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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Studi Komparatif Algoritma Random Forest dan Logistic Regression dalam Analisis Sentimen Ulasan Aplikasi E-Wallet Dana

Diah Fitriani[1*], Afril Efan Pajri[2], Aprillia Dwi Ardianti[3]

[1] Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
[2] Fakultas Sains dan Teknologi, Sistem Komputer, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
[3] Fakultas Sains dan Teknologi, Teknik Mesin, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
Email: [1,*]diafitria16@gmail.com, [2]afril@unugiri.ac.id, [3]aprilliadwia@unugiri.ac.id
Email Penulis Korespondensi: diafitria16@gmail.com
Submitted **16-01-2026;** Accepted **19-02-2026;** Published **28-02-2026**

Abstrak

Peningkatan penggunaan dompet digital di Indonesia mendorong semakin banyaknya opini yang disampaikan pengguna, termasuk pada platfom DANA di Play Store. Ulasan tersebut mencerminkan pengalaman serta tingkat kepuasan pengguna, sehingga dibutuhkan analisis sentimen dalam memahami pandangan masyarakat terkait kualitas layanan aplikasi. Penelitian melakukan studi komparatif antara algoritma Random Forest dan Logistic Regression dalam mengklasifikasikan sentimen aplikasi DANA. Data diperoleh melalui teknik scraping dan menghasilkan 2.068 ulasan setelah proses pembersihan. Tahapan analisis meliputi preprocessing teks, pelabelan berdasarkan skor ulasan, pembobotan menggunakan TF-IDF, serta pemodelan dengan kedua algoritma tersebut. Hasil evaluasi membuktikan Random Forest mendapatkan akurasi 86.23%, sementara Logistic Regression mendapatkan akurasi 84.54%. Kedua model mampu mengklasifikasikan sentimen dengan baik. Random Forest menunjukan kinerja lebih unggul dibandingkan Logistic Regression dalam tugas analisis sentimen ulasan aplikasi DANA. Sehingga bisa ditarik kesimpulan bahwa penggunaan algoritma Random Forest mampu menghasilkan analisis sentimen yang tepat dan dapat dijadikan acuan dalam pengambilan keputusan di penelitian selanjutnya.

Kata Kunci: Analisis Sentimen; Dana; Google Play Store; Logistic Regression; Random Forest.

Abstract

The increasing use of digital wallets in Indonesia has led to a growing number of user opinions expressed, including on the DANA platform in the Play Store. These reviews reflect users' experiences and satisfaction levels, necessitating sentiment analysis to comprehend public opinions about the application's service quality. The study conducts an analytical comparison between Random Forest and Logistic Regression methods in classification sentiments for DANA application. Data was obtained through scraping techniques, resulting in 2,068 reviews after the cleaning process. The analysis stages include text preprocessing, labeling based on review scores, weighting using TF-IDF, and modeling with both algorithms. The evaluation results demonstrate that Random Forest obtains an accuracy of 86.23%, while Logistic Regression obtains an accuracy of 84.54%. Both models are capable of classifying positive sentiments well but are less optimal in detecting negative sentiments. Random Forest shows higher performance compared to Logistic Regression within the task in sentiment analysis for DANA app reviews. Thus, we can conclude that using the random forest algorithm is able to produce accurate sentiment analysis and can act as a basis for making decisions in further research.

Keywords: Sentiment Analysis; Dana; Google Play Store; Logistic Regression; Random Forest

1. PENDAHULUAN

Kemajuan Teknologi di indonesia terus meningkat, salah satunya ditandai dengan meningkatnya penggunaan dompet digital. Salah satu layanan dompet digital yang banyak digunakan masyarakat adalah DANA. DANA termasuk platform dompet digital yang memberikan kemudahan, kecepatan, serta kenyamanan pada saat melakukan berbagai jenis transaksi [1], [2]. Berdasarkan penelitian yang dilakukan oleh YouGov, peningkatan pengguna dompet digital di Indonesia telah mencapai 87%, di mana sebagian besar penggunanya berasal dari kelompok usia 18 hingga 25 tahun. Namun, semakin bertambah banyaknya pengguna dompet digital, sehingga meningkat pula sebuah opini terhadap aplikasi tersebut, mengingat sebuah aplikasi pasti memiliki kelebihan dan kekurangan, terutama yang berkaitan dengan keamanan transaksi dan kinerja aplikasi. Kondisi ini menunjukkan adanya masalah terkait kepuasan pengguna aplikasi DANA [3].

Salah satu media utama bagi pengguna untuk menyampaikan pendapatnya adalah *Google Play Store*, platform yang memungkinkan pengguna untuk menulis dan membaca ulasan terkait aplikasi [4]. Ulasan ini menunjukkan pengalaman para pengguna sehingga menjadi sumber informasi yang penting untuk memahami opini pengguna, jika opini negatif pengguna tidak diatasi dengan baik, maka akan muncul penurunan rasa percaya terhadap aplikasi yang bisa menghambat peningkatan layanan dompet digital. Sehingga perlu dilakukan analisi sentimen untuk memahami persepsi pengguna [5]. Analisis sentimen adalah metode penggalian data yang mencerminkan sikap individu terhapadap suatu komentar dengan mengkategorikan suatu teks tersebut apakah positif atau negatif.

