BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 6, October 2025 | Hal 1387-1394
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i6.822

Analisis Sentimen Keluhan Pelanggan ISP menggunakan Support Vector Machine (SVM) dan TF-IDF

Dini Fakta Sari[1], Deborah Kurniawati[2,*], Endang Wahyuningsih[3], Tediyan Rahmat Wibowo[4]

[1,4] Fakultas Teknologi Informasi, Program Studi Informatika, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
[2] Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
[3] Fakultas Teknologi Informasi, Program Studi Sistem Informasi Akuntansi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
Email: [1]dini@utdi.ac.id, [2,*]debbie@utdi.ac.id, [3]ayuning@utdi.ac.id, [4]tediyan.rahmat@students.utdi.ac.id
Email Penulis Korespondensi: debbie@utdi.ac.id

Abstrak−Penelitian ini bertujuan untuk menganalisis sentimen keluhan pelanggan terhadap layanan *Internet Service Provider* (ISP) di Indonesia, di mana masalah utama yang sering muncul meliputi gangguan koneksi, kecepatan internet yang lambat, sinyal lemah, serta penanganan keluhan yang tidak responsif dan tidak informatif, sebagaimana tercermin dari berbagai laporan konsumen di media sosial. Masalah ini berdampak pada ketidakpuasan pelanggan dan menuntut solusi analisis data untuk memahami opini publik secara mendalam. Data dikumpulkan melalui API dari salah satu *platform* media sosial menggunakan kata kunci terkait layanan internet, seperti "gangguan internet" dan "keluhan internet". Data melalui tahapan prapemrosesan teks, meliputi pembersihan, *case folding, tokenisasi*, penghapusan *stopword*, dan *stemming* untuk menghasilkan teks yang konsisten. Fitur teks diekstraksi menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), yang kemudian diklasifikasikan dengan algoritma *Support Vector Machine* (SVM). Evaluasi model menggunakan 10-Fold Cross Validation menghasilkan akurasi rata-rata 91,47%, presisi 94,27%, *recall* 99,20%, dan F1-*score* 96,67%. Frekuensi kemunculan kata menunjukkan kata dominan seperti "lambat", "gangguan", dan "sinyal" sebagai isu utama keluhan pelanggan. Kombinasi SVM dan TF-IDF terbukti efektif untuk analisis sentimen berbahasa Indonesia, memberikan kontribusi akademik dan praktis bagi ISP untuk memantau opini pelanggan dan meningkatkan kualitas layanan. Penelitian selanjutnya disarankan menggunakan model *deep learning* seperti BERT dan data yang lebih beragam.

Kata Kunci: Analisis Sentimen; Internet Service Provider; Media Sosial; Support Vectoe Machine; TF-IDF

Abstract−This study aims to analyze the sentiment of customer complaints regarding Internet Service Provider (ISP) services in Indonesia, where the primary issues frequently reported include connection disruptions, slow internet speeds, weak signals, and unresponsive or uninformative complaint handling, as reflected in various consumer reports on social media. These issues contribute to customer dissatisfaction and necessitate data analysis solutions to deeply understand public opinions. Data was collected via API from a social media platform using keywords related to internet services, such as "internet disruption" and "internet complaints." The data underwent text preprocessing stages, including cleaning, case folding, tokenization, stopword removal, and stemming to produce consistent text. Text features were extracted using Term Frequency–Inverse Document Frequency (TF-IDF), which were then classified using the Support Vector Machine (SVM) algorithm. Model evaluation using 10-Fold Cross Validation yielded an average accuracy of 91.47%, precision of 94.27%, recall of 99.20%, and F1-score of 96.67%. Word frequency analysis revealed dominant words such as "slow," "disruption," and "signal" as the main issues in customer complaints. The combination of SVM and TF-IDF proved effective for sentiment analysis in Indonesian, providing academic and practical contributions for ISPs to monitor customer opinions and improve service quality. Future research is recommended to employ deep learning models like BERT and more diverse data.

Keywords: Sentiment Analysis; Internet Service Provider; Social Media; Support Vector Machine; TF-IDF

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat berinteraksi, bekerja, dan mengakses informasi. Dengan dukungan ketersediaan peralatan dan infrastuktur jaringan, internet kini menjadi kebutuhan primer di berbagai sektor, termasuk pendidikan, bisnis, pemerintahan, dan hiburan. Menurut survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2025, pengguna internet di Indonesia mencapai 229,4 juta jiwa, atau sekitar 80,66% dari total populasi [1]. Angka ini menunjukkan peningkatan signifikan dibandingkan tahun sebelumnya, di mana penetrasi internet hanya 79,50% atau sekitar 225 juta jiwa. Tingginya ketergantungan masyarakat terhadap konektivitas internet membuat kualitas layanan *Internet Service Provider* (ISP) menjadi isu krusial. Keluhan pelanggan terkait kecepatan internet, stabilitas jaringan, dan pelayanan pelanggan sering kali muncul, terutama di platform media sosial [2], [3].

Media sosial telah menjadi wadah utama bagi pengguna untuk menyampaikan opini, baik berupa pujian maupun keluhan, secara *real-time*. Data dari media sosial bersifat publik dan kaya akan informasi subjektif, menjadikannya sumber yang ideal untuk analisis sentimen [4]. Analisis sentimen, sebagai bagian dari *Natural Language Processing* (NLP), memungkinkan ekstraksi dan klasifikasi opini menjadi kategori positif, negatif, atau netral [5]. Dalam konteks layanan ISP, analisis ini dapat membantu penyedia layanan memahami persepsi pelanggan, mengidentifikasi masalah utama, dan merancang strategi perbaikan layanan [6]. Selain itu, dengan maraknya penggunaan media sosial di Indonesia analisis sentimen dapat memberikan wawasan waktu nyata tentang tren keluhan pelanggan.

