



DOI: <https://doi.org/10.38035/jimt.v7i3>
<https://creativecommons.org/licenses/by/4.0/>

Evaluasi Kualitas Layanan OTA dalam Perspektif Manajemen Operasi Jasa: Pendekatan *Text Mining* dan *Sentiment Analysis* pada Aplikasi Traveloka

Vanya Pinkan Maridelana¹, Setya Widyawan Prakosa², Tria Putri Noviasari³

¹Universitas Jember, Indonesia, vanyapinkan@unej.ac.id

²Universitas Jember, Indonesia, setyawp@unej.ac.id

³Universitas Jember, Indonesia, triaputri.feb@unej.ac.id

Corresponding Author: vanyapinkan@unej.ac.id¹

Abstract: *This study aims to analyze the service quality of the Traveloka mobile application based on user reviews on Google Play Store using a text mining and IndoBERT-based sentiment analysis approach. The dataset consists of 121,498 Indonesian-language reviews collected in 2025. The analysis includes preprocessing, sentiment classification, and mapping reviews into four E-S-QUAL dimensions: efficiency, fulfillment, system availability, and privacy. The results indicate that 48.8% of the reviews are positive and 23.7% are negative, with a 2:1 ratio. The IndoBERT model demonstrates strong performance, achieving 89.2% accuracy and an F1-score of 89.3%. The service quality score shows that efficiency obtains a positive score (+0.39), while fulfillment (−0.26), system availability (−0.44), and privacy (−0.71) receive negative scores. These findings suggest that although the application is perceived as easy to use, operational challenges remain in service fulfillment, system reliability, and data security. From a practical perspective, this study provides actionable insights for Traveloka by identifying priority areas for operational improvement based on real-time user-generated feedback. Specifically, the findings highlight the need to strengthen system integration with service partners, enhance technological infrastructure reliability, and improve data protection and transaction security mechanisms. The proposed approach can also serve as a continuous digital service quality monitoring tool to support strategic decision-making, customer satisfaction enhancement, and long-term user loyalty development.*

Keyword: *text mining, sentiment analysis, IndoBERT, digital service quality, E-S-QUAL, service operations management*

Abstrak: Penelitian ini bertujuan untuk menganalisis kualitas layanan aplikasi Traveloka berdasarkan ulasan pengguna di Google Play Store menggunakan pendekatan text mining dan sentiment analysis berbasis IndoBERT. Data yang digunakan terdiri dari 121.498 ulasan berbahasa Indonesia di sepanjang tahun 2025. Analisis dilakukan melalui tahapan preprocessing, klasifikasi sentimen, serta pemetaan ulasan ke dalam empat dimensi E-S-QUAL yaitu efficiency, fulfillment, system availability, dan privacy. Hasil analisis menunjukkan bahwa 48,8% ulasan bersentimen positif dan 23,7% bersentimen negatif dengan rasio 2:1.

Model IndoBERT menunjukkan performa yang baik dengan akurasi 89,2% dan skor F1 89,3%. Pengukuran skor kualitas layanan menunjukkan bahwa dimensi *efficiency* memperoleh skor positif (+0,39), sementara *fulfillment* (-0,26), *system availability* (-0,44), dan *privacy* (-0,71) menunjukkan skor negatif. Temuan ini mengindikasikan bahwa meskipun aplikasi dinilai mudah digunakan, masih terdapat permasalahan pada proses pemenuhan layanan, stabilitas sistem, dan keamanan data. Secara praktis, penelitian ini memberikan kontribusi bagi Traveloka dalam mengidentifikasi area prioritas perbaikan operasional berbasis data ulasan pengguna secara real-time, khususnya pada penguatan integrasi sistem dengan mitra layanan, peningkatan keandalan infrastruktur teknologi, serta optimalisasi perlindungan data dan keamanan transaksi. Pendekatan ini juga dapat dimanfaatkan sebagai mekanisme monitoring kualitas layanan digital secara berkelanjutan guna mendukung strategi peningkatan kepuasan dan loyalitas pengguna.

Kata Kunci: *text mining*, analisis sentimen, IndoBERT, kualitas layanan digital, E-S-QUAL, manajemen operasi jasa

PENDAHULUAN

Perkembangan teknologi digital yang pesat telah mendorong perubahan fundamental dalam cara konsumen mengakses dan memanfaatkan layanan jasa berbasis aplikasi. Fenomena ini juga terwujud dalam meningkatnya pemanfaatan teknologi aplikasi *mobile* sebagai sarana utama dalam penyampaian layanan jasa kepada konsumen, termasuk pada layanan jasa *travel agent*. Aktivitas seperti pemesanan tiket pesawat, hotel, transportasi, hingga atraksi wisata kini banyak dilakukan melalui aplikasi *mobile* yang menawarkan kemudahan dan kecepatan dalam proses transaksi. Transformasi ini tidak hanya mengubah perilaku konsumen, tetapi juga memengaruhi bagaimana perusahaan mengelola operasi jasanya. Pada tahun 2025, *Google Play Store* tercatat memiliki lebih dari 2 juta aplikasi aktif (*Google Play Store*, 2025), mencerminkan besarnya ekosistem aplikasi yang tersedia dan tingginya intensitas kompetisi dalam layanan digital. Pada sisi lain, tren penggunaan aplikasi *traveling* meningkat signifikan, di mana sekitar 93 juta pengguna aplikasi *traveling* secara global tercatat pada tahun 2021. Dalam konteks Indonesia dan Asia Tenggara, Traveloka adalah salah satu *online travel agent* (OTA) berbasis *website* dan aplikasi *mobile* yang sangat dominan. Berdasarkan data Traveloka (2025), aplikasi Traveloka telah diunduh lebih dari 140 juta kali secara global dan memiliki lebih dari 40 juta pengguna aktif setiap bulan. Di tengah intensitas kompetisi yang tinggi dan ekspektasi pelanggan yang semakin meningkat, perusahaan penyedia platform digital, seperti Traveloka, tidak hanya dituntut menghadirkan kemudahan akses, tetapi juga memastikan keandalan sistem, kecepatan respons, serta konsistensi proses layanan secara menyeluruh. Kondisi ini menjadikan evaluasi kinerja operasional pada platform digital, khususnya aplikasi *mobile*, sebagai isu penting yang perlu dianalisis secara sistematis dalam kerangka manajemen operasi jasa.

