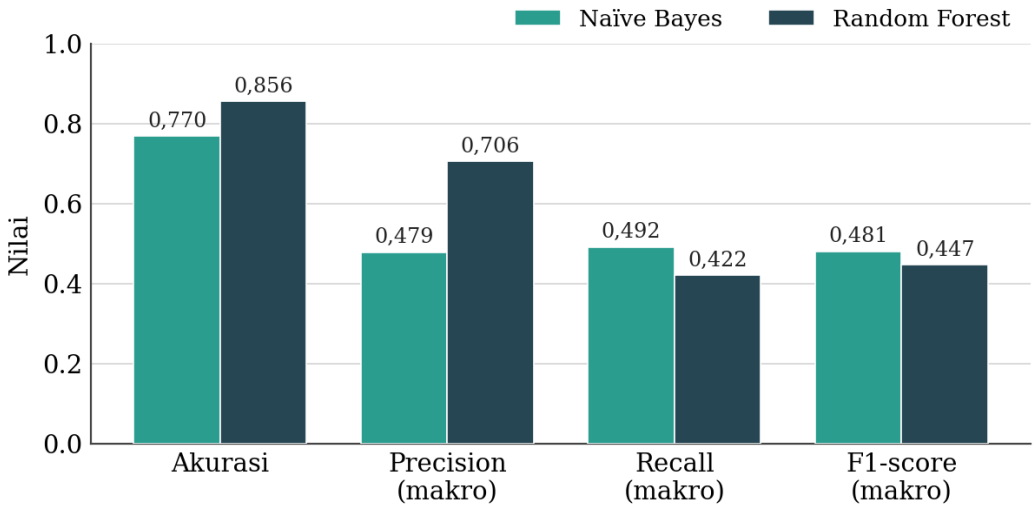
Tabel 4. Matriks kekeliruan Random Forest (baris menyatakan kelas sebenarnya, kolom menyatakan kelas hasil prediksi).

Aktual \ Prediksi	Negatif	Netral	Positif
Negatif	142	6	425
Netral	29	11	403
Positif	23	4	5126

Tabel 4 menampilkan sebaran prediksi Random Forest. Tampak kecenderungan model mengarahkan mayoritas ulasan minoritas ke kelas positif: dari 573 ulasan negatif, 425 diprediksi positif, dan dari 443 ulasan netral, 403 diprediksi positif, sehingga kelas minoritas nyaris tidak terwakili dengan benar.

Perbandingan keempat metrik makro tersebut divisualisasikan pada Gambar 3. Terlihat bahwa keunggulan Random Forest terpusat pada akurasi dan precision, sedangkan Naïve Bayes memimpin pada recall dan F1-score makro—dua ukuran yang lebih peka terhadap kinerja lintas kelas.

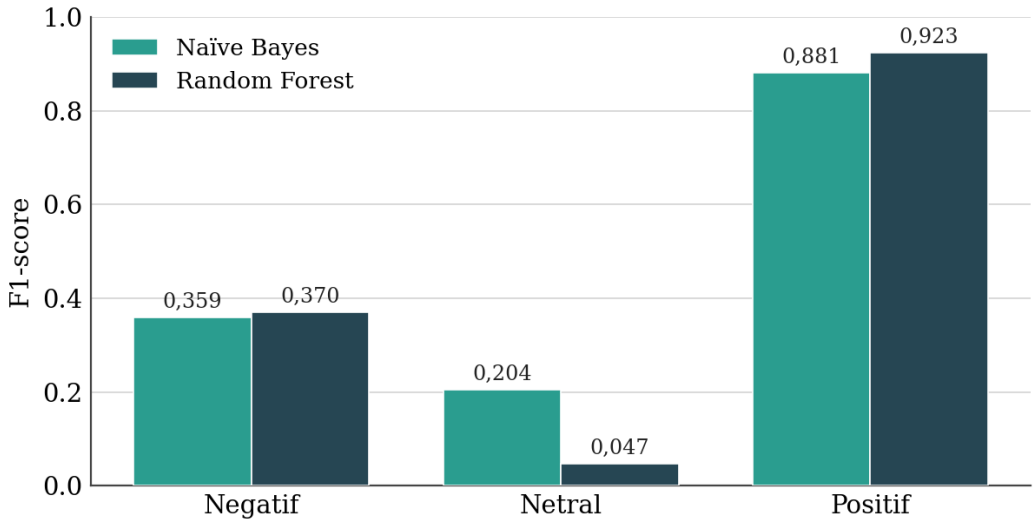
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315



Gambar 3. Perbandingan metrik makro Naïve Bayes dan Random Forest.

