TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

Analisis Klasifikasi Sentimen Neobank: Perbandingan Konfigurasi N-Gram pada TF-IDF Menggunakan Naive Bayes dan SVM

Fatha Amin Mujtahid*, Badroe Zaman, Galet Guntoro Aji

Fakultas Teknologi Informasi dan Komunikasi, Program Studi Teknik Informatika, Universitas Semarang, Semarang, Indonesia
Email: [1,*]fathaxrp@gmail.com, [2]badroezaman@usm.ac.id, [3]gallet@usm.ac.id
Email Penulis Korespondensi: fathaxrp@gmail.com

Abstrak–Meningkatnya jumlah pengguna aplikasi Neobank di Indonesia mengakibatkan pertumbuhan volume ulasan pengguna di Google Play Store yang dapat dimanfaatkan untuk mengetahui kepuasan dan pengalaman layanan. Analisis ulasan manual sangatlah tidak efisien, sehingga dilakukan pendekatan otomatis berbasis machine learning. Penelitian ini bertujuan mengevaluasi pengaruh konfigurasi N-Gram pada ekstraksi fitur TF-IDF terhadap performa klasifikasi sentimen ulasan pengguna aplikasi Neobank di Indonesia menggunakan Naive Bayes (NB) dan Support Vector Machine (SVM). Dataset berjumlah 3798 ulasan, diproses dari 5000 data awal yang diambil dari Google Play Store Indonesia, dengan 2385 ulasan positif dan 1413 ulasan negatif, dilabelkan berdasarkan rating bintang. Dataset dibagi menggunakan stratified five-fold cross-validation untuk memastikan proporsi pembagian data kelas positif dan negatif tetap seimbang di setiap fold. Fitur diekstraksi menggunakan TF-IDF dengan tiga konfigurasi N-Gram, yaitu unigram, bigram, dan unigram+bigram. Hasil menunjukkan bahwa konfigurasi N-Gram berpengaruh signifikan terhadap performa kedua model. Pada NB, konfigurasi unigram menghasilkan akurasi tertinggi sebesar 87,65%, sedangkan pada SVM, konfigurasi unigram+bigram menghasilkan akurasi terbaik 88,61% dengan F1-score 88,22%. Konfigurasi bigram konsisten menghasilkan performa terendah pada kedua model karena ulasan yang singkat dan informal menghasilkan fitur yang lebih sparse. Penelitian ini menyimpulkan bahwa pemilihan N-Gram perlu disesuaikan dengan karakteristik algoritma, dan kombinasi SVM dengan unigram+bigram merupakan pendekatan yang paling efektif untuk klasifikasi sentimen ulasan aplikasi Neobank di Indonesia.

Kata Kunci: TF-IDF; N-Gram; SVM; Naive Bayes; Neobank

Abstract–The increasing number of Neobank users in Indonesia has led to a growth in user reviews on the Google Play Store, which can be utilized to assess service satisfaction and user experience. Manual analysis of these reviews is inefficient, prompting the use of automated machine learning approaches. This study evaluates the effect of N-Gram configurations in TF-IDF feature extraction on the performance of sentiment classification of Neobank reviews using Naive Bayes (NB) and Support Vector Machine (SVM). The dataset consists of 3,798 reviews, preprocessed from 5,000 initial entries collected from Google Play Store Indonesia, with 2,385 positive and 1,413 negative reviews labeled based on star ratings. Data were split using stratified five-fold cross-validation to ensure balanced class proportions in each fold. Features were extracted with TF-IDF using three N-Gram configurations: unigram, bigram, and unigram+bigram. Results indicate that N-Gram configuration significantly affects the performance of both models. NB achieved the highest accuracy with unigram (87.65%), while SVM performed best with unigram+bigram (88.61% accuracy and 88.22% F1-score). Bigram consistently yielded the lowest performance due to short and informal reviews producing sparser features. This study concludes that N-Gram selection should align with algorithm characteristics, and SVM with unigram+bigram is the most effective approach for sentiment classification of Neobank reviews in Indonesia.

Keywords: TF-IDF; N-Gram; SVM; Naive Bayes; Neobank

1. PENDAHULUAN

Perkembangan pesat di sektor teknologi finansial dan perbankan digital telah mengubah cara masyarakat berinteraksi dengan lembaga keuangan secara mendalam. Saat ini, kegiatan bertransaksi bukanlah suatu hal yang sukar untuk dilakukan, mengingat layanan perbankan tidak lagi terbatas pada kantor fisik atau mesin ATM. Tanpa terkendala batasan ruang dan waktu, pengguna dapat mudah melakukan berbagai aktivitas finansial, mulai dari transfer dana, pembayaran tagihan, hingga pembukaan rekening baru, hanya melalui genggaman smartphone dan koneksi internet (Prayudya et al., 2025). Di Indonesia, penggunaan aplikasi perbankan digital berkembang pesat, termasuk aplikasi Neobank sebagai salah satu kanal layanan berbasis *mobile* (Rahmatulloh et al., 2024). Daya tarik aplikasi ini umumnya terkait pada kemudahan proses layanan melalui aplikasi serta pemberian insentif layanan tertentu, yang terbukti efektif meningkatkan intensitas transaksi harian pengguna (Helmi & Kristianto, 2024). Seiring dengan meningkatnya jumlah pengguna aplikasi Neobank, volume ulasan yang disampaikan pengguna di platform digital seperti Google Play Store mengalami pertumbuhan yang signifikan.

Ulasan pengguna tersebut berisi pengalaman, kepuasan, maupun keluhan terhadap layanan yang dapat menjadi indikator penting bagi keberlangsungan layanan. Dalam industri perbankan yang sangat mengandalkan kepercayaan, ulasan ini menjadi salah satu aset yang sangat berharga. Informasi tersebut memiliki potensi untuk menjadi sumber masukan yang berharga bagi penyedia layanan dalam mengevaluasi sistem, stabilitas aplikasi, serta kepercayaan pengguna. Namun dalam penerapan praktiknya, jumlah ulasan yang sangat besar dan terus bertambah setiap harinya dapat menjadikan proses analisis secara manual menjadi tidak efisien dan sukar dilakukan secara berkelanjutan. Selain itu, penggunaan Bahasa yang beragam, tidak baku, penuh dengan singkatan, serta penggunaan emoji, menimbulkan kendala dalam proses identifikasi sentimen secara konsisten dan objektif.

