

## Sentimen Analisis U-17 Pada Media Sosial X Dengan Metode *Support Vector Machine*

### *Sentiment Analysis of U-17 on Social Media X Using the Support Vector Machine Method*

Hendrawan M. Rorong<sup>1\*</sup>, Kristofel Santa<sup>2\*</sup>, Vivi Peggie Rantung<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknik, Universitas Negeri Manado

| Article Info                                                                                                                             | ABSTRAK                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Article history:</b><br/>Received: Feb 09, 2025<br/>Revised: Marc 20, 2025<br/>Accepted: Marc 28, 2025</p>                         | <p>Analisis sentimen merupakan proses penting untuk memahami opini publik mengenai berbagai topik yang dibahas di media sosial, khususnya di X (sebelumnya Twitter). Penelitian ini bertujuan untuk menganalisis sentimen dari tweet-tweet yang menggunakan tagar viral terkait tim nasional sepak bola U17. Dengan memanfaatkan teknik pemrosesan bahasa alami dan model analisis sentimen seperti VADER, serta algoritma <i>Support Vector Machine</i> (SVM), sistem ini dikembangkan untuk mengklasifikasikan sentimen menjadi positif dan negatif. Data diperoleh melalui scraping menggunakan Tweet Harvest, kemudian dilakukan praproses data untuk menghilangkan unsur-unsur yang tidak relevan. Untuk menangani ketidakseimbangan data, metode <i>random oversampling</i> diterapkan. SVM digunakan untuk meningkatkan akurasi klasifikasi sentimen setelah hasil skoring dari model VADER sebagai fitur input. Model yang dihasilkan menunjukkan akurasi sebesar 85%, sehingga diharapkan mampu memberikan <i>insight</i> yang bermanfaat tentang pola sentimen publik dalam menanggapi isu-isu terkait timnas U17. Hasil dari penelitian ini dapat dimanfaatkan oleh para pengambil keputusan, seperti pemasar atau pembuat kebijakan, untuk memahami preferensi serta kecenderungan opini publik di kalangan pecinta sepak bola remaja.</p> |
| <p><b>Kata kunci</b><br/>Sentimen Analisis,<br/>Sosial Media X,<br/>SVM ,<br/><i>Tweet Harvest</i>,<br/>U17,<br/><i>Vader</i>.</p>       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <p><b>Keywords</b><br/><i>Sentiment Analysis</i>,<br/><i>Social Media X</i>,<br/>SVM ,<br/><i>Tweet Harvest</i>,<br/>U17,<br/>VADER.</p> | <p><b>ABSTRACT</b><br/><i>Sentiment analysis is a crucial process for understanding public opinion on various topics discussed on social media, particularly on X (formerly Twitter). This study aims to analyze the sentiment of posts using viral hashtags related to the U17 national football team. Utilizing natural language processing techniques, sentiment analysis models such as VADER, and the Support Vector Machine (SVM) algorithm, this system is designed to classify sentiments as positive, negative, or neutral. Data were collected through scraping using Tweet Harvest, followed by data preprocessing to remove irrelevant elements. To address data imbalance, random oversampling was applied. SVM is utilized to improve sentiment classification accuracy, taking VADER's scoring as input features. The resulting model achieved an accuracy of 85%, providing valuable insights into public sentiment</i></p>                                                                                                                                                                                                                                                                                                                                                                                                            |

---

*patterns in response to issues surrounding the U17 national team. The findings of this research can be useful for decision-makers, such as marketers or policymakers, to understand preferences and trends in public opinion among youth football fans.*

---

***Corresponding Author:***

Kristofel Santa,  
Department of Informatics Engineering,  
Manado State University,  
Jl. Kampus Unima, Tounsar Village, South Tondano District, Tondano, North Sulawesi,  
Indonesia.  
Email: [kristofelsanta@unima.ac.id](mailto:kristofelsanta@unima.ac.id)

---

## **PENDAHULUAN**

Media sosial merupakan sarana bagi setiap penggunaannya untuk berinteraksi, bertukar ide dan mengemukakan pendapat. Media sosial sangat beragam salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia adalah Twitter. Twitter berubah nama menjadi X pada tahun 2023, pengguna X di Indonesia mencapai 24 juta dan menempati urutan ke 4 pengguna X terbanyak di dunia pada awal tahun 2023 (Annur, 2023). Pada media Sosial X pengguna bebas mengeluarkan aspirasi dan pendapat terkait isu yang viral saat ini. Isu yang viral biasanya ditandai dengan tagar (Alifia Seftin Oktriwina, 2022). Isu yang viral pada X mengandung opini publik dimana opini tersebut bisa jadi opini negatif maupun positif.