Pada prinsipnya, manajemen operasi jasa berfokus pada bagaimana organisasi merancang proses, mengelola kapabilitas layanan, dan menjaga kualitas layanan secara konsisten untuk memenuhi kebutuhan pelanggan (Johnston et al., 2021). Dalam konteks layanan berbasis digital seperti Traveloka, kualitas layanan aplikasi *mobile* (*mobile app service quality*) merupakan aspek kritis dalam manajemen operasi jasa modern karena aplikasi berfungsi sebagai *service delivery interface* utama antara perusahaan dan pelanggan. Hal ini didukung oleh penelitian Arcand et al. (2017) dan Shankar & Jebarajakirthy (2019) yang menyatakan bahwa platform digital, yang operasionalnya sangat ditentukan oleh pengalaman pengguna, perlu menempatkan kualitas layanan (*service quality*) sebagai faktor utama dalam membangun kepuasan serta loyalitas pelanggan. Dalam kerangka manajemen operasi jasa, kualitas layanan digital merefleksikan efektivitas integrasi antara *front-end digital interface* dan *back-end operational systems* (Chuang et al., 2015; Vukovic & Natarajan, 2012; Joshi et al., 2009), seperti keandalan server, integrasi data dengan maskapai/hotel, proses pembayaran, kecepatan respons layanan

pelanggan, dan mekanisme penanganan keluhan. Adanya gangguan kecil pada aplikasi merupakan indikasi masalah operasional dan secara langsung memengaruhi persepsi kualitas layanan oleh pengguna.

Mengukur kualitas layanan digital seperti pada aplikasi *mobile* Traveloka tidaklah mudah. Secara historis, kualitas layanan diukur menggunakan SERVQUAL yang menilai lima dimensi utama yaitu *tangibility*, *reliability*, *responsiveness*, *assurance*, dan *empathy* (Parasuraman et al., 1988). Namun, berbagai penelitian menegaskan bahwa model ini tidak cukup menggambarkan kualitas layanan digital, karena karakteristik operasional aplikasi *mobile* lebih menekankan *self-service operations*, automasi, dan kinerja teknologi (Carlson & O'Cass, 2011; Tate & Evermann, 2010; Parasuraman & Grewal, 2000). Sebagai respons terhadap kebutuhan pengukuran kualitas layanan digital, Parasuraman et al. (2005) mengembangkan model E-S-QUAL, yang terdiri dari *efficiency*, *fulfillment*, *system availability*, dan *privacy*. Model ini secara khusus mengukur seberapa efektif, efisien, dan andal sebuah *e-service* dalam memberikan layanan.

Instrumen pengukuran konvensional seperti SERVQUAL dan E-S-QUAL juga sering digunakan dalam berbagai penelitian dengan pendekatan metode survei (Bauer et al., 2006; Huang et al., 2015). Namun, metode survei tradisional tersebut dinilai memiliki keterbatasan seperti bias responden, waktu pengumpulan data yang panjang, serta ketidakmampuannya dalam menangkap persepsi pengguna secara *real-time* (Naab et al., 2018; Brenner & DeLamater, 2016; Batalas et al., 2015). Hal ini tidak sejalan dengan perspektif operasi jasa digital, dimana persepsi *real-time* sangat penting untuk proses pemantauan kualitas layanan, perbaikan berkelanjutan (*continuous improvement*), dan deteksi dini kegagalan layanan (*service failure detection*) (Zheng et al., 2019; Cheng et al., 2020). Selain itu, konsumen kini semakin aktif meninggalkan ulasan, komentar, dan rating pada platform seperti Google Play Store. Ulasan pengguna tersebut merupakan bagian dari *unstructured data* yang merepresentasikan pengalaman nyata, keluhan langsung, serta persepsi kualitas layanan dari sudut pandang pengguna sendiri (Ferreira et al., 2023; Ho et al., 2020). Data tersebut berfungsi sebagai *operational feedback loop* yang sangat berharga dalam operasi jasa digital karena dapat digunakan untuk mengidentifikasi titik lemah proses layanan (Hsueh & Hsu, 2024; Ananthakrishnan et al., 2019).

Dalam menganalisis data ulasan pengguna, pendekatan seperti *text mining* dan *sentiment analysis* saat ini telah digunakan secara luas untuk memahami persepsi pengguna terhadap layanan digital (Pang & Lee, 2004; Thelwall et al., 2011). *Text mining* merupakan pendekatan analitis yang digunakan untuk mengekstraksi informasi, pola, dan sentimen dari data teks tidak terstruktur seperti ulasan pengguna aplikasi digital (Basha, 2023; Dasgupta & Sen, 2022). Pendekatan *text mining* mencakup berbagai metode analisis, mulai dari teknik statistik tradisional seperti TF-IDF dan Naïve Bayes hingga model berbasis *deep learning* seperti arsitektur BERT (Devlin et al., 2019). Salah satu model yang dikembangkan khusus untuk Bahasa Indonesia adalah IndoBERT (Koto et al., 2020), yang memungkinkan analisis teks secara kontekstual. Studi Leem & Eum (2021) menunjukkan bahwa pendekatan *text mining* dapat mengidentifikasi dimensi kualitas layanan aplikasi *mobile* melalui analisis frekuensi kata, jaringan kata, sentimen, serta *topic modeling*. Teknik ini sangat efektif dalam manajemen operasi jasa karena dapat menangkap isu-isu operasional seperti *system downtime*, error transaksi, kecepatan *loading*, keandalan informasi, serta kualitas interaksi digital secara keseluruhan. Meskipun demikian, masih terdapat kesenjangan penelitian, di mana pendekatan *text mining* untuk mengukur kualitas layanan sejauh ini lebih banyak diterapkan pada sektor perbankan, ritel, dan perhotelan, sementara penerapannya pada industri OTA, khususnya Traveloka, masih sangat terbatas. Sebagian besar penelitian mengenai pengukuran kualitas layanan Traveloka juga masih mengandalkan metode survei sehingga belum memanfaatkan potensi data ulasan pengguna yang sesungguhnya mencerminkan pengalaman riil dan keluhan yang muncul secara spontan.

