

PERBANDINGAN SENTIMEN PENGGUNA SHOPEE DAN TOKOPEDIA MENGGUNAKAN NAÏVE BAYES

Grace Apriliani Puteri Santoso ¹⁾ Dedi Trisnawarman ²⁾ Agus Budiyantra ³⁾

¹⁾Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

²⁾Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

³⁾Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

email : grace.825220161@stu.untar.ac.id¹, dedit@fti.untar.ac.id², agusb@fti.untar.ac.id³

ABSTRAK

Pertumbuhan pesat ekosistem *e-commerce* di Indonesia menjadikan ulasan pengguna pada Google Play Store salah satu sumber data penting untuk membaca persepsi dan pengalaman pengguna secara langsung. Penelitian ini bertujuan menganalisis serta membandingkan sentimen ulasan pengguna terhadap aplikasi Shopee dan Tokopedia dengan menggunakan pendekatan klasifikasi Naïve Bayes. Data ulasan dikumpulkan dari halaman Google Play kedua aplikasi, kemudian diproses melalui tahapan prapemrosesan teks, meliputi pembersihan karakter, case folding, penghapusan stopword, dan stemming. Fitur teks direpresentasikan menggunakan Term Frequency Inverse Document Frequency (TF-IDF) dan diklasifikasikan ke dalam label sentimen positif, negatif, dan netral. Kinerja model dinilai menggunakan metrik akurasi, presisi, recall, dan F1-score, sedangkan perbedaan pola sentimen antara Shopee dan Tokopedia dianalisis berdasarkan distribusi kelas dan hasil klasifikasi yang dihasilkan. Temuan penelitian ini memberikan gambaran kuantitatif mengenai kecenderungan penilaian pengguna terhadap kedua platform dan dapat menjadi masukan bagi pengelola *e-commerce* dalam merumuskan strategi peningkatan kualitas layanan berbasis opini pengguna.

Key words

analisis sentimen, Naive Bayes, shopee, tokopedia

1. Pendahuluan

Literatur analisis sentimen menunjukkan bahwa opini dalam teks dapat dimodelkan sebagai polaritas yang terukur sehingga mendukung perancangan eksperimen klasifikasi yang sistematis, mulai dari prapemrosesan hingga evaluasi kinerja model [1]. Di sisi lain, pemanfaatan internet dan perangkat bergerak telah mendorong pergeseran perilaku belanja masyarakat ke

platform *e-commerce* [2]. Di Indonesia, Shopee dan Tokopedia menjadi dua *marketplace* besar sehingga ulasan dan penilaian pengguna di Google Play Store penting untuk menilai kualitas aplikasi mobile keduanya [3].

Jumlah ulasan yang sangat besar membuat pembacaan manual tidak efisien. *Text mining* ditujukan untuk mengekstraksi informasi dan pola yang relevan dari dokumen teks [1], sedangkan analisis sentimen secara khusus mengelompokkan opini pengguna ke dalam kategori positif, negatif, atau netral sehingga data kualitatif dapat diubah menjadi indikator kuantitatif yang berguna bagi pengambil keputusan [5]. Pada konteks ulasan aplikasi di Google Play, pendekatan ini membantu pengembang memahami persepsi pengguna dan mengidentifikasi aspek yang perlu ditingkatkan [3].

Dalam tugas klasifikasi sentimen berbasis teks, Naïve Bayes banyak digunakan sebagai baseline karena asumsi independensi fiturnya sederhana, komputasinya ringan, dan performanya cukup stabil pada korpus besar. Ramadhan et al. menunjukkan bahwa Naïve Bayes yang dikombinasikan dengan *feature extraction* Term Frequency-Inverse Document Frequency (TF-IDF) mampu mencapai akurasi di atas 80% untuk ulasan aplikasi *e-commerce* [4]. Implementasi serupa pada ulasan Shopee menghasilkan kinerja klasifikasi yang konsisten [5], dan pada ulasan produk kesehatan di Tokopedia juga terbukti efektif memisahkan ulasan positif dan negatif [6]. Hasibuan dan Heriyanto menemukan hasil sejenis pada ulasan aplikasi Amazon Shopping di Google Play Store [7]. Temuan-temuan ini menegaskan bahwa Naïve Bayes-TF-IDF layak digunakan sebagai model dasar analisis sentimen ulasan aplikasi *mobile*.

