

ANALISIS SENTIMEN ULASAN APLIKASI DANA DI PLAY STORE MENGGUNAKAN SUPPORT VECTOR MACHINE

Ilham Tri Cahya¹, Ade Irma², Irfan Ali³, Odi Nurdiawan⁴, Martanto⁵

STIMIK IKMI CIREBON, Jl. Perjuangan No. 10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat 45135

Keywords:

Analisis Sentimen, Support Vector Machine, Text Mining, Google Play Store, DANA.

Correspondent Email:

ilhamtrichahya05@gmail.com

Abstrak. Penelitian ini bertujuan menganalisis sentimen pengguna aplikasi DANA dari Google Play Store menggunakan Support Vector Machine (SVM). Data sebanyak 10.000 ulasan dikumpulkan melalui web scraping, menghasilkan 6.597 data unik setelah deduplikasi. Tahapan pra-pemrosesan meliputi cleaning, tokenisasi, stopword removal, dan stemming. Pelabelan dilakukan dengan pendekatan lexicon-based ke dalam kategori positif, negatif, dan netral. Hasil pengujian menunjukkan skenario pembagian data 80:20 menghasilkan akurasi tertinggi sebesar 94,2%, dibandingkan skenario 30:70 (90,5%) dan 10:90 (83,7%). Distribusi sentimen didominasi oleh sentimen positif (39,93%), diikuti netral (36,89%) dan negatif (23,18%). Temuan ini menunjukkan efektivitas SVM dalam klasifikasi teks dimensi tinggi untuk mendukung pengembangan fitur aplikasi fintech.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. This study aims to analyze the sentiment of DANA app users from the Google Play Store using Support Vector Machine (SVM). Data of 10,000 reviews were collected through web scraping, resulting in 6,597 unique data points after deduplication. Pre-processing stages included cleaning, tokenization, stopword removal, and stemming. Labeling was performed using a lexicon-based approach into positive, negative, and neutral categories. Test results showed that the 80:20 data split scenario produced the highest accuracy of 94.2%, compared to the 30:70 (90.5%) and 10:90 (83.7%) scenarios. The sentiment distribution was dominated by positive sentiment (39.93%), followed by neutral (36.89%) and negative (23.18%). These findings demonstrate the effectiveness of SVM in high-dimensional text classification to support the development of fintech application features

1. PENDAHULUAN

Perkembangan dompet digital di Indonesia meningkatkan ketergantungan pada transaksi mobile, menjadikan analisis sentimen instrumen vital untuk memahami persepsi pengguna. *State of the art* penelitian ini mencakup tantangan keragaman bahasa dan ambiguitas semantik pada ulasan aplikasi finansial. Meskipun model *deep learning* populer, algoritma tradisional seperti *Support*

Vector Machine (SVM) tetap relevan karena efektivitasnya pada dataset berlabel kecil dan efisiensi komputasi. *Gap analysis* menunjukkan adanya kebutuhan akan teknik pra-proses yang komprehensif untuk menangani variasi linguistik pada ulasan aplikasi DANA. Tujuan penelitian ini adalah mengevaluasi kinerja SVM dalam mengklasifikasikan sentimen pengguna guna memberikan wawasan strategis bagi pengembangan aplikasi.

Analisis sentimen merupakan proses penting dalam memahami persepsi pengguna terhadap suatu aplikasi, terutama dalam konteks aplikasi yang tengah berkembang pesat namun masih memiliki tantangan dalam mengenalkan layanannya kepada khalayak luas.[1]

Ketergantungan pada transaksi *mobile* di Indonesia memosisikan analisis sentimen sebagai instrumen vital untuk memetakan persepsi pengguna. Namun, implementasinya menghadapi hambatan teknis yang signifikan. Penggunaan BERTopic pada ulasan aplikasi dompet digital sering memicu tumpang tindih topik dan ambiguitas semantik yang menyulitkan ekstraksi informasi [2]. Dari sisi pemodelan, *Support Vector Machine* (SVM) menunjukkan kerentanan terhadap ketidakseimbangan kelas, yang berdampak pada rendahnya sensitivitas terhadap keluhan spesifik terkait fitur minoritas [3]. Sementara itu, model *deep learning* masih terkendala oleh rendahnya interpretabilitas dan kelangkaan data berlabel di domain *fintech* [4].

