Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

Analisa Sentimen Pengguna Aplikasi DANA Pada Ulasan Google Play Store Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors

Dian Ayu Sabillah*, M Afdal, Inggih Permana, Fitriani Muttakin

Fakultas Sains Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia
Email: [1,*]12050320317@students.uin-suska.ac.id, [2]m.afdal@uin-suska.ac.id, [3]inggihpermana@uin-suska.ac.id, [4]fitrianimuttakin@uin-suska.ac.id
Email Penulis Korespondensi: 12050320317@students.uin-suska.ac.id
Submitted: **03/07/2025**; Accepted: **01/09/2025**; Published**: 02/09/2025**

Abstrak−Penggunaan dompet digital seperti DANA di Indonesia terus meningkat seiring kebutuhan transaksi non-tunai yang cepat dan praktis. Ulasan pengguna di Google Play Store menjadi sumber informasi penting untuk menilai kepuasan dan permasalahan layanan. Penelitian ini bertujuan mengklasifikasikan sentimen pengguna terhadap aplikasi DANA menggunakan algoritma Naïve Bayes Classifier (NBC) dan K-Nearest Neighbor (KNN). Sebanyak 1.000 ulasan dikumpulkan dan diproses melalui pembersihan teks, tokenisasi, penghapusan stopword, dan stemming. Sentimen diklasifikasikan menjadi positif, netral, dan negatif menggunakan metode lexicon dan validasi pakar. Hasil menunjukkan NBC lebih unggul dibanding KNN, dengan akurasi tertinggi 71,83%, sedangkan KNN hanya mencapai 56,44%. NBC juga lebih efektif dalam mendeteksi sentimen negatif, meskipun keduanya kurang optimal pada sentimen netral. Visualisasi word cloud menampilkan kata-kata dominan pada setiap kategori sentimen. Kesimpulan penelitian ini menyatakan bahwa Naïve Bayes lebih efektif dalam analisis sentimen ulasan pengguna aplikasi dompet digital seperti DANA.

Kata Kunci: Aplikasi DANA; Naïve Bayes Classifier; K-Nearest Neighbor; Ulasan Pengguna; Google Play Store; TF-IDF; Pra-pemrosesan Teks; Pembelajaran Mesin; Dompet Digital

Abstract−The use of digital wallets such as DANA in Indonesia continues to increase along with the need for fast and practical non-cash transactions. User reviews on the Google Play Store are an important source of information to assess satisfaction and service problems. This study aims to classify user sentiment towards the DANA application using the Naïve Bayes Classifier (NBC) and K-Nearest Neighbor (KNN) algorithms. A total of 1,000 reviews were collected and processed through text cleaning, tokenization, stopword removal, and stemming. Sentiments were classified into positive, neutral, and negative using the lexicon method and expert validation. The results showed that NBC was superior to KNN, with the highest accuracy of 71.83%, while KNN only reached 56.44%. NBC was also more effective in detecting negative sentiment, although both were less than optimal for neutral sentiment. Word cloud visualization displays the dominant words in each sentiment category. The conclusion of this study states that Naïve Bayes is more effective in analyzing sentiment reviews of digital wallet applications such as DANA.

Keywords: DANA Application; Naïve Bayes Classifier; K-Nearest Neighbor; User Reviews; Google Play Store; TF-IDF; Text Preprocessing; Machine Learning; Digital Wallet

1. PENDAHULUAN

Perkembangan teknologi di Indonesia semakin memudahkan kehidupan sehari-hari, termasuk dalam hal bertransaksi. Kini, pembayaran non tunai menjadi pilihan yang praktis, efisien, dan makin lazim digunakan oleh masyarakat.[1]. E-money berperan penting dalam mendorong perkembangan digital marketing dan berbagai aspek lainnya, termasuk kemunculan dompet digital (e-wallet). E-wallet adalah aplikasi digital untuk menyimpan uang dan mempermudah transaksi keuangan secara praktis[2]. Kemajuan teknologi informasi dan komunikasi telah berperan besar dalam mengubah berbagai aspek kehidupan, terutama di bidang keuangan. Beragam inovasi digital yang muncul turut meningkatkan efisiensi, transparansi, dan kemudahan akses terhadap layanan keuangan bagi masyarakat[3].

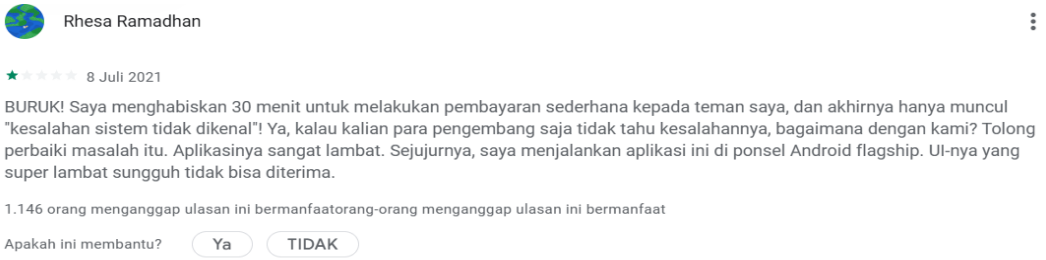
Salah satu perkembangan yang pesat saat ini adalah penggunaan aplikasi digital untuk sistem pembayaran, yang dikenal dengan istilah dompet digital atau *e-wallet*[4]. Aplikasi *e-wallet* memberikan kemudahan bagi pengguna dalam bertransaksi secara cepat, praktis, dan tanpa menggunakan uang tunai. Di Indonesia, penggunaan *e-wallet* terus meningkat seiring dengan tumbuhnya akses internet dan penggunaan smartphone yang semakin luas[5]. *E-wallet,* atau dompet digital, merupakan aplikasi yang memungkinkan pengguna melakukan pembayaran melalui komputer atau perangkat seluler, dengan memanfaatkan kartu bank yang terhubung, saldo, rekening tabungan, maupun layanan keuangan digital lainnya seperti produk likuiditas dan kredit[6].

Di Google Play Store, terdapat banyak aplikasi e-wallet yang memiliki jumlah unduhan dalam skala besar. Ulasan dari pengguna menjadi aspek penting yang perlu diperhatikan, mengingat beberapa aplikasi menunjukkan tingkat kepuasan yang serupa, sehingga membuat penilaian terhadap aplikasi terbaik menjadi kurang objektif jika hanya didasarkan pada jumlah unduhan. Berdasarkan data survei dari Katadata.go.id, peneliti mengidentifikasi tiga aplikasi e-wallet dengan jumlah unduhan terbanyak, masing-masing melebihi 10 juta unduhan, yaitu DANA, OVO, dan LinkAja [7].