Walaupun Shopee dan Tokopedia sudah menjadi objek berbagai kajian, sebagian besar penelitian fokus pada satu platform atau menggabungkan banyak aplikasi tanpa membahas secara spesifik perbedaan pola sentimen

antara kedua marketplace tersebut dalam satu kerangka pemodelan yang seragam [2]. Penelitian ini dimaksudkan untuk mengisi celah tersebut dengan melakukan analisis komparatif sentimen ulasan pengguna Shopee dan Tokopedia pada Google Play Store menggunakan klasifikasi Naïve Bayes, melalui pembangunan korpus ulasan, penerapan prapemrosesan teks dan TF-IDF, serta evaluasi dan perbandingan distribusi sentimen dan kinerja model berdasarkan akurasi, presisi, recall, dan F1-score.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen text mining untuk melakukan analisis komparatif sentimen pada dua korpus ulasan, yaitu Shopee dan Tokopedia. Seluruh tahapan mengikuti alur umum analisis sentimen berbasis *machine learning*, mulai dari pengumpulan data, *preprocessing* teks, ekstraksi fitur, pelatihan model klasifikasi Naïve Bayes, hingga evaluasi kinerja dan analisis perbandingan hasil.

2.1. Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini merupakan data primer berupa teks ulasan pengguna aplikasi Shopee dan Tokopedia pada Google Play Store. Data dikumpulkan dengan teknik *web scraping* mencakup 1.250 ulasan pengguna Shopee dan 1.250 ulasan Tokopedia.

2.2. Normalization

Pada tahap *normalization*, ulasan mentah hasil *web scraping* disusun kembali menjadi data yang siap dianalisis. Teks apa adanya diubah menjadi korpus yang formatnya seragam dan sudah memiliki kelas sentimen. Dalam penelitian ini, *normalization* terdiri dari dua kegiatan utama, yaitu *preprocessing* teks dan pelabelan data, mengikuti alur umum analisis sentimen pada ulasan aplikasi di Google Play Store

2.2.1 Preprocessing

Pada tahap ini, dataset menjalani proses pembersihan untuk mempersiapkan data agar dapat digunakan dalam pemodelan. Langkah yang dilakukan meliputi *Cleansing* dengan menghapus emoji, tautan, tag akun, angka yang tidak relevan, dan simbol lain yang tidak terkait isi ulasan. *Case Folding* yang mengubah seluruh huruf menjadi huruf kecil, serta *Tokenization* untuk memecah teks menjadi unit kata. Setelah itu dilakukan normalisasi guna menyeragamkan kata tidak baku dan singkatan, sekaligus merapikan pengulangan huruf. Langkah berikutnya adalah *Stopword removal* untuk menghilangkan kata fungsi yang frekuensinya tinggi tetapi tidak berkontribusi pada perbedaan sentimen. Tahap terakhir adalah *Stemming*, yang

mengembalikan kata berimbuhan ke bentuk dasarnya. Rangkaian langkah ini menghasilkan korpus yang bersih dan konsisten sehingga dapat direpresentasikan dengan TF-IDF dan diolah secara efektif oleh model Naïve Bayes [8]

2.2.2 Labeling

Sesudah *preprocessing*, setiap ulasan diberi kelas sentimen sebagai *ground truth* bagi model. Penelitian ini menggunakan tiga kelas, yaitu positif, negatif, dan netral. Ulasan yang didominasi pujian dan rasa puas dikategorikan positif, ulasan yang berisi keluhan atau kekecewaan yang jelas dikategorikan negatif, sedangkan ulasan yang cenderung deskriptif atau campuran tanpa kecenderungan emosi yang kuat dikategorikan netral [9]. Proses anotasi dilakukan oleh dua penelaah, masing-masing memberi label secara mandiri, lalu menyamakan hasil jika terjadi perbedaan