Variasi linguistik dan konteks layanan yang beragam, mulai dari kecepatan transaksi hingga kualitas layanan pelanggan, menuntut pendekatan praproses yang lebih komprehensif [5]. Di luar tantangan tersebut, ulasan pengguna memiliki peran strategis dalam siklus pengembangan produk dengan mengungkap isu kegunaan, keandalan, dan keamanan secara langsung [2]. Pola sentimen yang dihasilkan mampu mengidentifikasi kelemahan stabilitas sistem [5] serta mendorong perbaikan desain antarmuka melalui umpan balik konsumen yang [6]. Kepuasan pengguna terbukti berkorelasi kuat dengan inovasi desain yang didorong oleh masukan langsung [7], yang pada akhirnya memperkuat pengelolaan kepercayaan pengguna pada platform digital [8]. Integrasi analisis sentimen yang adaptif menjadi kunci dalam menghasilkan wawasan strategis bagi ekosistem dompet digital yang berkelanjutan.

2. TINJAUAN PUSTAKA

Bagian ini merangkum teori *Text Mining* dan algoritma *Support Vector Machine*. SVM bekerja dengan mencari *hyperplane* optimal untuk memisahkan kelas data dengan margin maksimal. Penelitian terkait menunjukkan SVM mampu mengungguli model kompleks dalam tugas analisis sentimen tingkat lanjut pada dataset tertentu.

3. METODE PENELITIAN

Penelitian ini dilakukan secara sistematis melalui lima tahapan utama: pengumpulan data, pra-pemrosesan, pelabelan sentimen, pemodelan menggunakan *Support Vector Machine* (SVM), dan evaluasi model. Alur penelitian ini dirancang untuk mengubah data ulasan tidak terstruktur menjadi informasi sentimen yang akurat.

3.1. Pengumpulan Data

Data dikumpulkan dari Google Play Store menggunakan teknik *web scraping* dengan *library* *google-play-scraper* pada aplikasi DANA (*app_id* = 'id.dana'). Sebanyak 10.000 ulasan terbaru diambil dan disimpan dalam format CSV dengan atribut *Review ID*, *Username*, *Rating*, *Review Text*, dan *Date*.

1. Pembagian Data: Data fitur dan label dibagi menjadi Data Latih (*Training*) dan Data Uji (*Testing*) untuk mengevaluasi kinerja model secara objektif.
2. Rasio Pembagian: Dalam penelitian ini digunakan beberapa skenario pembagian data untuk menguji pengaruh proporsi data pelatihan terhadap kinerja model klasifikasi. Pembagian data dilakukan dengan tiga rasio berbeda, yaitu 80% data latih dan 20% data uji, 30% data latih dan 70% data uji, serta 10% data latih dan 90% data uji. Variasi rasio pembagian data ini bertujuan untuk menganalisis bagaimana perbedaan jumlah data pelatihan mempengaruhi kemampuan model dalam mempelajari pola data serta menghasilkan prediksi sentimen yang akurat pada data pengujian.
3. Proses Pelatihan: Model SVM diinisialisasi dengan *kernel* linier dan dilatih menggunakan Data Latih. Proses ini memungkinkan model untuk mempelajari batas-batas pemisah sentimen berdasarkan fitur vektor yang telah disiapkan.

3.1.1. Text Processing

Tahap ini bertujuan membersihkan derau (*noise*) pada teks agar siap diproses oleh algoritma. Langkah-langkah yang dilakukan meliputi:

1. Deduplikasi

Menghapus ulasan ganda sehingga diperoleh 6.597 entri unik

Tabel 3. 1 Deduplikasi

Tahapan Data	Jumlah Data (N)	Keterangan
Data Hasil <i>Scraping</i>	10.000	Total ulasan yang berhasil diambil.
Data Setelah Deduplikasi	6.597	Data unik yang siap untuk pra-pemrosesan.

2. **Cleaning & Case Folding:**

Menghapus tanda baca, angka, dan mengubah teks menjadi huruf kecil (*lowercase*).

