



INNOVATIVE: Journal Of Social Science Research

Volume 5 Nomor 1 Tahun 2025 Page 2348-2357

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Analisis Sentimen Terhadap Review Produk Lazada Indonesia Menggunakan Random Forest Classifier

Widiyanto Hadi^{1✉}, Tommy Dwi Putra², Dewi Oktafiani³

STMIK AMIKOM Surakarta

Email: widiyantohadi@gmail.com^{1✉}

Abstrak

Penelitian ini menggunakan random forest classifier untuk analisis sentimen review produk harddisk di Lazada Indonesia dengan akurasi data latih sebesar 52,87% dan data uji sebesar 83,34%. Fokus penelitian ini adalah memahami opini konsumen terhadap produk elektronik melalui analisis, setelah melakukan pra-pemrosesan data yang meliputi pembersihan, pengemasan case, tanda baca, analisis dan pembobotan kata. Hasil penelitian diharapkan dapat memberikan pemahaman tentang penerapan algoritma Random Forest terkait analisis sentiment serta juga tentang kepuasan pelanggan, kelebihan dan kekurangan produk, serta meningkatkan pemahaman tentang persepsi pelanggan dalam e-commerce. Performa model diukur menggunakan teknik 10 fold cross validation untuk melihat sejauh mana performa dari algoritma Random Forest.

Kata Kunci : *Analisis Sentimen, E-Commerce, Lazada, Random Forest, 10 Fold Cross Validation*

Abstract

This study uses a random forest classifier for sentiment analysis of hard disk product reviews on Lazada Indonesia with an accuracy of 52.87% in training data and 83.34% in test data. The focus of this research is to understand consumer opinions on electronic products through analysis, after pre-processing data which includes cleaning, case packaging, punctuation, analysis and word weighting. The results of the study are expected to provide an understanding of the application of the Random Forest algorithm related to sentiment analysis as well as customer satisfaction, product advantages and disadvantages, as well as increase understanding of customer perception in e-commerce. The performance of the model was measured using the 10-fold cross validation technique to see the extent of the performance of the Random Forest algorithm.

Keyword: *Sentiment Analysis, E-Commerce, Lazada, Random Forest, 10 Fold Cross Validation*

PENDAHULUAN

Dengan semakin berkembangnya teknologi, media online dan media elektronik pun ikut berkembang. Media online tidak hanya sekedar untuk mencari informasi tetapi juga dapat dijadikan sebagai wadah bisnis atau biasa disebut dengan e-commerce. Mengubah sistem penjualan dan pembelian tradisional menjadi media online menghemat waktu dan biaya serta dapat mempercepat dan meningkatkan penjualan. Kehadiran e-commerce sebagai alat penjualan online dapat membawa manfaat baik bagi penjual maupun pembeli.

E-commerce telah melahirkan banyak marketplace di Indonesia seperti Tokopedia, Shopee, Lazada dan masih banyak lagi. Marketplace merupakan model bisnis yang memberikan kemudahan bagi pedagang untuk berjualan secara online. Konsep dasarnya mirip dengan pasar konvensional di mana penjual yang berbeda menjual produknya di satu lingkungan, namun pasar tersebut memfasilitasi perdagangan online dengan menggunakan media online. Dalam hal ini, pasar menyediakan berbagai promosi pemasaran kepada konsumen dengan bantuan Internet, produk yang dibeli dan dijual dapat dilihat kapan saja dan di mana saja, kecuali lokasi penjual, dan pelanggan juga dapat datang ke mana saja. negara lainnya .

Statistik menunjukkan bahwa e-commerce di Indonesia terus berkembang, dengan 139 juta pengguna pada tahun 2017 dan diperkirakan mencapai 193,2 juta pengguna pada tahun 2021. Perkembangan ini mengubah cara masyarakat berbelanja langsung melalui media internet. Oleh karena itu, masyarakat mulai memanfaatkan media digital untuk memenuhi kebutuhan sehari-hari, menjadikan e-commerce sebagai tempat mencari dan membeli produk. Belanja online telah menjadi cara alternatif bagi masyarakat untuk membeli atau mencari layanan seperti pengiriman online tanpa bertemu dan berkomunikasi secara langsung.