2.3. Modeling

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan sebagai tahap pemodelan awal untuk mengubah ulasan yang semula berbentuk teks bebas menjadi representasi numerik. Secara sederhana, TF-IDF memberi bobot pada setiap kata berdasarkan dua hal: seberapa sering kata tersebut muncul dalam satu ulasan (*term frequency*) dan seberapa jarang kata tersebut muncul di seluruh kumpulan ulasan (*inverse document frequency*). Kata yang sangat umum, seperti “aplikasi” atau “bagus”, akan mendapat bobot lebih rendah, sedangkan kata yang lebih spesifik, misalnya “lemot”, “error”, atau “responsif”, akan diberi bobot lebih tinggi karena dianggap lebih informatif untuk membedakan sentimen [10].

2.4. Confusion Matrix

Evaluasi kinerja model dilakukan menggunakan *confusion matrix* tiga kelas (positif, negatif, netral). *Confusion matrix* menyajikan jumlah prediksi benar dan salah untuk tiap kelas dalam bentuk tabel, sehingga tidak hanya menunjukkan seberapa sering model memberikan prediksi yang tepat, tetapi juga pola kesalahan yang muncul, misalnya ulasan negatif yang sering terbaca sebagai netral. Dari tabel ini kemudian dihitung beberapa ukuran kinerja utama, yaitu akurasi, presisi, *recall*, dan *F1score* untuk setiap kelas sentimen.

Akurasi menyatakan proporsi seluruh prediksi yang benar terhadap jumlah sampel uji. Presisi menunjukkan seberapa besar bagian dari prediksi suatu kelas yang benar, sedangkan *recall* menunjukkan seberapa besar bagian data sebenarnya pada kelas tersebut yang berhasil dikenali model. *F1score* menggabungkan presisi dan *recall* dalam satu nilai rata-rata harmonik sehingga lebih stabil ketika distribusi kelas tidak seimbang [11]. Secara matematis, keempat metrik tersebut dapat dihitung menggunakan rumus berikut:

$$Akurasi = \frac{\text{Jumlah prediksi benar}}{\text{Total prediksi}} \quad (1)$$

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

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 \text{ Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

dengan TP (*true positive*), TN (*true negative*), FP (*false positive*), dan FN (*false negative*) diperoleh dari nilai pada *confusion matrix*.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari ulasan pengguna aplikasi Shopee dan Tokopedia pada Google Play Store. Pengumpulan data dilakukan melalui proses *web scraping* menggunakan pustaka Python. Data yang diambil meliputi username pengguna nama aplikasi, *review*, *review id*, tanggal serta penilaian sentimen awal yang diperoleh berdasarkan *rating* bintang pengguna. Total data yang berhasil dikumpulkan berjumlah 2.500 ulasan, terdiri atas 1.250 ulasan untuk aplikasi Shopee dan 1.250 ulasan untuk aplikasi Tokopedia. Masing-masing ulasan kemudian dianalisis untuk menentukan kecenderungan sentimen pengguna terhadap aplikasi tersebut.

Hasil analisis awal menunjukkan bahwa distribusi sentimen cenderung didominasi oleh kelas positif, diikuti oleh netral, dan negatif. Distribusi ini memberikan indikasi bahwa pengguna umumnya memiliki persepsi yang cukup baik terhadap kedua aplikasi, namun tetap terdapat bagian ulasan yang mengandung kritik terhadap performa atau layanan. Proses pengumpulan data ini menjadi tahap awal untuk memahami pola opini pengguna dan menjadi dasar bagi proses *preprocessing* serta analisis sentimen lebih lanjut.

3.2 Normalization

Tahap *normalization* mencakup dua bagian, yaitu *preprocessing* teks dan *labeling* data sentimen. Hasil dari *preprocessing* ditampilkan pada Tabel 1 dan tahap *labeling* pada Tabel 2.