Tabel 3. 2 Cleaning

Indeks Kolom	Nama Kolom	Non-Null Count	Tipe Data
0	Review ID	10000	object
1	Username	10000	object
2	Rating	10000	int64
3	Review Text	10000	object
4	Date	10000	object

3. **Tokenization**

tahap fundamental dalam pra-pemrosesan teks yang bertujuan untuk memecah untaian teks yang telah dibersihkan menjadi unit-unit yang lebih kecil dan bermakna, yang dikenal sebagai token (umumnya berupa kata). Proses ini sangat penting karena algoritma machine learning seperti SVM memerlukan input dalam bentuk unit kata yang terpisah, bukan dalam format kalimat atau paragraf mentah. Selain itu, tokenisasi berfungsi sebagai jembatan yang mengubah data teks menjadi struktur data berbasis daftar (list) yang siap untuk manipulasi linguistik selanjutnya, seperti stopword removal dan stemming. Dalam penelitian ini, proses tokenisasi diterapkan pada kolom teks ulasan yang telah melalui tahap normalisasi.

4. **Stopword Removal**

Menghilangkan kata umum yang tidak memiliki bobot sentimen (seperti "dan", "yang", "di").

5. **Stemming:**

Mengubah kata berimbuhan menjadi kata dasar menggunakan *library* Sastrawi untuk mengurangi dimensi fitur. Stemming adalah proses reduksi morfologis yang bertujuan untuk mengubah kata-kata berimbuhan

menjadi bentuk kata dasarnya (root word) dan merupakan tahap terakhir dalam normalisasi teks sebelum pembobotan fitur. Tujuan utama stemming adalah menyamakan berbagai bentuk kata yang memiliki akar makna yang sama, seperti "mengevaluasi", "dievaluasi", dan "evaluasi" menjadi "evaluasi", sehingga dapat mengendalikan ukuran kosakata dan mengurangi variasi kata [9]. Dengan demikian, stemming berkontribusi pada peningkatan generalisasi model karena mampu mengurangi ukuran vocabulary dan kompleksitas fitur, sehingga model SVM dapat bekerja lebih efektif [10]. Dalam penelitian ini, proses stemming dilakukan menggunakan Sastrawi, yaitu library berbasis aturan untuk bahasa Indonesia, yang diterapkan pada token hasil stopword removal, dengan hasil akhir yang dirangkum pada Tabel 3.1 sebagai teks yang siap untuk ekstraksi fitur. Secara lebih luas, stemming merupakan teknik penting dalam preprocessing karena mampu mengonversi berbagai bentuk morfologis kata ke dalam bentuk yang lebih konsisten, sehingga mengurangi sparsitas kosakata dan meningkatkan keseragaman fitur, yang pada akhirnya membantu model dalam menangkap pola semantik secara lebih stabil dalam analisis sentimen. Penelitian oleh [11] menunjukkan bahwa penerapan stemming pada teks bahasa Indonesia nonformal secara signifikan meningkatkan akurasi klasifikasi menggunakan SVM dibandingkan data tanpa stemming, menegaskan manfaat reduksi variasi morfologis bagi generalisasi model. Selain itu, dalam pendekatan berbasis word embedding, stemming membantu menghasilkan representasi vektor yang lebih bersih dengan mengurangi derau infleksi tanpa menghilangkan makna inti kata [12], sehingga model menjadi lebih efisien dalam mempelajari hubungan semantik. Kajian literatur terkini juga menegaskan bahwa stemming berperan dalam meningkatkan robustitas model

dengan mengurangi risiko overfitting terhadap variasi morfologis, sehingga performa tetap stabil pada berbagai domain [13]. Penerapan stemming juga terbukti menguntungkan dalam studi analisis sentimen ulasan aplikasi mobile sebagai bagian penting dari pipeline preprocessing untuk meningkatkan performa klasifikasi dan generalisasi model [14]. Selain itu, menurut [15], stemming mampu menyeimbangkan reduksi derau dan generalisasi fitur sehingga model yang lebih sederhana seperti SVM dapat menghasilkan batas keputusan yang lebih jelas karena fitur berasal dari bentuk kata yang terstandardisasi.

3.1.2. Pelebelan data (Lexicon-based)

Data diberi label secara otomatis menggunakan pendekatan *Lexicon-Based Sentiment Analysis*. Setiap token dicocokkan dengan kamus sentimen positif dan negatif untuk menghitung skor akhir (*sentiment score*).

Tabel 3. 3 pelebelan lexicon-based

Kondisi Skor Sentimen	Kriteria	Label Sentimen
Skor Positif Lebih Dominan	Skor positif > Skor negatif (Skor Akhir > 0)	Positif
Skor Negatif Lebih Dominan	Skor negatif > Skor positif (Skor Akhir < 0)	Negatif
Skor Seimbang	Skor positif = Skor negatif (Skor Akhir = 0)	Netral

3.1.3. Pemodelan Support Vector Machine

Tahap ini mencakup pemilihan metode representasi fitur dan algoritma klasifikasi yang digunakan untuk membangun dan melatih model analisis sentimen.