Tabel 2 beserta confusion matrix pada Tabel 3 dan Tabel 4 menjelaskan penyebab perbedaan tersebut. Kedua model sama-sama kuat pada kelas mayoritas, yaitu positif (F1 Naïve Bayes 0,881; Random Forest 0,923). Perbedaan utamanya terletak pada kelas minoritas. Random Forest memaksimalkan akurasi dengan memprediksi hampir seluruh ulasan sebagai positif (5.954 dari 6.169 ulasan), sehingga recall positifnya nyaris sempurna (0,995) tetapi kinerjanya runtuh pada kelas minoritas: model hanya mengenali 11 dari 443 ulasan netral (recall 0,025) dan 142 dari 573 ulasan negatif (recall 0,248). Sebaliknya, Naïve Bayes menyebar prediksinya lebih merata sehingga mampu memulihkan lebih banyak instans kelas minoritas (F1 negatif 0,359; F1 netral 0,204), meskipun dengan akurasi keseluruhan yang lebih rendah.

Gambar 4 mempertegas perbedaan tersebut pada tataran per kelas. Pada kelas positif kedua model sama-sama tinggi, tetapi pada kelas netral capaian Random Forest nyaris runtuh (F1 hanya 0,047), jauh di bawah Naïve Bayes; selisih inilah yang menekan rata-rata makro Random Forest meskipun akurasinya tinggi.



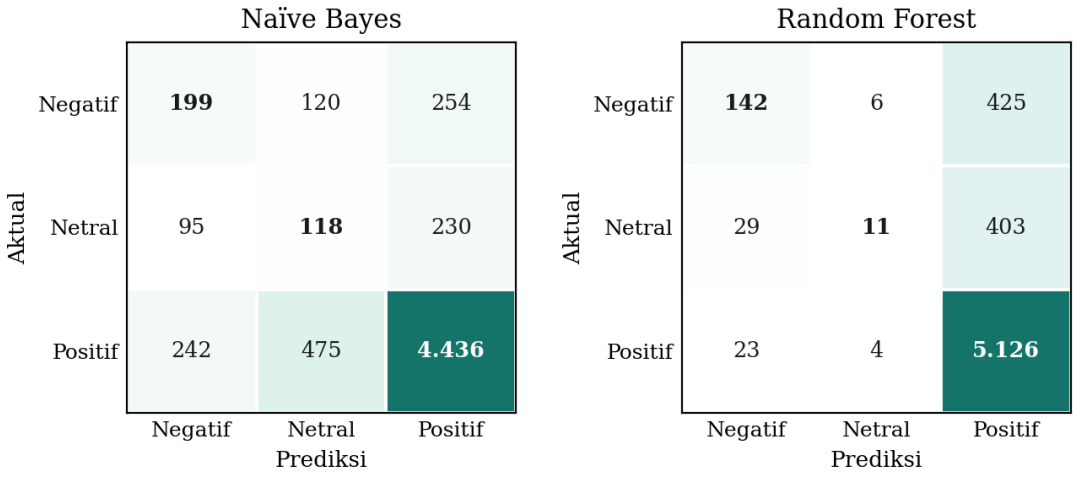
Gambar 4. Perbandingan F1-score tiap kelas pada kedua model.