Tabel 1. Hasil *Preprocessing*

Proses	Teks
Ulasan Asli	harga sii ok.. ongkir jg lebih murah
Case Folding	harga sii ok ongkir jg lebih murah
Stopword Removal	[harga, sii, ok, ongkir, jg, lebih, murah]
Tokenizing	[harga, sii, ongkir, jg, lebih, murah]

Stemming	harga sii ongkir jg lebih murah
----------	---------------------------------

Tabel 1 menampilkan hasil tahapan *preprocessing* pada data ulasan pengguna. Proses ini mengubah teks asli menjadi bentuk yang lebih sederhana dan seragam untuk analisis selanjutnya. Tahap *case folding* menstandarkan huruf menjadi huruf kecil, diikuti dengan *stopword removal* yang menghapus kata-kata umum tanpa makna penting. Selanjutnya, *tokenizing* memecah kalimat menjadi unit kata individual, dan stemming mengembalikan setiap kata ke bentuk dasarnya. Hasil akhirnya berupa teks yang bersih, efisien, serta siap digunakan dalam proses representasi fitur dan klasifikasi sentimen.

Tabel 2. Hasil *Labeling*

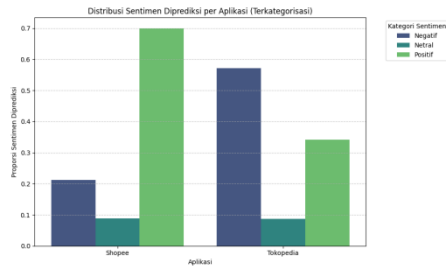
Ulasan	Sentimen
Pilihannya banyak ,bnyak gratis ongkir	Positif
Perlu pengoptimalan Untuk aplikasinya	Netral
Pengirim lambat. nggk jelas.	Negatif

Tabel 2 memperlihatkan hasil proses *labeling* yang mengelompokkan setiap ulasan pengguna ke dalam kategori positif, netral, atau negatif. Proses ini dilakukan dengan meninjau isi teks berdasarkan ekspresi emosi dan konteks kalimat. Ulasan seperti “Pilihannya banyak, banyak gratis ongkir” diklasifikasikan sebagai positif karena mengandung kepuasan terhadap layanan. Kalimat “Perlu pengoptimalan untuk aplikasinya” diberi label netral karena bersifat saran tanpa ekspresi emosi yang kuat, sedangkan “Pengirim lambat, nggk jelas” termasuk negatif karena menunjukkan ketidakpuasan pengguna. Hasil *labeling* ini menjadi dasar utama dalam proses pelatihan model klasifikasi sentimen.

3.3 Modeling

Tahap modeling dilakukan dengan menerapkan kombinasi TF-IDF dan *Multinomial Naïve Bayes* untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee dan Tokopedia yang diperoleh dari Google Play Store. Data dibagi menjadi 80% data latih dan 20% data uji. Model dievaluasi menggunakan *confusion matrix* serta metrik akurasi, presisi, *recall*, dan *F1 score*.

3.3.1 Distribusi Sentimen



Gambar 1. Prediksi distribusi sentimen

Berdasarkan hasil prediksi pada Gambar 4, ulasan pengguna Shopee didominasi oleh sentimen positif sebesar 70%, diikuti oleh negatif sekitar 21%, dan netral di bawah 10%. Hal ini menunjukkan bahwa mayoritas pengguna Shopee memberikan tanggapan yang baik terhadap layanan, terutama pada aspek harga dan promosi. Sementara itu, ulasan pengguna Tokopedia menunjukkan proporsi negatif yang lebih tinggi, yakni sekitar 57%, sedangkan positif sebesar 34% dan netral di bawah 10%. Hasil ini mengindikasikan adanya persepsi yang lebih kritis terhadap performa aplikasi dan proses pengiriman pada Tokopedia dibanding Shopee.