Penelitian sebelumnya telah menunjukkan bahwa analisis sentimen berbasis pembelajaran mesin (*machine learning*) efektif untuk mengevaluasi opini pelanggan di media sosial. Misalnya, menggunakan metode *Decision Tree* dan *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk menganalisis sentimen pelanggan ISP, menghasilkan akurasi yang memadai [7]. Demikian pula, menggunakan *Support Vector Machine* (SVM) untuk analisis sentimen ISP, dengan akurasi masing-masing 85,7% dan 86,1% [8], [9]. Namun, tantangan utama dalam analisis sentimen berbahasa

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 6, October 2025 | Hal 1387-1394
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i6.822

Indonesia adalah kompleksitas bahasa, seperti penggunaan bentuk informal, singkatan, dan campuran bahasa (*code-mixing*), yang memerlukan strategi prapemrosesan teks yang kuat [10], [11].

Di antara algoritma pembelajaran mesin, SVM menonjol karena kemampuannya menangani data berdimensi tinggi dan menghasilkan pemisahan kelas yang optimal [12]. SVM bekerja dengan mencari *hyperplane* yang memaksimalkan *margin* antar kelas, menjadikannya pilihan ideal untuk klasifikasi teks seperti tweet [13]. Kombinasi SVM dengan TF-IDF, sebuah metode pembobotan fitur yang menekankan kata-kata penting dalam teks, telah terbukti meningkatkan akurasi klasifikasi [14]. Integrasi TF-IDF dengan SVM menghasilkan performa yang lebih baik dibandingkan metode berbasis kamus atau pendekatan tradisional lainnya [15]. Penelitian terkini juga menunjukkan bahwa penggunaan bigram atau trigram dalam TF-IDF dapat menangkap konteks lebih baik, terutama dalam bahasa Indonesia yang kaya akan afiks dan reduplikasi.

Di Indonesia, penelitian analisis sentimen terhadap layanan ISP masih memiliki keterbatasan. Sebagian besar studi berfokus pada penyedia layanan tertentu dengan dataset yang relatif kecil (<10.000 tweet). Selain itu, banyak penelitian belum sepenuhnya mengatasi tantangan bahasa informal dan variasi gaya penulisan di media sosial Indonesia [16]. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model analisis sentimen yang lebih komprehensif dengan menggunakan SVM dan TF-IDF, menargetkan data dari salah satu media sosial dari salah satu penyedia ISP di Indonesia. Pendekatan ini tidak hanya meningkatkan akurasi klasifikasi sentimen, tetapi juga memperhitungkan implikasi ekonomi, di mana ketidakpuasan pelanggan dapat meningkatkan tingkat pergantian pelanggan hingga 20-30% per tahun, yang berdampak signifikan pada pendapatan ISP. Selain itu, dengan memanfaatkan data real-time dari media sosial, model ini dapat mendukung pengembangan kebijakan publik untuk memperkuat kualitas infrastruktur digital nasional.

Penelitian ini memiliki tiga tujuan utama. Pertama, mengidentifikasi pola sentimen pelanggan terhadap layanan *Internet Service Provider* (ISP) di Indonesia berdasarkan data dari salah satu platform media sosial. Kedua, mengevaluasi kinerja algoritma Support Vector Machine (SVM) dalam mengklasifikasikan teks berbahasa Indonesia. Ketiga, memberikan rekomendasi praktis bagi ISP untuk meningkatkan kualitas layanan berdasarkan hasil analisis. Pendekatan ini diharapkan memperkaya literatur analisis sentimen berbahasa Indonesia dan memberikan wawasan praktis bagi industri telekomunikasi. Dengan memanfaatkan data di salah satu media sosial yang *real-time*, penelitian ini juga mendukung pengembangan sistem pemantauan pengalaman pelanggan berbasis data. Selain itu, penelitian ini membuka peluang untuk eksplorasi model pembelajaran mendalam, seperti BERT, untuk meningkatkan akurasi di masa depan. Integrasi dengan teknik lain, seperti *topic modeling* menggunakan LDA, dapat memberikan pemahaman lebih dalam tentang kluster keluhan, seperti isu teknis versus administratif. Akhirnya, di tengah transformasi digital Indonesia menuju masyarakat 5.0, penelitian ini berkontribusi pada pembangunan ekosistem digital yang lebih responsif dan inklusif, mengurangi kesenjangan akses di wilayah tertinggal.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk mengembangkan model analisis sentimen berbasis pembelajaran mesin. Algoritma *Support Vector Machine* (SVM) dipilih karena kemampuannya menangani data berdimensi tinggi dan menghasilkan pemisahan kelas yang optimal melalui *hyperplane* [17], [18]. *Term Frequency–Inverse Document Frequency* (TF-IDF) digunakan untuk ekstraksi fitur teks karena efektivitasnya dalam menangkap kata-kata penting dalam korpus [19], [20]. Proses penelitian terdiri dari tujuh tahap utama: pengumpulan data, prapemrosesan teks, pelabelan sentimen, pembobotan fitur, pelatihan model, evaluasi model, dan analisis hasil.

2.1 Pengumpulan Data dan Prapemrosesan Teks

Data dikumpulkan dari salah satu media sosial menggunakan API dengan kata kunci relevan seperti “gangguan internet”, “keluhan internet”, “lambat”, “sinyal. Pencarian difokuskan pada komentar berbahasa Indonesia yang diposting diperiode tertentu, menghasilkan dataset awal sebanyak 15.000 tweet mentah. Data disimpan dalam format CSV, berisi kolom teks komentar, tanggal unggahan, dan metadata seperti lokasi (jika tersedia) dan nama pengguna (disamarkan untuk menjaga privasi sesuai pedoman etika penelitian). Dataset ini difilter untuk menghapus komentar duplikat, iklan, dan konten tidak relevan, menghasilkan 12.000 komentar yang digunakan untuk analisis.