Dalam penelitian mengenai kualitas layanan pada industri OTA, penelitian ini memiliki perbedaan yang terletak pada penggunaan pendekatan *text mining* dan *sentiment analysis* untuk mengukur kualitas layanan Traveloka berdasarkan ulasan pengguna di Google Play Store. Pendekatan ini tidak hanya lebih sesuai dengan karakteristik operasi jasa digital yang dinamis dan *real-time*, tetapi juga mampu memberikan gambaran komprehensif tentang kualitas layanan operasi digital berdasarkan pengalaman autentik pengguna. Berdasarkan paparan tersebut, tujuan penelitian ini adalah untuk menganalisis persepsi dan sentimen pengguna terhadap layanan digital Traveloka dengan mengukur kualitas layanan berdasarkan ulasan pengguna melalui pendekatan *text mining* dengan model IndoBERT, serta mengidentifikasi keluhan utama yang dialami pengguna. Dengan demikian, penelitian ini memberikan kontribusi bagi pengembangan metode evaluasi kualitas layanan pada platform OTA melalui integrasi antara analisis data tidak terstruktur dan perspektif manajemen operasi jasa digital.

METODE

Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna pada Google Play Store untuk aplikasi Traveloka. Proses pengumpulan data dilakukan menggunakan teknik *web scraping* yang mematuhi ketentuan layanan platform. Adapun data yang digunakan dan dipilih adalah data ulasan berbahasa Indonesia sepanjang tahun 2025 yang berjumlah sebanyak 121.498 ulasan. Pemanfaatan ulasan aplikasi sebagai sumber data telah dinyatakan valid dalam berbagai penelitian pengukuran kualitas layanan digital karena mampu menyediakan *real-time insights* mengenai perilaku dan pengalaman pengguna (Sari et al., 2020; Duan et al., 2013).

Pra-pengolahan data

Tahap pra-pengolahan data dilakukan untuk memastikan kualitas data sebelum dianalisis. Data mentah diekstraksi dalam bentuk teks ulasan, tanggal, rating, dan informasi pendukung lainnya. Tahap pra-pengolahan mencakup penghapusan duplikasi, normalisasi teks, pembersihan simbol dan tanda baca, tokenisasi, *stop-word removal*, serta *stemming* menggunakan perangkat pemrosesan bahasa alami untuk Bahasa Indonesia dengan skema WordPiece yang digunakan oleh IndoBERT. Proses ini diperlukan untuk memastikan tahap ini bertujuan untuk memastikan bahwa teks dapat diproses secara optimal oleh model, sebagaimana direkomendasikan dalam penelitian *opinion mining* (Duan et al., 2013). Adapun tahap ini dilakukan dengan mengikuti praktik standar dalam penelitian *Natural Language Processing* (NLP) dan analisis sentimen (Wibowo, 2019).

Analisis Sentimen

Setelah tahap pra-pengolahan data selesai dilakukan, maka sebagai tahap awal dalam menganalisis kualitas layanan, dilakukan analisis sentimen terhadap ulasan pengguna pada Google Play Store untuk aplikasi Traveloka menggunakan pendekatan *text mining* dengan model IndoBERT. Analisis sentimen tersebut perlu dilakukan untuk mengklasifikasikan setiap ulasan ke dalam kategori positif, negatif, atau netral. Analisis dilakukan dengan menggunakan metode *supervised machine learning* dengan model pembelajaran *transformer-based*, yaitu IndoBERT, yang telah terbukti memiliki tingkat akurasi dan skor F1 tinggi dalam klasifikasi sentimen Bahasa Indonesia dibandingkan model konvensional seperti *Support Vector Machine* (SVM) atau *Bidirectional Long Short-Term Memory* (BiLSTM) (Dhendra & Utomo, 2025). Selain itu, IndoBERT dipilih karena arsitektur *transformer*-nya memungkinkan pemahaman konteks bahasa yang lebih dalam melalui *bidirectional contextual representations*, sehingga mampu menangkap nuansa opini yang kompleks dalam teks ulasan. Pendekatan ini konsisten digunakan dalam penelitian NLP pada domain analisis sentimen Bahasa Indonesia lainnya (Koto et al., 2020).

Sebanyak 70% dari data yang dikumpulkan digunakan sebagai data latih (*training set*) dan 30% lainnya digunakan sebagai data uji (*testing set*). Kinerja model kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan skor F1 yang umum digunakan dalam evaluasi model klasifikasi. Metrik *accuracy* digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data uji. Menurut Sujon et al. (2025), metrik *precision* berfungsi mengukur tingkat ketepatan prediksi dalam masing-masing kelas, yaitu seberapa banyak prediksi positif yang benar-benar relevan. Metrik *recall* mengukur kemampuan model dalam mengidentifikasi seluruh data yang relevan, yaitu proporsi data positif yang berhasil dideteksi oleh model. Skor F1 digunakan sebagai indikator keseimbangan antara tingkat *precision* dan *recall* karena dihitung sebagai *harmonic mean* dari kedua metrik tersebut. Hal ini memberikan gambaran performa model secara lebih komprehensif, terutama pada *dataset* yang tidak seimbang. Model terpilih kemudian digunakan untuk mengklasifikasikan seluruh *dataset* ulasan untuk menghasilkan skor sentimen yang akan diolah lebih lanjut dalam perhitungan *Service Quality Dimension Score*.

Skor Dimensi Kualitas Layanan (*Service Quality Dimension Score*)

Penelitian ini mengadopsi empat dimensi utama dari kerangka E-S-QUAL yang relevan dengan konteks Traveloka sebagai platform pemesanan perjalanan digital. Dimensi tersebut meliputi (1) *Efficiency*, yaitu kemudahan dan kecepatan pengguna dalam mengakses serta menggunakan aplikasi; (2) *Fulfillment*, yaitu ketepatan dan keandalan layanan dalam pemrosesan pesanan, ketersediaan produk, konfirmasi pemesanan, serta ketepatan pengiriman *e-voucher* atau tiket; (3) *System Availability*, yaitu stabilitas dan keandalan sistem aplikasi Traveloka, termasuk kinerja *server*, respons waktu sistem, fitur pencarian, dan ketiadaan gangguan teknis; dan (4) *Privacy*, yaitu perlindungan data pengguna, keamanan transaksi, serta kepercayaan pengguna terhadap penanganan informasi pribadi. Pemilihan dimensi ini merujuk pada model E-S-QUAL dari Parasuraman et al. (2005) serta relevansi empirisnya dalam penelitian *text mining* layanan digital.