A. Pemilihan Algoritma

1. Representasi Fitur (Ekstraksi Fitur)

Teks yang telah melalui proses *stemming* harus diubah menjadi format numerik agar dapat diproses oleh model ML. Proses ini dikenal sebagai *Vectorization*. Dalam penelitian ini, *Vectorization* mengonversi *token-token* yang telah dibersihkan menjadi matriks hitungan (*count*) atau frekuensi, yang kemudian akan menjadi *input* (fitur) bagi model klasifikasi.

2. Pemilihan Algoritma Klasifikasi

Algoritma yang dipilih untuk klasifikasi sentimen adalah Support Vector Machine (SVM). Prinsip Kerja: SVM bekerja dengan mencari *hyperplane* optimal yang berfungsi sebagai batas pemisah terbaik antara kelas sentimen (Positif, Negatif, atau Netral) di ruang fitur. Efektivitas: SVM dikenal efektif dalam memecahkan masalah klasifikasi data teks dengan dimensi tinggi dan telah banyak digunakan dalam studi analisis sentimen [16].

B. Evaluasi Model

Dalam konteks analisis sentimen, evaluasi kinerja algoritma machine learning (ML) menjadi sangat penting untuk menentukan efektivitas model klasifikasi teks. Metode evaluasi ini umumnya mengukur seberapa baik model dapat memprediksi label sentimen (positif, negatif, netral) berdasarkan data teks yang tersedia. Metode yang paling sering digunakan adalah akurasi (*accuracy*), presisi (*precision*), *recall*, dan *F1-score*, yang masing-masing memberikan informasi berbeda terkait performa model.

a. Accuracy Akurasi

mengukur proporsi prediksi yang benar dibanding total data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Dalam analisis sentimen, True Positive (TP) adalah jumlah ulasan yang benar-benar memiliki sentimen positif dan berhasil diklasifikasikan dengan tepat sebagai positif oleh model. True Negative (TN) menggambarkan jumlah ulasan yang sebenarnya bersentimen negatif dan berhasil diprediksi dengan benar sebagai negatif. Sebaliknya, False Positive (FP) mencerminkan jumlah ulasan yang sebenarnya negatif tetapi secara keliru diprediksi sebagai positif, yang disebut juga sebagai *Kesalahan Tipe I*. Sementara itu, False Negative (FN) adalah kondisi ketika ulasan yang seharusnya positif justru salah diklasifikasikan sebagai negatif oleh model, yang dikenal sebagai *Kesalahan Tipe II*. Keempat metrik ini menjadi dasar dalam menghitung akurasi, *precision*, *recall*, dan *F1-score*, sehingga sangat penting untuk mengevaluasi kualitas sistem klasifikasi sentimen.

Namun, akurasi saja tidak cukup untuk menilai performa model pada dataset yang memiliki

distribusi kelas tidak seimbang. Oleh karena itu, metrik Precision dan Recall digunakan untuk memberikan pemahaman yang lebih komprehensif.

b. Precision (presisi)

Presisi mengukur seberapa tepat model dalam memprediksi kelas positif, atau seberapa banyak dari seluruh prediksi positif yang benar-benar

Kelas Sentimen	Jumlah Ulasan	Persentase (%)
Positif	3.993	39.93%
Netral	3.689	36.89%
Negatif	2.318	23.18%
Total	10.000	100.00%

positif

$$\text{Precision} = \frac{TP}{TP + FP}$$

Presisi tinggi menunjukkan bahwa model jarang melakukan kesalahan dengan memprediksi data negatif sebagai positif (*false positive*).

c. Recall

Recall, juga dikenal sebagai sensitivitas, mengukur kemampuan model dalam mengenali semua data yang benar-benar positif

$$\text{Recall} = \frac{TP}{TP + FN}$$

Nilai recall yang tinggi menunjukkan bahwa model mampu menangkap sebagian besar ulasan positif yang sebenarnya ada di dataset.

d. F1-Score

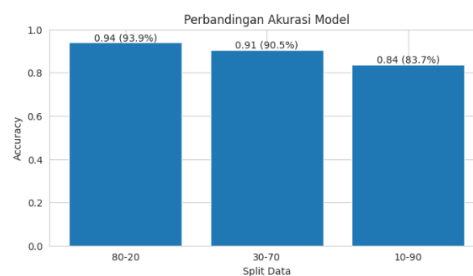
$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-score sangat berguna ketika distribusi kelas tidak seimbang, karena memberikan ukuran kinerja model yang lebih representatif dibandingkan hanya melihat akurasi.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan dari proses analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) serta pembahasan mendalam mengenai efektivitas model dan pola opini pengguna aplikasi DANA.