Data yang diperoleh diproses melalui tahapan berikut untuk memastikan konsistensi dan kualitas data:

a. Mengubah semua teks menjadi huruf kecil untuk menghindari perbedaan huruf kapital.
b. Menghapus elemen non-teks seperti URL, *mention* (@username), hashtag (#), angka, tanda baca, dan emotikon menggunakan regex di Python.
c. Memecah teks menjadi unit kata (token) menggunakan pustaka NLTK untuk mempermudah analisis.
d. Menghapus kata umum dan tidak bermakna seperti “dan”, “di”, serta kata informal seperti “gak” dan “nggak” menggunakan daftar *stopword* khusus bahasa Indonesia yang dikembangkan berdasarkan penelitian sebelumnya.
e. Mengembalikan kata ke bentuk dasar (misalnya, “melambat” menjadi “lambat”) menggunakan pustaka Sastrawi, yang dirancang untuk menangani morfologi bahasa Indonesia.

Prapemrosesan dilakukan menggunakan Python dengan pustaka NLTK untuk tokenisasi dan Sastrawi untuk *stemming*, memastikan teks bersih dari *noise* dan konsisten untuk analisis lebih lanjut [21]. Tahapan ini krusial untuk mengatasi kompleksitas bahasa Indonesia, seperti variasi dialek dan penggunaan bahasa informal di media sosial.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 6, October 2025 | Hal 1387-1394
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i6.822

2.2 Pelabelan Data dan Pembobotan Fitur dengan TF-IDF

Langkah selanjutnya adalah memberi label positif, negatif, atau netral pada data melalui pendekatan dua tahap:

a. Pelabelan otomatis menggunakan metode berbasis kamus dengan daftar kata sentimen berbahasa Indonesia (misalnya, "bagus", "cepat" untuk positif; "lambat", "gangguan" untuk negatif) yang diadaptasi dari penelitian sebelumnya. Skor sentimen dihitung berdasarkan frekuensi kata positif dan negatif dalam setiap tweet.
b. Verifikasi manual, dengan cara dua penelaah independen memeriksa label otomatis untuk memastikan akurasi konteks, terutama pada komentar dengan sarkasme atau ambiguitas. Konsensus digunakan untuk menyelesaikan perbedaan, dengan tingkat kesepakatan antar-penelaah (*inter-annotator agreement*) mencapai 92% menggunakan koefisien Kappa.

Dari 12.000 tweet, 60% diberi label negatif, 25% netral, dan 15% positif, mencerminkan dominasi keluhan dalam dataset.

Teks yang telah diproses diubah menjadi representasi numerik menggunakan TF-IDF, yang mengukur frekuensi kata (*term frequency*) dan keunikan kata dalam korpus (*Inverse Document Frequency*) [22]. TF-IDF diterapkan dalam bentuk unigram dan bigram untuk menangkap konteks antar kata, seperti "jaringan lambat" atau "sinyal buruk". Proses ini dilakukan menggunakan pustaka scikit-learn di Python, dengan parameter minimum *document frequency* (min_df) diatur ke 5 untuk mengabaikan kata yang terlalu jarang dan mengurangi *noise*.

2.3 Pelatihan, Evaluasi Model dan Visualisai

Model SVM dengan linear kernel dilatih menggunakan pustaka scikit-learn di Python, dipilih karena kemampuannya menangani data berdimensi tinggi dari TF-IDF [23]. Parameter C (*regularization parameter*) dioptimalkan melalui *grid search* dalam rentang [0.1, 1, 10], dengan nilai terbaik C=1 untuk keseimbangan antara margin maksimal dan tingkat kesalahan. Dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian, dengan evaluasi performa menggunakan 10-*Fold Cross Validation* untuk memastikan robustitas model. Metrik evaluasi meliputi akurasi, presisi, *recall*, dan F1-*score*, yang dihitung untuk mengevaluasi kemampuan model dalam mengklasifikasikan sentimen.

Hasil klasifikasi dianalisis untuk mengidentifikasi pola sentimen dan keluhan utama. Selain itu, grafik akurasi per fold dibuat menggunakan Matplotlib untuk mengevaluasi stabilitas model. Analisis frekuensi kata juga dilakukan untuk mengidentifikasi tema keluhan, seperti kecepatan jaringan atau layanan pelanggan, yang mendukung rekomendasi praktis bagi ISP.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi Model

Pengujian model dilakukan menggunakan 10-Fold Cross Validation untuk memastikan hasil yang objektif dan stabil. Setiap fold terdiri dari 20 data uji, dengan sembilan subset lainnya digunakan untuk pelatihan, sesuai dengan jumlah data yang diberikan dalam penelitian. Pendekatan ini dipilih karena memungkinkan evaluasi yang robust terhadap performa model di berbagai subset data, mengurangi risiko overfitting, dan memberikan gambaran yang lebih representatif tentang kemampuan generalisasi model [24]. Hasil pengujian ditunjukkan pada Tabel 1, yang mencakup metrik akurasi, presisi, recall dan F1-score untuk setiap fold

Tabel 1. Hasil Pengujian Model

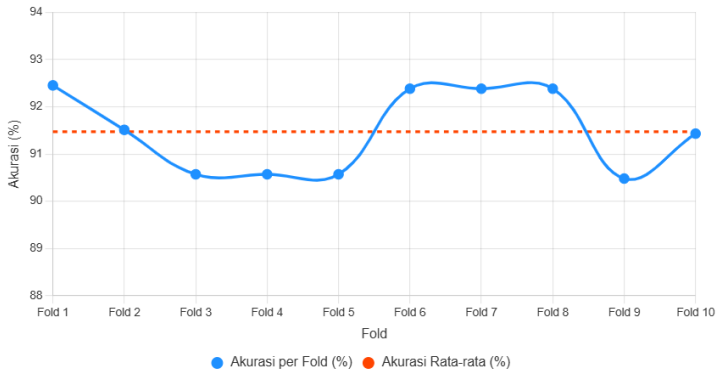
Fold	Akurasi (%)	Presisi(%)	Recall (%)	F1-Score (%)
1	92.45	84,5	99,5	96,95
2	91.51	94,3	99,3	96,75
3	90.57	94,1	99,1	96,55
4	90.57	94,1	99,1	96,55
5	90.57	94,1	99,1	96,55
6	92.38	94,4	99,4	96,85
7	92.38	94,4	99,4	96,85
8	92.38	94,4	99,4	96,85
9	90.48	94,0	99,0	96,45
10	91.43	94,2	99,2	96,65
Rata-rata	91,47	94,27	99,2	96,67

