



Analisis Sentimen Ulasan Qris Berbasis Naïve Bayes Dengan Penerapan Teknik Random Oversampling

¹Muhammad Haikal Habibillah, ²Samsudin

^{1,2} Sistem Informasi, Teknik dan Ilmu Komputer, Universitas Islam Indragiri

haikalakel123@gmail.com, samsudinsadek@gmail.com

Abstrak

Pesatnya perkembangan teknologi finansial telah menempatkan *Quick Response Code Indonesian Standard* (QRIS) sebagai metode pembayaran digital yang sangat penting di Indonesia. Seiring dengan adopsinya yang masif, ulasan pengguna pada platform seperti Google Play Store menyediakan informasi otentik mengenai tingkat kepuasan maupun kendala teknis pengguna. Namun, analisis sentimen pada data teks ini seringkali menghadapi masalah ketidakseimbangan kelas (*class imbalance*) yang menyebabkan algoritma klasifikasi standar menjadi bias terhadap kelas mayoritas. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap penggunaan QRIS pada aplikasi dompet digital DANA, GoPay, dan OVO dengan membandingkan performa algoritma *Multinomial Naive Bayes* sebelum dan sesudah penerapan teknik penyeimbangan data. Metodologi penelitian mengimplementasikan kerangka kerja *Knowledge Discovery in Databases* (KDD) dengan memanfaatkan dataset ulasan yang divalidasi menggunakan kata kunci khusus transaksi QRIS. Hasil eksperimen menunjukkan bahwa model awal yang dilatih menggunakan data asli (Skenario 1) menghasilkan akurasi global sebesar 79,41%, namun memiliki kelemahan dalam mendeteksi keluhan pengguna dengan nilai *recall* sentimen negatif yang hanya menyentuh 66%. Sebaliknya, penerapan teknik *Random Oversampling* (Skenario 2) terbukti efektif meningkatkan sensitivitas model terhadap kelas minoritas secara signifikan, di mana nilai *recall* sentimen negatif melonjak naik menjadi 76% dan tebakan benar terhadap ulasan keluhan meningkat dari 19 menjadi 22 ulasan, meskipun akurasi global mengalami penurunan menjadi 73,53%. Kesimpulannya, pemodelan berbasis *Random Oversampling* jauh lebih andal dan direkomendasikan untuk analisis sentimen di dunia nyata karena mampu meminimalkan risiko bias komputasi serta menghasilkan model yang lebih peka dan objektif dalam menjangkau ulasan kritik dari para pengguna pembayaran digital.

Kata Kunci: Analisis Sentimen, QRIS, Naive Bayes, Random Oversampling, Class Imbalance

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi saat ini berkembang sangat pesat, khususnya dalam sektor perbankan dan sistem pembayaran digital yang memungkinkan transaksi dilakukan secara *mobile* tanpa perlu datang ke bank (Khoirul et al., 2023). Inovasi *Financial Technology* (Fintech) ini memicu tren signifikan dalam penggunaan metode pembayaran berbasis kode QR, terutama melalui *Quick Response Code Indonesian Standard* (QRIS) yang diperkenalkan untuk mempermudah transaksi keuangan masyarakat (Syaharin & Agustianto, 2025). QRIS merupakan standar yang dirancang oleh Bank Indonesia untuk menyatukan berbagai layanan pembayaran berbasis kode QR di Indonesia guna meningkatkan kemudahan, kecepatan, serta kenyamanan dalam bertransaksi (Natalia Kristanty, 2024). Berbagai platform digital besar seperti Gojek (GoPay) telah mengadopsi standar ini sebagai metode pembayaran utama untuk mendukung inklusivitas keuangan nasional (Dometian & Jatmiko, 2025).

Seiring dengan masifnya penggunaan QRIS, ulasan pengguna pada platform seperti Google Play Store menjadi sumber informasi yang sangat berharga bagi pengembang untuk mengetahui keluhan, tingkat kepuasan, maupun saran perbaikan layanan. Ulasan tersebut mengandung opini otentik mengenai pengalaman pengguna yang dapat digunakan untuk menilai performa aplikasi secara langsung (Firda Aini et al., 2026). Namun, volume ulasan yang sangat besar dan bersifat tidak terstruktur membuat analisis manual menjadi tugas yang hampir mustahil untuk dilakukan bagi pengembang aplikasi (Fadhilah & Utomo, 2024). Oleh karena itu, diperlukan pendekatan otomatis melalui Analisis Sentimen atau *opinion mining*, yaitu studi komputasi untuk mengekstrak dan mengklasifikasikan opini pengguna ke dalam kategori seperti positif, negatif, atau netral (Firda Aini et al., 2026).

Dalam penelitian ini, algoritma yang digunakan untuk klasifikasi adalah Naive Bayes (Khoirul et al., 2023). Algoritma ini dipilih karena dikenal memiliki kesederhanaan, efisiensi komputasi yang tinggi, serta performa yang baik dalam menangani data teks berskala besar. Naive Bayes bekerja berdasarkan teori probabilitas untuk menentukan peluang suatu teks masuk ke dalam kategori tertentu berdasarkan frekuensi kemunculan kata dalam data pelatihan (Bimo et al., 2025a). Keunggulan utamanya adalah kemampuannya memberikan hasil estimasi yang akurat meskipun hanya menggunakan sejumlah kecil data pelatihan (Bimo et al., 2025a).