Sebagai salah satu E-wallet populer, DANA mendapatkan perhatian melalui ulasan positif dan negatif di Google Play Store. Oleh karena itu, analisis sentimen terhadap ulasan pengguna menjadi krusial untuk memahami respons masyarakat terhadap layanan E-wallet[8].Meskipun DANA telah mendapatkan ulasan positif, negatif, dan netral di Google Play Store, terdapat ulasan atau tanggapan dari pengguna mengenai permasalahan yang dialami nya

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

pada aplikasi DANA, seperti fitur yang masih kurang, seringnya terjadi error, kurangnya keterbukaan informasi, keamanan, respons layanan pelanggan, kompatibilitas dan bug, kesulitan pendanaan atau penarikan dana dan kasus uang yang tidak kembali[9].



Gambar 1. Ulasan Pengguna DANA

Berdasarkan ulasan pengguna aplikasi DANA pada Gambar 1 menunjukkan beberapa masalah utama, seperti kesalahan sistem yang tidak jelas, waktu proses transaksi yang lama, dan kinerja aplikasi yang lambat meskipun digunakan di perangkat premium. Pengguna juga mengeluhkan UI yang tidak responsif dan minimnya kejelasan dari pihak pengembang. Secara keseluruhan, ulasan ini mencerminkan ketidakpuasan yang tinggi dan menunjukkan bahwa masih ada banyak aspek teknis dan pelayanan yang perlu diperbaiki oleh tim pengembang DANA agar dapat memberikan pengalaman pengguna yang lebih baik dan terpercaya.

Ulasan pengguna menjadi sumber informasi yang populer dan efektif untuk menilai suatu produk atau layanan tertentu. Tujuan utamanya adalah untuk mengevaluasi kinerja aplikasi, memberikan umpan balik kepada pengembang, dan memfasilitasi perbaikan [10].Pengguna aplikasi umumnya cenderung membaca ulasan sebelum mengunduh aplikasi, membuat monitoring data teks atau yang dikenal sebagai *text mining* menjadi suatu kebutuhan[11]. Karena itulah, menganalisis sentimen dari ulasan pengguna menjadi hal penting untuk memahami bagaimana tanggapan masyarakat terhadap layanan E-wallet. Dalam hal ini, *text mining* digunakan untuk menggali informasi dari komentar pengguna dengan cara mengelompokkan dan menganalisis kata-kata. Prosesnya meliputi tokenisasi, kapitalisasi, stemming, dan filtering[12].

Analisis sentimen merupakan teknik yang digunakan untuk mendapatkan emosi seseorang dari teks[13] . Sentimen merupakan informasi dalam bentuk teks yang memiliki polaritas, yaitu positif, netral, atau negatif. Polaritas ini dapat digunakan sebagai dasar dalam pengambilan keputusan, terutama dalam mengevaluasi opini atau respons pengguna. Dalam proses analisis sentimen, terdapat dua pendekatan utama yang umum digunakan, yaitu lexicon-based dan machine learning. Pendekatan lexicon-based menggunakan kamus kata yang telah diberi nilai sentimen untuk membandingkan kata-kata dalam teks. Melalui kamus tersebut, sistem dapat mengidentifikasi apakah suatu teks mengandung opini dan menentukan jenis sentimennya berdasarkan kata-kata yang muncul [14].

Namun, banyaknya ulasan yang masuk setiap hari membuat pengembang kesulitan untuk menganalisis sentimen secara manual. Oleh karena itu, dibutuhkan cara yang lebih cepat dan efisien untuk memahami pendapat pengguna[15] . Tanpa bantuan sistem analisis otomatis, banyak informasi penting dari ulasan pengguna bisa terlewat dan tidak dimanfaatkan secara maksimal dalam pengembangan aplikasi. Oleh sebab itu, dibutuhkan pendekatan teknologi yang mampu mengklasifikasikan dan menganalisis sentimen pengguna dengan cepat dan akurat [16] . Salah satu solusi yang dapat diterapkan adalah dengan memanfaatkan metode machine learning, khususnya algoritma klasifikasi seperti *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (KNN), untuk membantu dalam mengelompokkan sentimen pengguna secara otomatis dan efisien [17] .

Algoritma *Naïve Bayes Classifier* adalah algoritma yang menggunakan kemungkinan atau probabilitas untuk mengklasifikasi statistik kelas data; teori probabilitas membuat klasifikasi ini dikelompokkan ke kelas tertentu [18]. Sedangkan *K-Nearest Neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut [19]. Penelitian ini menggunakan algoritma *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (KNN) untuk analisis sentimen. NBC menghitung probabilitas tiap fitur dan memilih kelas dengan kemungkinan tertinggi, sehingga cocok untuk klasifikasi teks, spam filtering, dan prediksi real-time. Sementara itu, KNN mengklasifikasikan data berdasarkan kedekatannya dengan data latih, bersifat sederhana, mudah digunakan, dan efektif untuk data dengan noise tinggi serta proses pengelompokan[20].

Berdasarkan penelitian yang dilakukan oleh Mustofa dkk(2022), mendapatkan hasil akurasi bahwa algoritma NBC lebih baik daripada algoritma KNN dengan hasil akurasi 79,67% berbanding 78,86% [21]. Pada penelitian yang berjudul Perbandingan Algoritma Naive Bayes Dan Knn Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Capcut mendapatkan hasil Naïve Bayes mengungguli KNN dengan akurasi 79.41% dibanding 75.63%.[22] Pada penelitian Amrillah dkk(2024) mendapatkan hasil Temuan menunjukkan bahwa Naïve Bayes mencapai akurasi 84%, melampaui KNN yang hanya mencapai 68% [23]. Pada penelitian selanjutnya mendapatkan hasil metode Naïve Bayes Classifier merupakan metode dengan tingkat akurasi yang lebih tinggi dibandingkan K-Nearst Neighbor dengan tingkat akurasi sebesar 80,00%.[24]. Menurut Syafrizal dkk(2024), hasilnya menunjukkan model NBC lebih baik dibandingkan KNN dengan akurasi sebesar 77,69%, recall 53,14%, precision 59,84% dan F1-Score 54,09%.[25] .