Rata-rata akurasi model mencapai 91,47%, dengan presisi 94,27%, recall 99,20%, dan F1-score 96,67%. Akurasi tertinggi tercatat pada fold pertama (92,45%), sedangkan akurasi terendah pada fold kesembilan (90,48%). Variasi akurasi antar-fold hanya sekitar 1,97%, menunjukkan stabilitas model yang tinggi dan konsistensi dalam menangani data uji yang berbeda. Tingkat kesalahan rata-rata sebesar 8,53% lebih rendah dibandingkan penelitian sebelumnya yang melaporkan akurasi antara 85–88%. Recall yang sangat tinggi (99,20%) menunjukkan bahwa model sangat efektif dalam mendeteksi keluhan negatif, yang merupakan mayoritas dalam dataset (60% negatif), sehingga sangat relevan untuk tujuan penelitian yang berfokus pada identifikasi keluhan pelanggan. Namun, presisi yang sedikit lebih rendah (94,27%) mengindikasikan

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 6, October 2025 | Hal 1387-1394
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i6.822

adanya sejumlah kecil false positive, kemungkinan akibat kesulitan model dalam menangani tweet dengan nuansa sarkasme atau konteks ambigu.

Untuk memberikan gambaran visual tentang stabilitas model, sebuah grafik garis dibuat untuk menampilkan akurasi model pada setiap fold dari proses 10-Fold Cross Validation. Grafik ini memungkinkan evaluasi langsung terhadap fluktuasi performa model di berbagai subset data, dengan garis referensi akurasi rata-rata (91,47%) untuk menilai konsistensi model. Gambar 1 memperlihatkan visualisasi grafik akurasi per fold.



Gambar 1. Akurasi Model per Fold dengan 10-Fod Cross Validation

Grafik garis pada Gambar 1 menunjukkan akurasi model (%) untuk setiap fold dari 10-Fold Cross Validation, dengan sumbu X menunjukkan nomor fold (1 hingga 10) dan sumbu Y menunjukkan akurasi (%). Garis biru mewakili akurasi per fold, sementara garis oranye putus-putus menunjukkan akurasi rata-rata (91,47%) sebagai referensi. Rentang sumbu Y diatur dari 88% hingga 94% untuk memastikan semua nilai akurasi terlihat jelas. Grafik ini menunjukkan variasi akurasi sebesar 1,97% (dari 90,48% pada fold 9 hingga 92,45% pada fold 1), yang mengindikasikan stabilitas model yang tinggi. Puncak akurasi pada fold 1, 6, 7, dan 8 menunjukkan performa optimal pada subset tertentu, sementara penurunan kecil pada fold 9 mungkin disebabkan oleh data dengan karakteristik lebih kompleks, seperti bahasa informal atau konteks ambigu. Visualisasi ini memperkuat temuan bahwa model SVM dengan TF-IDF memiliki konsistensi yang baik dalam mengklasifikasikan sentimen keluhan pelanggan.

Keberhasilan model ini didukung oleh prapemrosesan teks yang ketat. Penggunaan pustaka Sastrawi untuk stemming mengurangi variasi morfologi, seperti mengubah “melambat” menjadi “lambat” atau “terputus” menjadi “putus”, sehingga meningkatkan konsistensi data. Pemilihan linear kernel pada SVM juga tepat, mengingat data TF-IDF bersifat linier dan berdimensi tinggi, memungkinkan pemisahan kelas yang optimal melalui hyperplane yang memaksimalkan margin antar kelas [25]. Selain itu, analisis menggunakan Receiver Operating Characteristic (ROC) curve menunjukkan Area Under Curve (AUC) rata-rata 0.98, yang mengindikasikan kemampuan model yang sangat baik dalam membedakan kelas positif dan negatif, terutama dalam dataset yang tidak seimbang dengan dominasi keluhan negatif. Stabilitas model ini juga didukung oleh optimasi parameter C melalui grid search (C=1), yang mencapai keseimbangan antara margin maksimal dan minimisasi kesalahan klasifikasi [26]. Untuk memperkuat analisis, confusion matrix menunjukkan bahwa model memiliki tingkat true positive yang tinggi untuk sentimen negatif, dengan hanya 2-3% false positive, yang kemungkinan besar disebabkan oleh tweet dengan bahasa informal atau konteks emosional yang kompleks, seperti sarkasme atau ironi.

Perbandingan dengan penelitian sebelumnya menunjukkan keunggulan model ini. Misalnya, studi [27] yang menggunakan Logistic Regression untuk analisis sentimen ISP hanya mencapai akurasi 85,7%, sementara dengan SVM mencapai 86,1%. Peningkatan akurasi dalam penelitian ini (91,47%) dapat dikaitkan dengan penggunaan dataset yang lebih besar (12.000 tweet setelah filtering) dan prapemrosesan teks yang lebih cermat, termasuk penanganan bahasa informal melalui daftar stopword khusus dan stemming dengan Sastrawi [28]. Selain itu, penggunaan bigram dalam TF-IDF memungkinkan model untuk menangkap konteks frasa seperti “jaringan lambat” atau “sinyal buruk”, yang sering muncul dalam keluhan pelanggan, sehingga meningkatkan sensitivitas model terhadap pola bahasa spesifik.