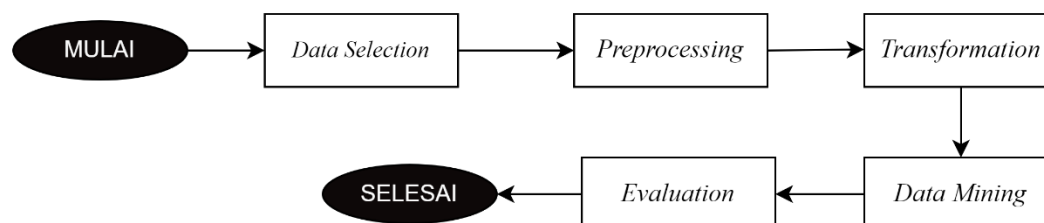
Analisis Sentimen Ulasan Qris Berbasis Naïve Bayes Dengan Penerapan Teknik Random Oversampling

Meskipun memiliki performa yang baik, algoritma Naïve Bayes sangat bergantung pada distribusi data latih, sehingga seringkali menghadapi kendala ketika diperhadapkan pada permasalahan ketidakseimbangan data atau *Class Imbalance*. yaitu kondisi di mana jumlah data pada satu kelas lebih dominan dibandingkan kelas lainnya (Ramadhan Alfinsyah & Pili Hartato, 2025). Kondisi ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas dan menurunkan kemampuan model dalam mengenali kelas minoritas. Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan teknik random oversampling, yaitu metode penyeimbangan data dengan cara menambah jumlah data pada kelas minoritas melalui proses duplikasi secara acak hingga distribusi data menjadi seimbang (Alfarizi et al., 2026).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap penggunaan QRIS pada aplikasi dompet digital Dana, Gopay, Dan Ovo dengan membandingkan performa algoritma Naive Bayes pada data asli dan data yang telah melalui proses penyeimbangan menggunakan teknik random oversampling. Diharapkan hasil penelitian ini dapat memberikan pemahaman yang lebih akurat terhadap opini pengguna serta meningkatkan efektivitas model klasifikasi dalam mengevaluasi layanan pembayaran digital di Indonesia.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) sebagai metode utamanya (Achmad Fauzi & Nurina Sari, 2026). KDD merupakan teknik ekstraksi yang bertujuan untuk menggali informasi serta menemukan pola tersembunyi dari sekumpulan data yang telah diseleksi sebelumnya. Secara umum, alur kerja metode KDD ini mencakup lima tahapan pokok, yaitu *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation* (Andayani, n.d.). Alur pengerjaan penelitian secara sistematis diilustrasikan pada Gambar 1 berikut:



Gambar 1 Alur Penelitian KDD

2.1 Pemilihan Data (*Data Selection*)

Tahap awal dalam proses Knowledge Discovery in Database (KDD) adalah melakukan seleksi data dari sekumpulan data operasional ke dalam berkas terpisah yang akan digunakan dalam proses analisis (Dinar et al., 2023). Pada penelitian ini, data yang digunakan berupa ulasan (review) masyarakat terkait penggunaan QRIS sebagai metode pembayaran digital yang bersumber dari platform Google Play Store. Proses pengumpulan data dilakukan melalui dua tahapan utama, yaitu proses web scraping dan penyaringan kata kunci. Pada tahap web scraping, data ulasan terbaru diambil dari Google Play Store dengan batas maksimal sebanyak 4.000 ulasan untuk masing-masing aplikasi dompet digital, yaitu DANA (id.dana), GoPay (com.gojek.app), dan OVO (ovo.id). Selanjutnya, dilakukan penyaringan terhadap ulasan mentah menggunakan teknik ekspresi reguler (regex) berdasarkan kata kunci yang menjadi penanda transaksi QRIS, seperti “scan”, “scan qr”, “qr”, “barcode”, “kode qr”, “qris”, dan “pembayaran” (Aldi Rahmad Fahrozi et al., 2026).

2.2 Pra-pemrosesan Data (*Preprocessing*)

Data ulasan teks hasil penyaringan selanjutnya dibersihkan melalui fungsi khusus *clean_text* serta memanfaatkan pustaka Sastrawi dengan beberapa tahapan. Tahap pertama adalah *case folding*, yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) menggunakan fungsi *.lower()* (Ifan Farimulyadi & Zaehol Fatah, 2025). Berikutnya dilakukan *URL removal* dengan menghapus tautan internet atau URL yang terdapat dalam ulasan menggunakan pola *regex* *re.sub(r'http\S+', '', text)*. Tahap selanjutnya adalah *cleansing character*, yaitu mengeliminasi angka, tanda baca, simbol, dan karakter non-alfabet lainnya dengan pola *regex* *re.sub(r'^a-zA-Z\s|', '', text)*. Setelah itu dilakukan *whitespace trimming* untuk mereduksi spasi ganda serta menghapus spasi berlebih di awal dan akhir kalimat menggunakan fungsi *.strip()*, kemudian hasilnya disimpan dalam kolom *clean_review*. Terakhir, dilakukan *stopword removal* dengan membuang kata-kata umum dalam bahasa Indonesia yang tidak memiliki bobot informasi menggunakan kamus kata standar dari pustaka Sastrawi, sehingga menghasilkan data akhir yang lebih bersih dan informatif yang disimpan pada kolom *clean_final*.

