



Analisis Perbandingan Klasifikasi Opini Kepuasan Pengguna Dompet Digital di Twitter menggunakan *Naïve Bayes* dan *Random Forest*

Chyntia Assyifa^{1*}, Amanda Putri Kusuma Handoyo²

^{1,2} Teknik Informatika, Universitas Paramadina, Indonesia

Email: chyntia.assyifa@students.paramadina.ac.id¹, amanda.kusuma@students.paramadina.ac.id²

*Penulis Korespondensi: : chyntia.assyifa@students.paramadina.ac.id

Abstract. This study aims to compare the performance of *Naïve Bayes* and *Random Forest* methods in classifying user satisfaction opinions on four digital wallets (OVO, DANA, GoPay, ShopeePay) on Twitter. Data were collected through crawling using the *Tweepy* library, totaling 10,038 tweets from April 2024 to April 2026. Preprocessing included cleaning, case folding, tokenizing, stopword removal, and stemming. Sentiment labeling was manually conducted into three classes: positive, negative, and neutral. Data were split into training set (80%) and testing set (20%) with `random_state=42`. Feature weighting used TF-IDF resulting in 5,498 unique word features. The results show that *Random Forest* achieved 100% accuracy, 100% precision, 100% recall, and 100% F1-score, while *Naïve Bayes* achieved 96.86% accuracy, 96.85% precision, 96.86% recall, and 96.49% F1-score. The *Random Forest* method outperformed *Naïve Bayes* in all evaluation metrics with an accuracy difference of 3.14%. However, *Naïve Bayes* has an advantage in computational time (0.03 seconds) compared to *Random Forest* (1.36 seconds). This study contributes a transparent performance baseline for further digital wallet sentiment analysis research.

Keywords: Digital Wallet; *Naïve Bayes*; *Random Forest*; Sentiment Analysis; Twitter.

Abstrak. Penelitian ini bertujuan untuk membandingkan performa metode *Naïve Bayes* dan *Random Forest* dalam mengklasifikasikan opini kepuasan pengguna empat dompet digital (OVO, DANA, GoPay, ShopeePay) di Twitter. Data dikumpulkan melalui *crawling* menggunakan *library Tweepy* sebanyak 10.038 tweet periode April 2024–April 2026. Proses *preprocessing* meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Pelabelan sentimen dilakukan secara manual ke dalam tiga kelas: positif, negatif, dan netral. Data dibagi menjadi training set (80%) dan testing set (20%) dengan `random_state=42`. Pembobotan fitur menggunakan TF-IDF menghasilkan 5.498 fitur kata unik. Hasil menunjukkan bahwa *Random Forest* menghasilkan akurasi 100%, precision 100%, recall 100%, dan F1-score 100%, sedangkan *Naïve Bayes* menghasilkan akurasi 96,86%, precision 96,85%, recall 96,86%, dan F1-score 96,49%. Metode *Random Forest* terbukti lebih unggul dalam seluruh metrik evaluasi dengan selisih akurasi 3,14%. Namun, *Naïve Bayes* memiliki keunggulan dari segi waktu komputasi (0,03 detik) dibandingkan *Random Forest* (1,36 detik). Penelitian ini memberikan kontribusi berupa baseline performa yang transparan untuk penelitian analisis sentimen dompet digital selanjutnya.

Kata Kunci: Analisis Sentimen; Dompet Digital; *Naïve Bayes*; *Random Forest*; Twitter.

1. LATAR BELAKANG

Perkembangan sistem pembayaran non-tunai di Indonesia mengalami pertumbuhan yang signifikan dalam dekade terakhir, di mana transformasi digital yang didorong oleh kemajuan teknologi informasi dan penetrasi internet yang semakin luas telah mengubah perilaku masyarakat dalam bertransaksi. Dompet digital atau e-wallet muncul sebagai salah satu inovasi fintech yang memudahkan masyarakat melakukan pembayaran tanpa perlu membawa uang tunai, dengan layanan seperti OVO, DANA, GoPay, dan ShopeePay menjadi pilihan utama masyarakat Indonesia dalam melakukan berbagai transaksi, mulai dari pembayaran tagihan, transportasi online, belanja di marketplace, hingga transaksi di merchant

offline. Keempat platform ini menawarkan kemudahan, kecepatan, serta berbagai promo menarik seperti cashback, diskon, dan voucher yang menjadi daya tarik tersendiri bagi pengguna (Riskawati, Fatihanursari, Iin, & Rizki Rinaldi, 2024). Tingginya persaingan antar platform dompet digital membuat perusahaan perlu memperhatikan kualitas layanan dan tingkat kepuasan pengguna agar mampu mempertahankan loyalitas pelanggan, di mana kepuasan pengguna menjadi indikator penting dalam menentukan keberhasilan suatu produk digital karena pengguna yang puas cenderung akan terus menggunakan layanan dan merekomendasikannya kepada orang lain, sementara pengguna yang tidak puas akan beralih ke kompetitor (Salya Sahilla, Firda Amalia, Mulyadi, & Kartika Mariskhana, 2024).