Pendekatan otomatis berbasis teknologi sangat diperlukan untuk mengatasi permasalahan tersebut, karena kemampuannya mengolah ulasan pengguna dalam jumlah besar secara cepat, konsisten dan akurat. Salah satu pendekatan yang banyak digunakan adalah analisis sentimen berbasis *machine learning* dan pemrosesan bahasa alami. Pendekatan ini telah banyak digunakan dalam penelitian analisis sentimen ulasan aplikasi digital dan terbukti efektif

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

dalam mengatasi keterbatasan analisis manual (Kusnawi et al., 2023)(Kadek et al., 2025). Dalam klasifikasi teks, performa model tidak hanya ditentukan oleh algoritma klasifikasi, tetapi juga oleh metode representasi fitur yang digunakan. Salah satu metode representasi teks yang banyak digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF), yang efektif dalam merepresentasikan kepentingan kata dengan mengurangi kontribusi kata-kata yang sering muncul namun memiliki nilai informasi rendah (Putra et al., 2023). TF-IDF kerap dikombinasikan dengan variasi N-Gram, seperti unigram dan bigram, untuk merepresentasikan makna kata dan konteks frasa dalam teks (Hadi & Utami, 2024).

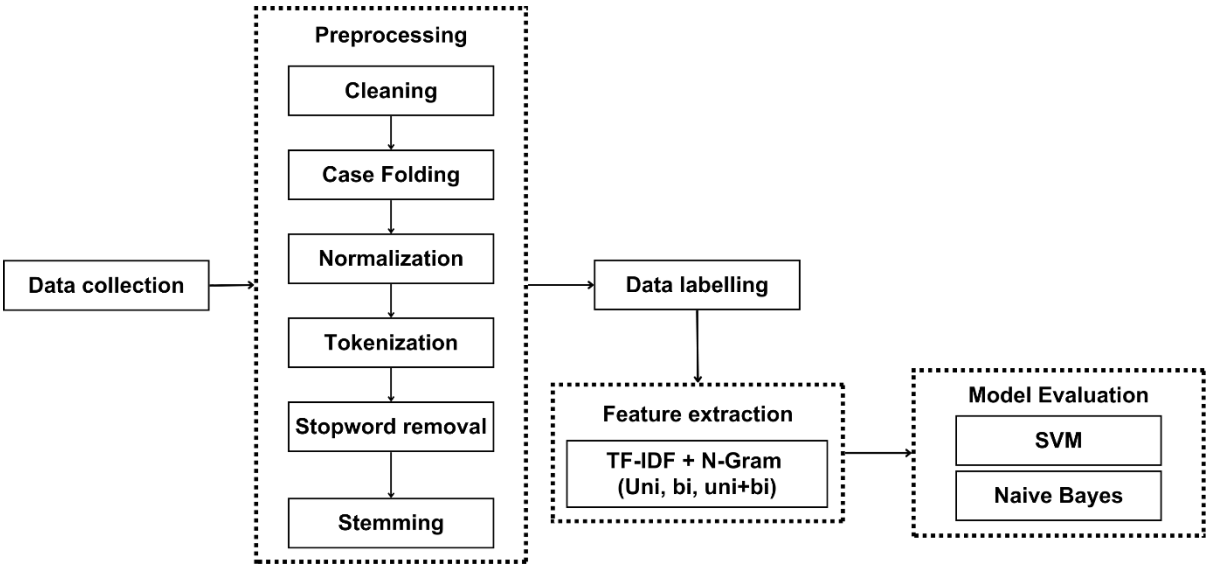
Pada suatu penelitian mengenai analisis sentimen ulasan aplikasi Gojek menunjukkan bahwa TF-IDF gabungan unigram dan bigram yang dikombinasikan dengan suatu algoritma dilaporkan mencapai akurasi 84% (Adyatma et al., 2023). Dalam penelitian analisis sentimen opini publik terkait kebijakan efisiensi anggaran pemerintah Indonesia pada media sosial X, penerapan algoritma *Support Vector Machine* (SVM) dengan ekstraksi fitur TF-IDF menghasilkan performa terbaik pada kernel linear dengan akurasi 75,79% (Nurhayati et al., 2026). Penelitian lain juga menunjukkan bahwa klasifikasi sentimen ulasan produk skincare di platform e-commerce Shopee dapat diotomatisasi menggunakan TF-IDF, variasi N-Gram (unigram dan bigram), dan Multinomial Naïve Bayes, dengan model mencapai akurasi 70,06% (Putri & Soeleman, 2026). Temuan tersebut menunjukkan bahwa konfigurasi fitur dan algoritma klasifikasi memiliki peran penting dalam menentukan performa analisis sentimen.

Meskipun demikian, efektivitas konfigurasi N-Gram tidak selalu konsisten pada setiap algoritma dan konteks data yang digunakan. Masih sangat sedikit studi yang secara khusus membandingkan pengaruh variasi konfigurasi N-Gram terhadap performa TF-IDF dalam menganalisis ulasan aplikasi Neobank berbahasa Indonesia. Ulasan berbahasa Indonesia memiliki perbedaan tersendiri, seperti penggunaan bahasa informal, campuran kata dari bahasa lain, dan singkatan yang tidak baku. Oleh karena itu, cara terbaik untuk merepresentasikan fitur pada domain lain belum tentu efektif di sini. Kondisi ini menimbulkan terbukanya celah penelitian, di mana belum ada studi sistematis yang secara langsung membandingkan pengaruh konfigurasi unigram, bigram, dan gabungan unigram+bigram pada dua algoritma dengan karakteristik yang berbeda, yaitu NB dan SVM, dalam analisis sentimen ulasan Neobank Indonesia.

Oleh karena itu, penelitian ini menggunakan algoritma NB dan SVM sebagai model klasifikasi untuk mengevaluasi pengaruh konfigurasi *N-Gram* pada ekstraksi fitur TF-IDF terhadap performa klasifikasi. Pemilihan kedua algoritma tersebut didasarkan pada perbedaan karakteristiknya dalam memodelkan data teks, sehingga memungkinkan evaluasi yang lebih komprehensif terhadap representasi fitur yang digunakan. Dengan demikian, Penelitian ini bertujuan untuk melakukan analisis pengaruh konfigurasi N-Gram pada ekstraksi fitur TF-IDF terhadap performa klasifikasi sentimen ulasan pengguna aplikasi Neobank di Indonesia menggunakan algoritma NB dan SVM. Hasil penelitian diharapkan dapat menjadi dasar pertimbangan pemilihan konfigurasi representasi fitur yang efektif untuk analisis sentimen, serta mendukung upaya peningkatan kualitas layanan perbankan digital.

2. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian eksperimen yang bertujuan untuk menganalisis pengaruh konfigurasi *N-Gram* pada ekstraksi fitur TF-IDF terhadap performa klasifikasi sentimen ulasan pengguna aplikasi Neobank menggunakan algoritma NB dan SVM.



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur metodologi penelitian secara keseluruhan. Tahap pertama dimulai dari pengumpulan data dari Google Play Store Indonesia. Selanjutnya data di proses melalui tahapan preprocessing untuk menghasilkan teks yang bersih dan siap untuk diolah. Tahap ketiga adalah pelabelan data berdasarkan rating bintang

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

untuk menentukan kelas sentimen positif dan negatif. Kemudian, fitur diekstraksi menggnuakn TF-IDF dengan tiga konfigurasi N-Gram. Tahap terakhir adalah evaluasi model menggunakan algoritma NB dan SVM dengan skema stratified *five-fold cross-validation.*

2.1 *Data Collection*

Pengumpulan data ulasan pengguna aplikasi Neobank diambil dari platform Google Play Store menggunakan *library google-play-scraper* yang tersedia di bahasa pemrograman Python (Arifin et al., 2025). Dataset awal berjumlah 5000 ulasan pada rentang 22 April 2025 hingga 30 Oktober 2025, dengan atribut utama teks ulasan dan rating bintang. Pemanfaatan ulasan pada Google Play Store sebagai sumber data untuk analisis sentimen telah banyak digunakan dalam berbagai penelitian terkini, termasuk penelitian yang mengkaji ulasan terhadap aplikasi layanan digital (Rahmaliyadi & Maridjan, 2025).

2.2 *Preprocessing*

Data ulasan yang telah diambil selanjutnya diproses melalui serangkaian tahapan untuk menghasilkan data teks yang bersih dan siap diekstraksi fiturnya. Dari data ulasan yang didapatkan, proses akan melalui tahapan :

a. *Cleaning* : untuk membersihkan elemen *noise* seperti angka, tanda baca, simbol, dan emoji,
b. *Case folding* : mengubah seluruh format teks menjadi huruf kecil (*lowercase)* untuk menyamakan representasi kata, sehingga kata yang sama meskipun berbeda kapitalisasi dianggap setara oleh model
c. *Normalization* : dilakukan normalisasi kata dilakuakan untuk memperbaiki ejaan kata-kata slang dan informal menjadi kata baku menggunakan kamus kata baku (Zafira et al., 2025).
d. *Tokenization* : untuk memisahkan teks menjadi potongan kata-kata terpisah.
e. *Stopword removal* : dilakukan untuk menghapus kata-kata umum yang sering muncul namun tidak memiliki kontribusi sentimen yang signifikan (seperti “yang”, “di”, “dan”) menggunakan *library Natural Language Toolkit* (NLTK) (Raharjo et al., 2022).
f. *Stemming* : untuk mengubah kata berimbuhan dan membuatnya menjadi kata dasar menggunakan *library* sastrawi yang dikhususkan untuk pemrosesan Bahasa Indonesia (Ulgasesa et al., 2022).

Sebagian hasil setiap tahap preprocessing ditampilkan pada Tabel 1.

Tabel 1. Tahapan *Preprocessing*

Tahap Preprocessing	Hasil
Ulasan	Terima kasih Neobank, sudah berkali kali ajukan pinjaman disini, bunga rendah, sangat membantu sekali
Cleaning	Terima kasih Neobank sudah berkali kali ajukan pinjaman disini bunga rendah sangat membantu sekali
Case Folding	terima kasih neobank sudah berkali kali ajukan pinjaman disini bunga rendah sangat membantu sekali
Normalization	terima kasih neobank sudah berkali kali ajukan pinjaman disini bunga rendah sangat membantu sekali
Tokenization	['terima', 'kasih', 'neobank', 'sudah', 'berkali', 'kali', 'ajukan', 'pinjaman', 'disini', 'bunga', 'rendah', 'sangat', 'membantu', 'sekali']
Stopword Removal	['terima', 'kasih', 'neobank', 'berkali', 'kali', 'ajukan', 'pinjaman', 'bunga', 'rendah', 'membantu']
Stemming	terima kasih neobank kali kali aju pinjam bunga rendah bantu

Tabel 1 menunjukkan sebagian hasil dari setiap langkah dari tahap preprocessing yang diterapkan pada satu ulasan. Terlihat hasil dari proses *cleaning* berhasil menghapus tanda baca, sementara *case folding* mengubah seluruh teks menjadi huruf kecil. Proses *stopword removal* menghilangkan kata-kata umum seperti *‘sudah’, ‘disini’, ‘sangat’ dan ‘sekali’.* Stemming mengubah kata berimbuhan menjadi bentuk dasarnya seperi *‘berkali’* menjadi *‘kali’*, *‘ajukan’* menjadi *‘aju’, ‘pinjaman’* menjadi *‘pinjam’* dan *‘pinjaman’* menjadi *‘pinjam’.* Rangkaian tahaoan ini menghasilkan teks yang lebih ringkas dan siap digunakan dalam proses ekstraksi fitur.

2.3 *Data Labelling*

Pelabelan sentimen dilakukan dengan menggunakan rating Bintang sebagai acuan pelabelan. Rating 1-2 dikategorikan sebagai sentimen negative dan rating 4-5 sebagai sentimen positif, sedangkan rating 3 dianggap netral dan tidak digunakan dalam pengujian utama agar focus penelitian berada pada klasifikasi dua kelas (positif dan negative). Setelah pelabelan, data dibersihkan dengan menghapus baris ulasan yang memiliki teks kosong (*missing value*) agar tidak mengganggu proses ekstraksi fitur dan pelatihan model. Sehingga dataset akhiryang digunakan berjumlah 3798 ulasan, yang terdiri dari 2385 ulasan posisitf dan 1413 ulasan negatif

2.4 *Feature Extraction*

Feature Extraction dilakukan menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) dan variasi konfigurasi N-Gram untuk menganalisis sentimen ulasan pengguna aplikasi Neobank. Proses ini mengubah teks ulasan menjadi representasi numerik yang digunakan dalam model klasifikasi (Bimantara & Zufria, 2024). Fitur diekstraksi

