

Systematic Literature : Algoritma, Dataset, Dan Aplikasi Pada Sentiment Analysis

Mayhendra Kusuma^{*1} Ramadhan Renaldy^{*2}

^{1,2} Program Studi Informatika, Universitas PGRI Semarang, Kota Semarang

*Email: Uzumakib04@gmail.com

Abstract

Sentiment analysis is the process of identifying, extracting, and evaluating whether the opinions and sentiments expressed in text are positive, negative, or neutral. As its use has become increasingly popular, much research has been conducted in this area to understand the attitudes or reactions of groups of people towards various topics, products, services or applications. This systematic literature review examines findings from 30 selected journals, revealing that the most commonly used algorithm is Naïve Bayes, accounting for 47.05% of studies. The main data set used for sentiment analysis research consists of tweets, which account for 36.67% of the reviewed literature. Additionally, the review highlights that applications related to sentiment analysis, particularly those focused on applications, account for 20% of all surveyed journals. The purpose of this review is to find out algorithms, datasets and topics that are often discussed in research.

Keywords: analysis sentiment, Algoritma, dataset, aplikasi

Abstrak

Analisis sentimen merupakan proses mengidentifikasi, mengekstraksi, dan mengevaluasi apakah opini dan sentimen yang diungkapkan dalam teks bersifat positif, negatif, atau netral. Seiring dengan semakin populernya penggunaannya, banyak penelitian dilakukan di bidang ini untuk memahami sikap atau reaksi sekelompok orang terhadap berbagai topik, produk, layanan, atau aplikasi. Tinjauan literatur sistematis ini mengkaji temuan dari 30 jurnal terpilih, mengungkapkan bahwa algoritma yang paling umum digunakan adalah Naïve Bayes, yang mencakup 47,05% penelitian. Kumpulan data utama yang digunakan untuk penelitian analisis sentimen terdiri dari tweet, yang mencakup 36,67% dari literatur yang diulas. Selain itu, tinjauan ini menyoroti bahwa aplikasi yang terkait dengan analisis sentimen, khususnya yang berfokus pada aplikasi, mencakup 20% dari seluruh jurnal yang disurvei. Tujuan dilakukan review ini untuk mengetahui algoritma, dataset, dan topik yang sering dibahas dalam penelitian.

Kata Kunci: sentimen analisis, Algoritma, himpunan data, aplikasi

1. Pendahuluan

Analisis sentimen memberikan pemahaman mendalam terhadap pendapat dan sentimen yang diungkapkan dalam data teks[1]. Metode ini efektif mengidentifikasi pola positif, negatif, dan netral dari setiap teks yang dianalisis[2]. Sentimen Hasil analisis memberikan wawasan berharga untuk pengambilan keputusan dalam berbagai bidang situasi, dari bisnis hingga penelitian akademis[3].

Dalam analisis sentimen ada beberapa alogaritma yang sering digunakan dalam melakukan penelitian. Alogaritma yang umum digunakan adalah Naive Bayes[4], Support Vector Machines (SVM)[5], dan algoritma Deep Learning seperti Convolutional Neural Networks (CNN)[6] dan Recurrent Neural Networks (RNN). Algoritma ini efektif dalam mengklasifikasikan teks ke dalam kategori sentimen positif, negatif, atau netral berdasarkan pola pada data teks yang dianalisis[7].

Kumpulan data yang umum digunakan dalam penelitian analisis sentimen bervariasi tergantung pada area penerapan dan tujuan penelitian[5]. Contoh kumpulan data umum

mencakup kumpulan data ulasan produk seperti ulasan pelanggan Amazon dan peringkat Yelp, kumpulan data media sosial seperti kumpulan data analisis sentiment.Twitter, dan kumpulan data umum seperti penggunaan aplikasi Shopee menggunakan algoritma Support Vector Machine (SVM) [8]. Kumpulan data ini memungkinkan peneliti untuk melatih dan menguji model analisis sentimen untuk berbagai konteks dan scenario.