Pola kekeliruan pada Tabel 3 dan Tabel 4 memperjelas asal perbedaan itu. Random Forest mengarahkan mayoritas ulasan minoritas ke kelas positif: dari 573 ulasan negatif, 425 diprediksi positif, dan dari 443 ulasan netral, 403 diprediksi positif. Naïve Bayes menyebarkan kekeliruannya lebih merata sehingga, meskipun turut salah mengarahkan sebagian ulasan ke kelas positif, model ini masih mengenali 199 ulasan negatif dan 118 ulasan netral secara benar—jauh lebih banyak daripada 142 dan 11 yang dicapai Random Forest. Dengan kata lain, tingginya akurasi Random Forest sebagian besar bersumber dari 5.126 ulasan positif yang tertebak benar, bukan dari kemampuan memilah ketiga kelas secara seimbang.

Visualisasi confusion matrix pada Gambar 5 menegaskan pola tersebut secara sekilas. Pada Random Forest, kolom prediksi positif tampak jauh lebih pekat dibandingkan kolom lain, memperlihatkan bahwa hampir seluruh ulasan—termasuk yang sebenarnya negatif dan netral—dialirkan ke satu kolom. Naïve Bayes menampilkan sebaran yang lebih merata pada ketiga kolom, sehingga meskipun diagonal utamanya tidak sepekat Random Forest pada kelas positif, model

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

ini lebih sering menempatkan ulasan minoritas pada kolom yang tepat. Perbedaan visual inilah yang secara angka terangkum pada selisih recall dan F1-score antar kelas.



Gambar 5. Confusion matrix Naïve Bayes dan Random Forest dalam bentuk heatmap.

Rincian per kelas pada Tabel 2 juga menyingkap pola precision-recall yang berbeda di antara kedua model. Pada kelas negatif, Random Forest mencatat precision tinggi (0,732) tetapi recall rendah (0,248); artinya, ketika model menandai sebuah ulasan sebagai negatif prediksinya kerap benar, namun sebagian besar ulasan negatif justru luput dan tergolongkan ke kelas lain. Naïve Bayes menunjukkan precision dan recall yang lebih berimbang pada kelas yang sama (0,371 dan 0,347). Pola serupa tampak pada kelas netral: precision Random Forest 0,524 tidak berarti banyak apabila recall-nya hanya 0,025, sebab model nyaris tidak pernah memilih kelas tersebut. Keseimbangan precision dan recall inilah yang diringkas oleh F1-score dan, pada gilirannya, oleh rata-rata makro.

Kontras ini menunjukkan mengapa akurasi semata dapat menyesatkan pada data tidak seimbang: akurasi tinggi Random Forest sebagian besar mencerminkan keberhasilannya pada kelas dominan, bukan kinerja yang seimbang antar kelas. F1-score makro, yang memberi bobot setara pada setiap kelas, menghukum kegagalan Random Forest pada kelas netral, sehingga Naïve Bayes—walau kurang akurat secara keseluruhan—dinilai lebih seimbang. Kelas netral menjadi kelas yang paling sulit bagi kedua model, kemungkinan karena jumlahnya paling sedikit dan secara semantik paling ambigu, yakni ulasan bintang tiga yang kerap tumpang tindih kosakata dengan ulasan positif maupun negatif.

Perbedaan perilaku kedua model membawa konsekuensi praktis bagi pemantauan opini pelanggan. Bagi pengelola merek, kemampuan mengenali ulasan negatif dan netral sering justru lebih bernilai daripada sekadar memastikan ulasan positif terdeteksi, sebab keluhan dan sinyal ketidakpuasan itulah yang menuntut tindak lanjut. Dalam konteks tersebut, model dengan akurasi tinggi tetapi nyaris buta terhadap kelas minoritas—sebagaimana Random Forest pada data ini—berpotensi menyesatkan pengambilan keputusan karena melewatkan sebagian besar umpan balik kritis. Naïve Bayes yang menyebarkan prediksinya lebih merata memberikan gambaran yang lebih seimbang, walau ketepatannya pada kelas mayoritas sedikit lebih rendah.