Sentimen analisis adalah salah satu jenis NLP atau *natural language processing*. Analisis ini menggali konteks untuk bisa mengidentifikasi suatu informasi dari materi sumber (Undap et al., n.d.). Sekarang ini, opini masyarakat semakin variatif. Tempat untuk menyalurkan pendapat pun semakin banyak jenisnya. Sentimen analisis bisa menjadi alat yang efektif untuk memantau dan memahami tiap opini dan *sentiment* yang ada (Kaparang et al., 2021).

Sentimen analisis merupakan bagian dari *machine learning* (Kambey et al., 2021), Beberapa metode klasifikasi digunakan saat membuat model pembelajaran, antara lain pengklasifikasi naïf Bayes, mesin vektor dukungan, dan regresi logistik. Sebuah studi oleh Ryo Benhard Dahlian dkk. tentang analisis sentimen transisi di televisi digital. Teknik yang digunakan adalah multinomial naïf Bayes, support vector machine, dan regresi logistik. Berdasarkan hasil penelitian, akurasi tertinggi ditemukan pada 94,00% untuk mesin vektor dukungan, 90,00% untuk regresi logistik, dan 88,00% untuk metode multinomial naïf Bayes (Dahlian & Sitanggang, 2023).

Penelitian yang dilakukan oleh Indra Kurniawan, dkk. Tentang Sentimen Analisis Marketplace. Metode yang digunakan adalah *Naïve Bayes* dan *Support vector machines*. Untuk menguji kinerja algoritma klasifikasi NB dan SVM menggunakan metode *Cross Validation* dan dari hasil hasil pengujian perbandingan tersebut bahwa algoritma SVM memiliki nilai accuracy yang terbaik dibandingkan dengan algoritma NB. Skor akurasi Tokopedia menggunakan algoritma NB sebesar 85,34% dan skor akurasi menggunakan SVM sebesar 86,82%, sedangkan skor akurasi Shopee menggunakan NB sebesar 80,04% dan skor akurasi menggunakan SVM sebesar 80,91%. Nilai akurasi pada Lazada dengan menggunakan

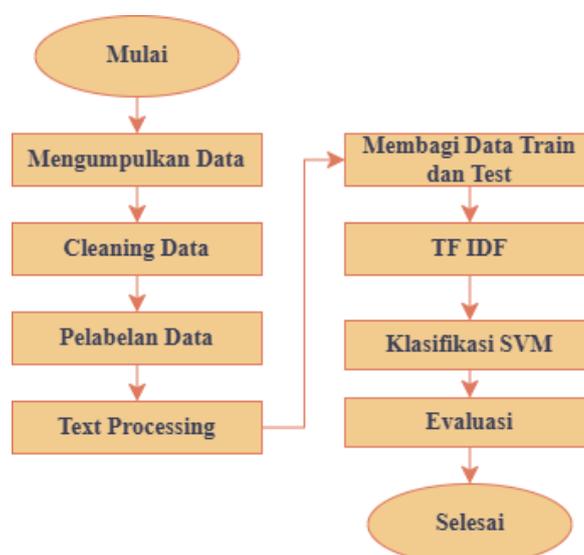
algoritma NB sebesar 83,52%, sedangkan nilai akurasi menggunakan SVM sebesar 88,93% yang berarti akurasi tertinggi menggunakan algoritma SVM (Kurniawan et al., 2023).