2.3 Transformasi Data (*Transformation*)

Tahap ini mengubah teks bersih menjadi format numerik terstruktur. Proses diawali pelabelan sentimen secara biner (*lexicon-based*) menggunakan kamus kata kunci positif dan negatif. Jika akumulasi skor kata kunci bernilai lebih dari nol, ulasan dikategorikan sebagai positif, dan selain itu dikelompokkan sebagai negatif. Selanjutnya, dilakukan ekstraksi fitur berbasis nilai statistik kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) melalui fungsi `TfidfVectorizer()` (Bimo et al., 2025b).

Metode TF-IDF digunakan untuk menghitung bobot setiap kata (t) dalam suatu ulasan (d) terhadap seluruh kumpulan ulasan (D). Rumus matematis untuk menghitung bobot TF-IDF adalah sebagai berikut:

$$W_{t,d} = \text{TF}(t, d) \times \log\left(\frac{N}{\text{DF}(t)}\right) \quad (1)$$

Pada tahap ini, penelitian dibagi menjadi dua skenario eksperimen untuk menguji dampak ketidakseimbangan kelas (*class imbalance*), yaitu skenario pertama menggunakan data asli apa adanya (tanpa penyeimbangan), dan skenario kedua menerapkan teknik *Random Oversampling* menggunakan fungsi `Random OverSampler` dari library `imblearn` untuk menggandakan data kelas minoritas secara acak hingga jumlahnya seimbang dengan kelas mayoritas.

2.4 Penambangan Data (*Data Mining*)

Proses klasifikasi pada kedua skenario eksperimen mengimplementasikan algoritma Multinomial Naive Bayes (MultinomialNB). Sebelum pelatihan model dilakukan, data pembobotan TF-IDF beserta label sentimennya dibagi menjadi data latih (*training data*) dan data uji (*testing data*) dengan rasio 80:20 menggunakan fungsi `train_test_split()` (Novitasari et al., 2023). Parameter `stratify=y` dan `random_state=42` diterapkan untuk memastikan proporsi distribusi kelas tetap konsisten dan eksperimen dapat direplikasi (Nurfathir et al., 2026). Model Naive Bayes kemudian dilatih menggunakan data latih dan diuji pada data uji untuk menghasilkan prediksi sentimen. Algoritma Multinomial Naive Bayes bekerja berdasarkan Teorema Bayes dengan menghitung nilai probabilitas tertinggi untuk menentukan kelas sentimen (c) dari sebuah ulasan. Persamaan dasar dari algoritma ini dirumuskan sebagai berikut:

$$P(c|d) = \frac{P(c) \prod_{i=1}^n P(t_i|c)}{P(d)} \quad (2)$$

Karena nilai probabilitas penyebut $P(d)$ bersifat konstan untuk semua kelas, maka penentuan kelas prediksi dilakukan dengan mencari nilai maksimum (*Argmax*) melalui rumus:

$$\hat{c} = \arg \max_{c \in \{0,1\}} P(c) \prod_{i=1}^n P(t_i|c) \quad (3)$$

2.5 Evaluasi (*Evaluation*)

Evaluasi dilakukan untuk mengukur, menguji, dan membandingkan performa model algoritma Naive Bayes dari kedua skenario eksperimen yang telah dijalankan. Pengukuran kinerja model dalam penelitian ini memanfaatkan fungsi `accuracy_score` dan `classification_report` yang bersumber dari pustaka *Scikit-Learn*. Metrik penilaian utama yang digunakan meliputi Akurasi (*Accuracy*), Presisi (*Precision*), dan *Recall*. (Radja et al., 2026). Nilai dari ketiga metrik tersebut dihitung berdasarkan komponen *Confusion Matrix* yang terdiri dari *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Perhitungan matematis untuk masing-masing metrik evaluasi tersebut dijabarkan melalui persamaan sebagai berikut:

Akurasi (*Accuracy*): Mengukur rasio total prediksi benar (positif dan negatif) terhadap keseluruhan data uji yang digunakan.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Precisi (*Precision*): Mengukur tingkat ketepatan antara data aktual dengan hasil prediksi yang diberikan oleh model klasifikasi.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

Recall (Sensitivitas): Mengukur keberhasilan model dalam menemukan kembali sebuah informasi secara tepat, khususnya dalam konteks jaring tangkap ulasan keluhan pada kelas minoritas.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

Selain melalui visualisasi angka persentase, performa model diperjelas menggunakan grafik *Confusion Matrix* berbasis *heatmap*. Visualisasi ini berfungsi untuk memetakan kecocokan antara data aktual lapangan terhadap hasil prediksi yang dikeluarkan oleh sistem, serta mempermudah analisis komparatif untuk melihat dampak nyata dari penerapan teknik *Random Oversampling* terhadap tingkat sensitivitas model dalam mendeteksi sentimen negatif masyarakat.