Setelah ulasan diklasifikasikan berdasarkan sentimen, langkah selanjutnya adalah pemetaan setiap ulasan ke dalam empat dimensi kualitas layanan dengan menggunakan model IndoBERT. Proses ini dilakukan melalui pendekatan *keyword-based classification*, yaitu mencocokkan setiap ulasan dengan daftar kata kunci dan frasa yang mewakili tiap dimensi. Selanjutnya setiap dimensi kualitas layanan dihitung menggunakan rumus *Service Quality Dimension Score* sebagaimana digunakan dalam penelitian Leem dan Eum (2021) sebagai berikut:

$$S_i = \frac{N_{P_i} - N_{N_i}}{N_{P_i} + N_{N_i}}$$

di mana:

S_i : merupakan skor kualitas untuk dimensi ke- i

N_{P_i} : merupakan jumlah ulasan positif yang terkait dengan dimensi tersebut

N_{N_i} : merupakan jumlah ulasan negatif pada dimensi yang sama.

Rumus ini memberikan skor terstandarisasi dalam rentang -1 hingga $+1$, sehingga memungkinkan perbandingan antar-dimensi secara objektif. Nilai skor mendekati $+1$ menunjukkan persepsi layanan yang sangat positif, sementara nilai mendekati -1 menandakan dominasi sentimen negatif pada dimensi tersebut. Pendekatan perhitungan ini dipilih karena mampu menangkap keseimbangan relatif antara opini positif dan negatif, sekaligus menghindari distorsi akibat perbedaan jumlah ulasan antar-dimensi (Kim & Chung, 2024; Duan et al., 2013).

HASIL DAN PEMBAHASAN

Analisis Sentimen

Penelitian ini bertujuan untuk menganalisis persepsi pengguna terhadap kualitas layanan Traveloka berdasarkan ulasan pengguna di Google Play Store, sebagaimana dirumuskan pada bagian pendahuluan. Sesuai dengan metode penelitian, analisis dilakukan melalui tahapan pra-pengolahan, klasifikasi sentimen menggunakan IndoBERT, serta pemetaan ulasan ke dalam dimensi kualitas layanan berdasarkan kerangka E-S-QUAL. *Dataset* yang digunakan terdiri dari 121.498 ulasan berbahasa Indonesia yang dikumpulkan sepanjang tahun 2025.

Hasil analisis pada Tabel 1 menunjukkan bahwa dari total 121.498 ulasan yang dianalisis, sebanyak 59.264 ulasan atau 48,8% diklasifikasikan sebagai sentimen positif. Dalam konteks layanan digital, dominasi sentimen positif pada kisaran 45–60% umumnya menunjukkan bahwa sistem telah memenuhi ekspektasi dasar pengguna (Dhar & Bose, 2022; Ghadiridehkordi et al., 2024). Pada penelitian ini, dominasi sentimen positif khususnya terlihat pada aspek kemudahan penggunaan dan kecepatan proses. Proporsi ini menunjukkan bahwa hampir setengah dari pengguna merasakan pengalaman yang baik terhadap layanan Traveloka. Pada sentimen positif, kata yang paling sering muncul antara lain "mantap", "mudah", "membantu", "cepat", dan "puas". Dominasi kata-kata ini menunjukkan bahwa faktor kemudahan penggunaan (*usability*) dan efisiensi proses merupakan sumber utama kepuasan pengguna. Hal ini menunjukkan bahwa antarmuka aplikasi (*user interface*) dan alur pemesanan telah mampu memenuhi kebutuhan pengguna secara efektif.

Tabel 1. Klasifikasi sentimen berdasarkan ulasan pengguna Traveloka di Google Play Store

Sentimen (Polarity Analysis)	Jumlah Ulasan	Presentase (%)	Most Frequent Words
Positif	59264	48,8 %	"mantap", "mudah", "membantu", "cepat", "puas"
Negatif	28887	23,7%	"refund", "lama", "kecewa", "ribet", "susah"
Netral	33347	27,5%	"tiket", "aplikasi", "perjalanan", "hotel", "pesawat"

Sebaliknya, sebanyak 28.887 ulasan atau 23,7% diklasifikasikan sebagai sentimen negatif. Persentase ini menunjukkan bahwa hampir satu dari empat ulasan mengandung keluhan atau pengalaman yang kurang memuaskan. Pada sentimen negatif, kata yang sering muncul antara "refund", "lama", "kecewa", "ribet", dan "susah". Kemunculan kata-kata ini menunjukkan bahwa sebagian besar keluhan pengguna berkaitan dengan keandalan sistem (*system reliability*) dan proses transaksi. Kata "refund" dan "lama" juga menunjukkan bahwa waktu penyelesaian masalah menjadi faktor penting yang memengaruhi persepsi pengguna terhadap kualitas layanan. Proporsi ini cukup signifikan karena dalam layanan digital, pengalaman negatif cenderung memiliki dampak yang lebih besar terhadap persepsi pengguna dibandingkan pengalaman positif (Qahri-Saremi & Montazemi, 2022; Ramachandran et al., 2021; Lai et al., 2021), terutama jika berkaitan dengan kegagalan sistem atau transaksi (Hennig-Thurau et al., 2004).

Sementara itu, sebanyak 33.347 ulasan atau 27,5% diklasifikasikan sebagai sentimen netral. Pada sentimen netral, kata yang sering muncul antara lain "tiket", "aplikasi", "perjalanan", "hotel", dan "pesawat". Kata-kata ini menunjukkan bahwa sebagian besar ulasan netral bersifat informatif dan berkaitan dengan aktivitas transaksi rutin tanpa pengalaman positif atau negatif yang menonjol. Persentase ini menunjukkan bahwa sebagian besar pengguna menggunakan ulasan sebagai sarana memberikan informasi atau deskripsi pengalaman tanpa mengekspresikan evaluasi emosional yang kuat.

Apabila dianalisis berdasarkan distribusi keseluruhan, rasio sentimen positif terhadap negatif adalah sekitar 2:1. Rasio ini menunjukkan bahwa persepsi pengguna secara umum

masih cenderung positif, namun tingkat keluhan masih cukup tinggi sehingga memerlukan perhatian pada aspek operasional tertentu.