Kesulitan pada kelas netral layak ditelaah lebih jauh. Di samping jumlahnya yang paling sedikit, ulasan bintang tiga secara semantik berada di antara dua kutub sehingga pilihan katanya sering beririsan dengan kelas positif maupun negatif; sebuah ulasan dapat memuji satu aspek produk sembari mengeluhkan aspek lain dalam kalimat yang sama. Pada Tabel 3 terlihat bahwa dari 443 ulasan netral, Naïve Bayes mengarahkan 230 ke kelas positif dan 95 ke kelas negatif, sedangkan hanya 118 yang dikenali sebagai netral. Representasi TF-IDF yang memperlakukan kata secara terpisah tanpa memperhitungkan konteks turut mempersulit pemisahan kelas peralihan ini, sehingga capaian pada kelas netral tetap jauh di bawah kelas positif bagi kedua model.

Kecenderungan peringkat model yang berbalik ketika metrik berganti sejalan dengan pemahaman umum pada klasifikasi data tidak seimbang, yaitu bahwa model condong ke kelas mayoritas apabila penilaian menekankan ketepatan keseluruhan. Random Forest pada data ini memaksimalkan akurasi dengan bermain aman menebak kelas positif, perilaku yang secara statistik menguntungkan ketika satu kelas menguasai lebih dari empat perlima data. Upaya meredam bias semacam ini, seperti penyeimbangan data melalui SMOTE [6] atau pemberian bobot kelas, umumnya diterapkan hanya pada data latih agar model lebih memperhatikan kelas minoritas tanpa mengganggu keutuhan data uji.

Dari sisi penerapan, pilihan algoritma sebaiknya mengikuti tujuan pemantauan. Apabila sasaran utamanya adalah menjaring keluhan dan ulasan negatif seakurat mungkin, model yang menyebarkan prediksi lebih merata seperti Naïve Bayes lebih sesuai meskipun akurasi keseluruhannya lebih rendah. Sebaliknya, bila kebutuhan hanya sebatas memperkirakan proporsi sentimen positif pada arus ulasan yang memang didominasi kelas positif, akurasi tinggi Random Forest masih dapat diterima dengan catatan keterbatasannya pada kelas minoritas dipahami betul. Dalam banyak kasus nyata, memadukan kedua sudut pandang—akurasi untuk gambaran umum dan F1-score makro untuk keseimbangan antarkelas—memberikan dasar keputusan yang lebih utuh daripada mengandalkan satu angka saja.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

Pelajaran metodologis dari perbandingan ini berlaku melampaui kasus Amazon Fire HD 7. Banyak persoalan klasifikasi di dunia nyata—deteksi ulasan palsu, penyaringan keluhan, hingga pemantauan isu—memiliki sebaran kelas yang timpang, sehingga model yang tampak unggul menurut akurasi belum tentu berguna secara operasional. Menyertakan metrik yang sensitif terhadap kelas minoritas sejak tahap evaluasi, bukan sekadar sebagai pelengkap, membantu memastikan bahwa model yang dipilih benar-benar menjawab kebutuhan dan bukan sekadar mengejar angka yang mengesankan.

Temuan ini sebagian sejalan dengan penelitian terdahulu. Pada aspek akurasi, hasil ini menggemakan studi ulasan Kredivo yang menempatkan Random Forest di atas Naïve Bayes (91% berbanding 82%) [6]; pada penelitian ini Random Forest juga lebih akurat. Namun, ketika dievaluasi dengan F1-score makro, peringkat keduanya berbalik. Hal ini menegaskan bahwa algoritma yang dianggap lebih baik sangat bergantung pada metrik yang dipilih dan tingkat ketidakseimbangan data [9]; sebuah angka akurasi tunggal dapat menyembunyikan buruknya kinerja pada kelas minoritas.

Menariknya, ketika penilaian bergeser ke F1-score makro, arah temuan ini lebih dekat dengan studi klasifikasi sentimen media sosial yang menempatkan Naïve Bayes di atas Random Forest [8]. Kesamaannya terletak pada kondisi data yang tidak seimbang: pada situasi demikian, kesederhanaan Naïve Bayes yang menyebarkan prediksi ke seluruh kelas dapat berbalik menjadi keunggulan dibandingkan kecenderungan Random Forest memusat pada kelas mayoritas. Dengan demikian, posisi penelitian ini tidak sepenuhnya berpihak pada salah satu kubu terdahulu, melainkan menunjukkan bahwa jawaban bergantung pada metrik dan tingkat ketidakseimbangan yang dihadapi.