Dalam proses analisis tersebut, diperlukan algoritma klasifikasi yang mampu mengolah data teks dengan efektif dan efisien. Beberapa algoritma yang digunakan antara lain adalah *Random Forest* dan *Logistic Regression*. *Random Forest* sebuah teknik dalam pembelajaran mesin yang diterapkan untuk tugas klasifikasi. Algoritma tersebut masuk ke dalam kategori pembelajaran *ensemble*, yakni pendekatan yang memadukan beberapa model dari seluruh pohon itu untuk menentukan hasil akhir sampel yang dipilih secara acak [6], [7]. Sedangkan *Logistic Regression* bekerja dengan cara memprediksi probabilitas suatu kelas berdasarkan fungsi logistik yang dapat memberikan akurasi tinggi pada data dengan distribusi yang seimbang [8].

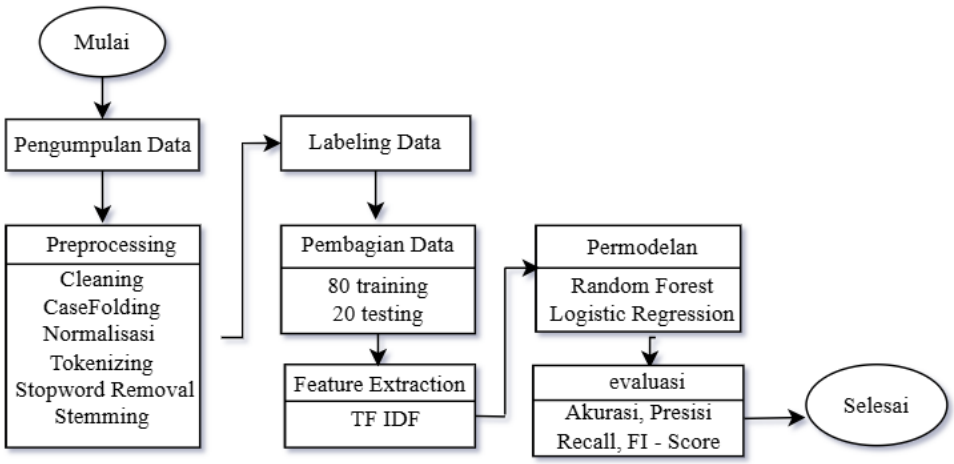
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Riset sebelumnya sudah menerapkan algoritma *Random Forest* dan *Logistic Regression* dalam analisis sentimen. Berdasarkan riset Bahtiar dkk [9]. Dalam jurnal *Sentiment Analysis on Marketplace Application* Reviews menunjukkan bahwa algoritma *Logistic Regression* memperoleh akurasi maksimal sebesar 90%. Selanjutnya riset tentang sentimen terhadap dompet digital di platform Twitter telah dilakukan oleh Hapsari dkk [3]. Dengan pendekatan klasifikasi *Random Forest*, temuan dari studi tersebut mengungkapkan tingkat akurasi sebesar 85,43% untuk Gopay, dan 85,50% untuk Shopeepay. Riset menganalisis komentar masyarakat mengenai kominfo pada pemblokiran situs non pse yang dilakukan oleh Catur Rahmawati dkk [10]. Hasil pengujian menunjukkan bahwa metode *Decision Tree* memperoleh akurasi sebesar 83,8%, *Random Forest* 85,8% dan *Logistic Regression* 85,4%. Riset oleh Rizky Saputra dkk [11]. Menganalisis sentimen masyarakat dalam menghadapi pembangunan (IKN) menggunakan algoritma *Logistic Regression*, dan *Naïve Bayes*. Hasil riset menunjukkan akurasi *Logistic Regression* dan *Naïve Bayes* masing-masing memperoleh akurasi 79%. Riset analisis pengguna i byond bsi pada *play store* dilakukan oleh Firdaus dkk [12]. Dengan metode SEMMA dan pembobotan TF-IDF. Hasil menunjukka bahwa algoritma *Random Forest* memiliki akurasi 90%. Riset n Barus dkk [13]. Akurasi *Random Forest* 91% dan *Logistic Regression* 87% ditemukan dalam analisis opini Tokopedia dengan TF-IDF sebagai ekstraksi fitur dan SMOTE.

Meskipun sudah banyak penelitian mengenai analisis sentimen menggunakan algoritma *Random Forest* dan *Logistic Regression*, penelitian yang secara khusus membahas ulasan pengguna aplikasi DANA di *Google Play Store* masih sangat terbatas. Sebagian besar penelitian sebelumnya menggunakan data dari media sosial seperti Twitter, bukan dari platform tempat pengguna aplikasi memberikan ulasan langsung. Selain itu, penelitian ini bertujuan untuk melakukan studi komparatif penerapan algoritma *Random Forest* dan *Logistic Regression* dalam menganalisis sentimen terhadap ulasan aplikasi DANA. Penelitian ini juga bertujuan untuk mengetahui algoritma yang paling efisien dalam mengklasifikasikan sentimen pada data teks. Dengan menggunakan dataset terbaru hasil *scraping di Play Store*. Melalui penelitian ini diharapkan mampu menyediakan informasi praktis dan terkini bagi pengembang DANA, dan menjadi acuan dalam memilih algoritma yang paling efektif untuk memonitor sentimen pengguna secara *real time.*

2. METODOLOGI PENELITIAN

Riset ini menggunakan pendekatan kuantitatif untuk Analisis Sentimen ulasan pengguna layanan dompet digital DANA di *Google Play Store*. Tahap penelitian tersebut seperti gambar 1, yang meliputi pengumpulan data, preprocessing, pelabelan, pembobotan fitur, pemodelan, serta evaluasi kinerja.