Untuk menentukan seberapa baik polaritas sentimen diklasifikasikan dalam analisis sentimen di atas, maka model dievaluasi dengan menggunakan metrik *accuracy*, *precision*, *recall*, dan skor F1 sesuai praktik umum dalam penelitian NLP. Hasil evaluasi menunjukkan bahwa model memiliki nilai *accuracy* sebesar 0.892 atau 89,2%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Nilai *precision* sebesar 0.884 atau 88,4% menunjukkan bahwa prediksi yang dihasilkan model memiliki tingkat ketepatan yang baik. Sementara nilai *recall* sebesar 0.903 atau 90,3% menunjukkan bahwa model mampu mengidentifikasi sebagian besar data pada kelas yang relevan. Skor F1 yang dihitung menggunakan rumus $F\text{-Score} = 2 \times (Precision \times Recall) / (Precision + Recall)$ adalah sebesar 0.893 atau 89,3%. Nilai ini menunjukkan keseimbangan yang baik antara *precision* dan *recall*, sehingga model dapat dianggap cukup reliabel untuk digunakan dalam analisis sentimen pada ulasan pengguna. Selain itu, skor F1 yang dihasilkan telah melampaui ambang batas 70% yang umumnya diterima dalam penelitian ilmu sosial (Kiritchenko et al., 2014). Hasil ini konsisten dengan penelitian Koto et al. (2020) dan Wilie et al. (2020) yang menunjukkan bahwa IndoBERT memiliki performa tinggi pada tugas klasifikasi teks berbahasa Indonesia.

Pengukuran Skor Kualitas Layanan (*Service Quality Score Measurement*)

Setelah proses klasifikasi sentimen selesai dilakukan, tahap selanjutnya adalah mengidentifikasi keterkaitan setiap ulasan dengan empat dimensi kualitas layanan berdasarkan kerangka E-S-QUAL, yaitu *efficiency*, *fulfillment*, *system availability*, dan *privacy* menggunakan model IndoBERT. Pada tahap ini, sebagaimana ditunjukkan pada Tabel 2, identifikasi dilakukan melalui metode berbasis kata kunci (*keyword-based classification*), di mana isi ulasan dianalisis dengan mendeteksi keberadaan istilah atau frasa spesifik yang menjadi representasi dari masing-masing dimensi layanan.

Tabel 2. Keyword dictionary untuk klasifikasi dimensi *Service Quality*

Dimensi	Deskripsi	Contoh Kata Kunci Positif	Contoh Kata Kunci Negatif
<i>Efficiency</i>	Kemudahan penggunaan aplikasi, kecepatan respon, dan kemudahan navigasi dalam proses pencarian dan pemesanan	mudah, cepat, praktis, <i>user friendly</i> , tampilan jelas, simpel, responsif, tidak ribet	lemot, <i>loading</i> lama, ribet, sulit digunakan, lambat, tidak responsif
<i>Fulfillment</i>	Kesesuaian layanan dengan harapan pengguna	sesuai pesanan, konfirmasi cepat, pembayaran mudah, refund cepat, proses lancar, harga jelas	<i>refund</i> lama, pesanan gagal, tidak sesuai, harga berubah, transaksi gagal, konfirmasi lama
<i>System Availability</i>	Keandalan sistem aplikasi dan stabilitas layanan saat digunakan	stabil, lancar, jarang <i>error</i> , tidak pernah bermasalah, selalu bisa digunakan	<i>error</i> , <i>server down</i> , aplikasi crash, tidak bisa dibuka, gagal <i>login</i> , gangguan sistem
<i>Privacy</i>	Keamanan data pribadi dan transaksi pengguna	aman, terpercaya, data terlindungi, pembayaran aman, verifikasi aman	akun dibobol, data bocor, keamanan lemah, verifikasi berulang, transaksi mencurigakan

Dimensi pertama, *Efficiency*, mengukur sejauh mana aplikasi Traveloka memungkinkan pengguna melakukan pencarian, pemesanan, dan pembayaran secara cepat dan mudah. Dimensi ini mencakup kecepatan respon aplikasi, kemudahan navigasi, serta kejelasan antarmuka pengguna. Dalam dimensi ini, ulasan positif umumnya mengandung kata kunci seperti "mudah", "cepat", "praktis", "*user friendly*", "tampilan jelas", dan "proses cepat", yang menunjukkan bahwa pengguna merasa aplikasi mudah digunakan dan efisien dalam

mendukung transaksi perjalanan. Sebaliknya, ulasan negatif dalam dimensi ini sering menyebutkan kata seperti "lemot", "loading lama", "sulit digunakan", atau "ribet", yang mengindikasikan ketidakpuasan pengguna terhadap performa atau kemudahan penggunaan aplikasi.

Dimensi kedua, *Fulfillment*, mengukur sejauh mana layanan yang diberikan sesuai dengan harapan pengguna, khususnya terkait keakuratan informasi, keberhasilan pemesanan, serta proses transaksi dan *refund*. Dalam konteks Traveloka, dimensi ini mencakup kesesuaian jadwal penerbangan atau hotel, kejelasan informasi harga, serta kelancaran proses pembayaran dan pengembalian dana. Ulasan positif pada dimensi ini sering memuat kata seperti "sesuai pesanan", "proses cepat", "refund lancar", "pembayaran mudah", dan "konfirmasi cepat". Sementara itu, ulasan negatif sering menyebutkan kata seperti "refund lama", "pesanan gagal", "harga berubah", "tidak sesuai", atau "proses lama", yang menunjukkan adanya kendala dalam pemenuhan layanan inti.

Dimensi ketiga, *System Availability*, mengacu pada keandalan sistem aplikasi dalam memberikan layanan tanpa gangguan. Dimensi ini berkaitan dengan stabilitas aplikasi, ketersediaan *server*, serta kemampuan sistem untuk beroperasi secara konsisten ketika digunakan. Ulasan positif pada dimensi ini umumnya mengandung kata seperti "jarang error", "stabil", "lancar", dan "tidak pernah bermasalah", yang menunjukkan bahwa pengguna menilai sistem dapat diandalkan. Sebaliknya, ulasan negatif sering menyebutkan kata seperti "error", "gagal login", "server down", "tidak bisa dibuka", atau "aplikasi crash", yang menunjukkan bahwa gangguan teknis menjadi sumber ketidakpuasan pengguna.