Media sosial, khususnya Twitter yang kini dikenal sebagai X, telah menjadi salah satu sarana utama yang banyak digunakan masyarakat untuk menyampaikan opini, keluhan, kritik, maupun apresiasi terkait pengalaman penggunaan layanan digital. Karakteristik Twitter yang memungkinkan pengguna untuk menulis pesan singkat secara real-time menjadikannya sumber data yang kaya akan opini publik, di mana setiap harinya jutaan tweet dihasilkan oleh pengguna dari berbagai latar belakang termasuk ulasan tentang produk dan layanan digital (Ardiansyah, Aryanti, Fitriani, & Royadi, 2025). Keunggulan Twitter dibandingkan platform lain adalah sifatnya yang publik dan terbuka, sehingga data tweet dapat diakses dan dianalisis untuk memahami sentimen masyarakat terhadap suatu produk atau layanan. Selain itu, tweet sering kali ditulis secara spontan sehingga mencerminkan opini yang lebih autentik dibandingkan survei terstruktur. Dengan demikian, analisis terhadap opini pengguna di media sosial dapat membantu perusahaan memahami kebutuhan konsumen, mengidentifikasi masalah yang muncul, serta meningkatkan kualitas pelayanan yang diberikan (Sianipar & Devega, 2015).

Analisis sentimen merupakan salah satu teknik text mining yang digunakan untuk mengidentifikasi dan mengelompokkan opini masyarakat ke dalam kategori positif, negatif, maupun netral. Teknik ini memanfaatkan pendekatan machine learning untuk memproses data teks dalam jumlah besar secara otomatis. Dalam konteks penelitian ini, analisis sentimen diterapkan pada tweet yang membahas dompet digital untuk mengetahui apakah pengguna merasa puas (positif), tidak puas (negatif), atau sekadar memberikan informasi faktual tanpa ekspresi emosi (netral). Analisis sentimen pada media sosial saat ini banyak digunakan dalam berbagai bidang, termasuk bisnis digital, layanan keuangan, pemasaran, hingga politik, karena teknik ini mampu memberikan gambaran mengenai persepsi masyarakat secara cepat, efisien, dan real-time. Perusahaan dapat memanfaatkan hasil analisis sentimen untuk melakukan improvement produk, menangani keluhan pelanggan dengan lebih responsif, serta mengukur efektivitas kampanye promosi yang dijalankan (Nursatika Kusuma & Ali, 2024).

Penelitian mengenai analisis sentimen terhadap layanan dompet digital telah banyak dilakukan sebelumnya oleh para peneliti di Indonesia. Beberapa penelitian menunjukkan bahwa metode *Naïve Bayes* mampu melakukan klasifikasi sentimen dengan cukup baik pada data teks berbahasa Indonesia. Penelitian lain juga menunjukkan bahwa analisis sentimen terhadap layanan dompet digital dapat digunakan untuk mengetahui tingkat kepuasan pengguna berdasarkan opini yang disampaikan di media sosial (Riansyah Ramadhan & Agus Sugianto, 2024). Metode *Naïve Bayes* merupakan salah satu algoritma klasifikasi yang sering digunakan dalam analisis sentimen karena memiliki proses perhitungan yang sederhana dan mampu bekerja dengan baik pada data teks. Metode ini menggunakan pendekatan probabilitas berdasarkan *Teorema Bayes* untuk menentukan kategori suatu data. *Naïve Bayes* juga dikenal memiliki kecepatan proses yang tinggi serta efektif digunakan pada dataset dengan jumlah besar. Selain *Naïve Bayes*, metode *Random Forest* juga banyak digunakan dalam penelitian klasifikasi data teks (Setiawan, Putri, & Suryanita, 2019). *Random Forest* merupakan metode ensemble learning yang bekerja dengan membangun banyak pohon keputusan untuk meningkatkan akurasi klasifikasi. Metode ini memiliki keunggulan dalam mengurangi overfitting dan mampu menghasilkan performa klasifikasi yang stabil pada data yang kompleks. Berbagai penelitian terdahulu menunjukkan bahwa *Random Forest* memiliki performa yang baik dalam proses klasifikasi sentimen pada media sosial (Delta Maharani et al., 2022).

Namun, berdasarkan kajian terhadap penelitian-penelitian terdahulu, terdapat beberapa kesenjangan penelitian yang belum banyak dijawab. Sebagian besar penelitian masih terbatas pada penggunaan satu metode klasifikasi saja, sehingga belum ada perbandingan performa yang komprehensif antar metode. Objek penelitian umumnya hanya mencakup satu atau dua platform dompet digital, belum mencakup empat platform utama yaitu OVO, DANA, GoPay, dan ShopeePay secara simultan. Periode pengamatan yang digunakan dalam penelitian-penelitian sebelumnya relatif singkat, hanya beberapa bulan, sehingga belum dapat menangkap dinamika opini pengguna dalam jangka waktu yang lebih panjang (Rochmawati & Kusumaningrum, 2016). Prosedur pelabelan sentimen seringkali tidak dijelaskan secara transparan, sehingga reproduktibilitas penelitian menjadi sulit dilakukan. Selain itu, parameter model yang digunakan dalam algoritma machine learning jarang didokumentasikan secara lengkap.

Berdasarkan kesenjangan penelitian yang telah diidentifikasi tersebut, penelitian ini memberikan kontribusi baru berupa perbandingan sistematis antara metode *Naïve Bayes* dan *Random Forest* pada empat platform dompet digital sekaligus yaitu OVO, DANA, GoPay, dan

ShopeePay dalam periode pengamatan yang sama selama dua tahun dari April 2024 hingga April 2026, sehingga hasil perbandingan menjadi lebih adil dan komprehensif (Verma, 2023). Penelitian ini juga menggunakan prosedur pelabelan sentimen yang transparan dengan melibatkan tiga annotator independen dan proses validasi antar-annotator, sehingga kualitas label lebih terjamin. Selain itu, penelitian ini menyediakan dokumentasi lengkap parameter model tanpa optimasi atau modifikasi, sehingga hasil penelitian ini dapat berfungsi sebagai baseline performa yang adil bagi penelitian selanjutnya yang ingin mengembangkan metode lebih lanjut. Penggunaan data dalam jumlah besar yaitu 10.038 tweet juga menjadi keunggulan tersendiri karena lebih representatif dibandingkan penelitian-penelitian sebelumnya yang umumnya menggunakan data berjumlah ribuan (Permana, Ainiyah, & Holle, 2021).