Gambar 1. Tahap Penelitian

2.1 Pengumpulan Data

Riset dimulai dengan pengumpulan informasi, di mana dataset yang diperlukan untuk studi diperoleh melalui pendekatan yang terstruktur dan metodis. Pada langkah ini, peneliti mengambil data dengan teknik *scraping* dari platform *Google Play Store*, yang berfokus pada ulasan pengguna mengenai aplikasi dompet digital DANA. Informasi yang dikumpulkan mencakup elemen seperti *Username*, *Content*, *Score*, *at*.

2.2 *Pre-processing* Data

Tahap *preprocessing* bertujuan untuk membersihkan serta merapikan data sehingga dapat dimanfaatkan secara maksimal pada tahapan analisis atau permodelan. Dalam analisis teks seperti ulasan di *Play Store*. Tahap tahap dilakukan untuk mengubah teks yang tidak tersetruktur menjadi lebih teratur dan mudah di analisis oleh algoritma pemrosesan bahasa alami. Proses ini juga membantu meningkatkan kualitas pada data [14].

a. *Cleaning*

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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Cleaning adalah langkah untuk membersihkan data dari kesalahan data yang berulang, agar data lebih terstruktur dan siap untuk membangun permodelan. Selain itu, pembersihan juga mencakup penghapusan data yang ganda, simbol tidak perlu, serta penyesuaian cara menulis agar memiliki format yang sama [15], [16]. Dengan demikian, hasil analisis atau pemodelan yang dilakukan setelah proses *cleaning* akan menjadi lebih akurat.

b. *Case Folding*

Suatu bagian teknik *preprocessing* adalah *case folding*, yaitu mengubah semua kata. Pada titik ini, sistem mengubah huruf kapital (A–Z) menjadi huruf kecil (a–z), sehingga kata dengan bentuk sama tidak dianggap sebagai entitas berbeda hanya karena adanya perbedaan penulisan [17]. Tujuannya adalah agar hasil analisis teks tidak dipengaruhi oleh perbedaan huruf (A-a).

c. Normalisasi

Normalisasi adalah proses menyamakan bentuk penulisan kata agar konsisten dan sesuai dengan bentuk baku. Tujuannya agar kata yang tidak baku atau singkatan bisa dikenali dengan benar oleh sistem analisis [18]. Dalam analisis teks, terutama pada data ulasan pengguna, sering ditemukan variasi penulisan seperti penggunaan singkatan (‘gk’, ‘tdk’), kata tidak baku (‘nggak’, ‘ga’), atau bentuk penulisan yang tidak konsisten. Melalui tahap normalisasi dapat membantu mengurangi noise pada data dan meningkatkan kualitas hasil analisis.

d. *Tokenizing*

Tahapan ini mencakup pemisahan teks yang panjang menjadi elemen-elemen kecil yang diberi nama token, yang umumnya terdiri dari kata atau frasa [19]. Tahap ini sangat penting untuk analisis teks karena model atau algoritma tidak mampu mengolah teks dalam bentuk kalimat secara langsung.

e. *Stopword Removal*

Proses Penghapusan *Stopword removal* diterapkan melalui penghilangan kata-kata yang sering ditemukan akan tetapi tidak berpengaruh besar terhadap makna, sehingga hanya kata penting saja yang digunakan dalam pemodelan [13].

f. *Stemming*

Proses mengembalikan kata ke bentuk dasarnya melalui penghilangan imbuhan awal maupun akhir seperti dikenal sebagai *stemming*. Istilah "berlari", "lari-lari", dan "pelari" digabungkan menjadi satu kata, "lari". Tujuannya adalah agar sistem dapat memahami bahwa setiap kata yang dibentuk dari satu bentuk dasar memiliki arti yang sama. [20].

2.3 Labeling Data

Perlabelan data adalah proses mengkategori setiap ulasan pengguna untuk menentukan jenis sentimen. Dalam penelitian ulasan DANA, perlabelan dilakukan untuk membedakan jenis ulasan positif dan negatif. Setiap komentar yang telah dikumpulkan dari pengguna kemudian dianalisis maknanya berdasarkan kata-kata yang digunakan [21]. Hasil perlabelan kemudian digunakan selama proses analisis sentimen.

2.3 Pembagian Data Training dan Testing

Dalam tahap pembagian dataset, pendekatan penelitian ini mengalokasikan data ke dalam dua kelompok utama 80% untuk tujuan latihan dan 20% untuk uji. Alokasi semacam ini biasanya menghasilkan performa yang lebih baik, sebab sebagian besar dataset disediakan untuk proses pembentukan model [22].

2.4 Pembobotan Kata

Frequency of Terms - Frequency of Inverse Documents (TF–IDF). Ini proses mengubah kalimat menjadi nilai numerik sehingga algoritma pembelajaran mesin dapat memahaminya [23]. TF-IDF adalah teknik yang memberikan nilai kepada setiap kata yang muncul pada dokumen, yang memberi bobot lebih kecil pada kata umum di seluruh dokumen. Metode ini biasanya berguna untuk memberi pikiran bahwa jika kata-kata muncul hanya sekali dalam satu dokumen, tetapi jarang atau sama sekali tidak muncul dalam dokumen lain, maka bobot kata dalam dokumen tersebut harus tinggi karena kata-kata tersebut dianggap sebagai konten utama dokumen tersebut. [24].