Analisis sentimen diterapkan pada berbagai topik dan industri untuk memahami opini dan persepsi peneliti. Contoh topik yang umumnya dianalisis untuk analisis sentimen mencakup produk[9], media sosial[10], perekonomian[11], transportasi[12], berita[13], mahasiswa[14], film[15], politik[16] dan aplikasi[17]. Analisis sentimen banyak digunakan dalam berbagai situasi untuk memahami opini dan sentimen dari data teks. Topik yang paling umum adalah memantau sentimen publik terhadap suatu website[18], produk[19], atau layanan[20]. Selain itu, analisis sentimen juga digunakan dalam penelitian akademis untuk mempelajari opini publik mengenai isu-isu sosial atau politik[21]. Dalam dunia bisnis[22], aplikasi ini membantu perusahaan memahami masukan pelanggan dan meningkatkan strategi pemasaran dan pengembangan produk mereka. Dengan kemajuan teknologi seperti media sosial[23], analisis sentimen juga digunakan untuk memprediksi tren pasar dan reaksi terhadap peristiwa penting secara real time[24].

Sistem rekomendasi berbasis pembelajaran mendalam (SLR) bertujuan untuk meningkatkan akurasi dan relevansi rekomendasi bagi pengguna[25]. Hal ini dicapai dengan menerapkan algoritma pembelajaran mendalam yang dapat memproses kumpulan data besar dan mengidentifikasi pola kompleks. Topik ini penting karena berdampak pada pengalaman pengguna di berbagai aplikasi seperti e-commerce[26], media sosial[27], dan platform streaming[21]. Dengan menggunakan kumpulan data yang luas dan beragam, SLR dapat mempersonalisasi rekomendasi untuk setiap pengguna, meningkatkan kualitas layanan dan kepuasan pengguna secara keseluruhan[28]. Penelitian dan pengembangan kamera SLR terus meningkatkan akurasi prediksi dan mengatasi tantangan baru dalam cakupan dan kecepatan proses rekomendasi.

2. Metode

Dalam penelitian ini, penulis menggunakan sistematik literatur review (SLR). SLR adalah metode penelitian yang bertujuan untuk melakukan tinjauan literatur secara sistematis dengan menggunakan prosedur yang direncanakan secara cermat. SLR terdiri dari tiga fase utama: perencanaan, pelaksanaan, dan pelaporan[29]. Di bawah ini penjelasan setiap langkah dalam SLR:

1) Planning

Perencanaan Langkah pertama dalam memulai SLR adalah menentukan topik. Penentuan topik yang sesuai akan memudahkan dalam pemilihan Research Question (RQ) dapat kita tetapkan berkaitan langsung dengan topik yang ditentukan. Pembuatan SLR memerlukan RQ, karena RQ berfungsi sebagai panduan untuk pencarian dan ekstraksi literatur. Suatu RQ dikatakan baik bila bermanfaat, terukur, dan konsisten dengan tema yang ditetapkan semula.

Langkah selanjutnya penulis memilih 30 jurnal dalam 3 sumber yang berbeda yaitu googlescholar (10), researchgate (10), dan semantic scholar (10).

2) Pelaksanaan

Pelaksanaan pada SLR adalah fase yang mencakup implementasi SLR. Mulailah dengan menentukan kata kunci yang akan digunakan untuk mencari jurnal, memahami penggunaan sinonim kata untuk meningkatkan akurasi pencarian, dan menentukan sumber pencarian jurnal.

Setelah mendapatkan 30 jurnal, langkah selanjutnya adalah memilih berdasarkan RQ. Untuk menyederhanakan proses direkomendasikan pemilihan jurnal berdasarkan RQ1, RQ2, dan RQ3 dan membuat kriteria untuk menghasilkan 30 jurnal untuk direview. Research Question pada penelitian ini adalah algoritma apa yang paling sering digunakan, apa dataset yang digunakan dalam penelitian sentiment analysis. Berikut merupakan *search sting* yang digunakan dalam pencarian tinjauan literatur:

Tabel 1. *Search String*

No	Sentimen analysis
1	sentiment analysis OR analisa sentimen

Tabel 2. *Tujuan dibentuk RQ*

ID	Research Question	Tujuan
RQ1	Alogaritma apa yang sering digunakan ?	Mengidentifikasi algoritma yang sering digunakan pada sentment analysis
RQ2	Apa dataset yang digunakan dalam penelitian sentiment analysis?	Mengidentifikasi dataset yang digunakan pada sentiment analysis
RQ3	Pada topik apa sentiment analysis di aplikasikan?	Mengidentifikasi topik yang sering digunakan dalam penelitian sentiment analysis