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

menggunakan N-Gram dengan konfigurasi unigram *(ngram_range = (1,1))*, bigram *(ngram_range = (2,2))*, dan gabungan unigram+bigram *(ngram_range = (1,2)). Term Frequency* (TF) dihitung berdasarkan jumlah kemunculan kata dalam dokumen.

$$TF(t,d) = f(t,d) \quad (1)$$

Pada rumus (1), t merepresentasikan kata tertentu, d merepresentasikan dokumen ulasan, dan $f(t,d)$ menunjukkan banyaknya kemunculan kata *t* pada dokumen *d.* Tingginya frekuensi suatu kata muncul dalam satu ulasan, maka semakin besar nilai TF yang diperoleh dan tinggi bobotnya dalam menggambarkan isi atau karakteristik ulasan tersebut.

$$DF(t) = |\{d: f(t,d) > 0\}| \quad (2)$$

Rumus (2) *Document Frequency* (DF) digunakan untuk mengukur berapa banyak dokumen yang mengandung term tersebut. *DF(t)* menunjukkan jumlah dokumen yang mengandung kata *t*, sedangkan himpunan *d:f(t,d)>0* menunjukkan sekumpulan dokumen dengan frekuensi kemunculan kata *t* lebih dari nol. Nilai DF digunakan untuk mengetahui seberapa sering suatu kata muncul dalam banyak dokumen. Kata yang muncul di banyak dokumen bersifat umum, sedangkan kata yang hanya muncul pada sebagian dokumen dapat menunjukkan konteks atau topik yang lebih khusus.

$$IDF(t) = log(\frac{N+1}{DF(t)+1}) + 1 \quad (3)$$

Rumus (3) digunakan untuk mendapatkan nilai *Inverse Document Frequency* (IDF), menghitung seberapa penting suatu kata dengan memberi bobot lebih pada kata yang jarang muncul di seluruh dokumen, menggunakan rumus dengan konfigurasi *smooth_id = True* untuk menghindari nilai IDF yang sangat kecil. *N* menunjukkan total dokumne ulasan yang digunakan, sedangkan *DF(t)* menunjukkan jumlah dokumen yang memuat kata *t.* Penambahan angka 1 pada pembilang dan penyebut merupakan bentuk *smoothing* agar perhitungan lebih stabil serta mencegah kemungkinan terjadinya pembagian dengan nol. Semakin sedikit suatu kata muncul dalam dokumen, semakin tinggi nilai *IDF* yang diperoleh, sehingga kata tersebut dinilai memiliki informasi yang lebih kuat dalam membedakan isi dokumen.

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \quad (4)$$

Rumus (4) menunjukkan bahwa nilai TF-IDF dihitung dengan perkalian antara niali TF dan IDF dari suatu *term* dalam dokumen. $TF - IDF(t,d)$ menunjukkan bobot kata *t* dalam dokumen *d*. Kata akan memiliki bobot tinggi apabila sering muncul dalam satu dokumen, tetapi tidak terlalu sering muncul dalam keseluruhan dokumen. Oleh karena itu, TF-IDF membantu model klasifikasi dalam mengenali kata-kata yang lebih penting, terutama kata yang dapat membedakan ulasan positif dan negatif

$$X_{(td)} = \frac{TFIDF_{(td)}}{\sqrt{\sum_{t=0}^{n} TFIDF_{(td)}^2}} \quad (5)$$

Rumus (5) digunakan untuk menormalisasikan vektor hasil TF-IDF dengan L2 *normalization* (*norm = "l2"). $X_{(td)}$* merupakan vektor fitur TF-IDF yang telah dinormalisasi. Bagian penyebut pada rumus menunjukkan panjang vektor, yang dihitung dari gabungan seluruh nilai bobot TF-IDF dalam satu dokumen. Proses normalisasi ini dilakukan agar setiap dokumen memiliki ukuran vektor yang seimbang. Sehingga, ulasan yang lebih panjang tidak otomatis dianggap lebih berpengaruh dibandingkan ulasan yang lebih pendek saat proses klasifikasi dilakukan.

2.5 *Model Evaluation*

Klasifikasi dilakukan menggunakan algoritma *Multinomial Naive Bayes* (NB) dan SVM kernel linear. *Multinomial Naive Bayes* sering menjadi metode andalan dalam klasifikasi teks karena efisiensi komputasinya dan kemampuannya menangani data dengan fitur *sparse* berdimensi tinggi (Enjelia et al., 2025). SVM dengan kernel linear dipilih karena efektivitasnya dalam menangani data teks berdimensi tinggi seperti TF-IDF, yang terbukti menghasilkan kinerja klasifikasi optimal untuk analisis sentimen ulasan aplikasi perbankan (Setiawan & Hasan, 2025).

Pengujian dilakukan menggunakan *k-fold cross-validation* dimana menggunakan skema stratified untuk menjaga proporsi kelas Positif dan Negatif pada setiap *fold* (Özyirmidokuz & Elmas, 2025). Skema *five-fold cross-validation* sering digunakan dalam studi pembelajaran mesin untuk memperoleh evaluasi yang lebih andal dan dapat digeneralisasi (Manjula et al., 2025). Setiap kombinasi algoritma dan konfigurasi N-Gram dievaluasi berdasarkan rata-rata metrik akurasi, presisi, *recall*, dan *F1-score* dari seluruh *fold*. Selain metrik evaluasi, performa model juga dianalisis menggunakan *confusion matrix* untuk melihat distribusi prediksi per kelas.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Rumus (6) digunakan untuk memperoleh nilai *accuracy*, yaitu perbandingan antara jumlah prediksi yang benar dengan total data yang diuji. *True Positive* (TP) dan *True Negative* (TN) menunjukkan jumlah prediksi benar untuk

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

kelas positif dan negatif, sedangkan *False Positive* (FP) dan *False Negative* (FP) menunjukkan jumlah prediksi yang salah.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Rumus (7) digunakan untuk menghitung tingkat ketepatan model dalam melakukan klasifikasi ke dalam kelas positif. Nilai dihitung dengan membandingkan jumlah TP dengan seluruh data yang diprediksi positif, yaitu gabungan TP dan FP. Semakin tinggi nilai yang didapat, menunjukkan bahwa model memiliki jumlah kesalahan prediksi positif yang rendah.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Rumus (8) digunakan untuk mencari nilai *recall*, yaitu kemampuan model dalam mengenali data yang sebenarnya termasuk ke dalam kelas positif. Nilai diperoleh dari pebandingan antara TP dengan seluruh data positif sebenarnya, yaitu TP ditambahdengan FN. Nilai recall yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data positif dan hanya sedikit data positif yang salah diklasifikasikan sebagai negatif