2.5 Permodelan Algoritma

Tahap selanjutnya yaitu permodelan dengan algoritma. Dalam riset ini menggunakan dua algoritma yaitu :

a. *Random Forest* adalah sebuah teknik dalam pembelajaran mesin yang diterapkan untuk tugas klasifikasi. Algoritma tersebut masuk ke dalam kategori pembelajaran ensemble, yakni pendekatan yang memadukan beberapa model guna mencapai hasil ramalan yang lebih tepat dan konsisten jika dibandingkan dengan penggunaan model tunggal. *Random Forest* berfungsi melalui pembentukan sejumlah besar pohon keputusan, kemudian menyatukan output dari seluruh pohon itu untuk menentukan hasil akhir[25]. Dalam analisis sentimen terkait aplikasi DANA, algoritma *Random Forest* dimanfaatkan untuk membagi opini pengguna ke dalam kelompok, yaitu positif serta negatif. Setelah teks ulasan dikonversi menjadi data numerik, informasi tersebut kemudian dimasukkan ke dalam model *Random Forest*. Secara matematis, hal ini dapat diungkapkan sebagai :

$$\hat{Y} = mode\ (h_1\ (X), h_2\ (X), \dots h_k\ (X)) \qquad (1)$$

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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Y^ = hasil prediksi akhir
mode = kelas yang paling banyak dipilih (voting mayoritas)
hi(X) = prediksi decision tree ke-1
k = jumlah pada model (forest)

b. *Logistic Regression* adalah algoritma yang dimanfaatkan untuk menyelesaikan masalah klasifikasi dengan cara memprediksi probabilitas suatu data masuk ke sebuah kelas. Algoritma ini mengubah hasil perhitungan menjadi nilai antara 0–1 menggunakan fungsi sigmoid, sehingga dapat menentukan apakah data termasuk kategori tertentu, seperti sentimen positif atau negatif. *Logistic Regression* fokus pada penentuan kelas berdasarkan probabilitas tersebut, sehingga banyak digunakan dalam berbagai tugas klasifikasi sederhana maupun kompleks [26]. Secara matematis, model logistic regression ditulis sebagai:

$$\rho(Y = 1) = \frac{1}{1+e^{2-(\beta_0+\beta_1 X)}} \quad (1)$$

$$p(Y = 0) = \frac{1}{1+e^{2-(\beta 0+\beta_1 X)}} \quad (2)$$

P(Y=1) menunjukkan peluang bahwa variabel terikat (Y) berada pada kelas 1, yaitu kategori positif.
P(Y=0) menunjukkan peluang bahwa variabel terikat (Y) berada pada kelas 0, yaitu kategori negatif.
e merupakan bilangan eksponensial atau bilangan Euler.
β0 merupakan intercept atau konstanta yang menggambarkan nilai dasar model ketika variabel X bernilai 0.
β1 merupakan koefisien regresi yang menunjukkan besar dan arah pengaruh variabel X terhadap probabilitas terjadinya kelas 1

2.6 Evaluasi

Proses menilai sistem, atau model, berhasil mencapai tujuan dikenal sebagai evaluasi. Dalam penelitian atau analisis data, evaluasi dilakukan untuk mengetahui seberapa baik kinerja, akurasi, dan keberhasilan model yang digunakan. Memberikan gambaran objektif tentang kemampuan model untuk prediksi atau klasifikasi, tahap ini sangat penting. Biasanya, dalam proses evaluasi, beragam metrik, seperti akurasi, precision, recall dan f1-score, masing-masing menunjukkan perspektif mengenai cara kerja model. Dengan mengevaluasi yang tepat, peneliti dapat mengetahui apakah model sudah bekerja dengan baik atau perlu dilakukan perbaikan untuk meningkatkan kualitas data.

3. HASIL DAN PEMBAHASAN

Dalam tahap ini, membahas secara menyeluruh terkait tentang pemrosesan data, *preprocessing* teks, pemodelan algoritma, dan evaluasi kinerja model. Hasil pemrosesan data menunjukkan jumlah ulasan yang berhasil dikumpulkan dari *Google Play Store*. Langkah persiapan awal melibatkan penyesuaian kapitalisasi huruf, pembersihan konten, standardisasi format, pemisahan kata-kata, penghapusan istilah umum, dan penyederhanaan terminologi, menghasilkan data yang bersih dan siap untuk dimodelkan. Pada bagian akhir, kinerja masing-masing model dievaluasi menggunakan performa setiap model diukur dengan memanfaatkan indikator seperti akurasi, *precission*, *recall*, dan *F1-score*. Tujuannya adalah mengetahui model yang paling efisien untuk mengklasifikasikan opini para pengguna aplikasi DANA.

3.1 Pengumpulan Data

Data terkumpul melalui *scraping* yang dilakukan pada halaman aplikasi Dana di *Play Store*. Langkah *Scraping* ini mengambil data ulasan pengguna, yang mencakup informasi seperti *Username*, *Content*, *Score*, *at*. Data tersebut dikumpulkan pada tanggal 6 Oktober 2025. Selanjutnya dilakukan pembersihan dengan *drop down* untuk menghilangkan data yang tidak relevan atau duplikat. Setelah proses ini, terdapat 2.068 data yang digunakan untuk analisis lebih lanjut. Selain itu, kolom atau baris yang kosong juga dihapus guna menunjukkan data yang diterapkan benar-benar bersih dan memberikan hasil analisis lebih akurat.