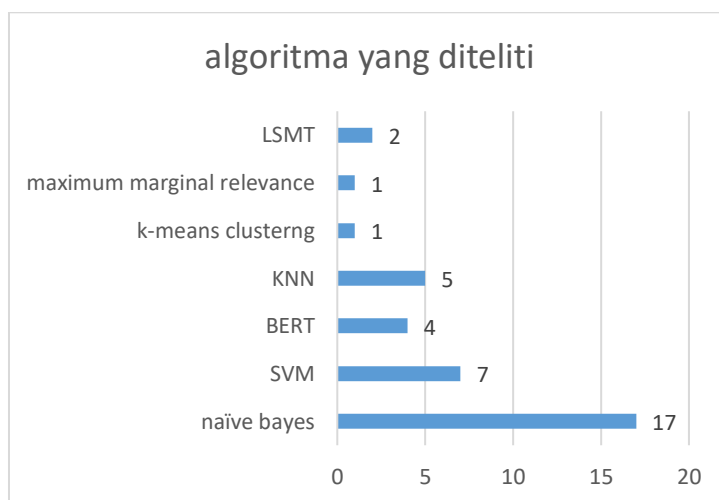
Tabel di atas menunjukkan peta pikiran dasar dari tinjauan literature sistematis. Tujuan utama dari tinjauan literatur sistematis ini adalah untuk mengetahui alogaritma, dataset, dan topik dari penelitian sentiment analysis.

3) Pelaporan

Langkah terakhir dalam penerapan SLR adalah pelaporan. Pada tahap ini penulis memulai penulisan laporan yang meliputi pendahuluan,metode, hasil dan pembahasan, dan kesimpulan. Ada juga absrak di awal laporan. Bagian absrak berisi review laporan SLR yang dibuat. Tujuan dari ringkasan adalah untuk membantu pembaca dengan mudah memahami isi laporan. Oleh karena itu, ringkasan harus relevan denga nisi laporan dan menggunakan Bahasa yang sederhana dan mudah dipahami. Pendahuluan mencakup filosofi topik yang sebelumnya telah dipilih. Metode membahas tentang langkah langkah SLR yang dilakukan oleh penulis. langkah langkah yang dilakukan antara lain Menyusun tema, mencari jurnal, menentukan RQ, dan meriview jurnal yang telah disesuaikan dengan RQ yang ditentukan. Untuk hasil dan pembahsan mencakup hasil dari jurnal yang diambil dan disaring. Hasil dan pembahasan juga menguraikan jawaban atas RQ yang teridentifikasi yang diperoleh dari jurnal. Bagian akhir adalah kesimpulan yang menyimpulkan laporan. Kesimpulannya mencakup ringkasan SLR yang dilakukan oleh penulis berdasarkan RQ yang telah ditetapkan sebelumnya.

3. Hasil dan Pembahasan

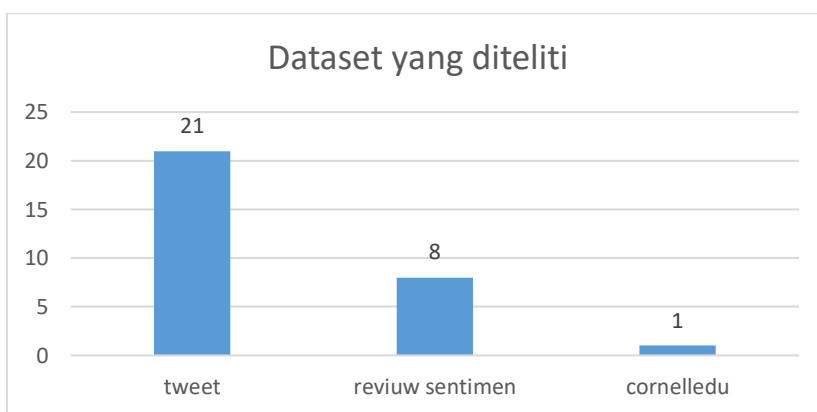
RQ1 (Alogaritma apa yang sering digunakan?)