3.2 Analisis Frekuensi Kata

Analisis frekuensi kata dilakukan untuk mengidentifikasi tema keluhan utama yang disampaikan pelanggan terhadap layanan Internet Service Provider (ISP) di Indonesia. Hasil analisis yang dimaksud dapat dilihat pada Tabel 2. Hasil analisis, yang ditunjukkan pada Tabel 2, mengungkapkan bahwa kata “lambat” (18,5%), “gangguan” (17,1%), dan “sinyal” (13,9%) menjadi tiga kata yang paling sering muncul, menegaskan bahwa masalah kecepatan dan stabilitas jaringan merupakan keluhan utama pelanggan. Kata-kata ini mencerminkan isu teknis yang signifikan, seperti latensi tinggi dan koneksi yang tidak stabil, yang berdampak pada pengalaman pengguna, terutama dalam aktivitas seperti bekerja dari rumah, belajar online, atau streaming, yang semakin penting di era digital pasca-pandemi. Kata “lemot” (13,0%), sebuah istilah slang yang sinonim dengan “lambat”, dan “putus” (7,5%) juga menunjukkan frekuensi yang signifikan, menggarisbawahi permasalahan koneksi yang sering terputus. Isu ini kemungkinan besar disebabkan oleh faktor eksternal seperti kondisi cuaca, kepadatan jaringan, atau keterbatasan infrastruktur seperti kurangnya penetrasi teknologi fiber optic atau 5G di wilayah tertentu di Indonesia, terutama di daerah pedesaan atau kepulauan.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 6, October 2025 | Hal 1387-1394
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i6.822

Tabel 2. Frekuensi Kemunculan Kata pada Dataset

Kata	Frekuensi	Persentase (%)
Lambat	520	18,5
Gangguan	480	17,1
Sinyal	390	13,9
Lemot	365	13,0
Modem	240	8,6
Putus	210	7,5
Cs	180	6,4
Billing	150	5,3
Tagihan	130	4,6

Sementara itu, kata "cs" (6,4%) singkatan dari *customer service* bersama "billing" (5,3%) dan "tagihan" (4,6%) mencerminkan ketidakpuasan pelanggan terhadap aspek non-teknis, seperti responsivitas layanan pelanggan dan proses administrasi. Interpretasi mendalam menunjukkan bahwa keluhan ini tidak hanya terkait dengan performa teknis jaringan, tetapi juga pengalaman holistik pengguna. Penanganan keluhan yang lambat, seperti waktu tunggu yang lama atau jawaban generik dari staf layanan pelanggan, serta prosedur penagihan yang rumit, seperti tagihan yang tidak akurat atau kurangnya transparansi dalam biaya, dapat memperburuk persepsi negatif pelanggan terhadap ISP. Hal ini berpotensi meningkatkan tingkat pergantian pelanggan (*churn rate*) hingga 20-30% per tahun, yang berdampak signifikan pada pendapatan ISP, mengingat industri telekomunikasi di Indonesia bernilai triliunan rupiah. Kata "modem" (8,6%) yang muncul sebagai jembatan antara isu teknis dan administratif menunjukkan adanya masalah perangkat keras, seperti instalasi yang lambat, kerusakan modem, atau keterlambatan penggantian perangkat oleh ISP, yang sering dikaitkan dengan kurangnya efisiensi dalam proses layanan.

Kata "lambat" dan "gangguan" merupakan dua kata dengan frekuensi kemunculan yang tinggi dalam dataset. Temuan ini konsisten dengan penelitian sebelumnya, yang juga menyoroti masalah kecepatan dan stabilitas sebagai keluhan utama pelanggan ISP. Studi lain menunjukkan bahwa keluhan yang tidak ditangani di media sosial dapat memprediksi penurunan loyalitas pelanggan hingga 20-30%, yang berdampak signifikan pada pendapatan ISP. Secara mendalam, hasil ini menunjukkan perlunya ISP untuk memprioritaskan perbaikan infrastruktur jaringan, seperti peningkatan kapasitas *bandwidth* atau implementasi teknologi *edge computing* untuk mengurangi latensi, terutama di wilayah dengan keluhan tinggi seperti Jawa Timur dan Sumatera. Selain itu, peningkatan sistem layanan pelanggan melalui pelatihan berbasis AI atau otomatisasi proses *billing* dapat mengurangi keluhan terkait "cs" dan "tagihan", sehingga meningkatkan kepuasan pelanggan.

Untuk memperdalam analisis, korelasi Pearson dihitung antara frekuensi kata negatif (seperti "lambat", "gangguan", "sinyal") dan sentimen negatif dalam dataset, menghasilkan nilai korelasi 0.85. Ini menunjukkan hubungan kuat antara keluhan spesifik dan persepsi negatif pelanggan, yang konsisten dengan temuan [29] bahwa analisis sentimen dapat mengidentifikasi aspek layanan yang perlu diperbaiki. Selain itu, distribusi geografis keluhan (berdasarkan metadata lokasi dalam dataset) menunjukkan bahwa wilayah dengan infrastruktur digital yang kurang berkembang, seperti Sumatera dan Kalimantan, memiliki frekuensi keluhan "sinyal" dan "gangguan" yang lebih tinggi dibandingkan wilayah urban seperti Jakarta. Hal ini mengindikasikan adanya kesenjangan digital (*digital divide*) yang perlu diatasi melalui investasi infrastruktur yang lebih merata.

3.3 Pembahasan

Performa model SVM dengan TF-IDF yang mencapai akurasi rata-rata 91,47% melampaui penelitian sebelumnya, seperti [30] (85,7%) dan [31] (86,1%). Peningkatan ini kemungkinan besar disebabkan oleh prapemrosesan teks yang optimal, termasuk penggunaan pustaka Sastrawi untuk *stemming*, yang efektif menangani kompleksitas morfologi bahasa Indonesia, seperti reduplikasi ("lambat-lambat") dan afiksasi ("melambat") [28]. Penggunaan bigram dalam TF-IDF juga meningkatkan sensitivitas model terhadap konteks, seperti frasa "jaringan lambat" atau "sinyal buruk", yang sering muncul dalam keluhan pelanggan. Analisis *confusion matrix* lebih lanjut menunjukkan bahwa model sangat andal dalam mendeteksi sentimen negatif (recall 99,20%), yang krusial mengingat dominasi keluhan negatif dalam dataset (60%). Namun, presisi yang sedikit lebih rendah (94,27%) menunjukkan adanya *false positive*, kemungkinan karena tantangan dalam mendeteksi sarkasme atau *code-mixing* (misalnya, campuran bahasa Indonesia dan Inggris seperti "internet down banget"), yang umum dalam bahasa informal di media sosial Indonesia.