Tabel 1. Hasil *Scraping* Data

UserName	Content	Score	At
Novitasofi	apaan sih mau kirim ke bank lain gak bisa!minta di upgrade terus,padahal udah di upgrade!tolong ini kenapa yaaa!!!	4	2025-10 05
Pilipus	agak resek	2	2025-10 05
Jack Funk	terbaik	5	2025-10 05
Angga	👍	5	2025-10 05
Pratama	apk tolol bayar transaksi tanpa konfirmasi.saldo dana saya tiba tiba ngilang minta pengajuan	1	2025-10 05
Bahrul	Sangat membantu	5	2025-10 05
Fauzen Adhi	ok	5	2025-10 05

3.2 *Pre-processing* Data

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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Setelah *scraping* data, tahapan *preprocessing* dilakukan untuk membuat data lebih terstruktur digunakan untuk Analisis Sentimen. Tahapan-tahap ini termasuk mengubah kata ke bentuk kecil semua, membersihkan, normalisasi, tokenisasi, menghilangkan kata umum, dan *stemming*.

a. *Cleaning*

Cleaning adalah langkah membersihkan data dari kesalahan data yang duplikat, Selain itu, pembersihan juga mencakup penghapusan data yang ganda, simbol tidak perlu, serta penyesuaian cara menulis agar memiliki format yang sama. Dengan demikian, hasil analisis atau pemodelan yang dilakukan setelah proses *cleaning* akan menjadi lebih akurat. Hasil *cleaning* seperti table 2.

Tabel 2. Hasil *Cleaning* Data

Cleaning
apaan sih mau kirim ke bank lain gak bisaminta di upgrade terus...
agak resek
terbaik
apk tolol bayar transaksi tanpa konfirmasisaldo dana saya hilang
saya punyak pengalaman buruk DI DANA Tidak ada solusi
sangat membantu

b. *Case Folding*

Menggabungkan semua huruf menjadi kecil. Contoh kata “DANA”, dan “dana” akan dianggap sama. Dalam tahap pra-pemrosesan teks, proses ini sangat penting karena membantu sistem menemukan kata yang sama tanpa terpengaruh perbedaan penulisan huruf sehingga analisis data menjadi lebih akurat. Hasil *Case Folding* seperti tabel 3.

Tabel 3. Hasil *Case Folding*

Case Folding
apaan sih mau kirim ke bank lain gak bisaminta di upgrade terus
agak resek
terbaik
apk tolol bayar transaksi tanpa konfirmasisaldo dana saya hilang
saya punyak pengalaman buruk di dana Tidak ada solusi
sangat membantu

c. Normalisasi

Normalisasi adalah proses menyamakan bentuk penulisan kata agar konsisten dan sesuai dengan bentuk baku. Tujuannya agar kata yang tidak baku atau singkatan bisa dikenali dengan benar oleh sistem analisis. Seperti contoh (‘gk’, ‘tdk’), kata tidak baku (‘nggak’, ‘ga’), atau bentuk penulisan yang tidak konsisten. Melalui tahap normalisasi dapat membantu mengurangi noise pada data dan meningkatkan kualitas data. Hasil *normalisasi* seperti tabel 4.

Tabel 4. Hasil *Nomalisasi* Data

Normalisasi
apaan sih mau kirim ke bank lain tidak bisaminta di upgrade terus
agak resek
terbaik
apk tolol bayar transaksi tanpa konfirmasisaldo dana saya hilang
saya punyak pengalaman buruk di dana tidak ada solusi
Sangat membantu

d. *Tokenizing*

Tokenizing merupakan proses pemisahan teks menjadi satuan kecil yang disebut dengan token, Dalam analisis sentimen, *tokenizing* berperan penting karena model klasifikasi tidak dapat mengolah teks dalam bentuk kalimat secara langsung, melainkan memerlukan representasi dalam bentuk daftar kata. Sebagai contoh, kalimat “Aplikasi DANA sangat bagus” akan diubah menjadi :
Contoh [“Aplikasi”, “DANA”, “sangat”, “bagus”].

Tabel 5. Hasil *Tokenizing*

Tokenizing
['apaan', 'sih', 'mau', 'kirim', 'ke', 'bank', 'lain', 'tidak', 'bisaminta', 'di', 'upgrade']
['agak', 'resek']
['terbaik']
[]

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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

['apk', 'tolol', 'bayar', 'transaksi', 'tanpa', 'konfirmasisaldo', 'dana',]....
['saya', 'punyak', 'pengalaman', 'buruk', 'di', 'dana', 'tidak', 'ada', 'solusi']
['sangat', 'membantu']

e. *Stopword Removal*

Stopword removal merupakan tahap penyaringan kata dengan cara menghilangkan kata-kata umum yang sering muncul dalam teks tetapi tidak memberikan kontribusi besar terhadap makna atau proses klasifikasi. kata tersebut biasanya berfungsi sebagai penghubung atau pelengkap dalam struktur kalimat, seperti, "dan", "yang", "di".

