

ANALISIS SENTIMEN ULASAN APLIKASI DANA DI GOOGLE PLAY STORE MENGUNAKAN ALGORITMA NAÏVE BAYES

Gery Riansyah Ramadhan, Castaka Agus Sugianto

Program Studi Teknik Informatika, Politeknik TEDC Bandung

Jl. Politeknik Jl. Pesantren No.2, Cibabat, Kec. Cimahi Utara, Kota Cimahi, Jawa Barat 40513

geryriansyah@gmail.com

ABSTRAK

Dalam era digital, aplikasi dompet *digital* seperti DANA menjadi penting terutama selama pandemi COVID-19 yang meningkatkan transaksi non-tunai. Penelitian ini menganalisis sentimen ulasan aplikasi DANA di *Google Play Store* menggunakan metode *Naïve Bayes Classifier*. Data sebanyak 2000 ulasan berbahasa Indonesia dikumpulkan melalui *web scraping* dan dilabeli sebagai positif, negatif, atau netral berdasarkan rating. Proses penelitian menggunakan metode *Knowledge Discovery from Data* (KDD) yang melibatkan tahapan seleksi data, *preprocessing*, pembobotan, klasifikasi, dan evaluasi. *Preprocessing* dilakukan untuk membersihkan dan menyiapkan data. *TF-IDF* untuk pembobotan kata, kemudian model *Naïve Bayes* digunakan untuk klasifikasi sentimen. Hasil evaluasi menggunakan *confusion matrix* menunjukkan performa model yang baik. Terdapat 1476 ulasan positif, 443 ulasan negatif, dan 81 ulasan netral dengan akurasi 86%, presisi 86%, recall 98%, dan *f1-score* 92%. Studi ini menyimpulkan bahwa model analisis sentimen ini mampu memberikan prediksi yang akurat, meskipun terdapat variasi dalam nilai presisi, *recall*, dan *f1-score*. Kata yang sering muncul pada ulasan negatif terkait dengan masalah masuk aplikasi, transaksi gagal, dan isu pengisian saldo, yang merupakan area yang perlu diperhatikan oleh pengembang aplikasi DANA.

Kata kunci : Analisis Sentimen, Ulasan, DANA, Google Play Store, Naïve Bayes

1. PENDAHULUAN

Di era industri 4.0, teknologi informasi dan komunikasi mengalami perkembangan pesat. Era ini ditandai dengan kompleksitas kegiatan, terutama dalam dunia bisnis. Salah satu kemudahan yang ditawarkan adalah transaksi *online*, yang membuat banyak orang beralih menggunakan uang elektronik (*E-Money*). *E-Money* memiliki pengaruh besar pada proses pemasaran *digital* dan berbagai bidang lainnya, termasuk transaksi *online* yang mempermudah jual beli secara *online*, sehingga terciptalah aplikasi dompet *digital* [1].

Di era *digital* saat ini, jumlah pengguna aplikasi *mobile* terus meningkat, sehingga penting untuk memahami tingkat kepuasan dan ketidakpuasan mereka saat menggunakan aplikasi. *Google* menyediakan layanan bernama *Play Store* yang menawarkan berbagai konten *digital*, seperti game, aplikasi, film, musik, dan buku dalam beragam kategori. Salah satu fitur di *Play Store* adalah *rating* dan ulasan, di mana pengguna dapat memberikan opini mereka tentang produk yang telah mereka gunakan [2].

Di tahun 2020, dunia mengalami sebuah wabah atau penyakit yaitu COVID-19, yang mengakibatkan adanya batasan kontak fisik antar manusia. Adanya aplikasi dompet *digital* atau *e-money* menjadi populer karena mampu mempermudah atau mengatasi masalah dengan transaksi yang dijalankan secara *online*. Aplikasi dompet *digital* yang populer disaat pandemi yaitu Dana [3]. Aplikasi Dana diluncurkan pada tahun 2018 memungkinkan pengguna melakukan transaksi keuangan non-tunai mulai dari pembayaran kebutuhan sehari-hari hingga *top up* saldo investasi. Berdasarkan

survei YouGov, DANA adalah dompet *digital* tercepat dalam peningkatan pengguna dengan pertumbuhan tercepat sekitar 4% di masa COVID-19 [4].

Aplikasi DANA memiliki banyak pengguna, sehingga sering kali terdapat ulasan positif, negatif dan netral yang tidak relevan dengan *rating* yang diberikan di *Google Play Store* [5].

Dalam penelitian ini, analisis sentimen dilakukan pada tingkat kalimat dengan menggunakan metode *Naïve Bayes Classifier*. Analisis sentimen biasanya berhubungan dengan perasaan, emosi, sikap, atau pendapat. Karena data yang digunakan adalah data tekstual, maka diperlukan analisis untuk mengekspresikan sentimen dan menghitung wawasan guna mengeksplorasi bisnis.

Metode *Naïve Bayes* adalah metode yang banyak digunakan karena metode ini dapat menghasilkan akurasi yang baik dan stabil [6].

Penelitian mengenai analisis sentimen sudah banyak dilakukan sebelumnya, seperti penelitian mengenai analisis sentimen yang membahas opini publik mengenai COVID-19 pada *twitter* menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbour*, *Naïve Bayes* memiliki nilai akurasi lebih tinggi [7][8].