Kesesuaian temuan ini dengan literatur terkini turut memperkuat simpulan. Studi yang juga menggunakan pelabelan berbasis rating menemukan bahwa klasifikasi tiga label cenderung berakurasi lebih rendah dibanding dua label [12], sejalan dengan sulitnya kelas netral pada penelitian ini. Perbandingan Naïve Bayes dengan algoritma lain pada sentimen produk gawai maupun ulasan aplikasi layanan juga menunjukkan bahwa peringkat antaralgoritma berpindah mengikuti karakteristik dan sebaran data [13], [14]. Secara umum, kajian menyeluruh mengenai analisis sentimen menegaskan bahwa pemilihan algoritma dan metrik perlu disesuaikan dengan konteks data dan tujuan analisis [15].

Penelitian ini memiliki sejumlah keterbatasan. Dataset hanya mencakup satu produk, yaitu Amazon Fire HD 7 berbahasa Inggris pada periode 2014–2015, sehingga generalisasi ke produk, bahasa, atau platform lain perlu diuji lebih lanjut. Selain itu, pelabelan otomatis berbasis peringkat bintang memperlakukan ulasan bintang tiga sebagai netral, yang tidak selalu mencerminkan opini netral secara semantik dan turut menyumbang sulitnya kelas tersebut. Pada alur kerja Orange, pembobotan TF-IDF juga dihitung atas seluruh korpus sebelum pemisahan data.

Beberapa keterbatasan lain juga perlu dicatat. Kedua algoritma dijalankan pada konfigurasi bawaan tanpa penyetelan hiperparameter, sehingga potensi kinerja terbaik masing-masing model belum tentu tercapai. Evaluasi pun bertumpu pada satu skema pembagian 80:20, bukan validasi silang, sehingga variasi hasil antarsampel belum terukur. Representasi fitur yang dipakai masih terbatas pada TF-IDF yang mengabaikan urutan dan konteks kata, sehingga nuansa seperti negasi berpeluang tidak tertangkap dengan baik.

4. KESIMPULAN

Studi ini menyandingkan algoritma Naïve Bayes dan Random Forest pada klasifikasi sentimen tiga kelas terhadap 30.846 ulasan produk Amazon Fire HD 7. Pengujian memperlihatkan Random Forest unggul dari sisi akurasi (0,856) atas Naïve Bayes (0,770), namun pada F1-score makro justru Naïve Bayes yang lebih tinggi (0,481) dibandingkan Random Forest (0,447). Karena sebaran kelas amat timpang, F1-score makro ditempatkan sebagai rujukan utama; menurut metrik ini Naïve Bayes berkinerja lebih merata di seluruh kelas, sedangkan keunggulan akurasi Random Forest sebagian besar lahir dari dominasi kelas positif dan dibarengi kelemahan yang menonjol pada kelas netral. Dengan demikian, untuk klasifikasi sentimen multikelas atas data ulasan yang tidak seimbang seperti ini, Naïve Bayes lebih layak dipilih apabila keseimbangan antarkelas menjadi prioritas, sementara Random Forest tetap bernilai bila fokus utamanya pada ketepatan keseluruhan. Temuan ini sekaligus menegaskan pentingnya menyelaraskan metrik evaluasi dengan karakteristik data, sebab pelaporan akurasi tunggal berisiko melebih-lebihkan kualitas model pada data yang timpang, sementara arah pengembangan berikutnya mencakup penyeimbangan data, penyetelan hiperparameter, penerapan validasi silang, serta pemanfaatan representasi berbasis konteks seperti word embedding untuk menangkap makna yang luput dari TF-IDF.

REFERENSI

[1] F. D. Sihaloho, Jasmir, dan Gunardi, "Klasifikasi sentimen ulasan produk olahraga di Tokopedia menggunakan metode machine learning dengan pendekatan TF-IDF," Prosiding Seminar Nasional Ilmu Teknik, vol. 2, no. 2, hlm. 976–989, 2025.