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Rumus (9) digunakan untuk menghitung *F1-Score*, yaitu nilai yang menggabungkan *precision* dan *recall*. *F1-Score* digunakan untuk melihat keseimbangan antara ketepatan prediksi positif dan kemampuan model dalam menemukan data positif.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Seluruh tahapan penelitian dilakukan menggunakan *Google Colab* dengan Bahasa pemrograman *Python*. Pengumpulan data dilakukan menggunakan library *google-play-scraper*, sedangkan ekstraksi fitur, pemodelan, dan evaluasi dilakukan menggunakan pustaka *scikit-learn* serta library NLP pendukung seperti *NLTK* dan *Sastrawi*. Penelitian ini menggunakan data ulasan pengguna aplikasi Neobank Indonesia berjumlah 5000 dengan atribut utama teks ulasan dan rating Bintang. Kemudian diproses melalui tahapan *preprocessing*, *labelling*, dan pembersihan data kosong sehingga didapatkan data akhir yang siap untuk digunakan adalah 3798 ulasan. Dari jumlah tersebut, didapatkan 2385 ulasan dengan ulasan berlabel positif dan 1413 berlabel negatif. Distribusi jumlah data setiap kelas sentimen terdapat pada Tabel 2.

Tabel 2. Data Label

Kelas Sentimen	Jumlah Data	Persentase (%)
Positif	2385	62,81%
Negatif	1413	37,19%
Total	3798	100%

Proses ekstraksi fitur dilakukan menggunakan TF-IDF dengan tiga konfigurasi yaitu unigram *(ngram_range = (1,1))*, bigram *(ngram_range = (2,2))*, dan gabungan unigram+bigram *(ngram_range = (1,2))*. Pada konfigurasi unigram yang terdapat pada Tabel 3, term dengan skor TF-IDF tertinggi adalah *"bagus"* dengan skor 0,0766 diikuti *"aplikasi"* sebesar 0,0742, *"bantu"* sebesar 0,0362, *"mudah"* sebesar 0,0354, dan *"bank"* sebesar 0,0331. Term *"aplikasi"* memiliki TF tertinggi dengan total 1316 dan muncul pada 1151 dokumen. Hasil menunjukkan bahwa pada konfigurasi unigram, ulasan pengguna banyak berupa kata yang memiliki unsur opini seperti "*bagus"*, "*mudah"*, dan "*mantap",* serta kata yang berkaitan dengan fitur layanan seperti "*bank"*, "*pinjam"*, dan "*tabung".*

Tabel 3. Hasil skor TF-IDF tertinggi konfigurasi unigram

Term	TF	DF	IDF	TF-IDF
bagus	620	602	2.8406	0.0766
aplikasi	1316	1151	2.1932	0.0742
bantu	348	336	3.4224	0.0362
mudah	366	352	3.376	0.0354
Bank	598	467	3.094	0.0331
mantap	200	198	3.9492	0.0317
pinjam	569	444	3.1444	0.0312
Neo	505	394	3.2636	0.0277
tabung	496	346	3.3932	0.0233
neobank	309	268	3.6478	0.0225

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

Pada konfigurasi bigram, term yang terbentuk berupa frasa dua kata. Berdasarkan Tabel 4, term dengan skor TF-IDF tertinggi Adalah “aplikasi bagus” sebesar 0,0276, diikuti “neo bank” sebesar 0,0176, “aju pinjam” sebesar 0,0100, “terima kasih” sebesar 0,0096, dan “bagus aplikasi” sebesar 0,0077. Term “neo bank” memiliki TF sebanyak 240 dan muncul pada 200 dokumen, sedangkan “aplikasi bagus” memiliki TF 178 dan muncul pada 178 dokumen. Frasa seperti buka rekening, top up, dan aju pinjam menunjukkan bahwa konfigurasi bigram mampu menangkap konteks layanan secara lebih spesifik dibandingkan unigram. Kata seperti “buka rekening”, “top up”, dan “aju pinjam” menunjukkan bahwa konfigurasi bigram mampu menangkap konteks layanan secara lebih spesifik dibandingkan unigram.

Tabel 4. Hasil skor TF-IDF tertinggi konfigurasi bigram

Term	TF	DF	IDF	TF-IDF
aplikasi bagus	178	178	4.0551	0.0276
neo bank	240	200	3.9392	0.0176
aju pinjam	110	106	4.5697	0.0100
terima kasih	95	94	4.6886	0.0096
bagus aplikasi	43	43	5.4583	0.0077
aplikasi bantu	63	62	5.0994	0.0076
bank neo	81	69	4.9940	0.0071
buka rekening	64	58	5.1650	0.0063
top up	106	101	4.6175	0.0059
aplikasi mudah	44	44	5.4358	0.0059

Berdasarkan Tabel 5, pada konfigurasi unigram+bigram, term dengan skor tertinggi masih banyak didominasi oleh unigram. Term “bagus” memiliki skor 0,0576, “aplikasi” 0,0510, “bantu” 0,0259, “mudah” 0,0242, “mantap” 0,0241, dan “bank” 0,0233. Hanya satu bigram yang masuk ke sepuluh besar, yaitu “aplikasi bagus” dengan skor 0,0198. Pola ini menunjukkan bahwa pada konfigurasi ini kontribusi utama berasal dari kata tunggal, sedangkan bigram menambahkan konteks tertentu.

Tabel 5. Hasil skor TF-IDF tertinggi konfigurasi unigram+bigram

Term	TF	DF	IDF	TF-IDF
bagus	620	602	2.8406	0.0576
aplikasi	1316	1151	2.1932	0.0510
bantu	348	336	3.4224	0.0259
mudah	366	352	3.376	0.0242
mantap	200	198	3.9492	0.0241
bank	598	467	3.0940	0.0233
pinjam	569	444	3.1444	0.0220
neo	505	394	3.2636	0.0200
aplikasi bagus	178	178	4.0551	0.0198
neobank	309	268	3.6478	0.0160

Tahap dilanjutkan dengan evaluasi model menggunakan algoritma NB dan SVM. Pengujian dilakukan dengan *stratified 5-fold cross-validation* untuk menjaga proporsi kelas pada setiap *fold*. Nilai yang disajikan pada Tabel 6 merupakan rata-rata hasil evaluasi dari seluruh *fold*.