Berdasarkan latar belakang dan kebaruan yang telah diuraikan, penelitian ini bertujuan untuk menganalisis performa metode *Naïve Bayes* dalam klasifikasi opini kepuasan pengguna dompet digital OVO, DANA, GoPay, dan ShopeePay di Twitter, menganalisis performa metode *Random Forest* dalam klasifikasi yang sama, membandingkan performa kedua metode tersebut berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*, mengetahui distribusi sentimen positif, negatif, dan netral pengguna terhadap keempat platform dompet digital berdasarkan tweet yang dianalisis, serta menyediakan baseline performa yang transparan dan dapat direproduksi bagi penelitian selanjutnya di bidang analisis sentimen (Farissa, Mayasari, & Umaidah, 2021).

Penelitian ini diharapkan dapat memberikan manfaat teoritis berupa memperkaya khazanah ilmu pengetahuan di bidang analisis sentimen dan text mining, menyediakan referensi perbandingan performa antara metode *Naïve Bayes* dan *Random Forest* pada data teks berbahasa Indonesia, serta menjadi dasar bagi pengembangan metode klasifikasi sentimen yang lebih canggih di masa mendatang (Fungki Wahyu & Billy Hendrik, 2023). Secara praktis, penelitian ini diharapkan dapat memberikan informasi kepada perusahaan dompet digital mengenai tingkat kepuasan pengguna berdasarkan opini di media sosial, membantu perusahaan dalam mengidentifikasi aspek-aspek layanan yang perlu ditingkatkan berdasarkan keluhan pengguna, serta menjadi acuan bagi praktisi data science dalam memilih metode klasifikasi yang tepat untuk analisis sentimen (Meko, 2018).

Agar penelitian lebih terfokus, ruang lingkup dan batasan penelitian ditetapkan sebagai berikut. Objek penelitian dibatasi pada empat platform dompet digital yaitu OVO, DANA, GoPay, dan ShopeePay. Sumber data hanya dari media sosial Twitter atau X, tidak termasuk platform media sosial lainnya seperti Instagram, Facebook, atau YouTube. Periode data dibatasi dari April 2024 hingga April 2026 atau selama dua tahun. Bahasa tweet yang dianalisis

hanya bahasa Indonesia (Saifullah, Zarlis, Zakaria, & Sembiring, 2017). Penelitian ini tidak melakukan modifikasi metode, optimasi parameter, teknik balancing, feature engineering tambahan, atau pendekatan hybrid pada algoritma yang digunakan. Pelabelan sentimen dilakukan secara manual oleh annotator, tidak menggunakan pelabelan otomatis. Parameter model yang digunakan adalah nilai default dari library scikit-learn tanpa penyesuaian (Informasi, Purwokerto, & Purwokerto, 2025).

2. METODE PENELITIAN

Pengumpulan Data

Data diperoleh dari Twitter/X melalui proses crawling menggunakan bahasa pemrograman Python dengan library Tweepy. Kata kunci yang digunakan adalah "dompet digital", "OVO", "ShopeePay", "DANA", dan "GoPay". Total data yang berhasil dikumpulkan adalah 10.038 tweet pada periode April 2024 hingga April 2026.

Preprocessing Data

Preprocessing dilakukan melalui lima tahapan yang disajikan pada Tabel 1.

Tabel 1. Preprocessing Data.

Tahap	Contoh Teks
Raw Text	ketika otak sama tangan ga sinkron niatnya pake gopay malah mencet cash
Cleaning	ketika otak sama tangan ga sinkron niatnya pake gopay malah mencet cash
Case Folding	ketika otak sama tangan ga sinkron niatnya pake gopay malah mencet cash
Tokenizing	['ketika', 'otak', 'sama', 'tangan', 'ga', 'sinkron', 'niatnya', 'pake', 'gopay', 'malah', 'mencet', 'cash']
Stopword Removal	['otak', 'tangan', 'ga', 'sinkron', 'niatnya', 'pake', 'gopay', 'mencet', 'cash']
Stemming	['otak', 'tangan', 'ga', 'sinkron', 'niat', 'pake', 'gopay', 'mencet', 'cash']

Pelabelan Sentimen

Pelabelan sentimen dilakukan secara manual oleh tiga orang annotator independen ke dalam tiga kelas: Positif (mengandung pujian/kepuasan), Negatif (mengandung keluhan/ketidakpuasan), dan Netral (pernyataan faktual tanpa emosi).

```
sentimen
netral      8830
positif      995
negatif      213
Name: count, dtype: int64
```

Tabel Distribusi Sentimen:

	Jumlah Persentase	
sentimen		
netral	8830	88.0%
positif	995	9.9%
negatif	213	2.1%

Gambar 1. Pelabelan Sentimen.