Adapun penelitian lainnya mengenai analisis sentimen yang membahas ulasan pada *e-commerce* *Shopee* menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*, akurasi yang di dapatkan untuk metode *Naïve Bayes* menghasilkan nilai *accuracy* sebesar 85%. Metode *Naïve Bayes* menghasilkan hasil akurasi dan *AUC* yang lebih baik dari metode *Support Vector Machine*. Metode *Support*

Vector Machine hanya mendapatkan nilai akurasi sebesar 81% [9].

Analisis sentimen adalah proses untuk menentukan apakah suatu ulasan bersifat positif, negatif, atau netral. Opini *mining* tidak memerhatikan topik teks, melainkan lebih fokus pada ekspresi yang disampaikan dalam teks opini tersebut. Proses ini digunakan untuk menentukan apakah ulasan atau review tentang aplikasi DANA di *Google Play Store* termasuk dalam kategori opini positif, negatif, atau netral [10].

2. TINJAUAN PUSTAKA

2.1. Text Mining

Text mining adalah penemuan dan ekstraksi pengetahuan menarik yang signifikan dari teks bebas atau tidak terstruktur. Dengan pengetahuan yang berasal dari pola dan hubungan dan digunakan untuk mengungkapkan fakta, tren atau ide. Selain itu, *text mining* juga diperlukan untuk mengubah teks menjadi data kemudian teks data mining lainnya digunakan untuk analisis. Dengan banyaknya sumber data yang mana untuk menganalisisnya secara manual maka diperlukan *text mining* untuk menangani data tersebut. Namun, perlu dilakukan identifikasi dan pemisahan jenis informasi tertentu dari teks mengingat perlunya teknik atau metode *text mining* untuk membantu dalam mengelompokkan data ke dalam kelompok yang berbeda sesuai dengan kebutuhan tertentu [11].

2.2. Analisis Sentimen

Menurut Rozi, dkk, analisis sentimen atau *opinion mining* adalah proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk memperoleh informasi emosional yang terkandung dalam kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan seseorang tentang suatu masalah atau objek. Tidak masalah jika orang tersebut memiliki opini negatif atau positif [12].

2.3. Ulasan

Menurut Kamus Besar Bahasa Indonesia (KBBI), ulasan adalah kupasan atau tafsiran atau komentar. Sedangkan menurut situs Kemdikbud, teks ulasan adalah suatu teks yang berisi ulasan, penilaian/*review* terhadap suatu karya seperti film, drama, atau sebuah buku. Teks ulasan berisi naskah yang memuat tulisan untuk memberikan komentar tentang suatu objek berdasarkan pandangan seseorang. Teks ulasan sering disebut juga dengan resensi [13].

2.4. DANA

Dana adalah aplikasi dompet digital berasal dari Indonesia yang terdapat di *Google Play Store* digunakan untuk pembayaran *digital cashless* dan *cardless* baik *online* maupun *offline*. Dana menyediakan fitur mulai dari pembelian pulsa, tagihan listrik, hingga *top-up* saldo investasi. Tujuannya adalah agar mempermudah pengguna dalam

bertransaksi dengan cepat, nyaman dan tetap aman [14].

2.5. Google Play Store

Google Play Store merupakan platform untuk konten *digital*, yang mengizinkan pengguna *smartphone android* pada mengunduh aplikasi dan produk *digital* lainnya, seperti *Google Play Store* adalah salah satu platform aplikasi android terbesar dan terpopuler. *Google Play Store* menyediakan mekanisme bagi pengguna aplikasi android untuk menilai aplikasi dan memberikan umpan balik dari ulasan pengguna. Ulasan pengguna sering kali menyertakan informasi penting seperti laporan *bug* dan permintaan fitur yang dapat digunakan pengembang untuk meningkatkan dan memelihara aplikasi mereka game, film, *e-book*, dan sebagainya, dari yang berbayar hingga yang gratis [15].

2.6. Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah algoritma pembelajaran mesin yang digunakan untuk masalah klasifikasi, terutama dalam klasifikasi teks yang melibatkan kumpulan data pelatihan berdimensi tinggi. Contoh penggunaannya termasuk analisis sentimen, penyaringan spam, dan klasifikasi teks. Algoritma ini dikenal karena kesederhanaannya dan efektivitasnya. Naïve Bayes memungkinkan pembangunan model dengan cepat, menjadikannya salah satu algoritma prediksi tercepat untuk dipelajari dan digunakan. Algoritma ini menghitung probabilitas suatu objek dengan asumsi bahwa kemunculan satu fitur tidak tergantung pada fitur lain, meskipun dalam kenyataannya fitur-fitur tersebut bisa saling bergantung. Inilah mengapa disebut "naïve". Algoritma ini didasarkan pada teorema dasar yang dikenal sebagai aturan Bayes atau Hukum Bayes, yang merupakan metode untuk menghitung probabilitas kondisional, yaitu probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya yang tersedia. Algoritma ini dinamai dari ahli statistik dan filsuf Thomas Bayes [16].

2.7. Web Scraping

Web scraping adalah proses pengambilan sebuah dokumen semi terstruktur dari internet, umumnya berupa halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen dari halaman tersebut untuk diambil data tertentu untuk digunakan bagi kepentingan lain [17].