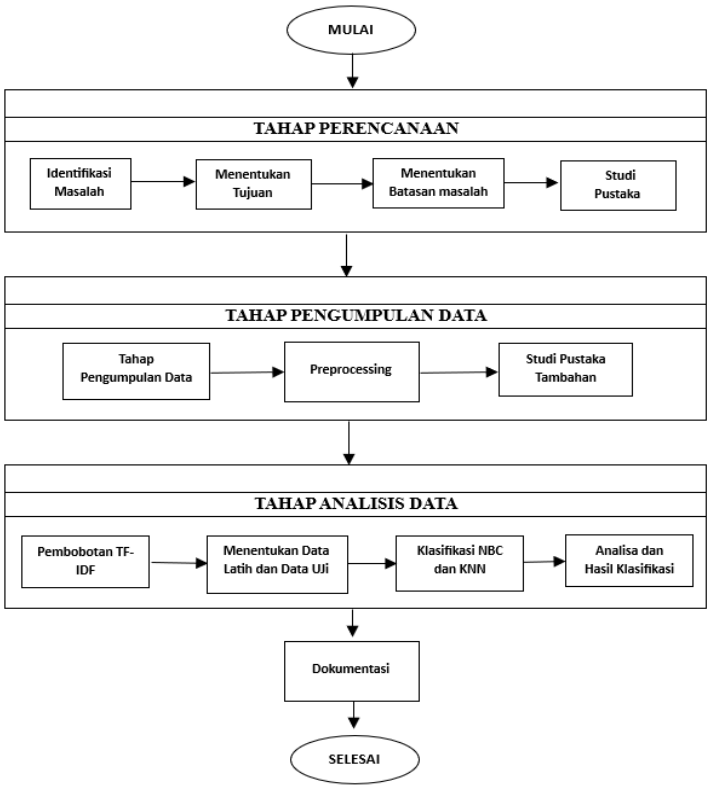
Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

Penelitian ini akan menguraikan tahapan-tahapan dalam melakukan analisis sentimen, yang mencakup proses pengumpulan data, pelabelan sentimen, preprocessing teks, klasifikasi sentimen, hingga evaluasi akurasi model. Penelitian ini dibatasi pada analisis ulasan aplikasi dompet digital DANA yang diambil dari Google Play Store, dengan total sebanyak 1.000 data ulasan. Pelabelan data dilakukan secara otomatis menggunakan kamus *Lexicon*, dan juga akan dibandingkan dengan pelabelan manual menggunakan pakar. Sentimen diklasifikasikan ke dalam tiga kategori, yaitu positif, netral, dan negatif. Data yang telah dilabeli kemudian dianalisis menggunakan algoritma Naïve Bayes Classifier *K-Nearest Neighbor* (KNN) dan .Tujuan dari penelitian ini adalah untuk mengetahui hasil klasifikasi komentar pengguna terhadap aplikasi DANA ke dalam kelas sentimen yang telah ditentukan, serta membandingkan akurasi algoritma Naïve Bayes dan K-Nearest Neighbor dalam mengklasifikasikan sentimen pada ulasan aplikasi tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk mencapai tujuan penelitian, tahapan metodologi yang diterapkan dalam penilitian ini terdapat pada Gambar 1.



Gambar 2. Metode Penelitian

Pada Gambar 2, penelitian menggunakan alur dari metodologi dimana dari tahap perencanaan yang dilanjutkan dengan tahap pengumpulan data. Pengumpulan data pada *Google Play Store* sebanyak 1000 ulasan dengan teknik *scraping* yang kemudian di olah dengan tahap *Pre-Processing,* ada pun tahapan nya seperti *cleaning*, *case folding*, *lowercase*, *tokenizing, stopword* dan *stemming.* Dilanjutkan dengan pembobotan TF-IDF yang setelahmya menntukan data latih dan data uji kemudian mengklasifikasi algoritma KNN dan NBC .

2.2 Tahap Pengumpulan Data

Penelitian ini berfokus pada ulasan pengguna aplikasi Dana di *Google Play Store.* Data yang dikumpulkan berupa ulasan ataupun informasi mengenai aplikasi DANA pada Platform *Google Play Store*. Data yang digunakan ialah selama tahun 2024 pada periode 1 Januari 2024 hingga 1 Desember 2024 sebanyak 1000 data ulasan menggunakan teknik *Scraping*.

2.3 Tahap *Pre-Processing*

Setelah proses pengumpulan data, maka selanjutnya adalah *pre-processing* yang dimulai dengan *cleaning* untuk menghilangkan tanda baca, angka, emoji, dan simbol. Kemudian dilakukan *case-folding*, yaitu mengubah seluruh huruf dalam teks menjadi *lowercase* atau huruf kecil semua agar konsisten. Pada proses *tokenizing* digunakan untuk memecah kalimat menjadi kata-kata terpisah. Selanjutnya, dilakukan *stopword* yaitu kata-kata umum yang dianggap

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

tidak memiliki makna penting dalam analisis. Tahap akhir adalah *stemming*, yang berfungsi untuk menghilangkan imbuhan sehingga setiap kata dikembalikan ke bentuk dasarnya.

2.4 Tahap Pengolahan Data

Tahapan ini bertujuan untuk mengelola, memahami, serta mengelompokkan data berbasis teks. Adapun langkah-langkah yang dilakukan meliputi:

a. Pelabelan Data
   Penentuan label sentimen (positif, negatif, atau netral) dilakukan secara manual oleh seorang pakar, terhadap data yang telah melewati proses *pre-processing* sebelumnya.
b. TF-IDF
   Tahapan ini digunakan untuk menghitung bobot pentingnya setiap kata dalam dokumen. Proses pembobotan ini dilakukan dengan memanfaatkan library *scikit-learn* pada bahasa pemrograman *Python*.
c. Membandingkan Hasil Lexicon dengan Pakar
   Pada tahap ini dilakukan untuk memperoleh hasil yang lebih akurat dengan membandingkan hasil pelabelan menggunakan metode Lexicon dan hasil pelabelan dari pakar.
d. Menentukan Data Latih dan Data Uji
   Pembagian data dilakukan dengan menerapkan metode K-Fold Cross Validation untuk memastikan evaluasi model yang lebih akurat dan menyeluruh. Setelah proses pembobotan dilakukan, data tersebut kemudian diproses menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor untuk melakukan klasifikasi sentimen.

2.5 Studi Pustaka Tambahan

a. Klasifikasi menggunakan Algoritma *Naïve Bayes Classifier* (NBC)
   Evaluasi dilakukan untuk menilai kinerja model klasifikasi yang digunakan. Salah satu metode evaluasi yang diterapkan adalah confusion matrix, yang menjadi dasar dalam menghitung metrik seperti akurasi, presisi, dan recall. Metrik-metrik ini digunakan untuk mengetahui seberapa sering model membuat prediksi yang benar terhadap data yang diuji. Algoritma Naïve Bayes dapat menarik kesimpulan berdasarkan klasifikasi data pelatihan sebelumnya. Meskipun sifat independen dari kata (istilah) atau parameter dalam dokumen tidak sepenuhnya terpenuhi, Naïve Bayes dapat diandalkan untuk klasifikasi dan bahkan lebih baik dalam hal realistis, kecepatan, dan akurasi [26]Bentuk umum teorema bayes adalah sebagai berikut:

$$\mathrm{P(H|X)} = \frac{P(H|\mathrm{X})\mathrm{P)H}}{P(X)} \quad (1)$$

   Dalam konteks klasifikasi menggunakan algoritma Naïve Bayes, terdapat beberapa komponen probabilistik yang penting untuk dipahami. Misalnya, X merepresentasikan data yang kelasnya belum diketahui, sedangkan H merupakan hipotesis bahwa data X termasuk ke dalam kelas tertentu. Tujuan utama dari metode ini adalah menghitung P(H|X), yaitu probabilitas hipotesis H berdasarkan kondisi X, yang disebut juga sebagai *posterior probability*. Untuk mendapatkan nilai ini, digunakan rumus Bayes yang memanfaatkan informasi P(H) atau *prior probability*, yakni probabilitas awal bahwa hipotesis H benar sebelum melihat data X—dan P(X|H), yaitu *likelihood*, yang menggambarkan seberapa besar kemungkinan data X muncul jika hipotesis H benar. Sementara itu, P(X) merupakan evidence atau probabilitas dari data X secara umum tanpa mempertimbangkan kelas tertentu. Semua komponen ini saling terkait dalam proses pengambilan keputusan klasifikasi oleh algoritma *Naïve Bayes*.
b. Klasifikasi Algoritma a*K-Nearest Neighbor* (KNN)
   Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut. KNN adalah sebuah metode untuk mencari kasusdengan menghitung kedekatan antara kasus baru dan kasus lama. Algoritma KNN adalah salah satu metode yang digunakan untuk analisis klasifikasi, namun metode KNN juga digunakan untuk prediksi. Jarak antara dua titik pada data training dan titik pada data testing dapat didefinisikan dengan rumus Euclidean. K-Nearest Neighbors (KNN) termasuk kedalam metode supervised learning yang mempunyai tujuan untuk memperoleh pola baru sedangkan metode unsupervised learning bertujuan untuk mendapatkan pola dalam sebuah kata, dalam hal ini berarti metode supervised learning sudah mempunya label dan yang dimanfaatkan untuk penelitian adalah mengklasifikan data (data testing) berdasarkan data yang sudah ada (data training)[27].

2.6 Tahap Hasil Dan Analisis

a. Pengujian *Confusion Matrix*
   Confusion matrix adalah metode yang digunakan untuk mengevaluasi kinerja algoritma klasifikasi dengan menyajikan hasil prediksi dalam bentuk matriks, yang memperlihatkan perbandingan antara kelas aktual dan kelas prediksiDalam confusion matrix, terdapat beberapa metrik evaluasi yang dapat dihitung, seperti akurasi, recall, precision, dan error. Metrik-metrik ini digunakan untuk mengukur seberapa baik kinerja suatu algoritma klasifikasi dalam mengklasifikasikan data. Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Untuk menghitung akurasi, digunakan confusion matrix sebagai dasar perhitungan [28]. Akurasi dihitung dengan rumus berikut:

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

Tabel 1. *Confusion Matrix*

No.		Positif	Field	Netral
1.	Positif	TPos	FposNeg	FposNet
2.	Negatif	FNegPos	Tneg	FnegNet
3.	Netral	FNetPos	FNetNeg	TNet

Tabel 1 diatas menggambarkan Confusion Matrix, yaitu alat untuk melihat performa model klasifikasi dalam memetakan data ke dalam tiga kategori sentimen: positif, negatif, dan netral. Misalnya, TPos menunjukkan jumlah data yang benar-benar berlabel positif dan berhasil diprediksi dengan benar oleh model. Sebaliknya, FposNeg berarti data yang sebenarnya positif namun salah diprediksi sebagai negatif. Dengan pola yang sama, kita bisa melihat seberapa sering model salah mengklasifikasikan data antar kelas. Untuk mengetahui seberapa baik model bekerja, digunakan tiga metrik utama:

1. Akurasi menunjukkan seberapa besar proporsi prediksi yang benar dari seluruh data. Rumusnya:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

Rumus ini memberikan gambaran umum seberapa sering model melakukan klasifikasi yang benar.

2. Presisi (Precision) mengukur ketepatan prediksi positif oleh model:

$$Presisi = \frac{TP}{FP+TP} \times 100\% \quad (3)$$

Presisi tinggi menunjukkan bahwa sebagian besar prediksi positif memang benar-benar positif.

3. Recall atau sensitivitas menunjukkan kemampuan model dalam menangkap semua data yang benar-benar positif. Ini penting jika kita ingin menghindari kesalahan karena meloloskan data positif.

$$Recall = \frac{TP}{FN+TP} \times 100\% \quad (4)$$

Metrik ini penting terutama ketika kesalahan mengabaikan data positif perlu dihindari.

Ketiga metrik ini bersama-sama memberikan gambaran menyeluruh tentang keakuratan dan ketepatan model dalam memahami sentimen pengguna.

b. Visualisasi *Wordcloud*

Hasil klasifikasi sentimen selanjutnya divisualisasikan menggunakan library WordCloud. WordCloud merupakan bentuk visualisasi data yang menampilkan kata-kata penting berdasarkan seberapa sering kata tersebut muncul dalam data.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dikumpulkan pada penelitian ini berasal dari komentar pengguna Dana di Google Playstore. Proses pengambilan data dilakukan dengan menggunakan teknik pengambilan data otomatis yaitu dengan teknik Crawling, yang diimplementasikan dengan bahasa pemrograman Python. Total jumlah data yang dikumpulkan adalah 1000 komentar. Adapun data yang diambil ialah data pada tahun 2024 (Januari - Desember 2024)

3.2 Proses *Pre-Processing*

Setelah data dikumpulkan, dilakukan tahapan pre-processing yang meliputi cleaning untuk menghilangkan karakter atau simbol tidak relevan, case-folding untuk mengubah teks menjadi huruf kecil, proses tokenizing untuk menguraikan teks ke bentuk satuan kata, stopword removal untuk menghapus kata-kata minim makna, dan stemming untuk mengembalikan kata ke bentuk dasar.

Tabel 2. *Pre-Processing*

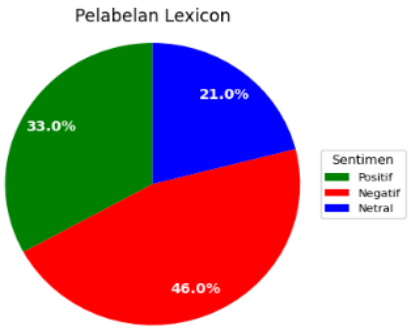
No.	Proses	Hasil
1.	Data awal	Makasih Dana Dengan Adanya Apk Ini Ngebantu Bgt Lebih Mudah Kalau Mau Tf Kemanapun Dan KeSesama EWallet Pun Bisa Itu Aku Senangnya Pakai Apk Ini Udah Gitu Ga Biaya Cashðÿ¤—Ðÿ¤—Ðÿ¤—Ðÿ™□
2.	*Cleaning*	Makasih Dana Dengan Adanya Apk ini Ngebantu Banget Lebih Mudah Kalau Mau Tf Kemanapun dan Kesesama Ewalletpun pun itu Aku Senengnya Pakai Apk Dah Gitu Ga Biaya Cash
3.	*Case Folding*	makasih dana dengan adanya apk ngebantu baget lebih mudah kalau mau tf kemanapun kesesama ewalletpun itu aku senengnya pakai apk dah gitu ga biaya cash
4.	*Tokenizing*	[makasih,dana,dengan,adanya,apk,ngebantu,baget,lebih,mudah,kalau,mau,tf,kekmanapun,kesesama,ewalletpun,itu,aku,senengnya,pakai,apk,dah,gitu,ga,biaya,cash