Tabel 6. Hasil *Stopword Removal*

Stopword Removal
['sih', 'kirim', 'bank', 'bisaminta', 'upgrade']
['resek']
['terbaik']
[]
['apk', 'tolol', 'bayar', 'transaksi', 'konfirmasisaldo', 'dana']
['punyak', 'pengalaman', 'buruk', 'dana']
['membantu']

f. *Stemming*

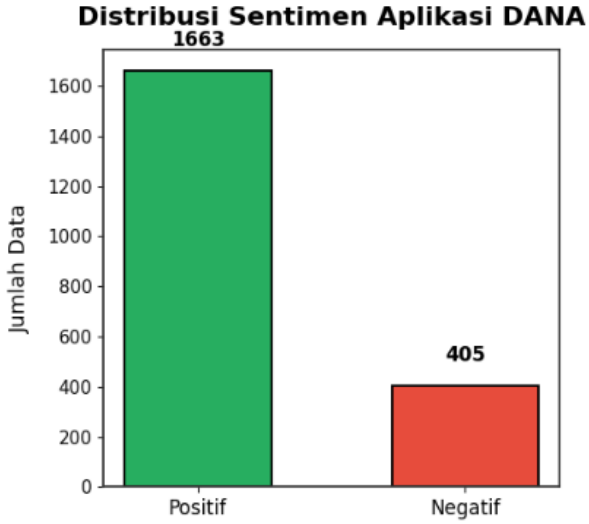
Stemming yaitu proses mengembalikan kata ke bentuknya yang paling dasar dengan menghapus imbuhan seperti awalan, sisipan, atau akhiran.

Tabel 7. Hasil *Stemming*

Stopword Removal
['sih', 'kirim', 'bank', 'bisaminta', 'upgrade']
['resek']
['terbaik']
[]
['apk', 'tolol', 'bayar', 'transaksi', 'konfirmasisaldo', 'dana']
['punyak', 'pengalaman', 'buruk', 'dana']
['bantu']

3.1 Perlabelan Data

Pelabelan data bertujuan untuk mengelompokkan data bersadasarkan dua kelompok yaitu positif serta negatif. Perlabelan sentimen diterapkan secara langsung melalui code python dengan pustaka pandas melalui fungsi apply(), yang mengklasifikasikan rating 1 dan 2 sebagai negatif, dan penilaian 3, 4, 5 sebagai positif. Seperti pada gambar 10, terdapat data positif sebanyak 1.663 dan data negatif sebanya 405, dengan total keseluruhan 2.068 data.



Gambar 2. Perlabelan Data

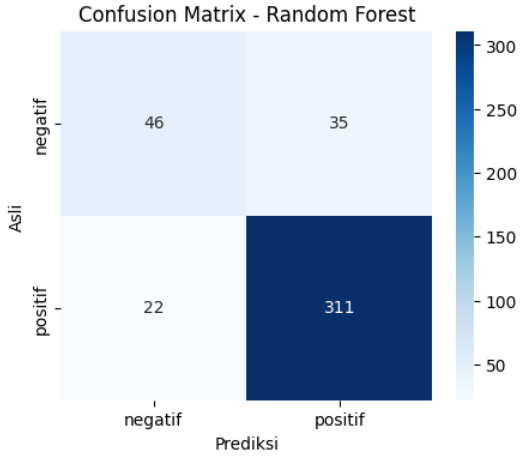
3.2 Permodelan Algoritma

Setelah perlabelan data peneliti kemudia memecah data diklasifikasikan ke dalam dua subbagian diantaranya data pelatihan dan data untuk pengujian, dengan presentase untuk pelatihan 80% yaitu sebanyak 1654 data, dan untuk pengujian 20% yaitu sebanyak 414 data. Langkah selanjutnya pembobotan kata diterapkan berdasarkan teknik TF-IDF,

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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

yaitu tahapan dalam mentransformasikan kalimat menjadi angka atau bilangan. Kemudian data dilatih menggunakan dua algoritma yaitu :

a. *Random Forest*



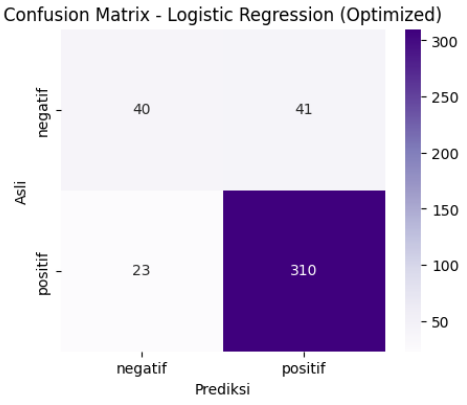
Gambar 3. *Confusion Matrix Random Forest*

Tabel 8. Laporan Klasifikasi *Random Forest*

	Precission	*Recall*	*F1 - score*	*Support*
Negatif	0.68	0.57	0.62	81
Positif	0.90	0.93	0.92	333
Accuracy			0.86	414
Macro avg	0.79	0.75	0.77	414
Weighted avg	0.86	0.86	0.86	414

Berdasarkan dari tabel 8, klasifikasi *Random Forest* menghasilkan akurasi sebesar 86% dalam mengklasifikasikan sentimen ulasan aplikasi Dana. Laporan klasifikasi menunjukkan bahwa nilai *precision* 0.68, *recall* 0.57, dan *f-I score* 0.62 dimiliki oleh kategori negatif, sedangkan nilai *precision* 0.90, *recall* 0.93 dan *f1- score* 0.92 dimiliki oleh kategori positif. Hasil macro avg dan wighted avg menunjukan performa keseluruhan model yaitu 0.77 dan 0.86 menandakan model mampu melakukan klasifikasi sentimen dengan baik dan konsisten pada seluruh kategori.

b. *Logistic Regression*



Gambar 5. *Confusion Matrix* Model *Logistic Regression*

Tabel 9. Hasil Klasifikasi *Logistic Regression*

	Precission	*Recall*	*F1- Score*	*Support*
Negatif	0.63	0.49	0.36	81
Positif	0.88	0.93	0.91	333
accuracy			0.85	414
Macro avg	0.76	0.71	0.73	414
Weighted avg	0.83	0.85	0.84	414