2.8. Google Colab

Google Colab adalah platform *cloud computing* yang mirip dengan *Jupyter Notebook* dan *Google Research*. *Google Colab* memungkinkan pengguna untuk menulis dan mengeksekusi kode program *Python* secara acak hanya dengan menggunakan browser web. Sehingga pengguna dapat memanfaatkan sumber daya server untuk memproses dan menganalisis data dengan cepat [18].

2.9. Python

Python adalah bahasa pemrograman populer yang bisa digunakan untuk berbagai tujuan, seperti membangun *website*, menganalisis data dalam *data science*, proses *scripting*, hingga pembuatan game. Python merupakan bahasa pemrograman *open source*, sehingga Anda bisa menggunakannya secara gratis dan bahkan ikut berkontribusi dalam pengembangannya. Banyak *programmer* sepakat bahwa Python adalah bahasa pemrograman yang interpretatif dan serbaguna. Sintaksnya yang mudah dibaca dan dipahami membuatnya cocok dipelajari oleh pemula [19].

2.10. API Google-Play-Scraper

API Google-Play-Scraper adalah metode untuk mengambil data dari Google Play Store tanpa dependensi eksternal menggunakan bahasa pemrograman *python*. Data yang diambil dapat berupa informasi aplikasi seperti judul aplikasi, *developer url*, kategori aplikasi, keseluruhan *rating* dan *review*, deskripsi, *thumbnail*, *rating* konten dan *screenshot* aplikasi. Selain informasi dari aplikasi API *google-play-scraper* juga dapat mengambil data ulasan pengguna seperti nama, foto, *rating*, tanggal, *comment likes*, dan *comment* [20].

Berikut ini langkah-langkah dalam Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 4308 Fakultas Ilmu Komputer, Universitas Brawijaya mengambil data ulasan di Google Play Store menggunakan API Google-Play-Scraper [5]:

- Melakukan install API Google-Play-Scraper dengan menjalankan 'pip install google-playscraper'.
- Melakukan import packages 'from google_play_scraper import app' 'import pandas as pd' 'import numpy as np'.
- Mengambil aplikasi ID yang terdapat di url seperti 'id.dana'
- Melakukan scraping review
- Memasukan hasil scraping ke dalam data frame

2.11. Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk memvisualisasikan kinerja model prediksi dalam pembelajaran supervised learning. Setiap data dalam masing-masing kelas pada tabel ini menunjukkan jumlah prediksi yang dibuat untuk mengklasifikasikan kelas dengan negatif atau positif. Untuk kasus dengan tiga kelas (negatif, netral, positif), confusion matrix akan berukuran 3x3, dengan baris mewakili kelas aktual dan kolom mewakili kelas prediksi [21].

a. Akurasi

Akurasi adalah metrik evaluasi yang mengukur seberapa sering model membuat prediksi yang benar dari semua prediksi yang dilakukan. Dalam klasifikasi, akurasi menunjukkan seberapa sering model memprediksi kelas yang tepat, baik positif maupun negatif. Untuk menghitung nilai akurasi, kita dapat menggunakan persamaan 1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

b. Presisi

Presisi adalah metrik evaluasi yang mengukur ketepatan model dalam membuat prediksi benar untuk kelas positif dari seluruh prediksi positif yang dibuat. Dalam klasifikasi, presisi menunjukkan seberapa sering model secara akurat memprediksi kelas positif di antara semua prediksi positif yang dihasilkan. Untuk menghitung nilai presisi, pertama-tama hitung presisi untuk setiap kelas, kemudian jumlahkan nilai-nilai tersebut dan cari rata-ratanya, Rumusnya bisa menggunakan persamaan 2:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Setelah menentukan nilai presisi untuk setiap kelas, selanjutnya adalah menghitung rata-rata dari semua nilai tersebut dengan menjumlahkan seluruh nilai presisi dan membaginya dengan jumlah kelas. Rumusnya adalah:

$$Precision = \frac{Precision A + B + C}{Jumlah Kelas} \quad (3)$$

c. Recall

Recall adalah metrik evaluasi yang menunjukkan kemampuan model dalam mengidentifikasi kelas positif secara akurat. Untuk menghitung nilai Recall, kita dapat menggunakan persamaan 4:

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

Setelah menentukan nilai recall untuk setiap kelas, selanjutnya adalah menghitung rata-rata dari semua nilai tersebut dengan menjumlahkan seluruh nilai recall dan membaginya dengan jumlah kelas. Rumusnya adalah:

$$Recall = \frac{Recall A + B + C}{Jumlah Kelas} \quad (5)$$

d. F1-Score

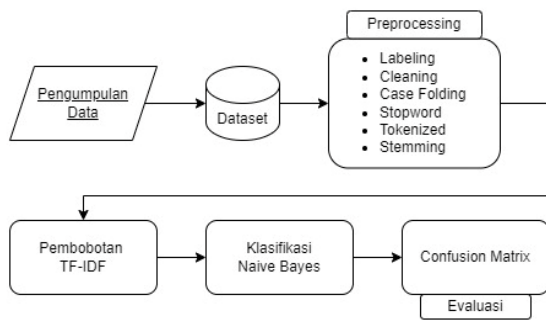
F1-Score adalah metrik evaluasi yang menggambarkan keseimbangan antara Presisi (Precision) dan Sensitivitas (Recall). Untuk menghitung nilai F1-Score, kita dapat menggunakan persamaan 5:

$$F1 Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (6)$$

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah Knowledge Discovery from Data (KDD), yang melibatkan lima tahapan: seleksi data, preprocessing, transformasi, data mining, dan evaluasi. Metode KDD ini digunakan untuk menggali data guna memperoleh informasi dan pola dari data yang telah dipilih. Untuk memberikan pemahaman

yang lebih jelas tentang alur penelitian, sebuah ilustrasi grafis disertakan dalam Gambar 1 [22].