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

No.	Proses	Hasil
5.	*Stopword*	[makasih,dana,dengan,apk,ngebantu,banget,mudah,kalau,mau,tf,mana,kesesama,ewalletpun,senang,pakai,apk,ga,biaya,cash]
6.	*Stemming*	makasih dana dengan apk ngebantu banget mudah kalau mau tf mana kesesama ewalletpun senang pakai apk ga biaya cash

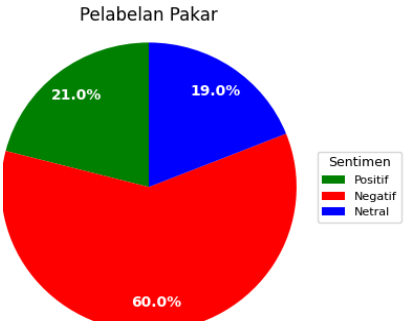
3.3 Pelabelan

Dari hasil pelabelan menggunakan dua cara yaitu, menggunakan bahasa lexicon dan manual oleh pakar sebanyak 1000 data. Pada program bahasa lexicon didapatkan pada komentar pada google playstore memiliki 542 sentimen negatif, 358 sentimen positif, dan 100 sentimen netral. Sedangkan Pada manual oleh pakar didapatkan pada komentar pada google playstore memiliki 600 sentimen negatif, 210 sentimen positif, dan 190 sentimen netral. persentase keberhasilan pelabelan setiap kategori dapat dilihat secara rinci pada Gambar 2 dan Gambar 3.



Gambar 3. Pelabelan Lexicon

Pada Gambar 3, dapat dilihat bahwa hasil dari pelabelan lexicon yang berwarna hijau atau sentimen positif mendapatkan hasil sebesar 33%, yang berwarna merah dengan sentimen negatif sebesar 46% dan yang berwarna biru hanya mendapatkan hasil sebesar 21%. Dimana pada pelabelan lexicon ini yang mendominasi dari hasil setiap sentimen adalah yang berwarna merah atau sentimen negatif.



Gambar 4. Pelabelan Pakar

Pada Gambar 4, dapat dilihat bahwa hasil dari pelabelan pakar yang berwarna hijau atau sentimen positif mendapatkan hasil sebesar 21%, yang berwarna merah dengan sentimen negatif sebesar 60% dan yang berwarna biru hanya mendapatkan hasil sebesar 19%. Dimana pada pelabelan pakar ini juga yang mendominasi dari hasil setiap sentimen adalah yang berwarna merah atau sentimen negatif.

3.4 Pembobotan TF-IDF

Tahap pembobotan TF-IDF untuk mendapatkan bobot dari kata. Hasil yang diperoleh setelah penerapan proses ini terdapat pada Tabel 3.

Tabel 3. Bobot TF-IDF

No	admin	akifitas	akun	apilkasi	dana	transaksi	uang
1	0	0	0	0	0.159916	0	0
2	0	0	0.219455	0	0	0	0
3	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…
1000	0	0	0	0	0.068357	0	0

Setelah semua data ulasan melalui tahap praproses, langkah selanjutnya adalah pemberian bobot kata menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*). Metode ini membantu mengukur

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

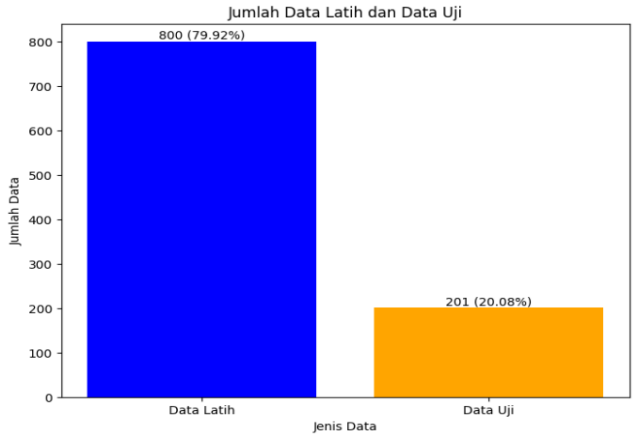
seberapa penting sebuah kata dalam satu ulasan dibandingkan dengan seluruh ulasan lainnya. Kata yang sering muncul dalam satu ulasan tapi jarang muncul di ulasan lain akan mendapatkan bobot lebih tinggi, karena dianggap lebih mewakili isi ulasan tersebut. Hasil dari proses ini ditampilkan.

Pada Tabel 3, yang menunjukkan bobot TF-IDF dari beberapa kata penting seperti "dana", "akun", dan "transaksi" di setiap ulasan. Sebagai contoh, pada ulasan pertama, kata "dana" memiliki bobot sebesar 0,1599, menandakan kata tersebut cukup penting dalam konteks ulasan tersebut. Sementara itu, kata-kata lain seperti "admin" atau "aktifitas" tidak memiliki bobot karena tidak muncul dalam ulasan tersebut atau dianggap kurang relevan. Dengan metode ini, sistem dapat lebih fokus pada kata-kata yang benar-benar mencerminkan isi dan sentimen dari masing-masing ulasan, sehingga hasil analisis menjadi lebih akurat.

3.5 Klasifikasi *Algoritma Naïve Bayes* Classifier *(NBC) dan* K-*Nearest Neighbor (KNN)*

3.5.1 Pembagian Data

Setelah dataset dilakukan proses pre-processing data, pelabelan, dan perhitungan TF-IDF, tahapan selanjutnya ialah melakukan proses klasifikasi menggunakan model NBC dan KNN. Sebelum tahapan klasifikasi dilakukan, Data dipisahkan menjadi dua bagian, yaitu data latih dan data uji, menggunakan metode hold-out validation. Dalam penelitian ini, pembagian data dilakukan dengan proporsi 80:20, yang berarti 80% (800) data dimanfaatkan sebagai data pelatihan untuk pembelajaran model, Sedangkan 20% (200) dari data digunakan untuk data uji dalam mengevaluasi performa model.



Gambar 5. Pembagian Data

3.5.2 Tahap Pengujian

Setelah melalui serangkaian proses preprocessing data, tahap selanjutnya adalah melakukan pengujian model untuk mengukur performa algoritma yang digunakan. Dalam penelitian ini, algoritma yang diuji meliputi Naïve Bayes dan KNN. Dengan membandingkan bahasa lexicon dan pakar. Dan proses pengujian, data dibagi menjadi dua bagian, yaitu 80% sebagai data latih (training data) dan 20% sebagai data uji (testing data). Tahapan pengujian dilakukan dengan Pengujian Confusion Matrix dan teknik K-Fold Cross Validation.