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.v13i1.9517
Hal 296-304
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

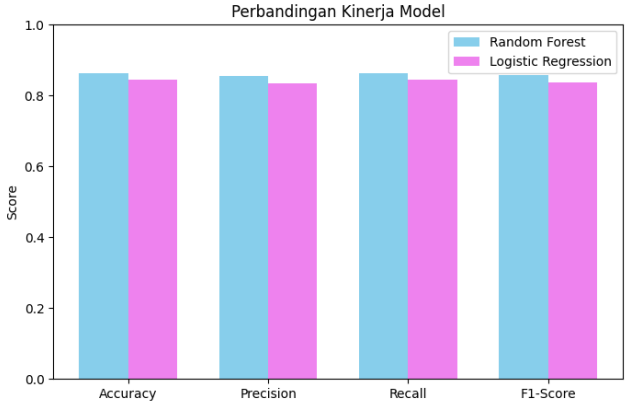
Berdasarkan dari tabel 9, hasil optimasi *Logistic Regression* menunjukkan akurasi sebesar 84% dalam mengklasifikasikan sentimen ulasan aplikasi Dana. Laporan klasifikasi menunjukkan bahwa nilai *precision* 0.62, *recall* 0.49 dan *f-1 score* 0.56 dimiliki oleh kategori negatif. Sementara itu, kategori positif memiliki nilai *precision* 0.88, *recall* 0.93 dan *f-1 score* 0.91. Hasil macro avg dan wighted avg menunjukan performa keseluruhan model yaitu 0.73 dan 0.84 menandakan model mampu melakukan klasifikasi sentimen dengan baik dan konsisten pada seluruh kategori.

3.3 Evaluasi

Hasil evaluasi pemodelan dua algoritma ini dapat dilihat hasil akurasi untuk algoratima *Random Forest* akurasi sebanyak 86.23%, dan *Logistic Regression* yaitu 84.58%. seperti pada Tabel 10. Pada grafik perbandingan kinerja model terlihat bahwa nilai kedua model sama-sama tinggi, namun *Random Forest* lebih unggul. Hal ini mengindikasikan Bahwa *Random Forest* mampu mengenali pola sentimen dalam ulasan aplikasi DANA dengan lebih stabil dan akurat. Sementara itu, *Logistic Regression* tetap menunjukkan performa yang kompetitif dan mendekati *Random Forest*, sehingga masih layak digunakan untuk analisis sentimen berbasis teks. Secara keseluruhan, *Random Forest* menjadi model yang memberikan hasil terbaik dalam penelitian ini, meskipun perbedaannya tidak terlalu besar.

Tabel 8. Perbandingan Algoritma

Algoritma	Akurasi
Random Forest	86.23%
Logistic Regression	84.54%



Gambar 7. Perbandingan Kinerja Algoritma

4. KESIMPULAN

Berdasarkan riset yang telah dilakukan terkait perbandingan performa algoritma *Random Forest* dan *Logistic Regression* dalam mengklasifikasikan sentimen opini para pengguna layanan aplikasi DANA. Data penelitian ini diperoleh dari *Google Play Store*. Melalui tahap *pre-processing*, perlabelan, pembobotan TF-IDF, dan tahap permodelan, diperoleh algoritma *Random Forest* memberikan performa lebih unggul dibandingkan *Logistic Regression.* Adapun riset yang telah diimplementasikan Hapsari dkk yaitu analisis terhadap dompet digital di platform Twitter dengan *Random Forest* mendapatkan akurari 85%, sementara riset yang dibangun pada penelitian ini dengan model *Random Forest* mendapatkan akurasi 86.23%, sedangkan *Logistic Regression* memperoleh akurasi 84.54%. Dengan demikian, algoritma *Random Forest* dinilai lebih unggul digunakan untuk analisis sentimen pada ulasan aplikasi DANA. Untuk penelitian selanjutnya, dapat mengambil data dari sumber lain seperti youtobe, instagram, facebook. Selanjutnya dapat menggunakan metode pelabelan lain, dan disarankan menambah variasi algoritma seperti *SVM* , *deep learning* dan lain sebagainya.

REFERENCES

[1] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, and B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," *Metik J.*, vol. 6, no. 2, pp. 97–103, 2022, doi: 10.47002/metik.v6i2.372.

[2] M. W. A. Putra, Susanti, Erlin, and Herwin, "Analisis Sentimen Dompet Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.

[3] N. Ambika Hapsari and A. Dwi Indriyanti, "Analisis Sentimen pada Aplikasi Dompet Digital Menggunakan Algoritma Random Forest," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 04, no. 03, pp. 186–192, 2023.

[4] A. Maulana, Inayah Khasnaputri Afifah, Asghafi Mubarrak, Kiagus Rachmat Fauzan, Ardhan Dwintara, and B. P. Zen, "Comparison of Logistic Regression, Multinomialnb, Svm, and K-Nn Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1487–1494, 2023, doi: 10.52436/1.jutif.2023.4.6.863.