Pada model NB, konfigurasi unigram menghasilkan *accuracy* 87,65%, *precision* 87,06%, *recall* 86,45%, dan *F1-score* 86,69%. Pada konfigurasi bigram, performa menurun menjadi *accuracy* 79,23%, *precision* 81,74%, *recall* 73,83%, dan F1-score 75,26%. Pada konfigurasi unigram+bigram, NB menghasilkan *accuracy* 86,49%, *precision* 86,48%, *recall* 84,34%, dan *F1-score* 85,16%. Hasil ini menunjukkan bahwa pada NB, konfigurasi unigram memberikan performa tertinggi, sedangkan bigram memberikan performa terendah.

Pada model SVM, konfigurasi unigram menghasilkan *accuracy* 88,84%, *precision* 87,98%, *recall* 88,39%, dan *F1-score* 88,13%. Ketika menggunakan bigram, performa turun menjadi *accuracy* 81,54%, *precision* 82,39%, *recall* 77,67%, dan *F1-score* 78,94%. Pada konfigurasi unigram+bigram, SVM menghasilkan *accuracy* 88,86%, *precision* 87,93%, *recall* 88,67%, dan *F1-score* 88,22%. Jika dibandingkan dengan unigram, pada unigram+bigram memberikan peningkatan performa sedikit lebih tinggi dengan selisih *accuracy* hanya 0,02%.

Tabel 6. Rata-rata hasil Evaluasi Model Berdasarkan Konfigurasi N-Gram

Model	N-Gram	Accuracy	Precision	Recall	F1-Score
NB	Unigram	87,65%	87,06%	86,45%	86,69%
	Bigram	79,23%	81,74%	73,83%	75,26%
	Uni+Bi	86,49%	86,48%	84,34%	85,16%
SVM	Unigram	88,84%	87,98%	88,39%	88,13%
	Bigram	81,54%	82,39%	77,67%	78,94%
	Uni+Bi	88,86%	87,93%	88,67%	88,22%

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

Tabel 7 merupakan *confusion matrix* model dengan nilai evaluasi paling tinggi pada SVM dengan konfigurasi unigram+bigram. Berdasarkan output pengujian, dari total 1413 data negatif, sebanyak 1242 berhasil diklasifikasikan dengan benar sebagai negatif dan 171 salah diprediksi sebagai positif. Pada kelas positif, sebanyak 2133 data diprediksi dengan benar sebagai positif dan 252 salah diprediksi sebagai negatif. Hasil ini menandakan bahwa model dapat membedakan kedua kelas dengan cukup baik, meskipun masih terdapat kesalahan prediksi, terutama karena distribusi data yang lebih besar pada kelas positif. Nilai *accuracy* yang mencapai 88,86% menunjukkan tingkat keakuratan model dalam memprediksi sentimen ulasan pengguna Neobank

Tabel 7. Confusion Matrix model SVM dengan unigram+bigram

	Negatif	Positif
Negatif	1242	171
Positif	252	2133
Accuracy		88.86%

3.2 Pembahasan

Hipotesis pada penelitian ini adalah bahwa konfigurasi *N-Gram* pada ekstraksi fitur TF-IDF dapat memberikan pengaruh performa klasifikasi sentimen ulasan pengguna aplikasi Neobank, dan pengaruh tersebut dapat berbeda pada setiap algoritma klasifikasi yang digunakan. Hasil pada Tabel 6 mendukung hipotesis tersebut, ketiga konfigurasi N-Gram menghasilkan nilai performa yang berbeda-beda pada model NB maupun SVM

Dari hasil pengujian yang dilakukan pada Tabel 6, konfigurasi bigram secara konsisten memberikan performa terendah pada kedua model. Pada NB, *accuracy* turun dari 87,65% pada unigram menjadi 79,23% pada *bigram*, dengan selisih 8,42%. Pada SVM *accuracy* turun dari 88,84% menjadi 81,54%, dengan selisih 7,3%. Pola ini menunjukkan bahwa hanya menggunakan frasa dua kata belum cukup stabil untuk untuk merepresentasikan variasi sentimen pada ulasan Neobank. Ulasan *Google Play Store* biasanya singkat, tidak baku, dan bervariasi, sehingga frasa dua kata muncul lebih jarang dan menghasilkan ruang fitur yang lebih *sparse* (jarang terisi). Kondisi tersebut membuat model lebih sulit mengenali pola sentimen dibandingkan ketika menggunakan unigram atau unigram+bigram.

Hasil ini sejalan dengan penelitian (Adyatma et al., 2023) yang menunjukkan bahwa kombinasi unigram dan bigram dapat meningkatkan performa klasifikasi sentimen pada ulasan aplikasi Gojek di *Play Store*, namun besar pengaruhnya juga disebabkan dengan algoritma yang digunakan. Pada peneletian tersebut menghasilkan akurasi 84% untuk model NB yang digabungankan dengan unigram dan bigram, sehingga dapat disimpulkan bahwa peran konfigurasi *N-Gram* memiliki sifat yang berbeda-beda menyesuaikan konteks data dan model yang digunakan.

Pada model NB, konfigurasi unigram menghasilkan performa tertinggi dari dua konfigurasi lainnya dengan nilai *accuracy* 0,8765. Hal ini menunjukkan bahwa kata tunggal masih menjadi representasi paling konsisten untuk model yang bekerja berdasarkan kemungkinan (*probability-based*). Pada konfigurasi unigram+bigram justru model mengalami penurunan *accuracy* menjadi 86,49%, selisih 1,16% terhadap unigram. Ini menunjukkan bahwa pada NB, penambahan konteks frasa belum memberikan kenaikan performa yang cukup untuk melampaui kestabilan representasi Tunggal. Pola ini juga sama dengan penelitian rujukan pada klasifikasi buku yang menunjukkan bahwa konfigurasi unigram menghasilkan akurasi pada NB sebesar 74,4%, lebih tinggi dibandingkan bigram dan trigram. Hal tersebut menunjukkan bahwa NB lebih unggul menggunakan fitur yang sederhana, padat, dan sering muncul secara konsisten dalam korpus, dibandingkan fitur berbentuk frasa yang cenderung lebih jarang. Dengan demikian, pada data ulasan Neobank yang umumnya pendek dan informal, representasi unigram sesuai dengan karakter kerja NB.