Tabel 4. Perbandingan Akurasi

Matriks	Naïve Bayes (Lexicon)	Naïve Bayes (Pakar)	KNN (Lexicon)	KNN (Pakar)
Accuracy	0.72	0.70	0.67	0.69
Sentimen Negatif				
Precision	0.73	0.74	0.64	0.68
Recall	0.90	0.90	0.96	0.97
F1- Score	0.81	0.81	0.77	0.80
Sentimen Positif				
Precision	0.71	0.66	0.78	0.67
Recall	0.62	0.64	0.35	0.42
F1- Score	0.66	0.65	0.49	0.52
Sentimen Netral				
Precision	0.00	0.36	0.00	0.00
Recall	0.00	0.12	0.00	0.00
F1- Score	0.00	0.18	0.00	0.00

Berdasarkan dari Tabel 4, semua algoritma menunjukkan tingkat akurasi yang bervariasi, berkisar antara 0,67 hingga 0,72. Penerapan pada pakar mengurasi akurasi pada model Naïve bayes dan meningkatkan pada KNN, Naïve Bayes (NB) mengalami penurunan dari 0.72 menjadi 0.70, sedangkan KNN meningkat dari 0.67 menjadi 0.69.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

Untuk sentimen negatif, accuracy tidak mengalami signifikan . Pada precision, recall, dan f1 score di semua algoritma. Misalnya, dalam NB meningkat dari 0.73 menjadi 0.74 pada precision, Recall tetap pada 0.90, dan F1-score tetap pada 0.81. sedangkan dalam KNN, meningkat dari 0.64 menjadi 0.68 pada precision, meningkat dari 0.96 menjadi 0.97 pada Recall, dan meningkat dari 0.77 menjadi 0.80 pada F1-score.

Untuk sentimen positif, accuracy mengalami penurunan. Pada NB meningkat dari 0.71 menjadi 0.66 pada precision, meningkat dari 0.62 menjadi 0.64 pada Recall, dan penurunan dari 0.66 menjadi 0.65 pada F1-score. sedangkan dalam KNN, penurunan dari 0.78 menjadi 0.67 pada precision, meningkat dari 0.35 menjadi 0.42 pada Recall, dan meningkat dari 0.49 menjadi 0.52 pada F1-score.

Untuk sentimen netral, accuracy tetap pada posisi 0.00. Pada precision, recall, dan f1 score di semua algoritma naïve bayes dan KNN. gambar di bawah ini adalah hasil K-Fold Cross Validation.

Tabel 5. Hasil Akurasi pada *K-Fold Cross Validation*

Nilai K	Nilai Akurasi			
	Naïve Bayes (Lexicon)	Naïve Bayes (Pakar)	KNN (Lexicon)	KNN (Pakar)
K = 1	0.7327	0.7327	0.6436	0.5347
K = 2	0.7327	0.6400	0.5000	0.4600
K = 3	0.7100	0.7000	0.5800	0.5100
K = 4	0.6500	0.6300	0.5200	0.4400
K = 5	0.7200	0.6100	0.5200	0.4400
K = 6	0.7100	0.6800	0.5800	0.4600
K = 7	0.7600	0.6900	0.5300	0.4700
K = 8	0.7300	0.7700	0.6500	0.5500
K = 9	0.6900	0.7700	0.5600	0.4800
K = 10	0.7400	0.7400	0.5600	0.5100
Rata – rata akurasi	0.7183	0.6893	0.5644	0.4855

Pada Tabel 5, Untuk mengukur keakuratan model dalam mengklasifikasikan sentimen, penelitian ini menggunakan metode *K-Fold Cross Validation* dengan nilai K dari 1 hingga 10. Hasilnya menunjukkan bahwa *Naïve Bayes* secara konsisten memberikan akurasi yang lebih tinggi dibandingkan K-*Nearest Neighbor* (KNN). Naïve Bayes dengan pelabelan *lexicon* mencatat rata-rata akurasi tertinggi sebesar 71,83%, sedangkan versi dengan pelabelanpakar berada di angka 68,93%. Akurasi terbaik muncul pada K = 7 dan K = 10, masing-masing mencapai 76% dan 74%. Sementara itu, KNN menunjukkan performa yang lebih rendah. Dengan pelabelan lexicon, rata-rata akurasinya hanya 56,44%, dan turun lagi menjadi 48,55% pada pelabelan pakar. Hal ini menunjukkan bahwa Naïve Bayes lebih efektif dan stabil dalam menganalisis sentimen dari ulasan pengguna aplikasi DANA.

Dari hasil ini, dapat disimpulkan bahwa pemilihan algoritma dan metode pelabelan berpengaruh besar terhadap akurasi klasifikasi. Naïve Bayes menjadi pilihan yang lebih unggul untuk tugas analisis sentimen berbasis teks. Hasil Akurasi pada K-Fold Cross Validation terdiri dari algoritma Naïve Bayes dan KNN pada penilaian lexicon dan pakar dapat di simpulkan rata-rata akurasi pada Naïve bayes lexicon memiliki nilai 0.7183 sedangkan pakar memiliki nilai 0.6893. sedangkan KNN rata-rata akurasi nilai sebesar 0.5644 sedangkan pakar 0.4855.

3.6 Visualisasi Data Sentimen

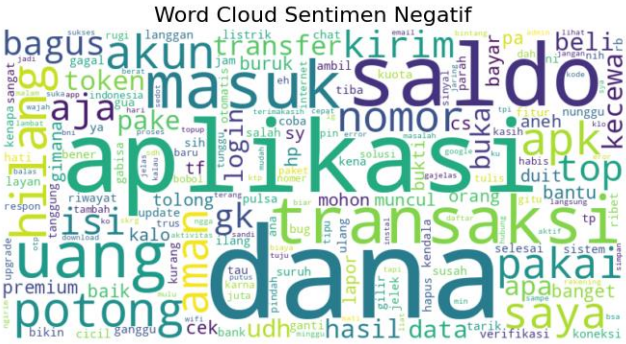
Visualisasi data bertujuan untuk mengidentifikasi topik yang sering dibahas oleh pengguna Dana. Dengan memanfaatkan teknik ini, kata yang tersembunyi dalam ribuan komentar dapat diungkap secara lebih jelas dan ringkas sehingga dapat menghasilkan informasi penting terhadap Kepuasan pengguna Dana.