[5] E. Ogi, I. Pratiwi, and W. Yustanti, "Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter (Studi

Kasus Ovo dan Dana)," *Jeisbi*, vol. 02, no. 03, p. 2021, 2021.
[6] H. Junianto, R. E. Saputro, B. A. Kusuma, and D. I. S. Saputra, "Comparison of Logistic Regression and Random Forest in Sentiment Analysis of Disdukcapil Application Reviews," *J. Tek. Inform.*, vol. 5, no. 6, pp. 1539–1547, 2024, doi: 10.52436/1.jutif.2024.5.6.1802.
[7] E. P. Efendi and M. A. Barata, "KOMPARASI ALGORITMA SVM DAN C4 . 5 DENGAN BACKWARD ELIMINATION UNTUK KLASIFIKASI TINGKAT STRES MAHASISWA COMPARATIVE ANALYSIS OF SVM AND C4 . 5 ALGORITHMS WITH BACKWARD," no. 3, pp. 539–549, 2024, doi: 10.26798/jiko.v9i3.2130.
[8] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *J. Inf. Syst. Informatics*, vol. 5, no. 3, pp. 915–927, 2023, doi: 10.51519/journalisi.v5i3.539.
[9] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.
[10] C. Rahmawati and P. Sukmasetya, "Sentimen Analisis Opini Masyarakat Terhadap Kebijakan Kominfo atas Pemblokiran Situs non-PSE pada Media Sosial Twitter," vol. 9, no. 5, pp. 1393–1400, 2022, doi: 10.30865/jurikom.v9i5.4950.
[11] N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 37–49, 2025, doi: 10.30591/jpit.v10i1.7106.
[12] M. Algoritma, S. V. M. Dan, R. Forest, and F. N. Firzatullah, "Analisis Sentimen Pengguna Aplikasi Byond BSI Pada Google Play Store," pp. 356–363, 2025, doi: 10.47002/metik.v9i2.1089.
[13] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF , SMOTE , and Traditional Machine Learning Models," vol. 9, no. 5, pp. 2552–2561, 2025.
[14] D. N. Agustia, R. R. Suryono, U. T. Indonesia, L. Ratu, and K. B. Lampung, "COMPARISON OF NAÏVE BAYES , RANDOM FOREST , AND LOGISTIC REGRESSION ALGORITHMS FOR SENTIMENT ANALYSIS ONLINE GAMBLING KOMPARASI ALGORITMA NAÏVE BAYES , RANDOM FOREST , DAN LOGISTIC REGRESION UNTUK ANALISIS," vol. 10, no. 1, pp. 284–295, 2025.
[15] R. T. Agustin, Y. Cahyana, K. A. Baihaqi, and T. Rohana, "Public Sentiment Analysis on the Boycott Israel Movement on Platform X Using Random Forest and Logistic Regression Algorithms," vol. 9, no. 3, pp. 938–945, 2025.
[16] A. Y. Pratama, G. A. Sanjaya, N. K. Lubis, and M. R. Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," 2025, doi: 10.47002/metik.v9i1.1055.
[17] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS," vol. 8, no. 3, pp. 207–212, 2023.
[18] Z. T. Zafira, K. D. Tania, W. K. Sari, S. Informasi, and U. Sriwijaya, "Sentiment-Based Knowledge Discovery of Wondr by BNI App Reviews Using SVM , KNN , and Naive Bayes for CRM Enhancement," vol. 9, no. 5, pp. 2498–2508, 2025.
[19] I. Yunanto and S. Yulianto, "TWITTER SENTIMENT ANALYSIS PEDULILINDUNGI APPLICATION USING NAÏVE BAYES AND SUPPORT VECTOR MACHINE ANALISIS SENTIMEN TWITTER APLIKASI PEDULILINDUNGI," vol. 3, no. 4, pp. 807–814, 2022.
[20] R. R. Pratama, R. R. Suryono, S. Informasi, and U. T. Indonesia, "PERFORMANCE COMPARISON OF NAIVE BAYES , SUPPORT VECTOR MACHINE AND RANDOM FOREST ALGORITHMS FOR APPLE VISION PRO SENTIMENT ANALYSIS PERBANDINGAN PERFORMA ALGORITMA NAIVE BAYES , SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN APPLE," vol. 6, no. 1, pp. 31–39, 2025.
[21] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," vol. 9, no. 3, pp. 224–233, 2024, doi: 10.30591/jpit.v9i3.6640.
[22] I. M. Sumertajaya, Y. Angraini, J. R. Harahap, and A. Fitrianto, "Sentiment Analysis on Covid-19 Vaccination in Indonesia Using Support Vector Machine and Random Forest," vol. 10, no. 1, pp. 1–8, 2022.
[23] S. Sumayah, F. Sembiring, W. Jatmiko, J. S. Informasi, U. Nusa, and P. Sukabumi, "ANALYSIS OF SENTIMENT OF INDONESIAN COMMUNITY ON METAVERSE ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP METAVERSE," vol. 4, no. 1, pp. 143–150, 2023.
[24] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, pp. 302–307, 2023.
[25] F. S. Pratiwi *et al.*, "IMPLEMENTASI METODE SMOTE DAN RANDOM OVER- SAMPLING PADA ALGORITMA MACHINE LEARNING UNTUK," vol. 8, no. 1, pp. 87–98, 2025.
[26] A. S. Putri, A. F. Rozi, U. Mercu, B. Yogyakarta, and D. I. Yogyakarta, "SENTIMENT ANALYSIS OF TELEGRAM APPLICATION USER SATISFACTION ON GOOGLE PLAY STORE USING NAÏVE BAYES , LOGISTIC REGRESSION AND SVM ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI TELEGRAM PADA GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES , LOGISTIC REGRESSION DAN SVM," vol. 10, no. 2, pp. 857–867, 2025.
[27] F. F. Wati and A. E. Widodo, "DEEPSEEK MENGGUNAKAN ALGORITMA RANDOM," vol. 5, no. 1, pp. 8–15, 2025.