Gambar 1. Alur penelitian

3.1. Pengumpulan Data

Data untuk penelitian ini diambil dari ulasan aplikasi DANA di *Google Play Store* menggunakan metode *web scraping*. Penelitian ini mengumpulkan 2000 ulasan berbahasa Indonesia. Data tersebut kemudian akan diproses melalui tahapan preprocessing..

3.2. Pelabelan

Setelah data dikumpulkan, langkah berikutnya adalah memberikan label berdasarkan rating. Dataset yang terkumpul akan dikategorikan menjadi tiga label untuk setiap ulasan: negatif, netral, dan positif. Ulasan dengan rating satu hingga dua akan diberi label sebagai sentimen negatif, ulasan dengan rating tiga akan diberi label sebagai netral, dan ulasan dengan rating empat hingga lima akan diberi label sebagai sentimen positif..

3.3. Preprocessing

Preprocessing adalah langkah berikutnya yang dilakukan setelah pengumpulan dataset yang terdiri dari ulasan. Tujuan *preprocessing* adalah untuk menghilangkan data yang tidak diperlukan sehingga menghasilkan dokumen yang lebih rinci dan terstruktur, serta menyiapkan data agar siap untuk proses selanjutnya. Proses ini sangat penting karena data yang dikumpulkan dari *Google Play Store* seringkali masih tidak sistematis atau tidak teratur. *Preprocessing* ini meliputi beberapa tahap, seperti pembersihan (*cleaning*), perubahan huruf besar menjadi kecil (*case folding*), penghapusan kata-kata umum (*stopword*), tokenisasi (*tokenized*), dan *stemming*.

3.4. Pembobotan

Pembobotan data adalah proses yang bertujuan untuk mengubah skala pengukuran data asli ke format lain agar sesuai dengan asumsi-asumsi yang mendasari analisis berbagai teknik data mining. Pada tahap ini, pembobotan kata akan dilakukan menggunakan metode *TF-IDF*.

3.5. Klasifikasi

Proses klasifikasi menggunakan Algoritma *Naïve Bayes* terdiri dari dua tahap: proses pelatihan (*training*) dan proses pengujian (*testing*). Pertama, dilakukan proses pelatihan untuk membangun model, kemudian dilanjutkan dengan proses pengujian yang mengacu pada probabilitas dari dataset pelatihan. Metode *Naïve Bayes Classifier* digunakan untuk mengelompokkan opini dengan baik, mampu mengklasifikasikan komentar menjadi positif, netral, atau negatif terhadap suatu produk atau isu yang sedang ramai diperbincangkan.

3.6. Evaluasi Model

Evaluasi kinerja model klasifikasi dilakukan dengan menggunakan hasil pengujian prediksi yang benar dan salah. Model klasifikasi ini menentukan akurasi, yang menunjukkan persentase rekaman yang diklasifikasikan dengan benar. Hasil perhitungan disajikan dalam bentuk matriks kebingungan (*confusion matrix*), yang mencakup jumlah nilai *true negative negative* (TNegNeg), *true neutral neutral* (TNetNet), *true positive positive* (TPP), *negative false neutral* (NegFNet), *negative false positive* (NegFP), *neutral false negative* (NetFNeg), *neutral false positive* (NetFP), *positive false negative* (PFNeg), *positive false neutral* (PFNet). Dalam penelitian ini, *confusion matrix* digunakan sebagai alat evaluasi kinerja model klasifikasi, yang memberikan gambaran sejauh mana model mampu melakukan klasifikasi dengan benar atau salah..

4. HASIL DAN PEMBAHASAN

4.1. Data Hasil Scraping

Data dalam penelitian ini diambil dengan metode scraping dari ulasan di Google Play Store, menggunakan bahasa pemrograman Python di Google Colab. Jumlah data yang diperoleh adalah 2000 ulasan berbahasa Indonesia.

[illegible]

Gambar 2. Hasil *scraping* data

Setelah scraping dilakukan, data awal yang tidak berurutan kemudian diurutkan secara kronologis menggunakan Excel. Hanya kolom ulasan dan skor (*rating*) yang diambil.

	content	score
110	Ku kira aman ternyata masih bisa di retass.. a...	1
58	Tgl 11 juni say melakukan pembayaran qris. Smp...	1
1818	Sangat tidak rekomend sama apk ini banyak kend...	1
254	Saya lakukan pembayaran sekolah tidak berhasil...	5
108	Maaf saya gak tahu ada apa dengan aplikasi ini...	1

Gambar 3. Data berurutan

4.2. Pelabelan

Setelah melakukan scraping data, berikutnya adalah memberi label ke setiap sentimen berdasarkan rating yang di dapatkan dari hasil scraping data. Sebelumnya akan dilakukan penghapusan kolom at (tanggal) sehingga yang tersisa hanya kolom ulasan dan score (rating). Ulasan dengan rating satu sampai dua dilabeli dengan sentimen negatif, ulasan dengan rating tiga dilabeli dengan sentimen netral dan ulasan rating empat sampai lima dilabeli dengan sentimen positif.