Gambar 1. Grafik Algoritma yang paling sering digunakan untuk melakukan sentiment analysis

Pada Gambar 1 diperlihatkan bahwa algoritma yang paling sering digunakan adalah naïve bayes dengan jumlah jurnal sebanyak 17 jurnal. support vector machine sebanyak 7 jurnal, bidirectional encoder representations from transformers sebanyak 4 jurnal, k-Nearest neighbors sebanyak 5 jurnal, k-means clustering dan , maximum marginal relevance sebanyak 1 jurnal. Dan Lexicon Sentiwordnet hanya digunakan pada 2 jurnal. Oleh karena itu presentase algoritma yang paling sering digunakan adalah naïve bayes sebesar 45,946% dari jurnal yang telah dipilih. Alasan naïve bayes sering digunakan karena penerapannya pun sangat cepat dan ada iterasi. Sebab, probabilitas dapat langsung dihitung. Jadi, algoritma ini berguna saat Anda membutuhkan kecepatan pelatihan yang tinggi. Terlebih jika asumsi independensi bersyarat berlaku, bisa dipastikan akan memberi hasil yang bagus.

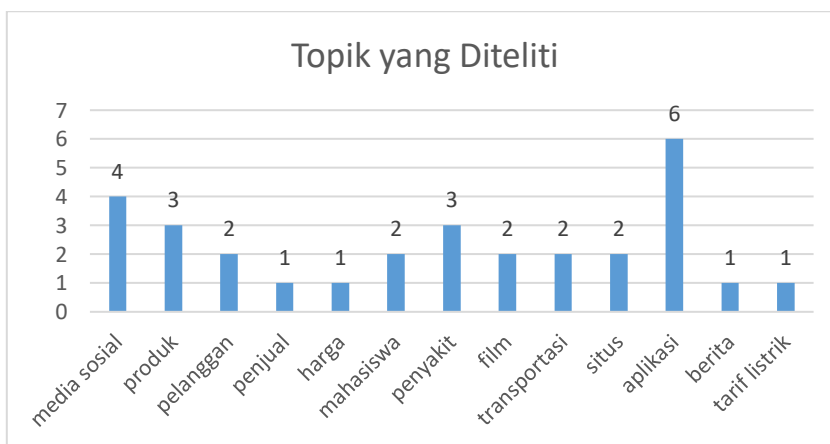
RQ2 (Apa dataset yang digunakan dalam penelitian sentiment analysis?)



Gambar 2. Grafik dataset yang digunakan dalam penelitian sentiment analysis

Pada Gambar 2 dijelaskan tentang metode ekstraksi fitur yang paling sering digunakan. Untuk metode yang paling sering digunakan adalah metode tweet sebanyak 21 jurnal, dataset review sentiment sebanyak 8 jurnal. Dataset cornelledu hanya digunakan 1 saja dalam jurnal. Presentase tweet dalam jurnal sebesar adalah 70%

RQ3 (Pada topik apa sentiment analysis di aplikasikan?)



Gambar 3 – Grafik Topik yang di aplikasin dalam sentiment analysis

Sentiment analysis diaplikasikan pada berbagai topik. Pada Gambar 3 dijelaskan bahwa Topik yang banyak diteliti adalah aplikasi sebanyak 6 jurnal. Topik yang membahas media sosial sebanyak 4 jurnal. Topik yang membahas produk dan penyakit sebanyak 3

jurnal. Topik yang membahas pelanggan, film, transportasi, situs sebanyak 2 jurnal. Dan topik yang membahas penjual, harga, berita, dan tarif listrik[30] sebanyak 1 jurnal saja. Presentase topik aplikasi dalam jurnal sebanyak 20%.

4. Kesimpulan

Dari 30 jurnal yang telah di review mendapatkan hasil atas RQ yang telah ditetapkan sebelumnya. Untuk RQ1, alogaritma yang sering digunakan untuk melakukan Analisa sentiment adalah naïve bayes sebesar 45,946% dari 30 jurnal. RQ2, dataset yang digunakan dalam penelitian sentiment analysis adalah tweet sebesar 70% dalam seluruh jurnal. RQ3 pada topik apa sentiment analysis diaplikasikan adalah topik aplikasi sebanyak 20% dalam semua jurnal.