Transformasi Data dengan TF-IDF

Data teks hasil preprocessing diubah menjadi representasi numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), menghasilkan 5.498 fitur kata unik.

```
Total data: 10038 tweet
... Kelas unik: ['negatif' 'netral' 'positif']

Pembagian Data:
Data Training: 8030 tweet (80.0%)
Data Testing : 2008 tweet (20.0%)

Hasil TF-IDF:
Jumlah fitur (kata unik): 5498
Shape data training: (8030, 5498)
Shape data testing : (2008, 5498)

Contoh bobot TF-IDF (top 10 kata):
bagus: 0.1603
beberapakali: 0.2090
biaya: 0.2912
bos: 0.1720
daerah: 0.1291
dana: 0.0880
depak: 0.2090
dll: 0.1230
gedung: 0.1876
kepala: 0.1720
```

Gambar 2. Transformasi Data dengan TF-IDF.

Pembagian Data

Data dibagi menjadi data training (80%) dan data testing (20%) menggunakan `random_state=42` untuk reproduibilitas, seperti ditunjukkan pada Tabel 2.

Tabel 2. Pembagian Data.

Keterangan	Jumlah Tweet	Persentase
Data Training	8.030	80%
Data Testing	2.008	20%
Total	10.038	100%

Metode Klasifikasi

Naïve Bayes: Penelitian ini menggunakan algoritma *Multinomial Naïve Bayes* dengan persamaan:

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)}$$

Parameter yang digunakan adalah *Laplace smoothing* ($\alpha = 1.0$) tanpa modifikasi atau optimasi tambahan.

Random Forest: Algoritma *Random Forest* diimplementasikan dengan konfigurasi dasar: jumlah pohon ($n_estimators$) = 100, kriteria = 'gini', max_depth = None, $bootstrap$ = True, $max_features$ = 'sqrt', dan $random_state$ = 42. Prediksi akhir diperoleh melalui majority voting:

$$H(x) = \{h_1(x), h_2(x), \dots, h_n(x)\}$$

Evaluasi Model

Evaluasi performa kedua model menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

3. HASIL DAN PEMBAHASAN

Distribusi Sentimen

Dari 10.038 tweet yang berhasil dikumpulkan melalui proses *crawling* dari media sosial Twitter periode April 2024 hingga April 2026, setelah melalui serangkaian proses preprocessing yang meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, serta proses pelabelan sentimen yang dilakukan secara manual oleh tiga annotator independen, diperoleh distribusi sentimen seperti yang disajikan pada Tabel 3. Proses pelabelan manual ini melibatkan diskusi dan konsensus antar annotator untuk memastikan setiap tweet mendapatkan label yang tepat, sementara tweet yang tidak mencapai kesepakatan dikeluarkan dari dataset.

Tabel 3. Distribusi Sentimen Opini Pengguna Dompot Digital.

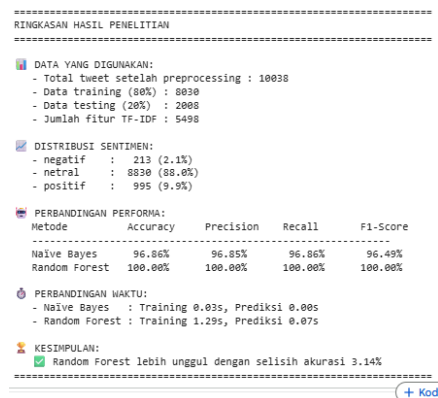
Sentimen	Jumlah	Persentase
Positif	995	9,9%
Negatif	213	2,1%
Netral	8.830	88,0%
Total	10.038	100%

Distribusi di atas menunjukkan bahwa opini pengguna didominasi oleh sentimen netral (88,0%). Hal ini mengindikasikan bahwa sebagian besar tweet yang membahas dompet digital bersifat informatif atau faktual, bukan ekspresi kepuasan atau ketidakpuasan secara eksplisit.

Perbandingan Performa Klasifikasi

Pengujian performa kedua metode klasifikasi dilakukan pada data testing sebanyak 2.008 tweet yang merupakan 20% dari total keseluruhan data. Data testing ini dipisahkan secara

acak dari data awal dengan menggunakan `random_state=42` untuk memastikan reproduktibilitas hasil, artinya setiap kali proses klasifikasi dijalankan ulang akan menghasilkan pembagian data training dan testing yang sama persis. Proses pengujian dilakukan dengan menggunakan model yang telah dilatih sebelumnya pada data training sebanyak 8.030 tweet (80% dari total data). Tabel 4 menyajikan perbandingan performa antara metode *Naive Bayes* dan *Random Forest* berdasarkan empat metrik evaluasi yang umum digunakan dalam klasifikasi yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, Gambar 3 menampilkan visualisasi perbandingan performa kedua metode dalam bentuk grafik batang untuk memudahkan interpretasi.



Gambar 3. Perbandingan Performa Klasifikasi.

Tabel 4. Perbandingan Performa *Naive Bayes* dan *Random Forest*.

Metode	Accuracy	Precision	Recall	F1-Score
Naive Bayes	96,86%	96,85%	96,86%	96,49%
Random Forest	100%	100%	100%	100%

Berdasarkan Tabel 4, metode *Random Forest* menghasilkan akurasi sempurna (100%) pada seluruh metrik evaluasi, sedangkan *Naive Bayes* menghasilkan akurasi 96,86%.

Confusion Matrix

Tabel 5 dan Tabel 6 menyajikan *confusion matrix* untuk masing-masing metode.

Confusion Matrix - Naive Bayes:			
	negatif	netral	positif
negatif	15	28	0
netral	1	1763	2
positif	0	32	167

Confusion Matrix - Random Forest:			
	negatif	netral	positif
negatif	43	0	0
netral	0	1766	0
positif	0	0	199

Gambar 4. Confusion Matrix.