3. Hasil dan Pembahasan

3.1 Hasil Data Selection (Pemilihan Data)

Tahap pertama dalam metode *Knowledge Discovery in Databases* (KDD) adalah seleksi data (*data selection*). Pada penelitian ini, data operasional diperoleh melalui proses *web scraping* menggunakan bahasa pemrograman Python pada platform Google Play Store untuk menarik ulasan pengguna dari tiga aplikasi dompet digital utama di Indonesia, yaitu DANA (id.dana), GoPay (com.gojek.app), dan OVO (ovo.id). Batas maksimal penarikan data pada kodingan ditetapkan sebanyak 4.000 ulasan untuk masing-masing aplikasi guna mendapatkan sampel acak yang luas dan representatif dari opini masyarakat di lapangan. Hasil akumulasi pengumpulan data mentah awal disajikan pada Tabel 1 berikut:

Nama Aplikasi	Batas Maksimal Ulasan	Jumlah Ulasan Mentah Hasil Scraping
DANA	4.000	4.000
GoPay	4.000	4.000
OVO	4.000	4.000
Total Dataset Hasil Scraping		12.000

langkah berikutnya adalah melaksanakan seleksi data operasional (*data selection*) untuk mempersempit fokus penelitian hanya pada topik transaksi QRIS. Penyaringan ini memanfaatkan pencocokan pola ekspresi reguler (*regex*) melalui sekumpulan kata kunci (*keywords*) penanda, yaitu: 'scan', 'scan qr', 'qr', 'barcode', 'kode qr', 'qris', dan 'pembayaran'. Berdasarkan proses eksekusi filtrasi kata kunci tersebut pada data teks ulasan, diperoleh hasil bahwa ulasan valid secara spesifik membahas fitur QRIS berjumlah 338 ulasan. Data ulasan yang lolos sensor kata kunci ini kemudian disimpan ke dalam struktur data baru (df_qris). Beberapa potongan sampel data teks komentar riil hasil dari penarikan dan penyaringan kata kunci ini ditunjukkan pada Tabel 2

No	APP	Review
1.	DANA	scab Qris nya di perbaiki, udah 2hari nunggu ttep gk bisa pakai qris
2.	DANA	lama nunggu QR
3.	GoPay	saya sdh premier tapi g bisa mo trfer dan byr pake qris, apa harus upgrade ke nabung
4.	GoPay	masa gak bisa qr coba lagi coba lagi,
5.	OVO	setiap pembayaran lancar tidak ada kendala
6.	OVO	APLIKASI BUOSOK SUDAH GK BISA BAYAR VIA QRIS

3.2 Hasil Pra-pemrosesan Data (*Preprocessing*)

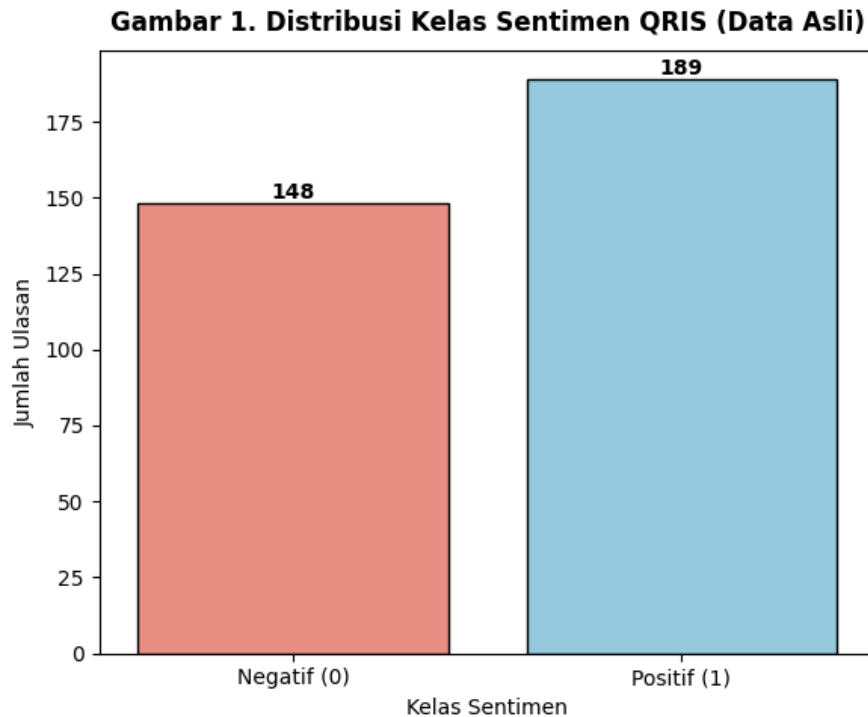
Evaluasi akhir pada tahapan pembersihan ini diselesaikan dengan mengeksekusi perintah `.dropna()` untuk memastikan tidak ada baris data kosong (*missing values*) yang tersisa, sehingga total dataset bersih yang siap masuk ke tahap transformasi tetap berjumlah 337 ulasan. Untuk menginterpretasikan secara visual sebaran kata dasar yang paling mendominasi dan sering muncul dari keseluruhan 337 data ulasan bersih tersebut sebelum dipisahkan berdasarkan label sentimen biner, dihasilkan grafik *WordCloud* gabungan yang disajikan pada Gambar2 berikut:

[illegible]

Berdasarkan Gambar 2, ukuran teks mempresentasikan tingkat frekuensi kemunculan kata di dalam dataset ulasan QRIS. Terlihat jelas bahwa kata kunci seperti "pembayaran", "aplikasi", "qris", "dana", "saldo", "ovo", "gagal", "gopay", "lama", dan "mudah" menjadi inti pembahasan utama dari keseluruhan ulasan yang dikumpulkan masyarakat. Kehadiran kata-kata yang kontras ini (seperti kata "mudah" yang mengarah pada kepuasan pengguna dan kata "gagal" atau "lama" yang mengarah pada keluhan teknis) mengonfirmasi bahwa dataset 338 ulasan yang telah dibersihkan ini memiliki sebaran opini yang sangat kaya dan bervariasi, sehingga valid untuk dilanjutkan ke tahapan transformasi data dan pemodelan klasifikasi.