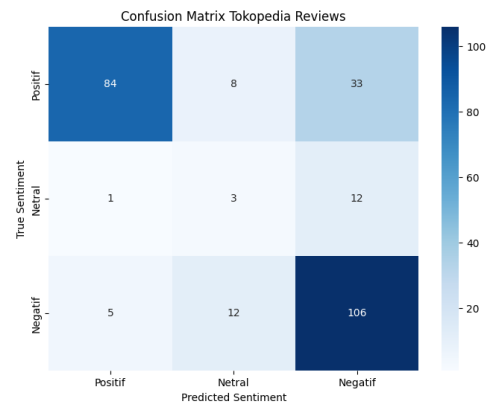
3.3.2 Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix* untuk mengukur performa klasifikasi pada masing-masing aplikasi. Gambar 2 memperlihatkan confusion matrix untuk Shopee, sedangkan Gambar 3 untuk Tokopedia.



Gambar 2. Confusion matrix Shopee

Berdasarkan Gambar 2. pada Shopee, model menghasilkan 160 prediksi benar pada kelas positif dengan kesalahan klasifikasi yang rendah, khususnya antar kelas positif dan negatif. Nilai akurasi yang diperoleh sebesar 0,88, dengan presisi 0,89, *recall* 0,86, dan *F1score* 0,87. Model mampu mengenali pola kata yang menggambarkan kepuasan pengguna dengan konsisten.



Gambar 3. Confusion matrix Tokopedia

Berdasarkan Gambar 3 untuk Tokopedia, model mencapai 106 prediksi benar pada kelas negatif. Akurasi yang dihasilkan sebesar 0,84, dengan presisi 0,85, *recall* 0,83, dan *F1score* 0,84. Kesalahan utama terjadi antara kelas positif dan negatif akibat kemiripan pola bahasa yang digunakan dalam ulasan campuran, misalnya antara pujian dan keluhan dalam satu kalimat.

Tabel 3. Perbandingan Hasil Pengujian

Aplikasi	Akurasi	Presesi	Recall	F1-Score
Shopee	0,88	0,89	0,86	0,87
Tokopedia	0,84	0,85	0,83	0,84

Tabel 3 menunjukkan bahwa model TF-IDF dan Naïve Bayes bekerja lebih baik pada *dataset* Shopee dibanding Tokopedia. Hal ini disebabkan oleh distribusi data Shopee yang lebih dominan pada kelas positif dan penggunaan bahasa yang lebih homogen, sehingga memudahkan model dalam membedakan polaritas sentimen. Sebaliknya, variasi ekspresi dan ambiguitas pada ulasan Tokopedia menyebabkan tingkat kesalahan prediksi lebih tinggi.

Secara keseluruhan, kedua model menunjukkan performa yang stabil dengan akurasi di atas 80%, membuktikan bahwa pendekatan TF-IDF dan *Multinomial* Naïve Bayes efektif untuk analisis sentimen teks berbahasa Indonesia, khususnya pada ulasan aplikasi *e-commerce*.

4. Kesimpulan

Penelitian ini bertujuan untuk membandingkan sentimen pengguna terhadap aplikasi Shopee dan Tokopedia berdasarkan ulasan pada Google Play Store menggunakan kombinasi metode TF-IDF dan algoritma *Multinomial* Naïve Bayes.

Hasil penelitian menunjukkan bahwa model mampu bekerja dengan baik pada kedua dataset. Model untuk Shopee mencapai akurasi 0,88 dan *F1-score* 0,87, sementara Tokopedia memperoleh akurasi 0,84 dan *F1-score* 0,84. Perbedaan ini menunjukkan bahwa pola bahasa dalam ulasan Shopee lebih mudah dikenali oleh model karena lebih homogen dan didominasi oleh sentimen positif. Sebaliknya, ulasan Tokopedia memiliki

variasi ekspresi dan konteks yang lebih beragam, sehingga tingkat kesalahannya sedikit lebih tinggi.

Dari distribusi sentimen yang dihasilkan, mayoritas ulasan Shopee bersifat positif, menggambarkan kepuasan pengguna terhadap aspek harga, promosi, dan kemudahan transaksi. Sementara itu, Tokopedia cenderung didominasi oleh ulasan negatif, yang banyak menyoroti masalah teknis seperti keterlambatan pengiriman dan performa aplikasi.