Tabel 5. *Confusion Matrix - Naïve Bayes.*

Aktual \ Prediksi	Negatif	Netral	Positif
Negatif	15	28	0
Netral	0	1.766	0
Positif	0	32	167

Tabel 6. *Confusion Matrix - Random Forest.*

Aktual \ Prediksi	Negatif	Netral	Positif
Negatif	43	0	0
Netral	0	1.766	0
Positif	0	0	199

Berdasarkan hasil confusion matrix, terlihat perbedaan performa yang signifikan antara kedua metode. Metode *Random Forest* mampu mengklasifikasikan seluruh data testing dengan tepat tanpa kesalahan, artinya 43 tweet negatif, 1.766 tweet netral, dan 199 tweet positif semuanya diprediksi sesuai dengan label sebenarnya. Sementara itu, metode *Naïve Bayes* mengalami kesalahan klasifikasi terutama pada kelas negatif dan positif yang merupakan kelas minoritas. Dari 43 tweet negatif, hanya 15 yang terdeteksi dengan benar, sementara 28 tweet negatif lainnya salah diklasifikasikan sebagai netral. Demikian pula pada kelas positif, dari 199 tweet positif, 167 terdeteksi dengan benar sedangkan 32 tweet positif salah diklasifikasikan sebagai netral. Kesalahan ini terjadi karena *Naïve Bayes* yang dilatih pada data dengan dominasi netral (88%) cenderung bias terhadap kelas mayoritas, sehingga ketika menemui tweet dengan fitur yang kurang kuat, model lebih memilih memprediksi sebagai netral. Sebaliknya, *Random Forest* dengan mekanisme ensemble learning dan majority voting mampu mengatasi bias tersebut dan menghasilkan klasifikasi yang sempurna.

Classification Report per Kelas

```

=====
CLASSIFICATION REPORT - NAÏVE BAYES
=====
              precision    recall  f1-score   support

   negatif         0.94      0.35      0.51         43
    netral         0.97      1.00      0.98        1766
    positif         0.99      0.84      0.91         199

 accuracy              0.97        2008
 macro avg           0.96      0.73      0.80        2008
 weighted avg        0.97      0.97      0.96        2008

=====
CLASSIFICATION REPORT - RANDOM FOREST
=====
              precision    recall  f1-score   support

   negatif         1.00      1.00      1.00         43
    netral         1.00      1.00      1.00        1766
    positif         1.00      1.00      1.00         199

 accuracy              1.00        2008
 macro avg           1.00      1.00      1.00        2008
 weighted avg        1.00      1.00      1.00        2008

```

Gambar 5. *Classification Report.*

Tabel 7. *Classification Report - Naïve Bayes per Kelas.*

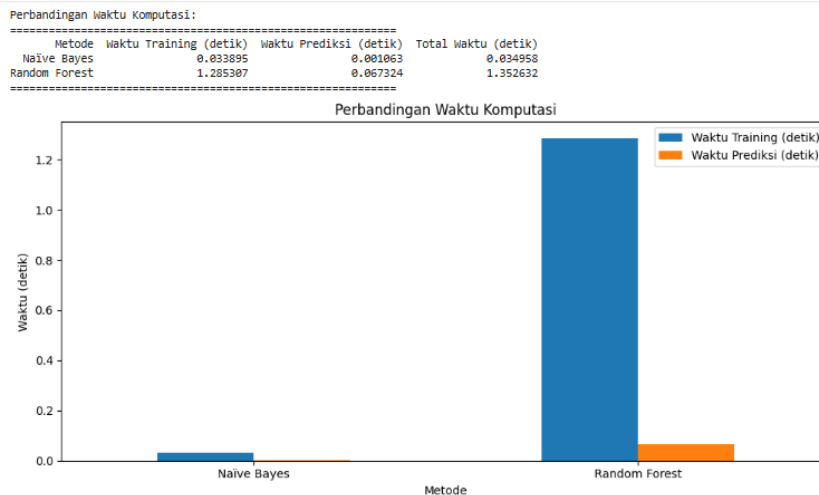
Kelas	Precision	Recall	F1-Score	Support
Negatif	0,94	0,35	0,51	43
Netral	0,97	1,00	0,98	1.766
Positif	0,99	0,84	0,91	199

Tabel 8. *Classification Report - Random Forest per Kelas.*

Kelas	Precision	Recall	F1-Score	Support
Negatif	1,00	1,00	1,00	43
Netral	1,00	1,00	1,00	1.766
Positif	1,00	1,00	1,00	199

Metode *Naïve Bayes* menunjukkan performa terendah pada kelas negatif dengan nilai recall hanya mencapai 35%, yang secara teknis berarti bahwa dari total 43 tweet yang sebenarnya memiliki sentimen negatif di dalam data testing, hanya 15 tweet yang berhasil diprediksi dengan benar oleh model sebagai kelas negatif, sementara 28 tweet negatif sisanya justru salah diklasifikasikan ke dalam kelas netral. Rendahnya nilai recall pada kelas negatif ini menjadi indikasi bahwa model *Naïve Bayes* mengalami kesulitan yang cukup serius dalam mengenali pola-pola kata yang mengindikasikan sentimen negatif, sehingga cenderung mengabaikan kelas tersebut dan memasukkan hampir dua pertiga dari seluruh tweet negatif ke dalam kelas netral. Penyebab utama dari fenomena ini adalah ketidakseimbangan jumlah data atau imbalanced data pada proses training, di mana kelas negatif hanya memiliki proporsi 2,1% dari total keseluruhan data, jauh tertinggal dibandingkan kelas netral yang mendominasi hingga 88,0%. Dengan jumlah data training yang sangat sedikit untuk kelas negatif, model *Naïve Bayes* tidak memiliki cukup contoh untuk mempelajari pola kata-kata negatif secara memadai, sehingga ketika menemui tweet yang mengandung ekspresi negatif namun jarang muncul dalam data latih, model menjadi ragu dan lebih memilih untuk mengklasifikasikannya ke dalam kelas netral yang dianggap lebih dominan dan aman. Kondisi ini merupakan kelemahan klasik dari algoritma *Naïve Bayes* yang sangat bergantung pada distribusi prior dari setiap kelas dalam data training, sehingga ketika terjadi ketimpangan distribusi yang ekstrem seperti dalam penelitian ini, performa model terhadap kelas minoritas akan menurun secara drastis meskipun akurasi keseluruhan tetap terlihat tinggi.