Lazada memberikan kemudahan bagi penggunanya dalam memasarkan produk dengan menggunakan gambar dan menulis keterangan produk. Sementara itu, sistem pencarian produk yang lengkap dengan berbagai kategori membuat hidup pembeli lebih mudah.. Sebagai perusahaan e-commerce, Lazada juga memberikan informasi lengkap mengenai reputasi penjual sehingga pelanggan dapat leluasa membandingkan dan memilih produk yang diinginkan. Untuk pelanggan online dan calon pelanggan, ulasan dari pembeli sebelumnya sangat membantu dalam membantu calon pelanggan membuat keputusan pembelian. Calon pelanggan dapat diyakinkan dengan mengajukan pertanyaan tentang produk atau menjawab keingintahuan mereka .

Membeli produk elektronik secara online menimbulkan tantangan baru karena

sedikitnya aksi antara pemasar dan konsumen di e-commerce. Oleh karena itu, calon konsumen harus jauh lebih berhati-hati saat ingin membeli. Tentu saja konsumen tidak bisa benar-benar menyentuh atau melihat langsung dan mencoba fitur-fitur produk tersebut. Tentu saja hal ini sangat menyulitkan konsumen produk elektronik. Tentu saja konsumen dapat mencoba produk elektronik jika ada kemungkinan masalah. Untuk itu perlu gambaran dari pembeli terdahulu, Analisis sentimen dapat membantu untuk memperoleh gambaran umum persepsi konsumen dengan cara mengelompokkan jenis opini menjadi kategori.

Analisis sentimen merupakan metode penggalan data opini, memahami data teks, dan mengolahnya secara otomatis untuk memvisualisasikan sentimen yang terkandung dalam opini. Latar belakang analisis sentimen yaitu untuk terhubung ke ribuan sumber internet, menyatukan argumen tentang merek atau produk, dan menelitinya menggunakan analisis teks. Hal ini memberikan wawasan baru tentang cara mengidentifikasi perubahan kesuksesan pasar dari waktu ke waktu dan bagaimana penyedia layanan dapat meningkatkan reputasi produk mereka. Konsumen seringkali mengandalkan review produk sebelum melakukan pembelian karena sulit menilai kualitas produk secara langsung. Namun, menjamurnya ulasan dan kepercayaan konsumen terhadap kejujuran dan keakuratan klaim dapat menjadi kendala. Selain kualitas produk, ulasan juga mencakup pengalaman transaksi, termasuk layanan, pengiriman, dan keandalan platform belanja online.

Random Forest adalah metode klasifikasi yang banyak digunakan untuk secara otomatis menentukan polaritas komentar pelanggan. Random Forest adalah teknik pembelajaran umum berdasarkan algoritma pohon keputusan (D-Tree), yang terdiri dari beberapa pohon keputusan sebagai pengklasifikasi. Kelas-kelas yang diperoleh dari proses klasifikasi diperoleh dari sebagian besar keputusan kelas yang dibuat oleh pohon keputusan yang ada. Memilih keputusan dari pohon keputusan yang ada meningkatkan keakuratan hutan acak. Random Forest adalah algoritma klasifikasi yang sangat akurat. Hutan acak menjadi lebih banyak digunakan dalam klasifikasi dalam beberapa tahun terakhir karena efisiensi algoritma jenis ini dibandingkan dengan SVM, Naive Bayes, dan algoritma pembelajaran mesin lainnya [8]. Random Forest memiliki hasil akurasi yang baik, tahan terhadap anomali dan noise, serta lebih cepat dibandingkan packing dan boosting .