Gambar 6. Hasil Visualisasi Sentimen Positif

Gambar word cloud di atas menunjukkan kata-kata yang paling sering muncul dalam ulasan positif pengguna aplikasi DANA. Kata seperti “mudah”, “transaksi”, “dana”, “bantu”, dan “uang” mendominasi, yang menandakan bahwa pengguna merasa aplikasi ini memudahkan aktivitas keuangan mereka. Ulasan juga banyak menyebut “baik”, “bagus”, dan “praktis”, mencerminkan kepuasan terhadap layanan. Munculnya kata “saldo”, “akun”, dan “cepat”

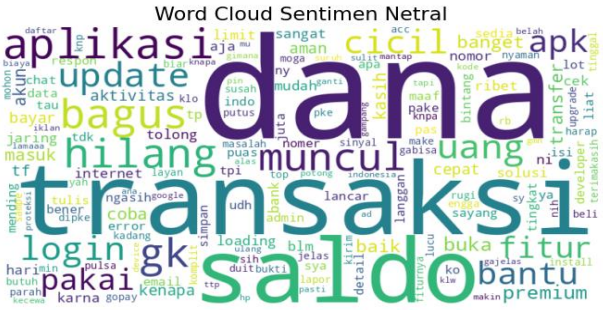
Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

menunjukkan bahwa fitur-fitur utama DANA dianggap efisien dan responsif. Word cloud ini menggambarkan kesan positif pengguna terhadap kemudahan, kecepatan, dan bantuan yang diberikan aplikasi.



Gambar 7. Hasil Visualisasi Sentimen Negatif

Pada gambar 7 word cloud diatas menunjukkan kata-kata yang paling sering muncul dalam ulasan negatif pengguna aplikasi DANA. Kata seperti “aplikasi”, “dana”, “saldo”, “masuk”, dan “transaksi” mendominasi, menandakan keluhan terkait saldo bermasalah, kesulitan login, dan transaksi gagal. Kata “potong”, “hilang”, dan “error” juga mencerminkan masalah teknis yang sering dialami. Selain itu, munculnya kata “lapor” dan “kecewa” menunjukkan rasa frustrasi pengguna terhadap layanan atau respons aplikasi. Visualisasi ini mengindikasikan bahwa ulasan negatif umumnya berkaitan dengan gangguan sistem dan pengalaman pengguna yang tidak memuaskan.



Gambar 8. Hasil Visualisasi Sentimen Netral

Pada Gambar 8 di atas menunjukkan kata-kata yang sering muncul dalam ulasan netral pengguna aplikasi DANA, seperti “dana”, “transaksi”, “saldo”, dan “aplikasi”. Ulasan ini umumnya bersifat deskriptif, menyampaikan pengalaman penggunaan tanpa penilaian emosional. Kata seperti “login”, “hilang”, dan “muncul” menandakan laporan kondisi aplikasi secara objektif, baik terkait fitur maupun kendala ringan.

Berdasarkan hasil keseluruhan analisis sentimen terhadap 1000 data komentar ditemukan pada Word Cloud Sentimen Positif Kata-kata yang menonjol adalah "dana", "mudah", "transaksi", "bantu", "uang", "tolong", dan "saldo". Ini menunjukkan bahwa pengalaman positif sering terkait dengan kemudahan transaksi, bantuan, atau pengelolaan dana/saldo. Sedangkan Word Cloud Sentimen Negatif Kata-kata yang sangat dominan adalah "dana", "aplikasi", "saldo", "transaksi", "masuk", "uang", "hilang", dan "potong". Kemunculan kata-kata seperti "hilang" dan "potong" sangat mengindikasikan masalah serius terkait dana, aplikasi, dan transaksi. Dan Word Cloud Sentimen Netral Kata-kata yang terlihat menonjol adalah "dana", "transaksi", "saldo", "aplikasi", "bagus", dan "gk". Kata-kata ini cenderung lebih umum dan tidak secara eksplisit mengungkapkan kepuasan atau ketidakpuasan yang kuat.

4. KESIMPULAN

Penelitian ini berhasil mengklasifikasikan sentimen pengguna aplikasi DANA berdasarkan ulasan di Google Play Store dengan menggunakan dua algoritma pembelajaran mesin, yakni Naïve Bayes Classifier (NBC) dan K-Nearest Neighbor (KNN). Data sebanyak 1.000 ulasan diproses melalui tahapan crawling, pre-processing teks, pelabelan sentimen, ekstraksi fitur menggunakan TF-IDF, serta pembagian data latih dan uji. Hasil analisis menunjukkan bahwa NBC secara konsisten memiliki performa lebih tinggi dibandingkan KNN, baik berdasarkan pelabelan lexicon maupun manual oleh pakar. NBC mencatat akurasi rata-rata sebesar 71,83% (lexicon) dan 68,93% (pakar), sementara KNN hanya mencapai 56,44% (lexicon) dan 48,55% (pakar). Pada klasifikasi sentimen negatif, NBC dan KNN menunjukkan presisi dan recall tinggi, tetapi keduanya belum optimal dalam mendeteksi sentimen netral. Visualisasi word cloud memperkuat hasil klasifikasi, dengan kata dominan seperti “mudah” dan “bantu” pada sentimen positif, serta “hilang” dan “potong” pada sentimen negatif. Penelitian ini menyimpulkan bahwa Naïve Bayes lebih unggul dalam klasifikasi sentimen berbasis teks untuk aplikasi e-wallet. Adapun keterbatasan terdapat pada distribusi data

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

sentimen yang tidak seimbang dan kinerja rendah pada kategori netral. Penelitian selanjutnya disarankan untuk mengeksplorasi metode ensemble atau deep learning serta memperbaiki teknik pelabelan otomatis.

REFERENCES

[1] B. Filemon, V. C. Mawardi, dan N. J. Perdana, "Penggunaan Metode Support Vector Machine untuk Klasifikasi Sentimen E-Wallet," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 10, no. 1, Mar 2022, doi: 10.24912/jiksi.v10i1.17824.

[2] Z. H. Pradana dan S. Aminah, "The Influence of Perceived Ease of Use and Promotion on Interest in Using OVO E-Wallet in Surabaya City," *Indonesian Journal of Business Analytics*, vol. 3, no. 5, hlm. 1827–1836, Okt 2023, doi: 10.55927/ijba.v3i5.5657.

[3] Nurian A dan Sari B.N, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, hlm. 2830–7062, 2023, doi: 10.23960/jitet.v11i3%20s1.3348.

[4] V. F. Calista dan H. Wandebori, "Ekombis Review-Jurnal Ilmiah Ekonomi dan Bisnis Proposed Marketing Strategy For E-Wallet To Increase The Growth Of Active Users (Case Study: Sap Cash) ARTICLE HISTORY," *Ekombis Review: Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 1, hlm. 129–144, 2024, doi: 10.37676/ekombis.v13i1.

[5] F. Farahdinna, P. Hari Wira Atmaja, J. Sawo Manila, dan P. Ps Minggu Jakarta, "Pemanfaatan Machine Learning dalam Analisis Sentimen Pengguna E-Wallet Menggunakan SVM," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 2, hlm. 3522–3526, Apr 2025, Diakses: 16 April 2025. [Daring]. Tersedia pada: https://ejournal.itn.ac.id/index.php/jati/article/download/13526/7436

[6] W. Bian Lin William Cong Yang Ji *dkk.*, "The Rise of E-Wallets and Buy Now Pay-Later: Payment Competition, Credit Expansion, and Consumer Behavior," *National Bureau Of Economic Research*, Mei 2023, [Daring]. Tersedia pada: https://www.juniperresearch.com/press/digital-wallet-users-exceed-5bn-globally-2026.