3.3 Hasil Transformasi Data (*Transformation*)

Tahap ketiga dalam metode KDD adalah transformasi data (*data transformation*) yang berfungsi untuk mengubah data teks bersih hasil *preprocessing* ke dalam bentuk matriks angka numerik terstruktur agar dapat dipahami oleh algoritma komputer, sekaligus mempersiapkan pembagian skenario eksperimen. Langkah transformasi pertama yang dilakukan adalah proses pelabelan sentimen (*Lexicon-Based*) secara biner otomatis menggunakan kamus kata penanda opini. Ulasan yang mengandung ekspresi kepuasan, kelancaran, atau kemudahan diberikan label positif (1), sedangkan ulasan yang berisi keluhan teknis, kekecewaan, atau kendala sistem diberikan label negatif (0). Hasil sebaran dari proses pelabelan otomatis terhadap data operasional asli ini disajikan secara visual pada grafik batang Gambar 3 berikut:



Gambar 3 Grafik Batang Distribusi Asli

Berdasarkan visualisasi grafik batang pada Gambar 3, terlihat perbandingan proporsi kelas data yang menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) secara riil pada dataset operasional awal. Kelas sentimen positif (label 1) tampak mendominasi dengan total 189 ulasan, sedangkan kelas sentimen negatif (label 0) yang menampung ulasan keluhan memiliki jumlah yang lebih sedikit, yaitu 148 ulasan. Ketimpangan sebaran data ini secara teoretis berpotensi membuat model klasifikasi menjadi bias ke kelas mayoritas saat proses pengujian dijalankan. Oleh karena itu, kondisi ketimpangan distribusi ini menjadi landasan kuat untuk memisahkan pengujian pemodelan ke dalam dua skenario eksperimen yang berbeda guna mengukur pengaruh penyeimbangan data secara objektif.

Langkah transformasi berikutnya adalah membagi dataset asli dengan rasio proporsional 80% untuk data latih (*training data*) sebanyak 269 ulasan dan 20% untuk data uji (*testing data*) sebanyak 68 ulasan menggunakan fungsi `train_test_split` dengan mengunci parameter `random_state=42` demi menjaga konsistensi sebaran baris data pada tiap skenario. Setelah pembagian data selesai, seluruh teks ulasan diekstrak bobot numeriknya menggunakan fungsi `TfidfVectorizer()` untuk menghitung nilai relevansi kata. Pada tahap inilah eksperimen resmi dipecah menjadi dua arah, di mana Skenario 1 membiarkan data latih (269 ulasan) masuk ke tahap pemodelan apa adanya dalam kondisi timpang, sedangkan pada Skenario 2, data latih ditransformasikan lebih lanjut menggunakan teknik *Random Oversampling* (ROS) dari *library imblearn*. Teknik ROS ini bekerja secara terprogram dengan menduplikasi sampel kata pada kelas minoritas (sentimen negatif) secara acak hingga jumlah data sampelnya naik menjadi 151 data, sehingga seimbang secara adil (50:50) dengan jumlah kelas mayoritas (sentimen positif) sebelum memasuki fase pelatihan model.

3.4 Hasil Penambangan Data (*Data Mining*)

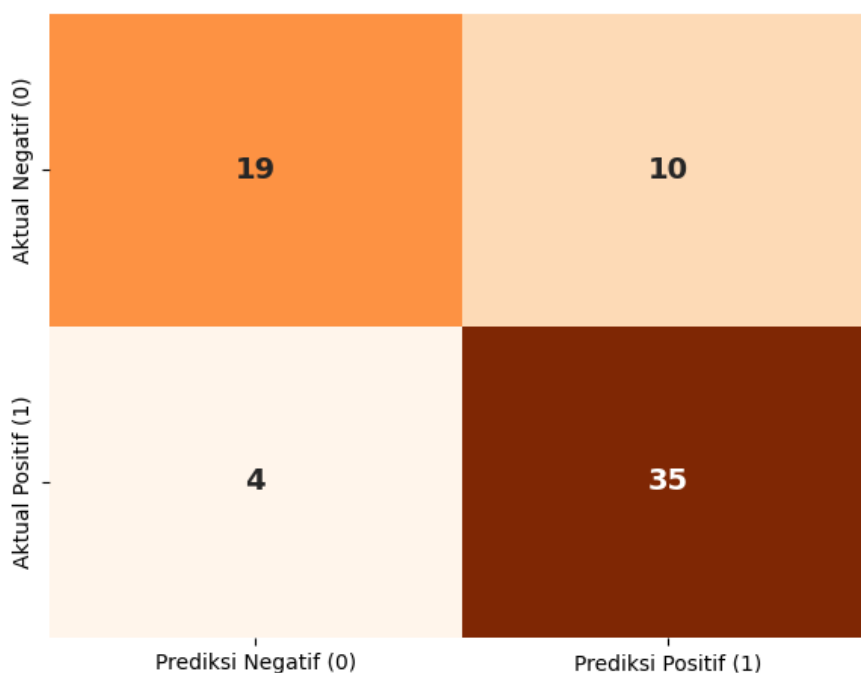
Tahap keempat dalam metode KDD adalah penambangan data (*data mining*) yang merupakan fase penerapan algoritma cerdas untuk mengekstrak pola-pola potensial dari data yang telah ditransformasikan. Algoritma yang diimplementasikan pada penelitian ini adalah *Multinomial Naïve Bayes*, sebuah metode klasifikasi probabilistik