1. Pra-pemrosesan dan Pelabelan Dataset awal yang berjumlah 10.000 ulasan berhasil direduksi menjadi 6.597 ulasan unik melalui tahap deduplikasi. Proses pelabelan otomatis menggunakan pendekatan *lexicon-based* menunjukkan distribusi sentimen yang bervariasi. Dari total data unik, sentimen positif mendominasi dengan persentase sebesar 39,93%, diikuti oleh sentimen netral sebesar 36,89%, dan sentimen negatif sebesar 23,18%. Dominasi sentimen positif menunjukkan bahwa secara umum pengguna merasa terbantu dengan fitur-fitur yang ditawarkan oleh aplikasi
2. Model SVM Pengujian dilakukan dengan membandingkan tiga skenario pembagian data untuk melihat pengaruh volume data latih terhadap akurasi prediksi. Hasil perbandingan evaluasi Model dalam gambar berikut.



Gambar 4. 1 Perbandingan

Pengujian model dilakukan untuk mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan pengguna aplikasi DANA. Pengujian ini menggunakan tiga skenario pembagian data pelatihan dan data pengujian, yaitu 80:20, 30:70, dan 10:90. Berdasarkan hasil pengujian, skenario pembagian data 80:20 menghasilkan tingkat akurasi tertinggi yaitu 0,94 (93,9%), yang menunjukkan bahwa model mampu melakukan klasifikasi sentimen dengan tingkat ketepatan yang sangat baik ketika didukung oleh jumlah data pelatihan yang lebih besar.

Pada skenario pembagian data 30:70, model memperoleh tingkat akurasi sebesar 0,91 (90,5%). Meskipun masih menunjukkan performa yang baik, nilai ini mengalami sedikit penurunan dibandingkan dengan skenario sebelumnya. Penurunan ini disebabkan oleh berkurangnya jumlah data pelatihan yang digunakan dalam proses pembentukan model, sehingga kemampuan model dalam mengenali

pola dan karakteristik sentimen pada data ulasan menjadi sedikit terbatas.

Sementara itu, pada skenario pembagian data 10:90, akurasi model menurun menjadi 0,84 (83,7%), yang merupakan nilai terendah di antara ketiga skenario pengujian. Hal ini menunjukkan bahwa jumlah data pelatihan yang terlalu sedikit dapat memengaruhi kemampuan model dalam mempelajari pola data secara optimal. Dengan demikian, dapat disimpulkan bahwa pembagian data 80:20 merupakan konfigurasi yang paling optimal dalam penelitian ini karena menghasilkan tingkat akurasi model yang paling tinggi.

3. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan dalam penelitian mengenai analisis sentimen ulasan aplikasi DANA menggunakan Support Vector Machine (SVM), diperoleh beberapa kesimpulan yang menjawab tujuan penelitian: Berdasarkan hasil penelitian dan pembahasan mengenai analisis sentimen ulasan aplikasi DANA menggunakan Support Vector Machine (SVM), diperoleh tiga kesimpulan utama yang menjawab secara langsung Rumusan Masalah yang diajukan:

1. Penerapan tahapan pra-pemrosesan Efektif dalam meningkatkan kualitas data tahapan *preprocessing* yang meliputi *cleaning text*, *tokenisasi*, *stopword removal*, dan *stemming* terbukti mampu mengatasi keragaman linguistik pada data ulasan yang bersifat tidak terstruktur. Proses ini berhasil menghilangkan noise, menstandarisasi teks, serta mengurangi kompleksitas data sehingga menghasilkan dataset yang lebih bersih dan siap digunakan dalam proses klasifikasi. Dengan demikian, tantangan linguistik dalam analisis sentimen dapat ditangani dengan baik.
2. Analisis sentimen mampu mengidentifikasi pola kepuasan pengguna dan aspek layanan Hasil pelabelan menunjukkan bahwa sentimen Positif (39,93%) dan Netral (36,89%) mendominasi dengan total 76,82%, yang mengindikasikan bahwa sebagian besar pengguna memiliki pengalaman yang cukup baik terhadap