Secara keseluruhan, pendekatan TF-IDF dan Naïve Bayes terbukti efektif dalam menganalisis teks ulasan berbahasa Indonesia. Hasil penelitian ini dapat menjadi dasar bagi pengembang aplikasi e-commerce untuk memahami persepsi pengguna secara lebih mendalam dan meningkatkan kualitas layanan sesuai kebutuhan mereka.

Sebagai langkah lanjutan, penelitian ini dapat dikembangkan dengan menerapkan algoritma yang lebih kompleks seperti *Support Vector Machine* (SVM), Random Forest, atau model berbasis *deep learning* (misalnya LSTM atau BERT). Selain itu, analisis sentimen berbasis aspek (*aspect-based sentiment analysis*) juga dapat dilakukan untuk mengidentifikasi bagian spesifik dari layanan yang paling memengaruhi tingkat kepuasan pengguna.

REFERENSI

- [1] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, 107134, 2021.
- [2] D. Ulhaq, "Comparative analysis of user experience in Shopee and Tokopedia e-commerce applications using the User Experience Questionnaire (UEQ) method," *Adopsi Teknologi dan Sistem Informasi (Atasi)*, vol. 3, no. 1, pp. 37–44, 2024.
- [3] S. Fransiska and A. I. Gufroni, "Sentiment analysis Provider by.U on Google Play Store reviews with TF-IDF and Support Vector Machine (SVM) method," *Scientific Journal of Informatics*, vol. 7, no. 2, 2020.
- [4] B.Z. Ramadhan, R. I. Adam, and I. Maulana, "Analisis sentimen ulasan pada aplikasi e-commerce dengan menggunakan algoritma Naive Bayes," *Journal of Applied Informatics and Computing*, vol. 6, no. 2, pp. 220–225, 2022.
- [5] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi algoritma Naive Bayes untuk analisis sentimen ulasan Shopee pada Google Play Store," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, 2022.
- [6] A. Ernawati, A. O. Sari, S. N. Sofyan, M. Iqbal, and R. F. W. Wijaya, "Implementasi algoritma Naïve Bayes dalam menganalisis sentimen review pengguna Tokopedia pada produk kesehatan," *Bulletin of Information Technology*, vol. 4, no. 4, pp. 533–543, 2023.
- [7] E. Hasibuan and E. A. Heriyanto, "Analisis sentimen pada ulasan aplikasi Amazon Shopping di Google Play Store menggunakan Naive Bayes classifier," *Jurnal Teknik dan Science*, vol. 1, no. 3, pp. 13–24, 2022.
- [8] A.F. Haikal and Y. Nuryaman, "Sentiment Analysis of Transjakarta App Reviews Using the Naive Bayes Algorithm," *Journal of Artificial Intelligence and Engineering Applications*, vol. 5, no. 1, pp. 1620–1626, 2025. doi:10.59934/jaiea.v5i1.1677.
- [9] D. S. Nurrochmah, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, "Penerapan Algoritma Naive Bayes dalam Analisis Sentimen Ulasan Aplikasi KitaLulus di Google Play Store," *Jurnal Informatika Terpadu*, vol. 11, no. 1, pp. 1–11, 2025. doi:10.54914/jit.v11i1.1544.
- [10] Helmayanti, S. A., Hamami, F., & Fa'rifah, R. Y. (2023). Penerapan algoritma TF-IDF dan Naïve Bayes untuk analisis sentimen berbasis aspek ulasan aplikasi Flip pada Google Play Store. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 4(3), 1822–1834. <https://doi.org/10.35870/jimik.v4i3.415>
- [11] S. Zhang, X. Li, and M. Zong, "Efficient text classification using improved evaluation metrics based on confusion matrix," *IEEE Access*, vol. 9, pp. 123456–123468, 2021. doi:10.1109/ACCESS.2021.3056789.

Grace Apriliani Puteri Santoso, mahasiswa program studi Sistem Informasi Universitas Tarumanagara.