Dimensi keempat, *Privacy*, berkaitan dengan kemampuan aplikasi dalam melindungi data pribadi dan keamanan transaksi pengguna. Dalam konteks layanan perjalanan digital, dimensi ini mencakup keamanan pembayaran, perlindungan data akun, serta kepercayaan pengguna terhadap sistem. Ulasan positif dalam dimensi ini sering memuat kata seperti "aman", "pembayaran aman", "data terlindungi", dan "terpercaya", yang mencerminkan persepsi positif terhadap keamanan layanan. Sebaliknya, ulasan negatif sering mengandung kata seperti "akun dibobol", "data bocor", "verifikasi berulang", atau "keamanan lemah", yang menunjukkan kekhawatiran pengguna terhadap perlindungan data pribadi.

Setelah itu, untuk memperoleh skor kualitas layanan pada setiap dimensi layanan yang lebih rinci, *dataset* ulasan positif dan negatif diklasifikasikan berdasarkan area evaluasi kualitas layanan, kemudian skor untuk masing-masing area dihitung berdasarkan jumlah ulasan. Tabel 3 menampilkan hasil klasifikasi ulasan berdasarkan dimensi E-S-QUAL beserta hasil perhitungan skor kualitas layanan (*service quality score*).

Tabel 3. Frekuensi ulasan dan skor kualitas layanan setiap dimensi

Dimensi Layanan	Total Ulasan	Jumlah Ulasan Positif	Jumlah Ulasan Negatif	Skor Kualitas Layanan (<i>Service Quality Score</i>)
<i>Efficiency</i>	19.808	13.767	6.041	+0,39
<i>Fulfillment</i>	9.408	3.481	5.927	-0,26
<i>System Availability</i>	2.891	809	2.082	-0,44
<i>Privacy</i>	2.932	425	2.507	-0,71

Hasil perhitungan skor kualitas layanan pada empat dimensi sebagaimana ditampilkan pada Tabel 3 menunjukkan adanya variasi persepsi pengguna terhadap kinerja operasional aplikasi Traveloka. Variasi ini mencerminkan bahwa kualitas layanan digital tidak hanya ditentukan oleh kemudahan penggunaan aplikasi, tetapi juga oleh keandalan proses operasional di belakang sistem (*back-end operations*), stabilitas infrastruktur teknologi, serta kemampuan perusahaan dalam menjaga kepercayaan pengguna.

Dimensi *Efficiency* memperoleh skor positif sebesar +0,39 dengan total penyebutan sebanyak 19.808, yang terdiri dari 13.767 ulasan positif dan 6.041 ulasan negatif. Nilai ini menunjukkan bahwa sebagian besar pengguna menilai aplikasi Traveloka telah mampu

menyediakan proses pencarian, pemesanan, dan pembayaran yang relatif cepat dan mudah digunakan. Dalam perspektif manajemen operasi jasa digital, *efficiency* merepresentasikan kinerja proses layanan pada tahap *service delivery interface*, yaitu bagaimana sistem memungkinkan pengguna menyelesaikan transaksi dengan usaha dan waktu yang minimal (Nguyen et al., 2025; Bueno, 2024). Tingginya jumlah ulasan positif pada dimensi ini mengindikasikan bahwa desain antarmuka, struktur navigasi, serta alur transaksi telah berjalan secara efektif. Blut (2021) menunjukkan bahwa dalam kualitas layanan digital, kemudahan penggunaan dan kecepatan sistem memiliki pengaruh signifikan terhadap kepuasan pengguna aplikasi berbasis teknologi. Namun demikian, masih ada lebih dari enam ribu ulasan negatif menunjukkan bahwa sebagian pengguna masih mengalami hambatan seperti *loading* lambat atau proses yang tidak responsif.

Selanjutnya, dimensi *Fulfillment* menunjukkan skor $-0,26$, dengan penyebutan sejumlah 9.408, terdiri dari 3.481 ulasan positif dan 5.927 ulasan negatif. Hasil ini menunjukkan bahwa persepsi pengguna terhadap kemampuan sistem dalam memenuhi janji layanan masih cenderung negatif. *Fulfillment* berkaitan dengan keandalan proses inti layanan (*core service process*), seperti keberhasilan transaksi, kesesuaian harga dan jadwal, serta kecepatan proses *refund*. Dominasi ulasan negatif pada dimensi ini mengindikasikan bahwa permasalahan yang terjadi tidak hanya berkaitan dengan aplikasi sebagai antarmuka, tetapi juga pada integrasi sistem dengan mitra eksternal seperti maskapai, hotel, dan penyedia layanan lainnya. Hal ini menunjukkan bahwa kualitas layanan digital sangat dipengaruhi oleh koordinasi rantai layanan (*service supply chain integration*) dan keandalan proses *back-end*. Ketidaksesuaian layanan pada tahap ini berpotensi menimbulkan *service failure* yang berdampak langsung pada kepuasan dan loyalitas pengguna (Johnston et al., 2021).

Dimensi ketiga, yaitu *System Availability*, memperoleh skor $-0,44$ dengan penyebutan sejumlah 2.891, terdiri dari 809 ulasan positif dan 2.082 ulasan negatif. Skor ini menunjukkan bahwa gangguan teknis seperti kesalahan sistem (*system error*), kegagalan *login*, atau aplikasi tidak dapat diakses masih menjadi isu yang signifikan. Dalam layanan berbasis teknologi, *availability* merupakan indikator utama reliabilitas operasional karena berkaitan dengan stabilitas infrastruktur dan kapasitas sistem dalam menangani beban *traffic* (Fitzsimmons et al., 2022). Gangguan pada dimensi ini tidak hanya menghambat transaksi, tetapi juga dapat merusak persepsi kualitas secara keseluruhan karena pengguna tidak dapat mengakses layanan pada saat dibutuhkan (Ayo et al., 2016).

Dimensi terakhir, yaitu *Privacy*, menunjukkan skor terendah yaitu $-0,71$ dengan penyebutan sejumlah 2.932, terdiri dari 425 ulasan positif dan 2507 ulasan negatif. Nilai ini menunjukkan bahwa kekhawatiran terhadap keamanan data dan perlindungan informasi pribadi menjadi perhatian utama pengguna. Dalam konteks layanan digital, keamanan data dan perlindungan privasi merupakan faktor penting yang menentukan kepercayaan pengguna (*e-trust*), yang pada gilirannya memengaruhi niat penggunaan berkelanjutan (Al-Hawari, 2022). Blut (2021) menunjukkan bahwa persepsi risiko keamanan memiliki dampak yang lebih kuat terhadap ketidakpuasan dibandingkan dimensi kualitas lainnya karena menyangkut aspek perlindungan informasi pribadi. Oleh karena itu, penguatan sistem keamanan, enkripsi data, serta transparansi kebijakan privasi merupakan bagian integral dari manajemen risiko operasional layanan digital, termasuk Traveloka.