279	Saya gak tau lagi dengan dana. Ada yang kirim ...	3	Netral
1918	Saya mau update premium gabisa apk kon Kalau u...	1	Negatif
102	2 kali kejadian, beli vocer belanja tapi gak b...	2	Negatif
22	ini apasi ya, pertama upgrade ke premium statu...	1	Negatif
1887	Ini kenapa sih kemarin masih bsa dipakek kok i...	4	Positif
154	Alhamdulillah bisa transaksi lebih aman dan ny...	5	Positif

Gambar 4. Hasil pelabelan data

4.3. Preprocessing

Preprocessing adalah langkah berikutnya yang dilakukan setelah pengumpulan dataset yang terdiri dari ulasan. Tujuan *preprocessing* adalah untuk menghilangkan data yang tidak diperlukan sehingga menghasilkan dokumen yang lebih rinci dan terstruktur, serta menyiapkan data agar siap untuk proses selanjutnya. Proses ini sangat penting karena data yang dikumpulkan dari *Google Play Store* seringkali masih tidak sistematis atau tidak teratur. *Preprocessing* ini meliputi beberapa tahap, seperti pembersihan (*cleaning*), perubahan huruf besar menjadi kecil (*case folding*), penghapusan kata-kata umum (*stopword*), tokenisasi (*tokenized*), dan *stemming*.

4.4. Cleaning

Proses *cleaning* atau pembersihan data adalah langkah yang bertujuan untuk menghilangkan data yang tidak bernilai, salah, tidak relevan, duplikat, dan tidak konsisten, karena keberadaan data tersebut dapat mengurangi kualitas atau keakuratan hasil. Pembersihan data juga mempengaruhi kinerja sistem data mining, karena mengurangi jumlah dan kompleksitas data yang diproses.

Jumlah data sebelum cleaning: 200			
Data sebelum cleaning:			
0	Pengembang anjiiing (Vivo Y53), setelah u...	content sentiment	negatif
1	Kenapa saldo saya tbtb terpotong otomatis pada...		negatif
2	Sekitar 1 thn yg lalu aplikasi sangat bagus, t...		netral
3	Tahun 2024 ini maaf saya sangat tidak puas den...		positif
4	Tolong sistem pembayaran otomatis di nonaktifk...		netral
Jumlah data setelah cleaning: 200			
Data setelah cleaning:			
0	Pengembang anjiiing (Vivo Y setelah update...	cleaned_review sentiment	negatif
1	Kenapa saldo saya tbtb terpotong otomatis pada...		negatif
2	Sekitar thn yg lalu aplikasi sangat bagus tapi...		netral
3	Tahun ini maaf saya sangat tidak puas dengan A...		positif
4	Tolong sistem pembayaran otomatis di nonaktifk...		netral

Gambar 5. Hasil data *cleaning*

4.5. Case Folding

Proses *case folding* menggabungkan semua karakter dalam dokumen menjadi satu format seragam, salah satu hasilnya adalah mengubah huruf menjadi kecil. Dalam penelitian ini, dataset ulasan akan diubah dengan mengonversi huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*). Tahap ini bertujuan untuk memastikan format huruf seragam sehingga tidak ada kata yang tidak terdeteksi atau *case sensitive*.

Tabel 1. Hasil data case folding

Sebelum	Sesudah
Bulan ini sudah 2 kai akun saya masuk dengn akun lain,saldo saya tiba2 tiba hilang ,dan setelah di cek histori ada transaksi yg tidak saya lakukan dan trasaksi Google	bulan ini sudah 2 kai akun saya masuk dengn akun lain,saldo saya tiba2 tiba hilang ,dan setelah di cek histori ada transaksi yg tidak saya lakukan dan trasaksi google

4.6. Stopword

Menghapus kata kunci berarti menghilangkan kata-kata yang tidak relevan dari teks, sehingga meningkatkan ketepatan bobot. Kata-kata seperti "dan" akan dihilangkan. Tahap ini dilakukan untuk mengoptimalkan data, dengan menghapus data yang tidak terlalu penting dan tidak akan mengurangi performa klasifikasi. "*Stopword*" adalah istilah yang digunakan untuk menyaring kata-kata yang sering muncul dan yang jarang muncul.

Tabel 2. Hasil data stopwords

Sebelum	Sesudah
bulan ini sudah kai akun saya masuk dengn akun lainsaldo saya tiba tiba hilang dan setelah di cek histori ada transaksi yg tidak saya lakukan dan trasaksi google	kai akun masuk dengn akun lainsaldo hilang cek histori transaksi yg lakukan trasaksi google

4.7. Tokenized

Memecah teks menjadi kata-kata tunggal dan menghilangkan tanda baca serta angka sesuai dengan kosakata yang digunakan disebut tokenisasi. Tahap ini memisahkan setiap ulasan menjadi kata-kata tunggal (*term*). Tujuan dari tahap ini adalah memisahkan setiap kata, yang biasanya dipisahkan oleh spasi.