aplikasi. Namun, keberadaan sentimen Negatif (23,18%) menunjukkan adanya permasalahan yang masih perlu diperhatikan, seperti kendala pada stabilitas aplikasi, transaksi, dan keamanan akun. Hal ini menunjukkan bahwa analisis sentimen mampu memberikan gambaran yang jelas mengenai tingkat kepuasan pengguna serta aspek layanan yang perlu ditingkatkan.

3. Metode Support Vector Machine (SVM) Terbukti efektif dengan performa terbaik pada pembagian data 80:20 Metode Support Vector Machine menunjukkan kinerja yang sangat baik dalam klasifikasi sentimen dengan nilai akurasi tertinggi sebesar 94,2% dan F1-score sebesar 0,942 pada skenario pembagian data 80:20. Sementara itu, pada skenario 30:70 diperoleh akurasi sebesar 90,5%, dan pada skenario 10:90 sebesar 83,7%. Hasil ini menunjukkan bahwa skenario 80:20 merupakan yang paling unggul karena memiliki proporsi data latih yang lebih besar, sehingga model mampu mempelajari pola data secara lebih optimal dan menghasilkan prediksi yang lebih akurat dibandingkan skenario lainnya. Dengan demikian, semakin besar data latih yang digunakan, maka semakin baik performa model dalam melakukan klasifikasi sentimen.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Analisis Sentimen Pada Ulasan Aplikasi Astro - Groceries In Minutes Di Google Play Store Menggunakan Metode Bidirectional Encoder Representation From Transformers (Bert) Ibrahim A, Mustikasari M, Bastian *Jurnal Informatika dan Teknik Elektro Terapan (2025) 13(1)*
- [2] Exploring user sentiments and adoption barriers in e-wallet services in Indonesia using BERTopic on Play Store reviews BT - IEEE International Conference on Emerging Research Areas

- (ICERA)Hanifah A, Saputra M, Fa'rifah R(2025)
- [3] Aspect-based sentiment analysis on user perceptions of OVO using Latent Dirichlet Allocation and Support Vector MachineAprilia E, Arifiyanti A, Sembilu NAVITEC (2025) 7(2)
- [4] Fintech sentiment analysis using deep learning approaches: A survey BT - IEEE AI Research Conference (AIRC)Anis S, Morsey M, Aref M(2024)
- [5] Assessment of mobile payment service based on user review in IndonesiaDewi I, Nurcahyo R, FarizalCommunications in Humanities and Social Sciences (2024) 4(2) 59-65
- [6] QUser satisfaction analysis of financial technology applications using the End User Computing Satisfaction modelErnawati I, Faroqi A, Aulia Vbit-Tech (2025) 8(1)
- [7] Sentiment analysis in finance using transformer models: A comparative studyGao Z, Liu Y, Chen JApplied Sciences (2023) 13(4) 2142
- [8] Analysis of user reviews on the use of fintech Dana SyariahLubis FInternational Journal of Science and Society (2023) 5(2) 678-685
- [9] Text stemming and lemmatization of regional languages: A systematic literature reviewAbidin ZJISEBI (2024) 8(1) 50341
- [10] Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiersSiino M, Tinnirello I, La Cascia MInformation Systems (2024) 121 102342
- [11] Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversationRianto M, Mutiara A, Wibowo EJournal of Big Data (2021) 8 26
- [12] Sentiment analysis of Indonesian datasets based on a hybrid modelLin C, collaboratorsJournal of Big Data (2023) 10 782
- [13] Generalizing sentiment analysis: a review of progress, challenges, and emerging directionsAlahmadi K, Alharbi S, Chen JSocial Network Analysis and Mining (2025) 15(1) 45
- [14] An improved sentiment classification approach for measuring user satisfaction toward governmental services' mobile apps using machine learning methods with feature engineering and SMOTE techniqueHadwan M, Al-Sarem M, Al-Hagery MApplied Sciences (2022) 12(11) 5547
- [15] Sentiment analysis: A survey on design framsework, challenges, and applicationsBordoloi MFuture Internet (2023) 15(5)