5. Referensi

- [1] Tri Vena Meiyanti, Muhammad Hatta, and Agus Seviana, "Analisis Sentimen Mahasiswa Dengan Dosen Menggunakan Metode Naïve Bayes Classifier Pada Kuesioner Dosen," *Jurnal Manajemen Sistem Informasi*, vol. Vol. 01 No. 02 Juni 2023 | ISSN (Online) 2987-0364, Jun. 2023.
- [2] Akhmad Rezki Purnajaya and Yonky Pernando, "Analisa Sentimen Informasi Hoaks Pasca Pandemi Covid-19 dengan Text Mining," *Journal of Computer System and Informatics (JoSYC)*, vol. Volume 4, No. 3, Mei 2023, Page 460–469.
- [3] Indra Muda Pulungan, "Analisa Sentimen Terhadap Penjualan Alat Pelindung Diri Pada Market Place Menggunakan Metode Maximum Marginal Relevance," *Informasi dan Teknologi Ilmiah (INTI)*, vol. Volume 9, No 3, Juni 2022, p. Hal: 64-68.
- [4] Komang Andi Sugiarta, Putri Anugrah Cahya Dewi, and Nengah Widya Utam, "ANALISA SENTIMEN MAHASISWA TERHADAP LAYANAN STMIK PRIMAKARA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR," *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, vol. Vol. 5 No. 3, Agustus 2023, hlm. 364 – 372.
- [5] Arya Damar Pratama and Hendry, "ANALISA SENTIMEN MASYARAKAT TERHADAP PENGGUNAAN CHATGPT MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. Vol. 9, No. 1, Maret 2024, Pp. 327-338.
- [6] Sri Hardani and Dinar Ajeng Kristiyanti, "Systematic Literature Review: Analisa Sentimen Penerimaan Masyarakat Terhadap Jenis Vaksin Covid-19 di Dunia," *Jurnal IKRAITH-INFORMATIKA*, vol. Vol 6 No 3 November 2022.
- [7] Andharini Dwi Cahyani, "Analisa Kinerja Metode Support Vector Machine untuk Analisa Sentimen Ulasan Pengguna Google Maps," *Journal of Computer System and Informatics (JoSYC)*, vol. Volume 4, No. 3, Mei 2023, Page 604–613.
- [8] Irma Surya Kumala Idris, Yasin Aril Mustofa*, and Irvan Abraham Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. Volume 5 Nomor 1 Januari 2023, Jan. 2023.
- [9] Roni Merdiansah, Siska, and Azhari Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. Volume 7, Nomor 1, Tahun 2024, Hal 221-228.
- [10] Ferdi Rejeki and Vina Ayumi, "Analisa Sentimen Mengenai Kenaikan Harga Bbm Menggunakan Metode Naïve Bayes Dan Support Vector Machine," *JSAI: Journal Scientific and Applied Informatics*, vol. Vol. 6, No. 1, Januari 2023, doi: 10.36085.
- [11] Didik Siswanto, Zamzami, Lasri Nijal, and Safrul Rajab, "ANALISA SENTIMEN PUBLIK MENGENAI PEREKONOMIAN INDONESIA PADAMASA PANDEMI COVID-19 DI