[2] B. Z. Ramadhan, I. Riza, dan I. Maulana, "Analisis sentimen ulasan pada aplikasi e-commerce dengan menggunakan algoritma Naïve Bayes," Jurnal Applied Informatics and Computing, vol. 6, no. 2, hlm. 220–225, 2022.

[3] K. S. Putri, I. R. Setiawan, dan A. Pambudi, "Analisis sentimen terhadap brand skincare lokal menggunakan Naïve Bayes classifier," Technologia: Jurnal Ilmiah, vol. 14, no. 3, hlm. 227–233, 2023.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1369-1378
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1315

[4] F. A. Larasati, D. E. Ratnawati, dan B. T. Hanggara, "Analisis sentimen ulasan aplikasi Dana dengan metode Random Forest," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 6, no. 9, hlm. 4305–4313, 2022.

[5] C. G. Indrayanto, D. E. Ratnawati, dan B. Rahayudi, "Analisis sentimen data ulasan pengguna aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan metode Random Forest," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 7, no. 3, hlm. 1131–1139, 2023.

[6] [nama penulis dan DOI dilengkapi dari Mendeley], "Perbandingan algoritma Naïve Bayes dan Random Forest dalam klasifikasi sentimen ulasan pengguna Kredivo di Play Store," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 13, no. 2, hlm. 297–308, 2026.

[7] F. F. Wati, Suleman, dan A. E. Widodo, "Analisis sentimen ulasan pengguna aplikasi DeepSeek menggunakan algoritma Random Forest dan Naïve Bayes," CONTEN: Computer and Network Technology, vol. 5, no. 1, hlm. 8–15, 2025, doi: 10.31294/hqpha267.

[8] T. D. Putra dan D. Oktafiani, "Klasifikasi sentimen postingan sosial media menggunakan machine learning Random Forest dan Naïve Bayes," Innovative: Journal of Social Science Research, vol. 5, no. 1, hlm. 2338–2347, 2025, doi: 10.31004/innovative.v5i1.17935.

[9] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, dan H. Pratiwi, "Perbandingan algoritma Random Forest, Naïve Bayes, dan Support Vector Machine pada analisis sentimen Twitter mengenai opini masyarakat terhadap penghapusan tenaga honorer," dalam Prosiding Seminar Nasional Official Statistics, vol. 2022, no. 1, hlm. 563–572, 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.

[10] A. Syah, F. Nurdiyansyah, dan A. Y. Rahman, "Analisis sentimen aplikasi Shopee, Tokopedia, Lazada dan Blibli menggunakan leksikon dan Random Forest," Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 3S1, hlm. 3576–3587, 2024, doi: 10.23960/jitet.v12i3S1.5155.

[11] Amazon, "Amazon Customer Reviews Dataset (produk Fire HD 7)," [dataset daring]. Tersedia: [URL sumber dilengkapi dari Mendeley] [Diakses: 2026].

[12] S. A. H. Bahtiar, C. K. Dewa, dan A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in sentiment analysis on marketplace reviews using rating-based labeling," Journal of Information Systems and Informatics, vol. 5, no. 3, hlm. 915–927, 2023, doi: 10.51519/journalisi.v5i3.539.

[13] J. Iskandar dan Y. Nataliani, "Perbandingan Naive Bayes, SVM, dan K-NN untuk analisis sentimen gadget berbasis aspek," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 6, hlm. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.

[14] G. Kanugrahan, V. H. C. Putra, dan Y. Ramdhani, "Analisis sentimen aplikasi Gojek menggunakan SVM, Random Forest dan Decision Tree," Jurnal Infortech, vol. 6, no. 2, hlm. 171–178, 2024, doi: 10.31294/infortech.v6i2.24594.

[15] P. Nandwani dan R. Verma, "A review on sentiment analysis and emotion detection from text," Social Network Analysis and Mining, vol. 11, no. 1, art. 81, 2021, doi: 10.1007/s13278-021-00776-6.