Penelitian sebelumnya membandingkan algoritma klasifikasi support vector machine, random forest, dan regresi logistik di review Shopee. 3.000 data digunakan untuk pengujian, 90% pelatihan, dan 10% pengujian. Hasil uji akurasi menggunakan matriks konfusi menunjukkan bahwa algoritma Random Forest memiliki akurasi tertinggi sebesar

94%, disusul Support Vector Machine (SVM) sebesar 91%, dan Regresi Logistik dari 91% menjadi 86%. Dalam penelitian ini, kami dapat memperoleh algoritma yang efektif dan terbaik untuk mengklasifikasikan sentimen ulasan positif, sentimen ulasan netral, dan sentimen ulasan negatif terkait aplikasi Shopee. Hasil algoritma optimal diintegrasikan ke dalam antarmuka web menggunakan bahasa pemrograman PHP dan Python untuk menentukan arah sentimen masukan di masa depan [10].

Penelitian sebelumnya terkait analisis sentimen review produk pada Tokopedia menggunakan algoritma Random Forest, Analisis sentimen ulasan produk Tokopedia dapat memberikan kesempatan untuk melihat bagaimana pelanggan Tokopedia menanggapi kualitas produk dan keramahan penjual. Hasil pengujian menunjukkan bahwa akurasi metode random forest dengan parameter akurasi sebesar 97,38% menunjukkan bahwa metode random forest dapat memprediksi review produk Tokopedia dengan baik. Semakin tinggi akurasinya, semakin baik kinerja model klasifikasi .

Penelitian terdahulu mengenai analisis sentimen dengan Random Forest mengenai Aplikasi Dompot Digital, dengan rincian sebanyak 1751 data untuk dompet digital gopay, 2982 data untuk dompet digital shopeepay, 3235 data untuk dompet digital linkaj. Dari hasil klasifikasi menggunakan Random Forest diperoleh hasil akurasi untuk dompet digital gopay sebesar 85,43%. Untuk dompet digital shopeepay diperoleh akurasi sebesar 85,50%. Kemudian untuk dompet digital linkaja diperoleh akurasi sebesar 89,02%.

Penelitian terdahulu menyoroiti pentingnya analisis sentimen dalam menanggapi ulasan film masyarakat di platform online. Tahap preprocessing data dilakukan sebelumnya, kemudian feature extraction menggunakan Word2Vec, dan metode klasifikasi menggunakan Random Forest. Penelitian terbaru menunjukkan penggunaan skip-gram Word2Vec 300 dimensi dengan Adaptive Boosting pada model dasar meningkatkan akurasi hingga 75.76%.

Kajian lainnya memaparkan analisis sentimen masyarakat terhadap program vaksinasi Sinovac di tengah pandemi COVID-19. Algoritme pengklasifikasi hutan acak digunakan untuk membuat model dan divalidasi dengan validasi silang K-fold dan matriks konfusi. Hasil penelitian menunjukkan bahwa model mencapai nilai akurasi dan presisi sebesar 79%, recall, dan nilai F1-score masing-masing sebesar 85%, 90%, dan 88% dengan dominasi sentimen positif .

METODE PENELITIAN

Dataset

Data yang digunakan untuk penelitian menggunakan data review produk e-commerce Lazada Indonesia pada kaggle dengan jumlah 6.000 baris. Dalam proses pengumpulannya data tersebut harus dilakukan Preprocessing. Berikut merupakan detail dataset yang dapat dilihat pada tabel 1.

Tabel 1. Detail Dataset

Item Id	Category	Name	Review Content
100002528	beli-harddisk-eksternal	Kamal U.	bagus mantap dah sesuai pesanan
100002528	beli-harddisk-eksternal	yofanca m.	Bagus, sesuai foto
100002528	beli-harddisk-eksternal	Lazada Customer	okkkkk mantaaaaaapppp ... good
100002528	beli-harddisk-eksternal	Lazada Customer	bagus sesuai
100003785	beli-harddisk-eksternal	Fadjar B.	baru 10 bulan layarnya dah bergaris
100003785	beli-harddisk-eksternal	agung p.	Pesan rabu sore,minggu sore sampe,,barang sesuai specs mantab,tkz lazada
100004505	beli-harddisk-eksternal	Andry M.	Pengirim barang tidak sesuai janji. Katanya express 1 hari. Dari hari jumat barang sampai hari rabu. SeMoga barangnya awet saja.
100004505	beli-harddisk-eksternal	Leonard	Kualitas ok