Tabel 3. Hasil data tokenized

Sebelum	Sesudah
kai akun masuk dengn akun lainsaldo hilang cek histori transaksi yg lakukan trasaksi google	['kai', 'akun', 'masuk', 'dengn', 'akun', 'lainsaldo', 'hilang', 'cek', 'histori', 'transaksi', 'yg', 'lakukan', 'trasaksi', 'google']

4.8. Stemming

Sufiks lain dapat diubah menjadi kata dasar melalui proses *stemming*. Sufiks dihilangkan dalam penelitian ini karena ada bentuk lampau dalam teks bahasa Inggris. Dalam tahap ini, data ulasan akan diubah dengan menghilangkan imbuhan sehingga setiap kata akan menjadi bentuk dasar. Dalam bahasa

Indonesia, proses ini kompleks karena harus menghilangkan semua imbuhan di setiap kata.

Tabel 4. Hasil data stemming

Sebelum	Sesudah
kai akun masuk dengan akun lainsaldo hilang cek histori transaksi yg lakukan trasaksi google	kai akun masuk dengan akun lainsaldo hilang cek histori transaksi yg laku trasaksi google

4.9. Visualisasi Word Cloud

Dalam ulasan negatif, netral, dan positif, tentu ada kata-kata yang paling sering muncul. Dalam analisis 2000 data ulasan, visualisasi *word cloud* digunakan untuk menemukan tren kata pada setiap label.



Gambar 6. Visualisasi *word cloud* negatif

Pada Gambar 6 merupakan hasil dari visualisasi word cloud pada kelas sentimen negatif. Kata yang paling banyak muncul diantaranya 'uang', 'saldo', 'tidak', 'masuk', 'aplikasi', 'dana'.



Gambar 7. Visualisasi *word cloud* netral

Pada Gambar 7 merupakan hasil dari visualisasi *word cloud* pada kelas sentimen netral. Kata yang paling banyak muncul diantaranya 'tolong', 'kenapa', 'padahal', 'transaksi', 'selalu'.



Gambar 8. Visualisasi *word cloud* positif

Pada Gambar 8 merupakan hasil dari visualisasi *word cloud* pada kelas sentimen positif. Kata yang paling banyak muncul diantaranya 'sangat', 'membantu', 'bagus', 'mantap', dan 'cepat'.

4.10. Pembobotan

Proses pembobotan kata dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Setiap kata dalam ulasan akan diberikan nilai bobot sesuai dengan perhitungan menggunakan algoritma TF-IDF ini. Tujuannya adalah untuk mengukur seberapa penting suatu kata dalam sebuah dokumen. Setelah proses TF-IDF dilakukan, dataset akan dibagi menjadi dua bagian: data *training* dan data *testing*. Data *training* digunakan untuk melatih model, sementara data *testing* digunakan untuk menguji model yang telah dilatih, guna mengetahui seberapa baik model tersebut bekerja..

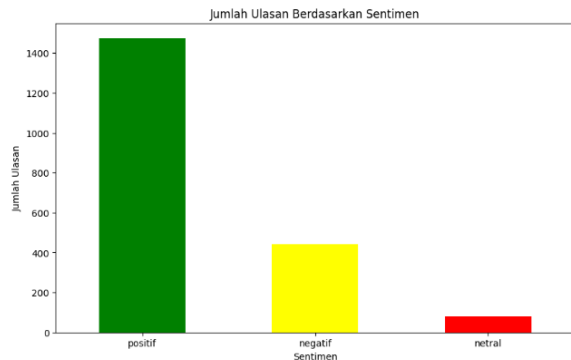
4.11. Klasifikasi

Tahap klasifikasi dengan Algoritma *Naïve Bayes* terdiri dari dua proses: proses pelatihan (*training*) dan proses pengujian (*testing*). Pertama, dilakukan proses pelatihan untuk membangun model, kemudian dilanjutkan dengan proses pengujian yang mengacu pada probabilitas dari dataset pelatihan. Metode *Naïve Bayes Classifier* digunakan untuk mengelompokkan opini dengan baik, mampu mengklasifikasikan komentar menjadi positif, netral, atau negatif terhadap suatu produk atau isu yang sedang ramai diperbincangkan.

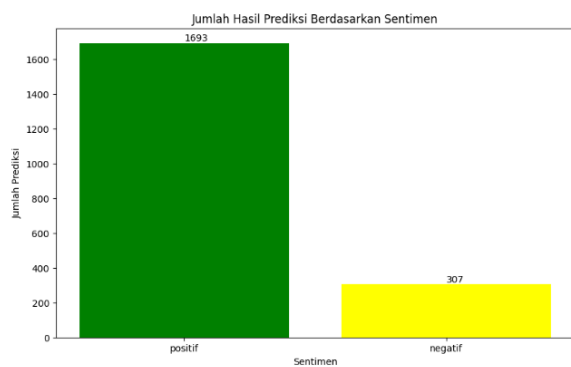
Akurasi: 0.86				
	precision	recall	f1-score	support
negatif	0.87	0.60	0.71	443
netral	0.00	0.00	0.00	81
positif	0.86	0.98	0.92	1476
accuracy			0.86	2000
macro avg	0.58	0.53	0.54	2000
weighted avg	0.83	0.86	0.84	2000

Gambar 9. Hasil klasifikasi

Dari Gambar 9 dapat dilihat bahwa hasil klasifikasi menggunakan algoritma Naïve Bayes menghasilkan 443 sentimen negatif, 81 sentimen netral, dan 1476 sentimen positif. Sentimen menunjukkan akurasi sebesar 86%, presisi 86%, recall 98%, dan nilai f1-score 92%. Hasil klasifikasi sentimen berdasarkan rating menunjukkan bahwa dari 2000 data, 1476 sentimen berlabel positif, 443 sentimen berlabel negatif, dan 81 sentimen berlabel netral.