[7] A. Rahman, E. Utami, dan S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, hlm. 60–71, Jul 2021, doi: 10.31603/komtika.v5i1.5188.

[8] J. Yuan Mambu, L. Mea, E. Sumanto, dan E. Lompoliu, "Quality Analysis of DANA Application on the Customer Satisfaction using MS-Qual," *Ministry of Education and Culture Accreditation*, vol. 24, no. 2, hlm. 144–151, Sep 2022, doi: 10.31294/p.v24i2.1383.

[9] N. A. Ningtyas dan A. Meiriza, "Penerapan Metode System Usability Scale Dalam Mengevaluasi User Experience Aplikasi DANA," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, hlm. 667, Apr 2023, doi: 10.30865/jurikom.v10i2.6083.

[10] M. P. Cristescu, R. A. Nerişanu, dan D. A. Mara, "Using Data Mining in the Sentiment Analysis Process on the Financial Market," *Journal of Social and Economic Statistics*, vol. 11, no. 1–2, hlm. 36–58, Des 2022, doi: 10.2478/jses-2022-0003.

[11] Siti Masturoh dan Achmad Baroqah Pohan, "Sentiment Analysis Against the DANA E-Wallet on Google Play Reviews Using the K-Nearest Neighbor Algorithm," *PILAR Nusa Mandiri*, vol. 17, no. 1, hlm. 53–57, Mar 2021, doi: 10.33480/pilar.v17i1.2182.

[12] F. A. Larasati, D. E. Ratnawati, dan B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, hlm. 4305–4313, 2022, [Daring]. Tersedia pada: http://j-ptiik.ub.ac.id

[13] S. Nada Apsariny, N. Chamidah, E. Ana, dan A. Kurniawan, "Sentiment Analysis of User Reviews Based on Naive Bayes Classifier Algorithm With Hyperparameter Optimization: A Case Study on Application 'KREDIT PINTAR,'" *Jurnal Ilmiah Indonesia*, vol. 7, no. 1, hlm. 1139–1150, Jan 2022.

[14] Y. Mao, Q. Liu, dan Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, Apr 2024, doi: 10.1016/j.jksuci.2024.102048.

[15] P. Warakmulty, D. Yeffry, dan H. Putra, "Optimalisasi Manajemen Sentimen di Media Sosial Universitas melalui Machine Learning dan AI: Studi Kasus pada Komentar Instagram," *Jurnal Tata Kelola dan Kerangka Kerja Teknologi informasi*, vol. 11, no. 1, hlm. 31–38, 2025, Diakses: 19 Juli 2025. [Daring]. Tersedia pada: https://ojs.unikom.ac.id/index.php/JTK3TI

[16] F. M. Pagi dan N. I. Widiastuti, "Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi TINDER dengan Metode Support Vector Machine," *Jurnal Ilmiah Komputer dan Informatika*, vol. 13, no. 2, hlm. 114–122, 2024, Diakses: 19 Juli 2025. [Daring]. Tersedia pada: https://ojs.unikom.ac.id/index.php/komputa/article/download/14078/4727

[17] Agus Suwarno, Andriani, dan Syaiful Rohman, "Analisis Sentimen Pada Media Sosial Twitter Mengenai Tanggapan Vaksinasi COVID-19 Menggunakan Metode Naive Bayes," *Jurnal Teknik Industri*, vol. 2, no. 1, hlm. 22–29, 2021, [Daring]. Tersedia pada: https://t.co/xGaLHsjizn

[18] A. Saepulrohman, S. Saepudin, dan D. Gustian, "Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *Accounting Information Systems and Information Technology Business Enterprise* , vol. 6, no. 2, hlm. 91–105, Des 2021, doi: 10.34010/aisthebest.v6i2.4919.

[19] A. R. Isnain, J. Supriyanto, dan M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 2, hlm. 121, Apr 2021, doi: 10.22146/ijccs.65176.

[20] S. S. Hasibuan, A. Angraini, E. Saputra, dan M. Megawati, "Sentimen Analisis Terhadap Fitur Tiktok Shop Menggunakan Naive Bayes dan K-Nearest Neighbor," *Jurnal Media informatika Budidarma*, vol. 8, no. 1, hlm. 303–311, Jan 2024, doi: 10.30865/mib.v8i1.7238.

[21] A. Mustofa dan R. Novita, "Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun 2022, doi: 10.47065/bits.v4i1.1628.

[22] S. N. S. Muslim, F. Nurdiyansyah, dan A. Y. Rahman, "Perbandingan Algoritma Naive Bayes dan KNN dalam Analisis Sentimen Ulasan Pengguna Aplikasi CapCut," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Okt 2024, doi: 10.23960/jitet.v12i3s1.5156.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1111−1121
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7861

[23] S. F. Amrillah, D. Krisbiantoro, dan A. Prasetyo, "Penerapan Metode K-Nearest Neighbors dan Naïve Bayes pada Analisis Sentimen Pengguna Aplikasi Bstation melalui Platform Playstore," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, hlm. 1281–1292, Des 2024, doi: 10.47065/bits.v6i3.5863.

[24] A. Oktian Permana dan Sudin Saepudin, "Perbandingan algoritma k-nearst neighbor dan naïve bayes pada aplikasi shopee," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, hlm. 25–32, Apr 2023, doi: 10.37859/coscitech.v4i1.4474.

[25] S. Syafrizal, M. Afdal, dan R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, hlm. 10–19, Des 2023, doi: 10.57152/malcom.v4i1.983.

[26] Muhammad Rayhan Elfansyah, Rudiman, dan Fendy Yulianto, "Perbandingan Metode K-Nearest Neighbor (KNN) DAN Naive Bayes Terhadap Analisis Sentimen Pada Pengguna E-Wallet Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF," *JURNAL TEKNOLOGI INFORMASI*, vol. 18, no. 2, hlm. 139–159, 2024, doi: https://doi.org/10.47111/JTI.

[27] K. Mustaqim, F. A. Amaresti, dan I. N. Dewi, "Analisis Sentimen Ulasan Aplikasi PosPay untuk Meningkatkan Kepuasan Pengguna dengan Metode K-Nearest Neighbor (KNN)," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, hlm. 11–20, Jun 2024, doi: 10.29408/edumatic.v8i1.24779.

[28] N. Nur, Asmawati, dan N. Syahra, "Perbandingan Metode k-NN dan Naïve Bayes dalam Klasifikasi Penentuan Calon Pendonor Darah," *Journal of Computer and Information System (J-CIS)*, vol. 1, no. 1, hlm. 21–28, Agu 2021, doi: 10.31605/jcis.v1i1.875.