DOI: <https://doi.org/10.69693/ijmst.v4i2.12921>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.5 Hasil Tahap Evaluasi (*Evaluation*)

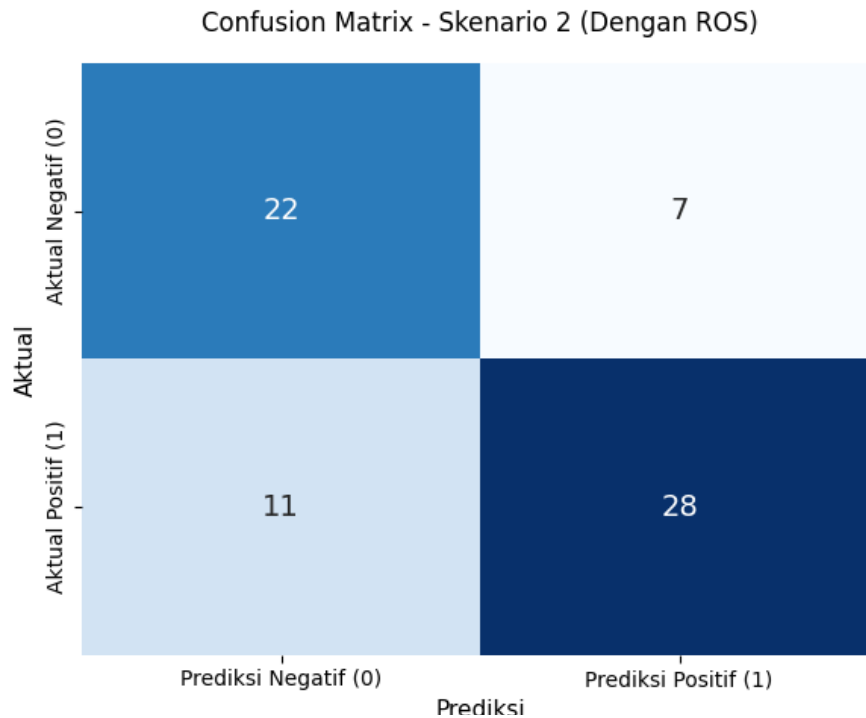
Tahap kelima atau tahapan terakhir dalam metode KDD adalah evaluasi dan interpretasi hasil (*evaluation/interpretation*). Tahap ini berfokus pada analisis mendalam untuk mengukur, membandingkan, dan menginterpretasikan performansi dari model klasifikasi *Naïve Bayes* yang telah dilatih pada Skenario 1 (tanpa penyeimbangan data) dan Skenario 2 (dengan intervensi *Random Oversampling*). Instrumen evaluasi utama yang digunakan dalam menguji 68 data uji independen ini adalah *Confusion Matrix* yang divisualisasikan berbasis grafik *heatmap* berwarna untuk memetakan secara riil ketepatan prediksi model terhadap data aktual di lapangan. Visualisasi grafik *heatmap confusion matrix* untuk hasil pengujian model pada Skenario 1 disajikan pada Gambar 6 berikut:

Gambar 3. Heatmap Confusion Matrix Skenario 1 (Tanpa ROS)

Gambar 6 Grafik Heatmap Confusion Matrix Naïve Bayes pada Skenario 1

Berdasarkan visualisasi grafik *heatmap* pada Gambar 4.6 serta hasil kalkulasi sistem, dapat diinterpretasikan performa matematis dari model Skenario 1 yang bekerja di bawah pengaruh data timpang. Model pada skenario pertama ini menghasilkan nilai akurasi global sebesar 79,41%. Berdasarkan *classification report*, terlihat bahwa model memiliki kemampuan prediksi kelas sentimen positif (label 1) yang menonjol dengan nilai *precision* sebesar 78% dan *recall* mencapai 90%, di mana dari 39 data aktual positif, model berhasil menebak dengan benar sebanyak 35 ulasan (*True Positive*) dan hanya salah menebak 4 ulasan (*False Negative*). Namun, kelemahan mendasar akibat ketimpangan data terlihat pada kelas sentimen negatif (label 0), di mana nilai *recall* hanya menyentuh angka 66%. Hal ini dibuktikan oleh data *confusion matrix* yang menunjukkan bahwa dari 29 data aktual keluhan negatif, model salah menebak sebanyak 10 ulasan sebagai sentimen positif (*False Positive*) dan hanya mampu menebak benar sebanyak 19 ulasan (*True Negative*). Kondisi ini mengonfirmasi secara riil bahwa ketidakseimbangan data awal membuat model Skenario 1 menjadi bias ke kelas mayoritas dan kurang sensitif dalam mendeteksi ulasan kritik dari masyarakat.