- TWITTER MENGGUNAKAN METODE KLASIFIKASI K-NN DAN SVM,” vol. Vol. 2No. 1(2022) 01–09E ISSN: 2809-4069, Jun. 2022.
- [12] Ismia Iwandini, Agung Triayudi, and Gatot Soepriyono, “Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor,” *Journal of Information System Research (JOSH)*, vol. Volume 4, No. 2, Januari 2023, pp 543–550.
- [13] Alifiyah Rohmatul Hidayati, Arief Senja Fitriani, and Mochamad Alfian Rosid, “Analisa Sentimen Pemilu 2019 Pada Judul Berita Online Menggunakan Metode Logistic Regression,” *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. Vol. 4, No. 2, April (2023), pp. 298–305, Terakreditasi Nomor /E/KPT/2022 204AD.
- [14] Moh Khoirul Insan, Umi Hayati, and Odi Nurdiawan, “ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. Vol. 7 No. 1, Februari 2023.
- [15] Diah Fatma SJORaida, Bucky Wibawa Karya Guna, and Dudi Yudhakusuma, “Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers),” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, Mar. 2024.
- [16] Icha Adellia Safra and Eri Zuliarso, “ANALISA SENTIMEN PERSEPSI MASYARAKAT TERHADAP PEMINDAHAN IBUKOTA BARU DI KALIMANTAN TIMUR PADA MEDIA SOSIAL TWITTER,” *Proceeding SENDIU 2020*.
- [17] kevin, Margareta Enjeli, and Andri Wijaya, “Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes,” *JURNAL ILMIAH COMPUTER SCIENCE (JICS)*, vol. Volume 2, Nomor 2, Januari 2024, Page 89-98, doi: <https://doi.org/10.58602/jics.v2i2.24>.
- [18] Muhammad Abdillah, Muhammad Fikry, Yusra, Alwis Nazir, and Fitri Insani, “Analisa sentimen terhadap kenaikan bbm di twitter (x) menggunakan naive bayes classifier,” *Jurnal Computer Science and Information Technology (CoSciTech)*, vol. Vol. 5, No. 1, April 2024, hal. 65–74, Apr. 2024.
- [19] Anugerah Tri Siswanto, Rani Rotul Muhima, and Septiyawan Rosetya Wardhana, “Analisa Sentimen Review Produk Handphone Pada Situs Amazon Menggunakan Pendekatan Lexicon Berdasarkan Sentiwordnet,” *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, vol. Vol. 3, No. 1, Juli 2022.
- [20] Dita Novita Sari, Firda Adelia, Fita Rosdiana, Belasana Butar Butar, and Muhadi Hariyanto, “ANALISA SENTIMEN TERHADAP REVIEW PRODUK KECANTIKAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” *JIKA (Jurnal Informatika) Universitas Muhammadiyah Tangerang*, Nov. 2020.
- [21] Edi Tohidi, Reza Perdana Herdiansyah, Edi Wahyudin, and Kaslani, “ANALISA SENTIMEN KOMENTAR VIDEO YOUTUBE DI CHANNEL TVONENEWS TENTANG CALON PRESIDEN PRABOWO SUBIANTO MENGGUNAKAN SUPPORT VECTOR MACHINE,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. Vol. 8 No. 1, Februari 2024, Jul. 2023.
- [22] Karianta Ginting, “Ulasan Algoritma Naïve Bayes Untuk Analisis Sentimen Terhadap Ulasan Produk,” Program Studi Magister Teknik Informatika, Fasilkom-TI Universitas Sumatera Utara, Medan, 20155, Sumatera Utara, Indonesia..
- [23] Farah Syadza Mufidah, Sri Winarno, Farrikh Al Zami, Erika Devi Udayanti, and Ramadhan Rakhmat Sani, “Analisis Sentimen Masyarakat terhadap Layanan ShopeeFood Melalui Media Sosial Twitter dengan Algoritma Naïve Bayes Classifier,” vol. *Journal of Information System* Vol. 7, No. 1, Mei 2022: 14-25.
- [24] Muljono, Dian Putri Artanti, Abdul Syukur, Adi Prihandono, and De Rosal I. Moses Setiadi, “Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes,” *STMIK Atma Luhur Pangkalpinang*, 8 – 9 Maret 2018.
- [25] Laila Ma’rifatul Azizah, Dimas Bagas Ajipratama, Nisrina Akbar Rizky Putri, and Cahya Damarjati, “Analisis Sentimen Masyarakat terhadap Kebijakan Vaksinasi Covid-19 di

- Indonesia pada Twitter Menggunakan Algoritma LSTM,” *Jurnal IPTEK-KOM (Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi)*, vol. Vol. 24 No. 2, Desember 2022: 161-172.
- [26] Nurul Amalia, Nur Ika Royanti, Indrayanti, and Bambang Ismanto, “Analisa Sentimen Pelanggan pada Review Belanja Online Berbasis Text Mining Menggunakan Metode K-Means,” *Journal of Information System Research (JOSH)*, vol. Volume 4, No. 4, Juli 2023, pp 1441–1447, doi: 10.47065/josh.v4i4.3781.
- [27] Nur Amalia, Tati Suprpti, and Gifthera Dwilestari, “ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAKSANAAN KURIKULUM MBKM,” *Jurnal Teknik Elektro dan Informatika*, vol. Volume 18 Nomor 1 (2023) 57-64.
- [28] Cindy Alifia Putri, Adiwijaya, and Said Al Faraby, “Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. Vol. 6, No. 2, Maret 2020, Hal. 181-193.
- [29] Elly Indrayuni, , Acmad Nurhadi, and Dinar Ajeng Kristiyanti, “Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors Untuk Analisa Sentimen Aplikasi Halodoc,” *Faktor Exaca*, vol. Vol. 14, No. 2, June 2021, p. 64~71.
- [30] Adi Kusuma and Agung Nugroho, “Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. Vol.15, No. 2, Tahun 2021.