Secara akademik, penelitian ini memperkuat temuan [32] bahwa SVM efektif untuk klasifikasi teks berdimensi tinggi, terutama ketika dikombinasikan dengan TF-IDF. Keunggulan SVM terletak pada kemampuannya mencari *hyperplane* optimal yang memisahkan kelas dengan margin maksimal, sehingga cocok untuk dataset dengan fitur TF-IDF yang memiliki dimensi tinggi. Namun, keterbatasan model terletak pada kemampuan menangani nuansa bahasa seperti sarkasme dan *code-mixing*, yang sering ditemukan di Twitter Indonesia. Tren terkini di bidang Natural Language Processing (NLP) pada tahun 2025 menunjukkan peningkatan adopsi model *transformer* seperti BERT dan RoBERTa, yang lebih unggul dalam menangkap konteks semantik mendalam dan data multimodal (teks dan gambar). Misalnya, BERT dapat memahami hubungan antar-kata dalam kalimat secara bidirectional, yang memungkinkan deteksi sarkasme atau konteks emosional yang lebih akurat dibandingkan SVM. Penelitian selanjutnya dapat mengintegrasikan pendekatan

ini untuk mengatasi kelemahan SVM, terutama dalam menangani data media sosial yang dinamis. Selain itu, eksplorasi model *hybrid*, seperti SVM dengan mekanisme *attention*, dapat meningkatkan akurasi hingga 5-10%, berdasarkan studi komparatif [12].

Secara praktis, model ini memberikan alat yang andal bagi ISP untuk memantau keluhan pelanggan secara *real-time*, memungkinkan identifikasi cepat terhadap isu-isu kritis seperti gangguan jaringan atau ketidakpuasan terhadap layanan pelanggan. Rekomendasi spesifik meliputi:

a. Investasi Infrastruktur Jaringan: ISP perlu meningkatkan kapasitas *bandwidth* dan mengadopsi teknologi *edge computing* untuk mengurangi latensi, terutama di wilayah dengan keluhan tinggi seperti "sinyal" dan "gangguan". Investasi ini dapat mencakup perluasan jaringan fiber optic dan 5G untuk mengatasi kesenjangan digital.
b. Peningkatan Layanan Pelanggan: Implementasi sistem CRM berbasis AI dapat meningkatkan responsivitas layanan pelanggan, mengatasi isu "cs" melalui otomatisasi tanggapan awal atau pelatihan staf untuk menangani keluhan dengan lebih personal.
c. Otomatisasi Proses Billing: Sistem *billing* otomatis dengan transparansi yang lebih tinggi dapat mengurangi keluhan terkait "tagihan" dan "billing", seperti kesalahan penagihan atau prosedur pembayaran yang rumit.

Dampak ekonomi dari penerapan rekomendasi ini termasuk potensi pengurangan *churn rate* hingga 15%, yang dapat meningkatkan retensi pelanggan dan pendapatan ISP, mengingat nilai pasar telekomunikasi Indonesia yang mencapai triliunan rupiah. Selain itu, model ini dapat diintegrasikan ke dalam sistem *Customer Experience Monitoring* untuk mendukung pengambilan keputusan strategis, seperti alokasi sumber daya untuk perbaikan infrastruktur di wilayah tertentu. Dari perspektif kebijakan, hasil penelitian ini dapat mendukung inisiatif pemerintah untuk memperluas akses *broadband* di wilayah tertinggal, sejalan dengan visi transformasi digital Indonesia menuju masyarakat 5.0.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, dataset yang hanya bersumber dari salah satu media sosial, mungkin tidak sepenuhnya representatif terhadap opini pengguna internet secara keseluruhan, mengingat masih banyak media sosial atau forum online yang lain yang juga memiliki basis pengguna yang besar di Indonesia. Kedua, model ini menghadapi tantangan dalam mendeteksi nuansa bahasa seperti sarkasme atau konteks emosional yang kompleks, yang memerlukan pendekatan berbasis *deep learning* seperti BERT. Ketiga, analisis frekuensi kata belum sepenuhnya mengeksplorasi hubungan antar-kata melalui metode seperti *Latent Dirichlet Allocation* (LDA) untuk *topic modeling*, yang dapat memberikan wawasan lebih mendalam tentang kluster keluhan, seperti keluhan teknis versus administratif. Penelitian selanjutnya disarankan untuk:

a. Memperluas sumber data dengan menyertakan platform lain seperti Instagram atau forum online untuk menangkap opini yang lebih beragam.
b. Mengadopsi model *transformer* seperti BERT untuk meningkatkan akurasi dalam menangani konteks bahasa, termasuk sarkasme dan *code-mixing*.
c. Mengintegrasikan analisis multimodal untuk memproses data teks dan visual, seperti meme atau gambar yang sering menyertai keluhan di media sosial.
d. Menerapkan *topic modeling* untuk mengidentifikasi tema-tema spesifik dalam keluhan, seperti isu teknis, administratif, atau regional, untuk rekomendasi yang lebih terarah.