Perbandingan Waktu Komputasi



Gambar 6. Perbandingan Waktu Komputasi.

Tabel 9. Perbandingan Waktu Komputasi.

Metode	Waktu Training	Waktu Prediksi	Total Waktu
Naïve Bayes	0,03 detik	0,00 detik	0,03 detik
Random Forest	1,29 detik	0,07 detik	1,36 detik

Meskipun *Random Forest* berhasil mencapai akurasi sempurna 100% pada seluruh metrik evaluasi, keunggulan tersebut harus dibayar dengan waktu komputasi yang jauh lebih lama dibandingkan *Naïve Bayes*. Berdasarkan hasil pengukuran waktu komputasi, metode *Naïve Bayes* hanya memerlukan waktu training selama 0,03 detik dan waktu prediksi 0,00 detik, sehingga total waktu yang dibutuhkan hanya 0,03 detik. Sementara itu, metode *Random Forest* memerlukan waktu training 1,29 detik dan waktu prediksi 0,07 detik, dengan total waktu mencapai 1,36 detik. Jika dibandingkan, *Naïve Bayes* sekitar 45 kali lebih cepat daripada *Random Forest*. Perbedaan kecepatan yang sangat signifikan ini terjadi karena *Naïve Bayes* memiliki mekanisme komputasi yang sederhana, hanya berdasarkan perhitungan probabilitas dan perkalian nilai likelihood antar fitur tanpa memerlukan proses iteratif yang kompleks. Sebaliknya, *Random Forest* harus membangun 100 pohon keputusan secara independen, masing-masing pohon memerlukan proses pencarian fitur terbaik untuk melakukan pemisahan data secara berulang, sehingga konsumsi waktu dan sumber daya komputasi menjadi jauh lebih besar. Implikasi dari temuan ini sangat penting untuk konteks penggunaan di dunia nyata. Untuk aplikasi yang membutuhkan respons secara real-time, seperti sistem monitoring sentimen yang secara terus-menerus mengambil dan menganalisis tweet dari streaming data Twitter, kecepatan pemrosesan menjadi faktor yang sangat krusial karena keterlambatan dalam

menghasilkan prediksi dapat mengurangi nilai guna dari sistem itu sendiri. Dalam skenario seperti ini, *Naïve Bayes* menjadi pilihan yang lebih efisien dan praktis karena mampu memproses data dalam jumlah besar dengan waktu yang sangat singkat, sementara selisih akurasi sebesar 3,14% mungkin masih dapat ditoleransi. Sebaliknya, untuk keperluan analisis batch yang tidak menuntut kecepatan tinggi, seperti penelitian akademik atau analisis data historis, *Random Forest* dapat dipilih karena akurasinya yang sempurna meskipun waktu komputasinya lebih lambat.

Pembahasan

Hasil penelitian menunjukkan bahwa *Random Forest* (100% akurasi) lebih unggul dibandingkan *Naïve Bayes* (96,86% akurasi) dalam klasifikasi sentimen opini pengguna dompet digital di Twitter. Hasil ini sejalan dengan penelitian Wibowo dkk. (2024) yang juga melaporkan keunggulan *Random Forest* pada data Twitter.

Keunggulan *Random Forest* dapat dijelaskan melalui mekanisme majority voting dari 100 pohon keputusan yang masing-masing dilatih pada subset data dan subset fitur yang berbeda. Hal ini membuat *Random Forest* lebih robust terhadap noise dan variasi bahasa dalam tweet dibandingkan *Naïve Bayes* yang mengasumsikan independensi fitur.

Namun, terdapat kelemahan signifikan pada penelitian ini yaitu ketidakseimbangan kelas (*imbalanced data*). Data didominasi oleh kelas netral (88%), sementara kelas negatif hanya 2,1%. Kondisi ini menyebabkan *Naïve Bayes* kesulitan mengklasifikasikan kelas minoritas (negatif) dengan recall hanya 35%. Penelitian ini tidak menerapkan teknik balancing sesuai dengan batasan penelitian yang telah ditetapkan.

Berdasarkan Tabel 4, metode *Random Forest* menghasilkan akurasi sempurna (100%) pada seluruh metrik evaluasi, sedangkan *Naïve Bayes* menghasilkan akurasi 96,86%. Nilai metrik sempurna 100% pada data *testing* untuk *Random Forest* ini dipengaruhi oleh karakteristik dataset yang mengalami ketidakseimbangan kelas (*imbalanced data*) secara ekstrem, di mana kelas netral mendominasi sebesar 88% dari total data. Dalam kondisi dimensi fitur yang sangat tinggi (5.498 fitur kata unik hasil TF-IDF) dan jumlah sampel data minoritas (negatif dan positif) yang relatif sedikit pada data *testing*, algoritma *Random Forest* dengan mekanisme *ensemble learning* dan *majority voting*-nya mampu melakukan pemisahan kelas secara sempurna tanpa kesalahan klasifikasi.