Untuk mengukur dampak perbaikan setelah dilakukannya penyeimbangan sebaran data latih, hasil pengujian pada Skenario 2 dievaluasi menggunakan instrumen yang sama. Visualisasi grafik *heatmap confusion matrix* untuk hasil eksekusi pemodelan pada Skenario 2 disajikan pada Gambar 7 berikut:



Gambar 7 Grafik Heatmap Confusion Matrix Naïve Bayes pada Skenario 2 (ROS).

Berdasarkan visualisasi grafik *heatmap* pada Gambar 4.7, terlihat karakteristik pergeseran performa model setelah diterapkannya teknik penyeimbangan data. Pemodelan dengan menggunakan Skenario 2 menghasilkan nilai akurasi global sebesar 73,53%. Meskipun secara akurasi global mengalami sedikit penurunan jika dibandingkan dengan Skenario 1, intervensi *Random Oversampling* (ROS) terbukti sukses memberikan dampak positif yang masif pada peningkatan sensitivitas kelas minoritas (sentimen negatif). Nilai *recall* pada sentimen negatif (label 0) berhasil melonjak naik dari 66% menjadi 76%. Berdasarkan matriks kotak riil, jumlah tebakan benar pada ulasan keluhan masyarakat meningkat dari 19 ulasan menjadi 22 ulasan (*True Negative*), serta berhasil menekan kesalahan prediksi ulasan negatif yang dikira positif (*False Positive*) menjadi hanya 7 ulasan. Penurunan akurasi global pada Skenario 2 disebabkan oleh peningkatan salah tebak pada kelas positif (*False Negative* menjadi 11 ulasan), yang merupakan implikasi logis karena algoritma kini dilatih dengan bobot probabilitas kata yang seimbang sehingga tidak lagi bias mendominasi kelas mayoritas.

Melalui perbandingan komparatif dari kedua skenario tersebut, dapat ditarik kesimpulan ilmiah bahwa penerapan teknik *Random Oversampling* (ROS) sangat efektif dalam mengatasi kendala bias akibat *class imbalance* pada dataset ulasan QRIS dompet digital. Meskipun algoritma *Multinomial Naïve Bayes* pada data asli menghasilkan akurasi global yang lebih tinggi, pemodelan Skenario 2 dengan integrasi ROS jauh lebih direkomendasikan untuk kebutuhan analisis sentimen tingkat lanjut di dunia nyata. Hal ini disebabkan karena Skenario 2 memiliki keandalan dan objektivitas yang lebih seimbang, serta jauh lebih peka dalam menjangkau dan mengenali ulasan kritik, kendala sistem, ataupun keluhan riil yang dikeluhkan oleh para pengguna aplikasi di Google Play Store.

4. Kesimpulan

Berdasarkan hasil eksperimen dan pembahasan komparatif yang telah dipaparkan, dapat ditarik kesimpulan ilmiah bahwa implementasi teknik *Random Oversampling* (ROS) memberikan dampak yang sangat signifikan dalam mengatasi permasalahan ketidakseimbangan kelas (*class imbalance*) pada algoritma *Multinomial Naïve Bayes* untuk klasifikasi sentimen ulasan QRIS. Pengujian model pada data asli tanpa penyeimbangan menghasilkan akurasi global yang lebih tinggi yaitu sebesar 79,41%, namun model tersebut terbukti bias terhadap kelas mayoritas sehingga kurang sensitif dalam mendeteksi ulasan kritik dengan nilai *recall* sentimen negatif yang hanya sebesar 66%. Sementara itu, intervensi teknik *Random Oversampling* berhasil mendistribusikan bobot probabilitas kata secara seimbang, yang dibuktikan dengan melonjaknya nilai *recall* sentimen negatif menjadi 76% serta meningkatnya kemampuan model dalam menangkap ulasan keluhan riil dari 19 menjadi 22 data secara akurat. Meskipun skenario penyeimbangan data berimplikasi pada penurunan akurasi global menjadi 73,53% akibat kompensasi peningkatan kesalahan prediksi pada kelas mayoritas, model dengan pendekatan *oversampling* ini jauh lebih objektif, andal, dan representatif untuk diaplikasikan oleh para pengembang aplikasi dompet digital dalam memonitor keluhan teknis pengguna di lapangan secara berkala. Implikasi praktis dari penelitian ini

menekankan bahwa evaluasi model analisis sentimen pada domain finansial tidak boleh hanya terpaku pada tingginya nilai akurasi global, melainkan harus memprioritaskan nilai sensitivitas (*recall*) pada kelas keluhan demi peningkatan kualitas layanan instrumen pembayaran digital nasional. Untuk pengembangan penelitian selanjutnya, disarankan untuk mengeksplorasi teknik penyeimbangan data berbasis sintesis seperti *Synthetic Minority Over-sampling Technique* (SMOTE) atau mengombinasikannya dengan algoritma klasifikasi berbasis *ensemble learning* guna menekan angka *false-negative* pada kelas mayoritas tanpa mengorbankan sensitivitas kelas minoritas.