Pemrosesan Data

Proses pertama mempersiapkan text mining untuk mengolah teks untuk diproses lebih lanjut. Keuntungan dari pra-pemrosesan data ini adalah mengumpulkan kata-kata penting dan berguna untuk analisis sentimen. Berikut ini adalah langkah-langkah pra-pemrosesan :

1. Cleaning, tugasnya adalah menghilangkan kata-kata yang tidak perlu seperti pola implisit, kalimat ganda, kolom tabel kosong, dan menghilangkan tanda baca.
2. Case Folding, mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar. Dalam tutorial ini, semua huruf diubah menjadi huruf kecil karena sebagian besar teks ditulis dalam huruf kecil.
3. Tokenizing, memisahkan kata demi kata. Tujuannya adalah untuk mengelompokkan

kemunculan kata dalam proses berikut.

4. Filtering, Setelah menyaring atau menyelesaikan kata-kata dalam proses ini, kata-kata yang tidak memiliki hubungan dengan klasifikasi emosional dihilangkan, dengan itu mengurangi ukuran teks tanpa mengurangi esensi teks.
5. Word Weighting, pembobotan kata menentukan skor setiap kemunculan kata dengan menggunakan metode TF-IDF.

Klasifikasi

Random forest classifier merupakan metode klasifikasi yang terdiri dari kumpulan pohon keputusan yang nantinya akan dijadikan vote untuk mendapatkan hasil terakhir dari pendeteksian sarkasme dengan pendukung berupa data latih dan fitur acak yang independen dengan fitur yang berbeda-beda. Pohon keputusan dibuat dengan menentukan node akar dan berakhir dengan beberapa node daun untuk mendapatkan hasil akhir[15]. Dengan Persamaan sebagai berikut:

$$(D) = 1 - \sum_{i=p}^m p_1^2$$

Information Gain:

$$(D, A) = \text{entropy}(D) - \sum_{v \in \text{values}(A)} \frac{|Dv|}{|D|} \cdot \text{entropy}(Dv)$$

Entropi:

$$\text{entropy}(D) = - \sum_{i=1}^m p_i \log_2(p_i)$$

Voting Mayoritas:

$$y = \text{argmax}_c \sum_{t=1}^T I(h_t(x) = c)$$

Evaluasi

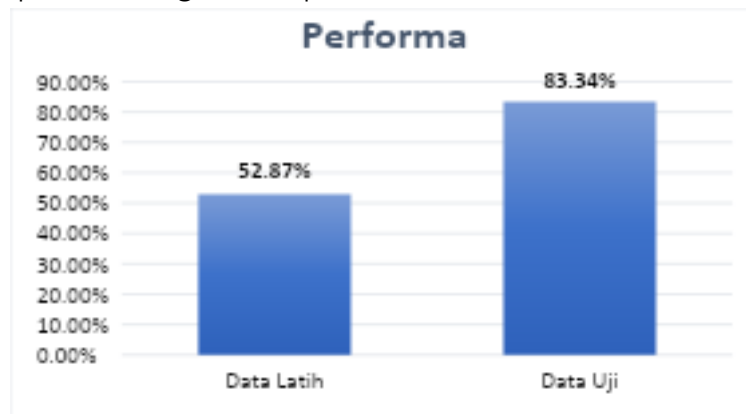
Evaluasi untuk mengetahui nilai kegunaan dari model yang telah berhasil dibuat pada Langkah sebelumnya. Dataset diuji menggunakan pembagian 70% data latih dan 30% data uji. Ukuran penilaian diberikan dengan persentasi dari 1 sampai 100 persen. Sebuah model bisa dianggap baik jika nilai Tingkat akurasi data uji lebih tinggi dari data latih. Untuk tahap evaluasi akan menggunakan teknik *10 fold cross validation*. Perhitungan akurasi dilakukan sesuai dengan persamaan dibawah ini.