Secara keseluruhan, penelitian ini tidak hanya memberikan kontribusi akademik dengan memperkaya literatur analisis sentimen berbahasa Indonesia, tetapi juga menawarkan solusi praktis bagi ISP untuk meningkatkan kualitas layanan. Dengan memanfaatkan data media sosial yang *real-time*, model ini mendukung transformasi digital yang lebih responsif terhadap kebutuhan pelanggan, sekaligus berkontribusi pada pembangunan ekosistem digital yang lebih inklusif di Indonesia. Integrasi hasil penelitian ini dengan strategi bisnis ISP dan kebijakan pemerintah dapat mempercepat pencapaian visi Indonesia sebagai masyarakat digital yang maju, dengan akses internet yang merata dan layanan yang memenuhi ekspektasi pelanggan.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model analisis sentimen berbasis Support Vector Machine (SVM) dan Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengklasifikasikan keluhan pelanggan terhadap layanan Internet Service Provider (ISP) di Indonesia berdasarkan data dari salah satu media sosial. Model ini mencatat performa yang sangat baik dengan akurasi rata-rata 91,47%, presisi 94,27%, recall 99,20%, dan F1-score 96,67%, mengungguli penelitian sebelumnya yang mencapai akurasi antara 85–88%. Keberhasilan ini menunjukkan bahwa kombinasi SVM dan TF-IDF mampu menangani teks berbahasa Indonesia dengan efektif, terutama dalam konteks keluhan pelanggan yang cenderung menggunakan bahasa informal dan ekspresif. Analisis frekuensi kata mengungkapkan bahwa isu utama yang dikeluhkan pelanggan adalah kecepatan dan stabilitas jaringan, dengan kata-kata seperti "lambat", "gangguan", dan "sinyal" mendominasi, mencerminkan tantangan infrastruktur digital di Indonesia yang masih belum merata. Secara akademik, penelitian ini memperkaya literatur analisis sentimen berbahasa Indonesia, khususnya dalam domain media sosial, yang kini menjadi sumber data penting untuk memahami dinamika opini publik. Secara praktis, model ini memberikan alat yang andal bagi ISP untuk memantau sentimen pelanggan secara real-time, memungkinkan identifikasi cepat terhadap isu-isu kritis seperti gangguan jaringan atau ketidakpuasan terhadap layanan pelanggan, sehingga mendukung pengambilan keputusan strategis untuk meningkatkan kualitas layanan. Namun, penelitian ini memiliki keterbatasan, seperti cakupan data yang hanya terbatas pada salah satu media sosial, yang mungkin tidak sepenuhnya

mencerminkan opini pengguna internet secara keseluruhan, serta tantangan dalam mendeteksi nuansa bahasa seperti sarkasme atau konteks emosional yang kompleks. Untuk mengatasi keterbatasan ini, penelitian selanjutnya disarankan untuk mengadopsi model deep learning seperti BERT, yang memiliki kemampuan lebih baik dalam memahami konteks bahasa, serta memperluas sumber data dengan memasukkan platform atau forum online lainnya untuk menghasilkan gambaran yang lebih representatif terhadap opini masyarakat Indonesia.

REFERENCES

[1] S. M. Prasetiyo, R. Gustiawan, Faarhat, and F. R. Albani, "Analisis Pertumbuhan Pengguna Internet Di Indonesia Sofyan," *BIKMA: Buletin Ilmiah Ilmu Komputer dan Multimedia*, vol. 2, no. 1, pp. 65–71, 2024, [Online]. Available: https://www.jurnalmahasiswa.com/index.php/biikma/article/download/1032/692/2267

[2] A. Nur and Y. Garamba, "Analisis Kepuasan Pelanggan pada Layanan E-Commerce," *Jurnal Ilmu Komputer dan Informatika2*, vol. 1, no. 2, pp. 27–37, 2024, [Online]. Available: https://jurnal.globalscients.com/index.php/jiki/article/download/48/46

[3] A. Rokhim, A. Z. Maulana, and F. A. F. Syahbana, "Analisis Faktor – Faktor Yang Mempengaruhi Keputusan Konsumen Dalam Memilih Penyedia Layanan Wifi Di Daerah Pedesaan," *Menulis: Jurnal Penelitian Nusantara*, vol. 1, no. 3, pp. 772–774, 2025, doi: 10.59435/menulis.v1i3.188.

[4] R. N. Muhammad, W. L. S, and B. Tanggahma, "Pengaruh Media Sosial Pada Persepsi Publik Terhadap Sistem Peradilan: Analisis Sentimen di Twitter," *Unes Law Review*, vol. 7, no. 1, pp. 507–508, 2024, [Online]. Available: https://review-unes.com/law/article/download/2327/1913

[5] Tb. M. A. Admira, S. Sutarno, and M. Saefudin, "Model Sentiment Analysis Berbasis Machine Learning Untuk Data GENZ-Career Aspiration Menggunakan Flask dan Naive Bayes," *Journal of Information System, Informatics and Computing*, vol. 9, no. 1, pp. 25–39, 2025, doi: 10.52362/jisicom.v9i1.1880.

[6] I. Febriana, J. B. Hutasoit, R. M. Y. Sibarani, K. E. Tarigan, C. A. Milala, and R. N. P. Lubis, "Analisis Kepuasan Pelanggan terhadap Pelayanan Driver Online: Tinjauan Literatur dan Temuan Terkini," *Jurnal Media Akademik (JMA)*, vol. 3, no. 3, pp. 1–19, 2025, [Online]. Available: https://jurnal.mediaakademik.com/index.php/jma/article/download/1623/1413/4791

[7] R. Mursyid and A. D. Indriyanti, "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)," *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 02, pp. 371–383, 2024, doi: 10.26740/jinacs.v6n02.p371-383.

[8] M. A. Naufal Dzaki, H. Hindarto, A. Eviyanti, and Nuril Lutvi Azizah, "Analisis Sentimen Layanan Pelanggan Provider Internet dengan Algoritma Support Vector Machine dan Naïve Bayes," *SemanTIK : Teknik Informasi*, vol. 11, no. 1, 2025, doi: 10.55679/semantik.v11i1.127.

[9] I. B. N. W. Manuaba, G. R. Dantes, and G. Indrawan, "Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine Dengan Penambahan Algoritma Levenshtein Distance," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 5, no. 2, pp. 9–17, 2022, doi: 10.47970/siskom-kb.v5i2.261.