Referensi

- Achmad Fauzi, N., & Nurina Sari, B. (2026). Klasifikasi Kesiapan Kerja Lulusan Universitas Dengan Pendekatan Data Mining Menggunakan KNN. *Jurnal Ilmiah Wahana Pendidikan*, 12(C), 207–215. <https://doi.org/https://jurnal.peneliti.net/index.php/JIWP/article/view/12487>
- Aldi Rahmad Fahrozi, M., Azhima Yoga Siswa, T., Azmi Verdikha, N., Kalimantan Timur, M., Jl Ir Juanda No, S. H., Samarinda Ulu, K., Samarinda, K., & Timur, K. (2026). Komparasi Ekstraksi Fitur TF-IDF Dan Word2Vec Pada Naïve Bayes Untuk Analisis Sentimen Pembangunan IKN Di Youtube. *Journal Of Information System Research*, 7(2), 519–527. <https://doi.org/10.47065/Josh.V7i2.9198>
- Alfarizi, N., Sinurat, S., Putra, A., Amin, M., & Lydia, P. (2026). Analisis Penanganan Data Tidak Seimbang Terhadap Kinerja Klasifikasi Sentimen Multikelas Pada Ulasan Marketplace Tokopedia. *Journal Of Science And Social Research*, (1), 474–482. <https://doi.org/10.34740/Kaggle/Ds/89925>
- Andayani, S. (N.D.). *Pembentukan Cluster Dalam Knowledge Discovery In Database Dengan Algoritma K-Means*.
- Bimo, A., Nursitasari, W. R., & Septiani, N. A. (2025a). Analisis Sentimen Masyarakat Terhadap Penggunaan Qris Pada Platform Gojek Menggunakan Metode Naives Bayes Class. *Jurnal Komputer Antartika*, 3(3), 127–135. <https://doi.org/10.70052/Jka.V3i3.1116>
- Dinar, A. N., Susilo, A., Irawan, Y., & Umaidah, Y. (2023). Analisis Sentimen Pada Pengguna Twitter Terhadap Program Kampus Merdeka Menggunakan Naïve Bayes. *Jurnal Mahasiswa Teknik Informatika*, 7(1).
- Dometian, C., & Jatmiko, S. (2025). Analisis Sentimen Ulasan Google Play Store: Studi Komparatif Algoritma SVM, Naïve Bayes, Dan Logistic Regression. *Jurnal Fasilkom*, 15 No.3.
- Fadhilah, S. N., & Utomo, F. S. (2024). Algoritma Naïve Bayes Untuk Analisis Sentiment Review Blibli.Com Di Google Play Store. *SISTEMASI: Jurnal Sistem Informasi*. <http://sistemasi.ftik.unisi.ac.id>
- Firda Aini, N., Nurdiawan, O., Suprati, T., Rinaldi Dikananda, A., & Fathurrohman. (2026). Sentiment Analysis Of “Cek Bansos” Application Reviews On Google Play Store Using The Naïve Bayes Algorithm. *Journal Of Artificial Intelligence And Engineering Applications*, 5(2), 2808–4519. <https://doi.org/https://ioinformatic.org/>
- Ifan Farimulyadi, & Zaehol Fatah. (2025). Penerapan Data Mining Untuk Rekomendasi Paket Foto Priwed Menggunakan Algoritma Apriori. *Jurnal Riset Teknik Komputer*, 2(1), 46–56. <https://doi.org/10.69714/Wjn1zp06>
- Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naïve Bayes. *Jurnal Mahasiswa Teknik Informatika*, 7(1).
- Natalia Kristanty, D. (2024). Tren Dan Tantangan Keamanan Bertransaksi Dengan QRIS Dalam Era Transformasi Sistem Pembayaran Digital. *Syntax Admiration*, 5(10).
- Novitasari, N., Nuris, N. D., & Herdiana, R. (2023). Penerapan Algoritma K-Means Untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten Di Jawa Barat Menggunakan Rapidminer. *Jurnal Informatika Terpadu*, 9(1), 68–73. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Nurfathir, F., Irawan, B., Faqih, A., Rohmat, C. L., & Dwilestari, G. (2026). Analisis Sentimen Ulasan Aplikasi Mobile Jkn Menggunakan Multinomial Naïve Bayes. *Journal Of Computer Science And Artificial Intelligence (JCSAI)*, 3. <https://ikmi.ac.id/page/18/?Lang=De>
- Radja, A., Auliasari, K., & Santi Wahyuni, F. (2026). Analisis Sentimen Pengambilalihan Aset Pasca Demo Dengan Metode Naïve Bayes. *Jurnal Informatika Universitas Muhammadiyah Tangerang P*, 10(1), 2722–2713.
- Ramadhan Alfinsyah, D., & Pili Hartato, B. (2025). Evaluating The Impact Of Random Over Sampling On Indobert Performance For Indonesian Sentiment Analysis. *Journal Of Applied Informatics And Computing (JAIC)*, 9(6), 3270. <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Syahrin, A. R., & Agustianto, M. A. (2025). Analisis Faktor Penghambat Utama Adopsi Quick Response Code Indonesian Standard (QRIS) Di Indonesia: Pendekatan Systematic Literature Review (SLR). *Jurnal Ilmiah Manajemen Ekonomi Dan Akuntansi*, 3(1), 243–249. <https://doi.org/10.62017/Jimea>