$$\text{akurasi} = \frac{\sum \text{data uji benar}}{\sum \text{jumlah total data uji}} 100$$

HASIL DAN PEMBAHASAN

Menerapkan model Random Forest Classifier untuk menganalisis sentimen dari review produk harddisk di e-commerce Lazada Indonesia. Data sampel yang digunakan mencakup 6.000 baris review. Tahapan preprocessing meliputi cleaning untuk menghapus kata-kata yang tidak relevan, case folding untuk menyamakan huruf kapital dan kecil, tokenizing untuk memecah teks menjadi kata-kata individual, filtering untuk menghilangkan kata-kata yang tidak berkaitan dengan sentimen, serta word weighting menggunakan metode TF-IDF untuk memberikan bobot pada kata-kata berdasarkan frekuensinya.

Model ini mengklasifikasikan sentimen berdasarkan review yang tersedia, dengan hasil evaluasi menunjukkan performa akurasi data latih sebesar 52,87% dan data uji sebesar 83,34%. Hasil ini menunjukkan bahwa model Random Forest dapat mengidentifikasi sentimen dari review produk dengan cukup baik.



Gambar 1. Grafik Perbandingan

Evaluasi dilakukan dengan skema dataset menjadi 70% data pelatihan dan 30% data pengujian. Evaluasi kinerja dilakukan menggunakan *10 fold cross validation* untuk menghitung performa algoritma Random Forest, dengan menunjukkan kemampuan algoritma ini dalam membuat prediksi yang diharapkan akurat dan sensitif terhadap perubahan data. Performa untuk evaluasi dapat dilihat tabel di bawah ini.

Tabel 2. Performa Data Latih

Accuracy: 52.87% +/-3.11% (micro average: 52.7%)				
True	Negatif	Netral	Positif	Class Precision
Netral	13264	2179	10112	51.90%
Negatif	0	0	0	0%
Positif	5528	1745	8680	54.41%
Class recall	70.58%	0.00%	46.19%	

Tabel 3. Performa Data Uji

Accuracy: 83.34% +/-3.74% (micro average: 83.34%)				
True	Netral	Negatif	Positif	Class Precision
Netral	3351	619	3014	82.28%
Negatif	0	0	0	0.00%
Positif	1530	376	2466	85.09%
Class recall	71.82%	0.00%	46.06%	

Random Forest Classifier pada penelitian ini menunjukkan keunggulan dalam beberapa aspek kinerja, terutama ketika diterapkan pada dataset besar dan data yang tidak seimbang. Metode ini menggunakan teknik ensemble learning yang menggabungkan beberapa pohon keputusan (decision trees) untuk memberikan hasil klasifikasi akhir. Dengan cara ini, Random Forest terbukti mampu mengurangi risiko overfitting yang sering terjadi pada pohon keputusan tunggal, meningkatkan keandalan dan generalisasi model. Penggunaan berbagai pohon keputusan membantu dalam menangani fitur yang beragam dan kompleks, karena setiap pohon dalam hutan acak dapat menangani subset berbeda dari data dan fitur. Hal ini memungkinkan model untuk mengidentifikasi pola yang lebih kompleks dalam data, yang mungkin tidak dapat ditangkap oleh model klasifikasi lainnya. Random Forest juga memiliki kemampuan untuk menilai pentingnya fitur secara otomatis, memberikan wawasan tentang fitur mana yang paling berkontribusi terhadap keputusan klasifikasi.

SIMPULAN

Akurasi data latih sebesar 52,87% dan data uji sebesar 83,34% menunjukkan bahwa ada tantangan dalam pengklasifikasian sentimen, kemungkinan disebabkan oleh variasi dalam bahasa atau kekayaan kosakata yang digunakan dalam review. Penerapan teknik lain seperti hyperparameter tuning atau integrasi dengan algoritma lain dapat menjadi langkah berikutnya untuk meningkatkan akurasi model. Selain itu, analisis lebih mendalam terhadap fitur-fitur yang digunakan dalam model bisa membantu dalam memahami faktor-faktor yang mempengaruhi performa model.