Kelebihan metode *Naive Bayes* memiliki performa yang baik, mudah digunakan, dan menghasilkan akurasi yang tinggi pada saat diaplikasikan ke dalam jumlah data yang banyak. Kelemahan metode *Naive Bayes* tidak berpengaruh terhadap atribut pada data sehingga dapat mengakibatkan besarnya dimensi fitur saat proses klasifikasi (Salsabila et al., 2023). Kelebihan metode *Support vector machines* terkenal dengan hasil akurasi yang tinggi dalam melakukan klasifikasi serta dapat menghindari kesulitan dari permasalahan dimensionalitas. Kekurangan metode *Support vector machines* dalam pemilihan fitur sekaligus pengaturan parameter di SVM yang secara signifikan mempengaruhi hasil akurasi klasifikasi (Solecha & Irnawati, 2023). Kelebihan metode *Logistic Regression* penjabaran linier yg telah teruji membangun penjabaran yg powerful menggunakan statistik probabilitas dan bisa menanggulangi permasalahan penjabaran multi kelas. Kekurangan dari *Logistic Regression* merupakan rentan terhadap *underfitting* pada dataset yg kelasnya tidak *balance*, sebagai akibatnya memiliki akurasi yg rendah (Alfarobi et al., 2023).

Tim nasional sepak bola U-17 Indonesia mewakili Indonesia dalam kompetisi sepak bola internasional di tingkat remaja. Tim ini berpartisipasi dalam berbagai turnamen, termasuk Piala Dunia U-17 FIFA dan Kejuaraan U-16 AFC. Seperti halnya negara-negara lain, Indonesia juga memiliki program pengembangan pemuda untuk menumbuhkan bakat-bakat muda bagi tim nasional. Tim U-17 berperan penting dalam mengidentifikasi dan melatih pemain-pemain berbakat untuk tim nasional senior di masa depan. Bertanding dalam kompetisi seperti Piala Dunia U-17 FIFA memberikan pengalaman berharga bagi pemain-pemain muda Indonesia, memungkinkan mereka menguji kemampuan mereka melawan tim-tim terbaik dari seluruh dunia. Ini juga membantu meningkatkan profil sepak bola remaja di Indonesia dan berkontribusi pada pertumbuhan olahraga secara keseluruhan di negara ini (Natasa Kumalasah Putri, 2023).

Penelitian ini menggunakan tagar viral U17. Dengan menggunakan *natural language processing* untuk *text processing* (Jimly Hanif et al., 2023) dan *machine learning* menggunakan algoritma SVM sebagai klasifikasi sentimen (Muhammad Furqan, Sriani, 2023). Evaluasi performa model menggunakan *confusion matrix* (Nurkholis et al., 2022).

## METODE PENELITIAN



Gambar 1 Diagram Penelitian

Penelitian ini dilakukan untuk melakukan klasifikasi tweet terkait tagar U17 menggunakan metode support vector machine. Tahapan penelitian dimulai mengumpulkan data menggunakan metode *crawling data* pada media sosial X menggunakan *library tweet harvest*, setelah data berhasil dikumpulkan tahap selanjutnya adalah *cleaning* dengan menghapus tagar, emotikon dan karakter spesial. Setelah data telah selesai di *cleaning* tahap selanjutnya melakukan pelabelan data menggunakan *vader sentiment*, data diberi label berdasarkan nilai *coumpon score* yang telah didapatkan berdasarkan *vader sentimen*, data yang memiliki *score < 0* akan di labeli dengan sentimen negatif, sedangkan *score >= 0* akan diberi label sentimen positif. Setelah dilabeli selanjutnya masuk ke tahap *text processing* (*case folding, stopword, tokenizing* dan *stemming*), membagi data train dan test dengan perbandingan 80:20 (Gholamy et al., 2018), melakukan pembobotan kata dengan TF-IDF, melakukan klasifikasi dengan support vector machine dan diakhiri dengan melakukan evaluasi performa model klasifikasi.

## HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data hasil dari *crawling* tweet pada X dengan total 2356 data. Data tweet yang di unduh adalah data yang menggunakan berbahasa indonesia dan mengacu dari beberapa tweet yang mengandung kata kunci U17 yang populer dibicarakan di Twitter. Rentang waktu data yang di dapatkan yaitu pada bulan 10 November 2023 hingga 07 Desember 2023. Pengumpulan data dilakukan dengan *crawling* menggunakan *library tweet harvest* menggunakan bahasa pemrograman python dan google collaboratory sebagai penulisan source code program. Data yang dikumpulkan yang berkaitan dengan U-17. Selanjutnya dilakukan *cleaning* dan pelabelan data hasil *crawling* dapat dilihat pada tabel berikut:

Tabel 1 Hasil *Cleaning* dan pelabelan

| <i>Cleaning</i>                                                                                                                 | Pelabelan |
|---------------------------------------------------------------------------------------------------------------------------------|-----------|
| Radja Nainggolan Beri Semangat Para Pemain Timnas U PialaDuniaU radjanainggolan TimnasDay                                       | Positif   |
| Penampilan gugup Timnas Indonesia U dalam laga debut Piala Dunia U membangkitkan memori Mundari Karya Ia teringat Piala Dunia U | Negatif   |

Setelah melakukan pelabelan pada data maka selanjutnya langkah untuk text preprocessing yaitu *case folding, stopword removal, tokenizing* dan *stemming*.