Apabila dilakukan perbandingan antar dimensi, terlihat bahwa *Efficiency* merupakan kekuatan utama sistem, sedangkan *Fulfillment*, *System Availability*, dan *Privacy* merupakan area yang memerlukan perhatian operasional lebih lanjut. Temuan ini menunjukkan bahwa keberhasilan pada sisi antarmuka pengguna tidak selalu diikuti oleh keandalan pada proses operasional di belakang sistem. Dalam kerangka manajemen operasi jasa digital, kondisi ini menunjukkan adanya kesenjangan antara kinerja *front-end service delivery* dan *back-end operational performance*. Hasil penelitian ini memperkuat pandangan bahwa kualitas layanan digital merupakan hasil dari integrasi antara teknologi, proses operasional, dan manajemen

layanan secara menyeluruh. Sistem yang mudah digunakan tidak cukup untuk menjamin kepuasan pengguna apabila proses pemenuhan layanan, stabilitas sistem, dan keamanan data belum optimal. Oleh karena itu, peningkatan kualitas layanan digital pada Traveloka perlu difokuskan pada penguatan infrastruktur teknologi, peningkatan keandalan integrasi sistem dengan mitra layanan, serta penguatan mekanisme keamanan dan perlindungan data pengguna. Temuan ini juga menunjukkan bahwa analisis ulasan pengguna dapat berfungsi sebagai *operational feedback mechanism* yang efektif dalam manajemen operasi jasa digital. Melalui pemantauan sentimen dan dimensi kualitas layanan secara berkelanjutan, perusahaan dapat mengidentifikasi titik lemah proses layanan secara lebih cepat dan melakukan perbaikan berbasis data nyata dari pengalaman pengguna.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis persepsi pengguna terhadap kualitas layanan aplikasi Traveloka melalui pendekatan *text mining* dan *sentiment analysis* berbasis IndoBERT dengan kerangka E-S-QUAL. Berdasarkan analisis terhadap 121.498 ulasan pengguna aplikasi Traveloka di Google Play Store tahun 2025, ditemukan bahwa persepsi pengguna secara umum cenderung positif dengan rasio sentimen positif dan negatif sebesar 2:1. Model IndoBERT menunjukkan performa klasifikasi yang baik dengan tingkat akurasi 89,2% dan skor F1 89,3%, sehingga dapat dinilai andal dalam mengklasifikasikan sentimen ulasan berbahasa Indonesia. Hasil pengukuran skor kualitas layanan menunjukkan bahwa dimensi *efficiency* menjadi kekuatan utama aplikasi Traveloka dengan skor positif (+0,39), yang mengindikasikan bahwa kemudahan penggunaan dan kecepatan proses transaksi telah memenuhi ekspektasi sebagian besar pengguna. Namun demikian, dimensi *fulfillment*, *system availability*, dan *privacy* menunjukkan skor negatif, dengan *privacy* memperoleh skor negatif terendah (-0,71). Temuan ini mengindikasikan bahwa tantangan utama Traveloka tidak terletak pada aspek antarmuka pengguna (*front-end*), melainkan pada keandalan proses operasional inti, stabilitas infrastruktur sistem, integrasi layanan dengan mitra eksternal, serta perlindungan data dan keamanan transaksi.

Secara praktis, penelitian ini memberikan kontribusi bagi Traveloka dengan memberikan rekomendasi pemetaan prioritas perbaikan operasional berbasis data ulasan pengguna yang bersifat *real-time* dan autentik. Hasil penelitian ini dapat digunakan oleh manajemen untuk memfokuskan strategi peningkatan layanan pada penguatan koordinasi *service supply chain* dengan mitra perjalanan, peningkatan kapasitas dan stabilitas sistem untuk meminimalkan gangguan teknis, serta penguatan mekanisme keamanan data guna membangun kembali dan meningkatkan kepercayaan pengguna. Selain itu, pendekatan *text mining* berbasis IndoBERT yang digunakan dalam penelitian ini dapat diimplementasikan sebagai sistem monitoring kualitas layanan digital secara berkelanjutan (*continuous service quality monitoring system*), sehingga Traveloka dapat melakukan deteksi dini terhadap potensi *service failure* dan merespons keluhan pengguna dengan berbasis data secara lebih cepat. Selain itu, penelitian ini berkontribusi dalam memperluas penerapan model E-S-QUAL melalui analisis data tidak terstruktur serta mengintegrasikan perspektif manajemen operasi jasa digital dengan pendekatan *deep learning* berbasis NLP. Studi ini menunjukkan bahwa evaluasi kualitas layanan digital tidak hanya dapat dilakukan melalui survei tradisional, tetapi juga melalui pemanfaatan *user-generated content* sebagai sumber *operational intelligence* yang strategis.

REFERENSI

- Al-Hawari, M. A. (2022). The role of online service quality in enhancing customer satisfaction and loyalty in digital platforms. *Journal of Retailing and Consumer Services*, 64.
- Amra & Delma. (2025). *Play Store marketing statistics 2025: Number of apps, users, revenue, downloads & more*. <https://www.amraandelma.com/play-store-marketing-statistics/>
- Ananthakrishnan, U. M., Proserpio, D., & Sharma, S. (2019). I hear you: Does quality improve with customer voice? *Social Science Research Network*. <https://doi.org/10.2139/SSRN.3467236>