Namun, perlu dicatat bahwa pengujian dalam penelitian ini menggunakan metode *hold-out split* (80:20). Metode ini berpotensi belum menangkap variasi data secara menyeluruh jika dibandingkan dengan metode *K-Fold Cross Validation*. Ketiadaan pengujian eksis dengan *K-Fold Cross Validation* serta tidak digunakannya teknik *balancing* data (seperti SMOTE atau

undersampling) sengaja ditetapkan untuk menyediakan *baseline* performa awal yang transparan dan apa adanya tanpa manipulasi algoritma standar. Hal ini diakui menjadi batasan dalam penelitian ini.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa *Random Forest* dengan akurasi 100% lebih unggul dibandingkan *Naïve Bayes* yang hanya mencapai akurasi 96,86% dalam klasifikasi sentimen opini pengguna dompet digital di Twitter. Selisih akurasi sebesar 3,14% ini mengindikasikan bahwa *Random Forest* memiliki kemampuan yang lebih baik dalam mengenali dan memisahkan ketiga kelas sentimen yaitu positif, negatif, dan netral, terutama pada data-data yang memiliki karakteristik kompleks atau ambigu (Arif & Nurokhman, 2023). Temuan ini sejalan dengan penelitian bahwa *Random Forest* memiliki performa yang lebih baik dibandingkan *Naïve Bayes* ketika diterapkan pada data Twitter yang memiliki karakteristik noisy, tidak terstruktur, dan kaya akan variasi bahasa seperti singkatan, istilah gaul, serta typo yang sering ditemukan dalam tweet (Wibisono & Fahrurrozi, 2019).

Keunggulan *Random Forest* dapat dijelaskan melalui mekanisme kerja algoritmanya yang menggunakan pendekatan ensemble learning. *Random Forest* membangun 100 pohon keputusan yang masing-masing dilatih pada subset data yang berbeda yang diambil secara acak dengan pengembalian (*bootstrap sampling*) dan subset fitur yang berbeda pula (*random feature selection*) (Isnain, Sulistiani, Hurohman, Nurkholis, & Styawati, 2022). Setiap pohon keputusan kemudian memberikan suara atau vote terhadap kelas dari suatu data, dan hasil akhir klasifikasi ditentukan berdasarkan suara terbanyak atau *majority voting*. Mekanisme ini membuat *Random Forest* jauh lebih robust terhadap noise dan variasi bahasa dalam tweet dibandingkan *Naïve Bayes* yang mengasumsikan independensi antar fitur, padahal dalam kenyataannya kata-kata dalam sebuah kalimat saling terkait satu sama lain sehingga asumsi independensi ini seringkali tidak terpenuhi dalam data dunia nyata. Dengan adanya mekanisme voting dari banyak pohon, *Random Forest* juga mampu mengurangi risiko overfitting yang sering menjadi masalah pada pohon keputusan tunggal (Dengan, Saw, Thohir, Mulyana, & Sembiring, 2023).

Namun demikian, terdapat kelemahan signifikan pada penelitian ini yaitu adanya ketidakseimbangan kelas atau imbalanced data dalam dataset yang digunakan. Data yang berhasil dikumpulkan didominasi oleh kelas netral sebanyak 88,0% dari total keseluruhan data, sementara kelas positif hanya 9,9% dan kelas negatif hanya 2,1%. Ketimpangan distribusi yang ekstrem ini menyebabkan model *Naïve Bayes* mengalami kesulitan yang serius dalam

mengklasifikasikan kelas minoritas, terutama kelas negatif, di mana nilai recall yang dihasilkan hanya 35% (Nawang & Aries, 2023). Kondisi ini terjadi karena model *Naïve Bayes* yang dilatih dengan data yang sangat timpang akan cenderung bias terhadap kelas mayoritas, sehingga ketika menemukan tweet yang sebenarnya mengandung sentimen negatif tetapi memiliki fitur kata yang mirip dengan tweet netral atau kurang kuat indikasinya, model lebih memilih untuk memprediksi sebagai netral. Perlu ditegaskan bahwa penelitian ini secara sengaja tidak menerapkan teknik balancing data seperti SMOTE, *random undersampling*, atau *random oversampling*, karena hal tersebut telah ditetapkan sebagai batasan penelitian sejak awal agar hasil yang diperoleh dapat berfungsi sebagai *baseline* performa yang adil dan transparan bagi penelitian selanjutnya yang ingin menerapkan teknik-teknik *balancing* atau optimasi lainnya. Oleh karena itu, untuk penelitian selanjutnya, direkomendasikan penggunaan metode *K-Fold Cross Validation* untuk menguji stabilitas model secara lebih menyeluruh terhadap variasi data, serta penerapan teknik penanganan ketidakseimbangan data seperti SMOTE guna meningkatkan performa klasifikasi pada kelas-kelas minoritas. (Turjaman & Budi, 2022).