Meskipun Random Forest Classifier menunjukkan akurasi yang baik dengan 83.34%, ada beberapa kekurangan yang perlu diperhatikan. Pertama, meskipun akurasi model cukup memadai, masih terdapat ruang untuk perbaikan, terutama dalam mengidentifikasi sentimen yang lebih kompleks atau subtle. Kualitas review, termasuk kejelasan dan konsistensi bahasa, dapat mempengaruhi kinerja model. Variasi dalam cara penulis menyampaikan sentimen, seperti penggunaan bahasa yang ambigu atau istilah slang, dapat

menambah kesulitan dalam klasifikasi yang akurat.

DAFTAR PUSTAKA

- A. Farhan Hasrul, Suharyati, and R. Sembiring, "Analisis pengaruh online customer review dan rating terhadap minat beli produk elektronik di tokopedia," vol. 2, no. 1, 2021.
- A. Hendra and F. Fitriyani, "Analisis Sentimen Review Halodoc Menggunakan Naïve Bayes Classifier," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 2, pp. 78–89, 2021, doi: 10.14421/jiska.2021.6.2.78-89.
- A. Saepudin, A. Faqih, and G. Dwilestari, "Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee," *J. TEKNO KOMPAK*, vol. 18, no. 1, pp. 178–192, 2024.
- B. B. Baskoro, I. Susanto, and S. Khomsah, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *INISTA (Journal Informatics Inf. Syst. Softw. Eng. Appl.)*, vol. 3, no. 2, pp. 21–29, 2021, doi: 10.20895/INISTA.V3.
- B. Zulfikar Ramadhan, I. Riza, and I. Maulana, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *J. Appl. Informatics Comput.*, vol. 2, pp. 220–225, 2022.
- E. Putra, S. Tinggi, and I. E. Pasaman, "Pengaruh Promosi Melalui Sosial Media Dan Review Produk Pada Marketplace Shopee Terhadap Keputusan Pembelian (Studi Pada Mahasiswa Stie Pasaman) the Influence of Promotion Through Social Media and Product Review on the Marketplace Shopee on Purchase Dec," *J. Apresiasi Ekon.*, vol. 8, no. 3, pp. 467–474, 2020.
- F. M. Mewoh, J. R. E. Tampi, and D. D. S. Mukuan, "Pengaruh Digital Marketing Terhadap Keputusan Pembelian Pada Matahari Department Store Manado Town Square," *J. Adm. Bisnis*, vol. 9, no. 1, p. 35, 2019, doi: 10.35797/jab.9.1.2019.23529.35-42.
- F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>
- Izzah Nur Masyithoh and Ivo Novitaningtyas, "Pengaruh Digital Marketing Terhadap Minat

- Beli Konsumen Pada Marketplace Tokopedia," *J. Manaj. Bisnis Kreat.*, vol. 7, no. 1, 2021, doi: 10.36805/manajemen.v7i1.1951.
- M. Y. Aldean, P. Paradise, and N. A. Setya Nugraha, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 64–72, 2022, doi: 10.20895/inista.v4i2.575.
- N. Ambika Hapsari and A. Dwi Indriyanti, "Analisis Sentimen pada Aplikasi Dompot Digital Menggunakan Algoritma Random Forest," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 04, no. 03, pp. 186–192, 2023.
- Stephenie, B. Warsito, and A. Prahutama, "Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method," *E3S Web Conf.*, vol. 202, pp. 1–10, 2020, doi: 10.1051/e3sconf/202020216006.
- T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, p. 32, 2021, doi: 10.24014/coreit.v7i1.12945.
- Y. Surya, S. Al Faraby, and M. Dwifabri, "Analisis Sentimen Terhadap Ulasan Film Menggunakan Word2Vec dan SVM," *e-Proceeding Eng.*, vol. 8, no. 4, pp. 4136–4144, 2021.