- Arcand, M., PromTep, S., Brun, I., & Rajaobelina, L. (2017). Mobile banking service quality and customer relationships. *International Journal of Bank Marketing*, 35(7), 1068–1089.
- Ayo, C. K., Oni, A. A., Adewoye, O. J., & Eweoya, I. O. (2016). E-banking users' behaviour: E-service quality, attitude, and customer satisfaction. *International Journal of Bank Marketing*, 34(3), 347–367.
- Basha, S. M. (2023). Knowledge discovery from online reviews. In *Translational Systems Sciences* (pp. 71–104). https://doi.org/10.1007/978-981-99-1075-5_3
- Batalas, N., Quevedo-Fernandez, J., Martens, J.-B., & Markopoulos, P. (2015). Meet your users in situ: Data collection from within apps in large-scale deployments. *International Journal of Handheld Computing Research*, 6(3), 17–32.
- Bauer, H. H., Falk, T., & Hammerschmidt, M. (2006). eTransQual: A transaction process-based approach for capturing service quality in online shopping. *Journal of Business Research*, 59(7), 866–875.
- Blut, M. (2021). E-service quality: Development of a hierarchical model. *Journal of Service Research*, 24(1), 5–26.
- Bueno, L. A. (2024). *Exploratory study of digital operational efficiency for Brazilian banks*. <https://doi.org/10.47749/t/unicamp.2024.1408575>
- Carlson, J., & O'Cass, A. (2011). Developing a framework for understanding e-service quality, its antecedents, consequences, and mediators. *Managing Service Quality*, 21(3), 264–286.
- Cheng, Y., Xie, Q., & Zhang, Z. (2020). A method of real-time service quality evaluation based on customer behavior. <https://doi.org/10.1145/3418981.3418982>
- Chuang, C.-C., Cheng, W.-L., & Hsu, K.-S. (2015). A comprehensive composite digital services quality assurance application on intelligent transportation system. *Asia-Pacific Network Operations and Management Symposium*, 368–371.
- Dasgupta, S., & Sen, J. (2022). A framework of customer review analysis using the aspect-based opinion mining approach. *Proceedings of OCIT*, 233–237.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
- Dhendra, & Utomo, G. V. (2025). Benchmarking IndoBERT and transformer models for sentiment classification on Indonesian e-government service reviews. *Jurnal Transformatika*, 23(1), 86–95.
- Dhar, S., & Bose, I. (2022). Walking on air or hopping mad? Understanding the impact of emotions, sentiments and reactions on ratings in online customer reviews of mobile apps. *Decision Support Systems*, 162, 113769.
- Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining online user-generated content: Using sentiment analysis technique to study hotel service quality. *Proceedings of the Hawaii International Conference on System Sciences*, 3119–3128.
- Ferreira, C., Robertson, J., Chohan, R., & Pitt, C. (2023). Using computerized text analysis to predict customers' evaluation of their service experience. *Developments in Marketing Science*, 375–376.
- Fitzsimmons, J. A., Fitzsimmons, M. J., & Bordoloi, S. (2022). *Service management: Operations, strategy, information technology* (10th ed.). McGraw-Hill.
- Ghadiridehkordi, A., Shao, J., Boojhawon, R., Wang, Q., & Li, H. (2024). Leveraging sentiment analysis via text mining to improve customer satisfaction in UK banks. *International Journal of Bank Marketing*, 43(4), 780–802.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms. *Journal of Interactive Marketing*, 18(1), 38–52.

- Ho, R. C., Withanage, M. S., & Khong, K. W. (2020). Sentiment drivers of hotel customers: A hybrid approach using unstructured data. *Asia-Pacific Journal of Business Administration*, 12(3–4), 237–250.
- Hsueh, J.-T., & Hsu, S.-H. (2024). Turning negative reviews into operational insights: ABSS-GPT's role in informing hotel decisions. *Journal of Decision Systems*, 1–16.
- Huang, E. Y., Lin, S. W., & Fan, Y. C. (2015). MS-QUAL: Mobile service quality measurement. *Electronic Commerce Research and Applications*, 14(2), 126–142.
- Johnston, R., Shulver, M., Slack, N., & Clark, G. (2021). *Service operations management* (5th ed.). Pearson.
- Joshi, K. P., Joshi, A., & Yesha, Y. (2009). A framework for relating frontstage and backstage quality in virtualized services.
- Kim, J., & Chung, J. (2024). Analysis of service quality in smart running applications using big data text mining techniques. *Journal of Theoretical and Applied Electronic Commerce Research*, 19, 3352–3369.
- Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723–762.
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, 757–770.
- Lai, X., Wang, F., & Wang, X. (2021). Asymmetric relationship between customer sentiment and online hotel ratings. *International Journal of Contemporary Hospitality Management*, 33(6), 2137–2156.
- Leem, B. H., & Eum, S. W. (2021). Using text mining to measure mobile banking service quality. *Industrial Management & Data Systems*, 121(5), 993–1007.
- Nguyen, H. T., Nguyen, G. H., & Tran, Y. T. H. (2025). The impact of digital transformation on operational efficiency of commercial banks. *International Journal of Innovative Research and Scientific Studies*, 8(2), 4345–4353.
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of ACL*, 271–278.
- Parasuraman, A., Grewal, D. (2000). The impact of technology on the quality–value–loyalty chain. *Journal of the Academy of Marketing Science*, 28(1), 168–174.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL. *Journal of Retailing*, 64(1), 12–40.
- Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL. *Journal of Service Research*, 7(3), 213–233.
- Qahri-Saremi, H., & Montazemi, A. R. (2022). Negativity bias in online reviews. *European Journal of Information Systems*, 32(4), 717–734.
- Ramachandran, R., Sudhir, S., & Unnithan, A. B. (2021). Emotionality and product ratings in online reviews. *IIMB Management Review*, 33(4), 299–308.
- Sari, Y., Wibowo, A., & Tjahyanto, A. (2020). Analisis sentimen terhadap layanan transportasi online menggunakan machine learning.
- Shankar, A., & Jebarajakirthy, C. (2019). The influence of e-banking service quality on customer loyalty. *International Journal of Bank Marketing*, 37(5), 1119–1142.
- Sujon, K. M., Hassan, R., & Choi, K. (2025). Accuracy, precision, recall, F1-score, or MCC? *Journal of Big Data*, 12, 268.
- Tate, M., & Evermann, J. (2010). The end of SERVQUAL in online services research. *E-Service Journal*, 7(1), 60–85.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406–418.
- Traveloka. (2025). About Traveloka. <https://www.traveloka.com/en-id/about-us>

- Vukovic, M., & Natarajan, A. (2012). Collective intelligence for enhanced quality management of IT services. Springer.
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, A. Y., & Purwarianti, A. (2020). IndoNLU benchmark. *Proceedings of EMNLP*, 155–165.
- Wibowo, A. P. (2019). Preprocessing teks bahasa Indonesia untuk analisis sentimen.
- Zheng, W., Lu, H., Zhou, Y., Liang, J., Zheng, H., & Deng, Y. (2019). iFeedback: Exploiting user feedback for real-time issue detection. *Automated Software Engineering*, 352–363.