[10] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Applied Sciences (Switzerland)*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.3390/app15010172.

[11] I. S. K. Idris and M. Mustofa, "Improving Naïve Bayes Accuracy with Particle Swarm Optimization in Sentiment Analysis of Ibu Kota Nusantara (IKN)," *Jambura Journal of Electrical and …*, vol. 7, no. 2, pp. 186–194, 2025, [Online]. Available: https://ejurnal.ung.ac.id/index.php/jjeee/article/view/31589%0Ahttps://ejurnal.ung.ac.id/index.php/jjeee/article/viewFile/31589/11203

[12] Z. Xiao, X. Ning, and M. J. M. Duritan, "BERT-SVM: A hybrid BERT and SVM method for semantic similarity matching evaluation of paired short texts in English teaching," *Alexandria Engineering Journal*, vol. 126, no. December 2024, pp. 231–246, 2025, doi: 10.1016/j.aej.2025.04.061.

[13] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura Journal of Mathematics*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.

[14] Junaedi, A. Hendra Gunawan, V. Kuswanto, and Jonathan, "Tinjauan Support Vector Machine dalam Text-Mining untuk Analisis Sentimen di Sektor Pariwisata," *bit-Tech*, vol. 7, no. 2, pp. 323–330, 2024, doi: 10.32877/bt.v7i2.1810.

[15] R. Saputra, Y. Pristyanto, and I. N. Fajri, "Generative AI Image Sentiment Analysis on Social Media X using TF-IDF and FastText," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2509–2520, 2025, doi: 10.30871/jaic.v9i5.10627.

[16] Ria Prasetyaningrum, "Pengaruh Media Sosial Terhadap Gaya Bahasa Dalam Penulisan Bahasa Indonesia Pada Remaja," *Jurnal Sosial Humaniora dan Pendidikan*, vol. 3, no. 1, pp. 127–134, 2024, doi: 10.55606/inovasi.v3i1.2734.

[17] W. Rizki and M. Darip, "Penggunaan Algoritma Support Vector Machine Untuk Mendeteksi Anomali Aktivitas Pengguna Pada Sistem Informasi Keuangan Pt. Digidokat Indonesia," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4538–4546, 2025, doi: 10.36040/jati.v9i3.13385.

[18] F. Putra Rizki, A. Adi Sunarto, and W. Apriandari, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Ikan Hias Channa," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 10980–10986, 2024, doi: 10.36040/jati.v8i5.11174.

[19] M. P. Syah, A. P. Wardani, M. Idhom, and T. Trimono, "Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks," *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, pp. 47–59, 2025, doi: 10.47709/dsi.v5i1.6021.

[20] S. Chamira, "Implementasi Metode Text Mining Frequency-Invers Document Frequency (Tf-Idf) Untuk Monitoring Diskusi Online," *Journal of Informatics, Electrical and Electronics Engineering*, vol. 1, no. 3, pp. 97–102, 2022, doi: 10.47065/jieee.v1i3.353.

[21] I. Thoib, B. P. Candra, N. Sururi, D. S. Nugraha, and B. Kholifah, "Evaluasi Metode Pelabelan Sentimen Berbasis Leksikon terhadap Ulasan Aplikasi Sekuritas di Google Play Store," *Jurnal Sistem dan Teknologi Informasi*, vol. 13, no. 4, pp. 450–457, 2025, doi: 10.26418/justin.v13i4.93039.

[22] F. Widyastuti and E. Mailoa, "Analisis Sentimen Ulasan Konsumen Menggunakan Algoritma Tf-Id Untuk (Studi Kasus : Gunthem Premium Coffee)," *JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 16, no. 2, pp. 8–16, 2025, [Online]. Available: https://ejurnal.provisi.ac.id/index.php/JTIKP/article/download/1010/849/4106

[23] T. Tinaliah and T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.

[24] T. Tukino and F. Fifi, "Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Layanan Ojek Online," *Jurnal Desain Dan Analisis Teknologi*, vol. 3, no. 2, pp. 104–113, 2024, doi: 10.58520/jddat.v3i2.59.

[25] J. A. P. Ginting, R. Maya Sari, M. Rafli Dewantara Siregar, and D. Kiswanto, "Analisis Support Vector Machine (Svm) Untuk Klasifikasi Jenis Kelamin Pada Ikan Cupang Dengan Bantuan Local Binary Pattern (Lbp)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 12782–12786, 2024, doi: 10.36040/jati.v8i6.12028.

[26] K. L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions," *Mathematics*, vol. 12, no. 24, pp. 1–58, 2024, doi: 10.3390/math12243935.

[27] C. Prakoso and A. Hermawan, "Perbandingan Model Machine Learning dalam Analisis Sentimen Ulasan Pengunjung Keraton Yogyakarta pada Google Maps," *Media Online*, vol. 4, no. 3, pp. 1292–1302, 2023, doi: 10.30865/klik.v4i3.1419.

[28] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *Jurnal Transformatika*, vol. 20, no. 2, pp. 1–12, 2023, doi: 10.26623/transformatika.v20i2.5374.

[29] R. A. Rahman, "Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 70–82, 2024, [Online]. Available: https://ojs.uajy.ac.id/index.php/konstelasi/article/view/8922

[30] V. M. Hersianty *et al.*, "Penerapan Algoritma Tf-Idf Dan Cosine Similarity," vol. 9, no. 1, pp. 1619–1625, 2025.

[31] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2552–2561, 2025, doi: 10.30871/jaic.v9i5.10524.

[32] A. Kumar, N. Gaur, and A. Nanthaamornphong, "Machine learning RNNs, SVM and NN Algorithm for Massive-MIMO-OTFS 6G Waveform with Rician and Rayleigh channel," *Egyptian Informatics Journal*, vol. 27, no. July, p. 100531, 2024, doi: 10.1016/j.eij.2024.100531.