Pada model SVM, konfigurasi unigram+bigram menghasilkan nilai tertinggi, tetapi keunggulannya sangat tipis dengan unigram, nilai *accuracy* naik dari 88,84% menjadi 88,86%. Hal ini menunjukkan bahwa SVM dapat memanfaatkan tambahan konteks dari bigram, namun kontribusinya tidak terlalu besar karena unigram sudah dapat merepresentasikan sebagaian besar informasi penting. Meskipun hanya mengalami peningkatan sebanyak 0,02%, hasil ini tetap memperlihatkan bahwa kombinasi kata tunggal dengan frasa dua kata memberikan ruang representasi yang lebih luas bagi SVM dalam melakukan klasifikasi sentimen.

Hasil pengujian dengan model SVM juga sejalan dengan penelitian (Nurhayati et al., 2026) yang menunjukkan bahwa TF-IDF berbasis *N-Gram* dengan model SVM *kernel linear* menghasilkan performa terbaik dengan akurasi 75,79% pada kasus analisis sentimen opini publik terkait kebijakan efisiensi anggaran pemerintah Indonesia dalam platform media sosial X. Kesamaan temuan ini menunjukkan bahwa SVM memiliki kemampuan yang cukup baik dalam memanfaatkan representasi fitur teks berbasis TF-IDF dan *N-Gram*, meskipun konteks data yang digunakan berbeda.

Dari representasi fitur yang dihasilkan, tabel 3 hingga tabel 5 menunjukkan bahwa konfigurasi unigram dan unigram+bigram masih didominasi oleh kata tunggal seperti *"bagus", "aplikasi", "bantu", "mudah",* dan *"bank".* Sementara itu, pada konfigurasi bigram mulai muncul konteks yang lebih spesifik seperti *"aplikasi bagus", "neo bank", "aju pinjam", "buka rekening",* dan *"top up".* Perbedaan ini menunjukkan bahwa bigram dapat menangkap konteks layanan dengan lebih spesifik, tetapi belum tentu menghasilkan performa klasifikasi yang lebih baik. Dengan kata lain, fitur yang lebih kaya tidak selalu lebih efektif jika frekuensi kemunculannya rendah dan penyebarannya dalam data tidak merata.

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

Pada Tabel 7, *confusion matrix* model SVM dengan konfigurasi *unigram+bigram* menunjukkan bahwa model mampu mengenali kedua kelas dengan cukup baik. Dari 1413 data negatif, sebanyak 1242 diklasifikasikan benar sebagai negatif, sedangkan 171 salah diklasifikasikan sebagai positif. Pada kelas positif, 2133 data berhasil diprediksi benar sebagai positif dan 252 salah diklasifikasikan sebagai negatif. Kondisi ini menunjukkan bahwa model dapat melakukan prediksi dengan seimbang, meskipun jumlah prediksi benar pada kelas positif lebih tinggi. Hasil ini kemungkinan dipengaruhi oleh data yang sebagian besar didominasi oleh kelas positif seperti yang tertera di Tabel 2. Ketidakseimbangan distribusi kelas berdampak pada tingginya kesalahan prediksi pada kelas negatif, di mana sebanyak 171 data negatif salah diprediksi sebagai sebagai positif. Hal ini menunjukkan bahwa model cenderung lebih sering memprediksi kelas mayoritas (positif), sehingga kelas minoritas (negatif) lebih mudah mengalami salah klasifikasi.

Secara keseluruhan dari bagian pembahasan ini menunjukkan bahwa hasil penelitian mendukung hipotesis bahwa konfigurasi N-Gram dapat memengaruhi performa klasifikasi sentimen. Pengaruh tersebut juga tidak bersifat sama pada setiap model, karena pada NB konfigurasi *unigram* memberikan hasil yang paling baik, sedangkan pada SVM konfigurasi unigram+bigram menghasilkan hasil sedikit lebih tinggi. Hasil ini juga konsisten dengan penelitian-penelitian sebelumnya, bahwa performa analisis sentimen berbasis teks tidak hanya ditentukan oleh konfigurasi fitur, akan tetapi juga dipengaruhi oleh algoritma klasfikasi yang digunakan

4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa konfigurasi N-Gram berpengaruh signifikan terhadap performa klasifikasi sentimen ulasan pengguna aplikasi Neobank. Model Naïve Bayes menunjukkan performa terbaik ketika menggunakan konfigurasi unigram, dengan akurasi mencapai 87,65%, sedangkan konfigurasi bigram menurunkan performa secara signifikan menjadi 79,23%. Sebaliknya, model Support Vector Machine kernel linear memperoleh hasil tertinggi dengan kombinasi unigram+bigram, meskipun peningkatan akurasinya hanya sebanyak 0,02% dibanding pada unigram, menunjukkan bahwa SVM mampu memanfaatkan konteks tambahan dari bigram untuk memisahkan kelas sentimen dengan lebih baik. Analisis *confusion matrix* menunjukkan bahwa kedua model dapat membedakan kelas positif dan negatif dengan seimbang, meskipun jumlah data yang lebih banyak pada kelas positif tetap memengaruhi frekuensi prediksi yang benar. Temuan ini mendukung hipotesis bahwa konfigurasi *N-Gram* mempengaruhi performa klasifikasi, serta pentingnya menyesuaikan pemilihan fitur teks dengan karakteristik algoritma. Terdapat keterbatasan dalam penelitian ini, seperti hanya menggunakan dua algoritma klasifikasi dan tiga konfigurasi *N-Gram*, serta data yang hanya bersumber dari satu platform (*Google Play Store*), sehingga hasil belum tentu dapat digeneralisasi pada konteks yang lebih luas. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi konfigurasi N-Gram yang lebih tinggi seperti trigram, menerapkan teknik penyeimbangan data untuk menangani ketidakseimbangan kelas, serta menggunakan pendekatan berbasis deep learning atau model bahasa berbasis transformer seperti IndoBERT yang dirancang khusus untuk pemrosesan bahasa Indonesia guna memperoleh akurasi klasifikasi yang lebih tinggi dan representasi makna yang lebih kaya.