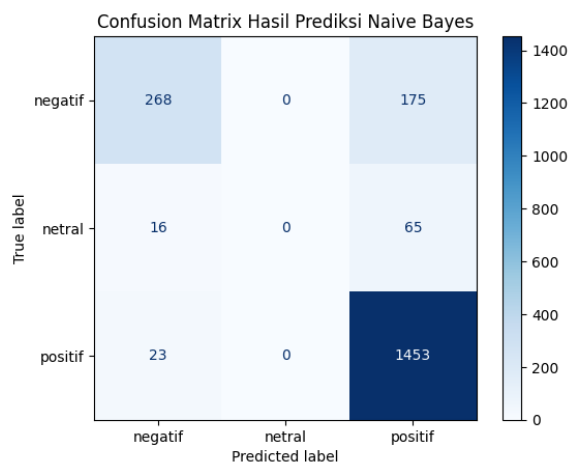
Gambar 10. Grafik sentimen berdasarkan rating

Gambar 11. Grafik hasil prediksi *naïve bayes*

Setelah dianalisis menggunakan *Naïve Bayes*, dari 2000 data, dihasilkan 1693 sentimen berlabel positif dan 307 sentimen berlabel negatif.

4.12. Evaluasi

Dalam penelitian ini, *confusion matrix* digunakan sebagai alat evaluasi kinerja model. *Confusion matrix* adalah metode pengujian yang digunakan untuk menghitung tingkat kinerja atau keakuratan suatu proses klasifikasi. *Confusion matrix* dapat menganalisis seberapa baik pengklasifikasi dapat mengenali rekaman data dari berbagai kelas.

Gambar 12. *Confusion matrix*

Dari Gambar 12, beberapa kesimpulan yang dapat diambil adalah sebagai berikut:

- TNegNeg (*True Negative Negative*) = Data berlabel negatif yang diprediksi benar sebagai label negatif sebanyak 268
- TNetNet (*True Neutral Neutral*) = Data berlabel netral yang diprediksi benar sebagai label netral sebanyak 0
- TPP (*True Positive Possitive*) = Data berlabel positif yang diprediksi benar sebagai label positif sebanyak 1453
- NegFNet (*Negative False Neutral*) = Data berlabel negatif yang diprediksi salah sebagai label netral sebanyak 0
- NegFP (*Negative False Positive*) = Data berlabel negatif yang diprediksi salah sebagai label positif sebanyak 175
- NetFNeg (*Neutral False Negative*) = Data berlabel netral yang diprediksi salah sebagai label negatif sebanyak 16
- NetFP (*Neutral False Positive*) = Data berlabel netral yang diprediksi salah sebagai label positif sebanyak 65
- PFNeg (*Positive False Negative*) = Data berlabel positif yang diprediksi salah sebagai label negatif sebanyak 23
- PFNet (*Possitive False Neutral*) = Data berlabel positif yang diprediksi salah sebagai label netral sebanyak 0

Berdasarkan hasil klasifikasi pada dataset uji yang berjumlah 2000, ditemukan bahwa 1721 ulasan diprediksi benar, sedangkan 279 ulasan diprediksi salah. Performa ini dapat dikatakan sangat baik.

5. KESIMPULAN DAN SARAN

Analisis sentimen ulasan terhadap aplikasi DANA di *Google Play Store* yang terdiri dari total dataset berjumlah 2000 data, ditemukan jumlah ulasan sentimen positif sebanyak 1476 data, jumlah ulasan sentimen negatif sebanyak 443 data, sedangkan jumlah ulasan sentimen netral sebanyak 81 data. Hasil analisis sentimen ulasan aplikasi DANA di *Google Play Store* didominasi oleh sentimen positif yang diperoleh dari scraping data menggunakan aplikasi *Google Colab*. Kata yang sering muncul pada ulasan negatif terhadap aplikasi DANA adalah 'uang', 'saldo', 'tidak', 'masuk', 'aplikasi', 'dana'. Maka dapat disimpulkan bahwa pengguna kerap mengeluhkan kinerja dari aplikasi DANA terkait gagal masuk ke dalam aplikasi, gagal melakukan transaksi, kendala dalam melakukan isi saldo, dan keluhan lainnya. Hasil evaluasi dari algoritma *Multinomial Naïve Bayes* berdasarkan segi performa hasil perhitungan menunjukkan Tingkat akurasi 86%, presisi 86%, *recall* 98%, dan *f1-score* 92%. Adapun saran untuk penelitian berikutnya adalah dapat melakukan komparasi dengan berbagai algoritma klasifikasi lainnya, seperti *K-Nearest Neighbour*, *Support Vector Machines*, dan lainnya, agar memiliki hasil komparasi yang beragam.