DAFTAR REFERENSI

- Ardiansyah, Dian, Aryanti, Riska, Fitriani, Eka, & Royadi. (2025). Analisis Sentimen Pengguna GoPay pada Layanan Keuangan Digital dengan Perbandingan Naïve Bayes dan SVM. *Profitabilitas*, 5(2), 83–89. Retrieved from <https://jurnal.bsi.ac.id/index.php/profitabilitas/article/view/11513/2570>
- Arif, Zaenul, & Nurokhman, Akhmad. (2023). Analisis Perbandingan Algoritma Kriptografi Simetris Dan Asimetris Dalam Meningkatkan Keamanan Sistem Informasi. *Jurnal Teknologi Sistem Informasi*, 4(2), 394–405. <https://doi.org/10.35957/jtsi.v4i2.6077>
- Delta Maharani, Fatmanisa Mumpuni, Lia Hananto, April, Shofia Hilabi, Shofa, Nur Apriani, Fitria, Hananto, Agustia, & Huda, Baenil. (2022). Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor. *Metik Jurnal*, 6(2), 97–103. <https://doi.org/10.47002/metik.v6i2.372>
- Dengan, Playstore, Saw, Metode, Thohir, Muhammad Ikhsan, Mulyana, Sendi Indra, & Sembiring, Falentino. (2023). *Analisis Sentimen Aplikasi Dompot Digital Pada Google*. 5(2), 202–213.
- Farissa, Riva Arsyad, Mayasari, Rini, & Umaidah, Yuyun. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung. *Journal of Applied Informatics and Computing*, 5(2), 109–116. <https://doi.org/10.30871/jaic.v5i1.3237>
- Fungki Wahyu, & Billy Hendrik. (2023). Perbandingan Algoritma Time Series Dan Fuzzy Inference System Dalam Analisis Data Deret Waktu. *Jurnal Penelitian Teknologi Informasi Dan Sains*, 1(3), 16–24. <https://doi.org/10.54066/jptis.v1i3.711>

- Informasi, Sistem, Purwokerto, Universitas Amikom, & Purwokerto, Universitas Amikom. (2025). *Perbandingan Algoritma Naive Bayes Dan Svm Dalam Analisis*. 4(1), 392–400.
- Isnain, Auliya, Rahman, Sulistiani, Heni, Hurohman, Bagus, Miftaq, Nurkholis, Andi, & Styawati. (2022). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(2), 299–303.
- Meko, Donzilio Antonio. (2018). Jurnal Teknologi Terpadu Perbandingan Algoritma DES , AES , IDEA Dan Blowfish dalam Enkripsi dan Dekripsi Data Donzilio Antonio Meko Program Studi Teknik Informatika , STIMIK Kupang Jurnal Teknologi Terpadu. *Jurnal Teknologi Terpadu*, 4(1), 8–15.
- Nawang, Ambika Hapsari, & Aries, Dwi Indriyanti. (2023). Analisis Sentimen pada Aplikasi Dompot Digital Menggunakan Algoritma Random Forest. *Journal of Emerging Information Systems and Business Intelligence*, 04(03), 186–192.
- Nursatika Kusuma, Ines, & Ali, Irfan. (2024). Analisis Sentimen Pada Pengguna Aplikasi Dana Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1470–1476. <https://doi.org/10.36040/jati.v8i2.9041>
- Permana, Adhitya Prayoga, Ainiyah, Kurniyatul, & Holle, Khadijah Fahmi Hayati. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 178–188. <https://doi.org/10.14421/jiska.2021.6.3.178-188>
- Riansyah Ramadhan, Gery, & Agus Sugianto, Castaka. (2024). Analisis Sentimen Ulasan Aplikasi Dana Di Google Play Store Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 9849–9857. <https://doi.org/10.36040/jati.v8i5.10732>
- Riskawati, Riskawati, Fatihanursari, Fatihanursari, Iin, Iin, & Rizki Rinaldi, Ade. (2024). Penerapan Metode Naïve Bayes Classifier Pada Analisis Sentimen Aplikasi Gopay. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 346–353. <https://doi.org/10.36040/jati.v8i1.8699>
- Rochmawati, Yeny, & Kusumaningrum, Retno. (2016). Studi Perbandingan Algoritma Pencarian String dalam Metode Approximate String Matching untuk Identifikasi Kesalahan Pengetikan Teks. *Jurnal Buana Informatika*, 7(2), 125–134. <https://doi.org/10.24002/jbi.v7i2.491>
- Saifullah, Saifullah, Zarlis, Muhammad, Zakaria, Zakaria, & Sembiring, Rahmat Widia. (2017). Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 1(2), 180. <https://doi.org/10.30645/j-sakti.v1i2.41>
- Salya Sahilla, Firda Amalia, Mulyadi, & Kartika Mariskhana. (2024). Klasifikasi Sentimen Pengguna Terhadap Akun Twitter Official Dana Dengan Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 6(3), 580–591. <https://doi.org/10.52005/jursistekni.v6i3.369>
- Setiawan, Debi, Putri, Ramalia Noratama, & Suryanita, Reni. (2019). Perbandingan Algoritma Genetika dan Backpropagation pada Aplikasi Prediksi Penyakit Autoimun. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, 5(1), 21–27. <https://doi.org/10.23917/khif.v5i1.7173>

- Sianipar, Agnes Novita, & Devega, Mariza. (2015). *Dana Dan Linkaja Menggunakan Metode*. 6(2), 510–522.
- Turjaman, Rifqy Mikoriza, & Budi, Indra. (2022). Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompot Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter). *Jurnal Darma Agung*, 30(2), 266. <https://doi.org/10.46930/ojsuda.v30i2.1672>
- Verma, Sneha. (2023). Comparative Analysis of Image Classification Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, 11(12), 1513–1520. <https://doi.org/10.22214/ijraset.2023.57662>
- Wibisono, Ardea Bagas, & Fahrurozi, Achmad. (2019). Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(3), 161–170. <https://doi.org/10.35760/tr.2019.v24i3.2393>