REFERENCES

Adyatma, A. D., Afuan, L., & Maryanto, E. (2023). The Effect Of Unigram And Bigram In The Naïve Bayes Multinomial For Analyzing Of Comment Sentiment Of Gojek Application In Google Play Store. *Jurnal Teknik Informatika (JUTIF)*, *4*(6), 1535–1540. https://doi.org/10.52436/1.jutif.2023.4.6.1310

Arifin, M. N., Hamzah, A., Huda, M. A., & Hasanah, N. (2025). Analysis of Google Play Store User Sentiment Towards Application X Using the SVM Algorithm. *Brilliance*, *5*(1), 249–258. https://doi.org/10.47709/brilliance.v5i1.6024

Bimantara, M. D., & Zufria, I. (2024). Text Mining Sentiment Analysis On Mobile Banking Application Reviews Using TF-IDF Method With Natural Language Processing Approach. *JINAV: Journal of Information and Visualization*, *5*(1), 115–123. https://doi.org/https://doi.org/10.35877/454RI.jinav2772

Enjelia, L., Cahyana, Y., & Wahiddin, D. (2025). Comparison of K-Nearest Neighbors and Naive Bayes Classifier Algorithms in Sentiment Analysis of 2024 Election in Twitter (X). *Journal of Applied Informatics and Computing (JAIC)*, *9*(3), 946–954. https://doi.org/10.30871/jaic.v9i3.9593

Hadi, K., & Utami, E. (2024). Analysis of K-NN with the Integration of Bag of Words , TF-IDF , and N-Grams for Hate Speech Classification on Twitter. *JUITA: Jurnal Informatika*, *12*(2), 289–298. https://doi.org/10.30595/juita.v12i2.23829

Helmi, A. Y., & Kristianto, A. H. (2024). Sistem RGEC Dalam Analisis Tingkat Kesehatan Bank Digital Yang Terdaftar Di BEI Periode 2019-2022. *MARGIN ECO: Jurnal Ekonomi Dan Perkembangan Bisnis*, *8*(1), 75–98. https://doi.org/10.32764/margin.v8i1.4511

Kadek, N., Puspita, F., Sudipa, I. G. I., Sunarya, I. W., Wayan, N., & Kusuma, J. (2025). Sentiment Analysis of Roblox Game Reviews on Google Play Store Using Lexicon-SVM Integration. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, *9*(4), 1863–1876. https://doi.org/10.33395/sinkron.v9i4.15272

Kusnawi, Rahardi, M., & Pandiangan, V. D. (2023). Sentiment Analysis of Neobank Digital Banking Using Support Vector Machine Algorithm in Indonesia. *JOIV*, *7*(June), 377–383. https://doi.org/10.30630/joiv.7.2.1652

Manjula, S., Rajini, N. H., & Chokkanathan, K. (2025). Enhanced chronic kidney disease detection using XGBoost with

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2375-2383
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9702

improved brainstorm optimization for hyperparameter tuning. *Discover Applied Sciences*, *7*, 1181. https://doi.org/10.1007/s42452-025-07633-7

Nurhayati, Tanti, L., & Triandi, B. (2026). Optimasi Support Vector Machine Menggunakan Particle Swarm Optimization pada Analisis Sentimen Program Efisiensi Anggaran Pemerintah. *Jurnal Minfo Polgan (JMP)*, *15*(1), 130–144. https://doi.org/https://doi.org/10.33395/jmp.v15i1.15929

Özyirmidokuz, E. K., & Elmas, B. M. (2025). AI-Based Sentiment Analysis of E-Commerce Customer Feedback : A Bilingual Parallel Study on the Fast Food Industry in Turkish and English. *Journal of Theoritical and Applied Electronic Commerce Research*, *20*, 294. https://doi.org/10.3390/jtaer20040294

Prayudya, D. R., Ikhwan, I., Nugroho, T., & Ramdiania, R. G. N. (2025). Comparing Neo and Traditional Banking Efficiency : A Three-Stage DEA Analysis in Indonesia. *Jurnal Ekonomi Malaysia*, *59*(December 2024). https://doi.org/http://dx.doi.org/10.17576/JEM-2025-5901-5 Comparing

Putra, K. T., Hariyadi, M. A., & Crysdian, C. (2023). Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis SVM. *Jurnal Cahaya Mandalika*, *3*(2), 1449–1463. https://www.ojs.cahayamandalika.com/index.php/jcm/article/view/2292

Putri, D. W., & Soeleman, M. A. (2026). Penerapan Algoritma Naïve Bayes Terhadap Sentimen Ulasan Produk Skincare Pada E-Commerce Shopee. *Building of Informatics, Technology and Science (BITS)*, *7*(4), 2218–2228. https://doi.org/10.47065/bits.v7i4.9209

Raharjo, R. A., Sunarya, I. M. G., & Divayana, D. G. H. (2022). Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin. *Jurnal Ilmiah Elektronika Dan Komputer*, *15*(2), 456–464. https://doi.org/10.51903/elkom.v15i2.918

Rahmaliyadi, V., & Maridjan, M. M. (2025). Sentiment Analysis of Indonesian-Language Plantix Application Reviews for Plant Disease Diagnosis Using Naive Bayes Methods. *Journal of Intelligent Systems Technology and Informatics*, *1*(2), 62–66. https://doi.org/10.64878/jistics.v1i2.12

Rahmatulloh, F., Sumarwan, U., Hartoyo, & Sartono, B. (2024). Unveiling Factors Influencing Neobanking Adoption With An Extended UTAUT-3 Model To Improve Neobank Marketing Strategy. *International Journal Of Economics And Finance Studies*, *16*(03), 203–228. https://doi.org/10.34109/ijefs.202416310

Setiawan, A., & Hasan, F. N. (2025). Analisis Sentimen Tanggapan Pengguna Aplikasi Bale By Btn Menggunakan Metode Support Vector Machine (SVM). *STORAGE – Jurnal Ilmiah Teknik Dan Ilmu Komputer*, *4*(4), 315–326. https://doi.org/10.55123/storage.v4i4.6469

Ulgasesa, R., Negara, A. B. P., & Tursina. (2022). Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal. *JUSTIN : Jurnal Sistem Dan Teknologi Informasi*, *10*(3), 286–293. https://doi.org/10.26418/justin.v10i3.53880

Zafira, Z. T., Tania, K. D., & Sari, W. K. (2025). Sentiment-Based Knowledge Discovery of Wondr by BNI App Reviews Using SVM , KNN , and Naive Bayes for CRM Enhancement. *Journal of Applied Informatics and Computing (JAIC)*, *9*(5), 2498–2508. https://doi.org/10.30871/jaic.v9i5.10323