DAFTAR PUSTAKA

- [1] E. S. Basryah, A. Erfina, and C. Warman, "Analisis Sentimen Aplikasi Dompot Digital Di Era 4.0 Pada Masa Pandemi Covid-19 Di Play Store Menggunakan Algoritma Naive Bayes Classifier," *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, vol. 1, no. 1, pp. 189–196, Sep. 2021.
- [2] A. Nurian and B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naive Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 2830–7062, 2023, doi: 10.23960/jitet.v11i3%20s1.3348.
- [3] A. Rahardyan, "Riset Dompot Digital Terpopuler: OVO dan GoPay Dominan," *Bisnis.com*. Accessed: Jul. 11, 2024. [Online]. Available: <https://finansial.bisnis.com/read/20220103/563/1484554/riset-dompot-digital-terpopuler-ovo-dan-gopay-dominan>
- [4] A. M. Damar, "Survei: DANA Jadi Dompot Digital dengan Pertumbuhan Jumlah Pengguna Paling Pesat," *Liputan6.com*. Accessed: Jul. 11, 2024. [Online]. Available: <https://www.liputan6.com/tekno/read/4709582/survei-dana-jadi-dompot-digital-dengan-pertumbuhan-jumlah-pengguna-paling-pesat>
- [5] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, Sep. 2022.
- [6] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1136–1144, Jul. 2023, doi: 10.47065/josh.v4i4.3803.
- [7] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan KNN," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- [8] H. Taufiqurrahman, F. T. Anggraeny, and M. M. Al Haromainy, "Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor Pada Analisis Sentimen Ulasan Aplikasi MyPertamina," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 6, pp. 3934–3939, Dec. 2023.
- [9] T. P. R. Sanjaya, A. Fauzi, and A. F. N. Masruriyah, "Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine," *INFOTECH: Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 16–26, Jun. 2023, doi: 10.37373/infotech.v4i1.422.
- [10] Surohman, S. Aji, Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K-Nearest Neighbor," *Jurnal Sains dan Manajemen*, vol. 8, no. 1, pp. 93–105, 2020.
- [11] C. A. Sugianto and S. Tresnawati, "A Covid-19 Sentiment Analysis on Twitter Using K-Nearest Neighbours," *Journal of Applied Intelligent System*, vol. 7, no. 1, pp. 58–69, Apr. 2022.
- [12] M. R. Nadhif, D. W. Brata, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi TIX ID di Indonesia Pada Google Play Store Menggunakan Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 8, pp. 3932–3937, Sep. 2022.
- [13] K. A. Mutiarasari, "Pengertian Teks Ulasan Beserta Struktur dan Ciri-cirinya," *Detik.com*. Accessed: May 11, 2024. [Online]. Available: <https://news.detik.com/berita/d-6606471/pengertian-teks-ulasan-beserta-struktur-dan-ciri-cirinya>
- [14] Dana, "Dana." Accessed: Dec. 18, 2023. [Online]. Available: <https://www.dana.id/help-center/tentang-dana/apakah-dana-itu>
- [15] A. S. P. Braja and A. Kodar, "Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store," *J I M P - Jurnal Informatika Merdeka Pasuruan*, vol. 7, no. 3, pp. 120–128, Sep. 2023, doi: 10.51213/jimp.v7i3.779.
- [16] A. Ridwan, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *Jurnal Sistem Komputer dan Kecerdasan Buatan*, vol. 4, no. 1, pp. 15–21, Oct. 2020.
- [17] L. S. Hasibuan, A. Fariqi, L. Prayitno, and M. B. Bangun, "KLIK: Kajian Ilmiah Informatika dan Komputer Integrasi data Protein-Protein Interactions dan Pathway untuk Menentukan Score pada pathway Menggunakan Analisis Graf," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 719–727, 2023, doi: 10.30865/klik.v3i6.932.
- [18] S. Ranti, "Mengenal Google Colab, Fungsi dan Manfaatnya," *Kompas.com*. Accessed: Jun. 11, 2024. [Online]. Available: <https://tekno.kompas.com/read/2023/03/26/17150047/mengenal-google-colab-fungsi-dan-manfaatnya->
- [19] A. Puguh, "Apa Itu Python? Pengertian, Fungsi dan Contohnya," *rumahweb*. Accessed: Jun. 12, 2024. [Online]. Available: <https://blog.rumahweb.com/python-adalah/>
- [20] D. Zub, "Scrape Google Play Store App in Python," *SerpApi*. Accessed: Jun. 12, 2024. [Online]. Available: <https://serpapi.com/blog/scrape-google-play-store-app-in-python/>

- [21] A. M. Khalimi, "Perhitungan Confusion Matrix Multi-Class Clasification 3x3," pengalaman-edukasi. Accessed: Aug. 01, 2024. [Online]. Available: https://www.pengalaman-edukasi.com/2020/11/menghitung-confusion-matrix-3-kelas.html#google_vignette
- [22] Riskawati, Fatihanursari, Iin, and A. R. Rinaldi, "Penerapan Metode Naïve Bayes Classifier Pada Analisis Sentimen Aplikasi GoPay," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 346–353, Feb. 2024.