Tabel 2 *Text Processing*

|                                                                                          |
|------------------------------------------------------------------------------------------|
| <i>Case folding</i>                                                                      |
| radja nainggolan beri semangat para pemain timnas u pialadunia radjanainggolan timnasday |
| <i>Stopword Removal</i>                                                                  |
| radja nainggolan semangat pemain timnas u pialadunia radjanainggolan timnasday           |
| <i>Tokenizing</i>                                                                        |
| radja,nainggolan,semangat,pemain,timnas,u,pialadunia,radjanainggolan,timnasday           |
| <i>Stemming</i>                                                                          |
| radja nainggolan semangat main timnas u pialadunia radjanainggolan timnasday             |

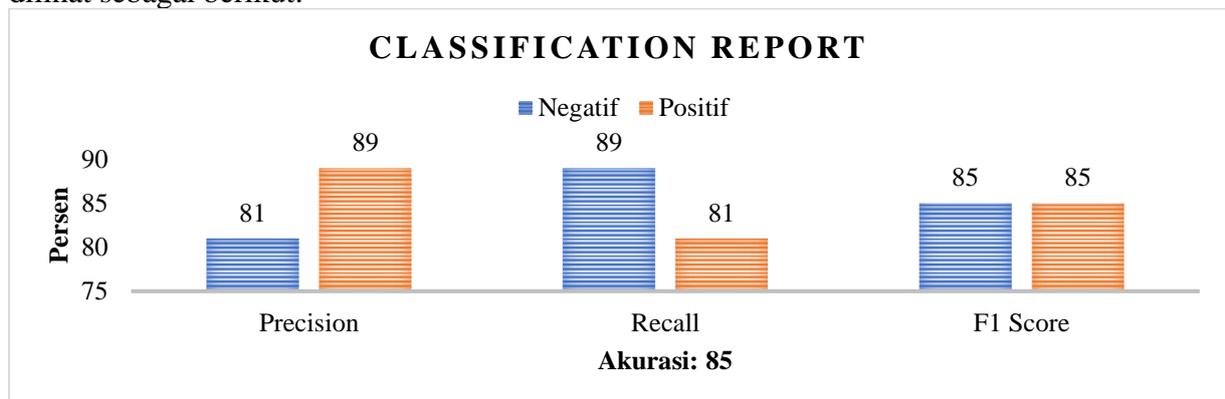
Pada proses pembagian data terdapat beberapa tahap yaitu: Membuat vektorisasi pada data text menggunakan fungsi *vectorizer* untuk mengubah text menjadi numerik, agar memudahkan dalam proses klasifikasi data. Melakukan penyeimbangan data menggunakan

*random oversampling* dengan hasil label sentimen positif dan negatif masing-masing berjumlah 1288. Melakukan pembagian data latih dan uji dengan perbandingan 80:20. Selanjutnya dilakukan pembobotan kata menggunakan TF-IDF jumlah kata pada penelitian ini berjumlah 3969 kata. Berikut Merupakan TF-IDF dari teks sample yang dapat dilihat pada tabel berikut:

Tabel 3 Hasil TF IDF

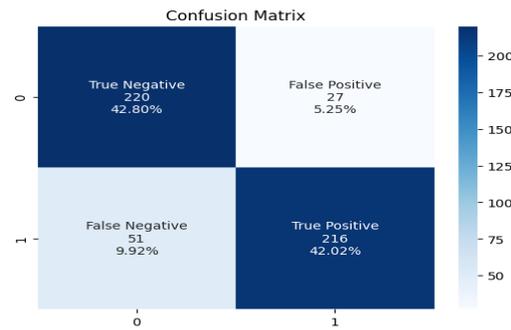
| <b>Stemming</b>                | <b>Kata</b>     | <b>Tf-Idf</b>       |
|--------------------------------|-----------------|---------------------|
| Radja nainggolan semangat main | Timnasday       | 0.40688039965061645 |
| timnas u pialadunia            | Radjanainggolan | 0.5031070118916848  |
| radjanainggolan timnasday      | Pialadunia      | 0.22849821527834296 |
|                                | Timnas          | 0.17342087552283372 |
|                                | Main            | 0.22065790002078953 |
|                                | Semangat        | 0.32395212454966105 |
|                                | Nainggolan      | 0.42409773997442096 |
|                                | Radja           | 0.40688039965061645 |

Pada tahap ini dilakukan pembuatan model menggunakan algoritma support vector machine dengan kernel yang digunakan adalah kernel linear. Berdasarkan hasil *training* dan *testing* pada model SVM yang dibuat hasil akurasi yang didapatkan yaitu 85%. Setelah membuat model svm maka langkah terakhir model yang telah dibuat dilakukan evaluasi berdasarkan *classification report* dan *confusion matrix*. Hasil *classification report* dapat dilihat sebagai berikut:



Gambar 2 *Classification Report*

Model klasifikasi SVM yang digunakan dalam analisis sentimen U17 menunjukkan kinerja yang baik dalam mendeteksi dan mengklasifikasikan sentimen positif dan negatif. Dengan *precision* dan *recall* yang tinggi untuk kedua kelas, serta *F1 score* yang seimbang, model ini dapat diandalkan untuk memprediksi sentimen pengguna dengan akurasi 85%. Adapun hasil confusion matrix disajikan sebagai berikut berikut:



Gambar 3 Confusion Matrix

Pada Gambar diatas hasil *confusion matrix* damenunjukkan bahwa sebanyak 27 data atau 5.25 % sentimen negatif yang di prediksi sebagai sentimen positif dan sebanyak 51 data atau 9.92% sentimen positif diprediksi sebagai sentimen negatif. Dalam analisis sentimen, *word cloud* dapat membantu mengidentifikasi kata-kata yang paling sering digunakan dalam konteks positif atau negatif, memberikan wawasan tambahan tentang sentimen keseluruhan dari teks. Adapun hasil *word cloud* pada sentimen positif dan negatif adalah sebagai berikut:



Gambar 4 Wordcloud

**SIMPULAN**

Hasil sentimen analisis pada tagar U17 adalah negatif mendapat dengan nilai *f1 score* 85% dan *f1 score* positif 85% artinya sentimen tidak condong kepositif maupun negatif karena hasilnya imbang. *F1 score* digunakan sebagai tolak ukur sentimen dikarenakan memberikan gambaran yang lebih seimbang mengenai performa model. Kinerja model svm pada penelitian ini diukur berdasarkan nilai akurasi, hasil akurasi model SVM yang didapat pada penelitian ini yaitu 85%. Berdasarkan akurasi tersebut model memiliki kinerja sebesar 85% dalam melakukan analisis sentimen pada tagar U17.

**DAFTAR PUSTAKA**

Alfarobi, I., Wirahadi, S., & Widiyanto, K. (2023). *MENGGUNAKAN HYPERPARAMETER TUNNING SVM DAN LOGISTIC REGRESSION*. 7(3), 854–861. <https://doi.org/10.52362/jisamar.v7i3.771>

- Alifia Seftin Oktriwina. (2022). *Hashtag: Apa Itu, Cara Kerja, Manfaat, dan Tips Menggunakannya*. Glints. <https://glints.com/id/lowongan/hashtag-adalah/>
- Annur, C. M. (2023). *Pengguna Twitter di Indonesia Capai 24 Juta hingga Awal 2023, Peringkat Berapa di Dunia?* Databooks. <https://databoks.katadata.co.id/datapublish/2023/02/27/pengguna-twitter-di-indonesia-capai-24-juta-hingga-awal-2023-peringkat-berapa-di-dunia>
- Dahlia, R. B., & Sitanggang, D. (2023). *Analisis Sentimen Migrasi Televisi Digital pada Twitter Menggunakan Perbandingan Algoritma Multinomial Naïve Bayes , Support Vector Machines , dan Logistic Regression*. 12, 280–288.
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *A Pedagogical Explanation A Pedagogical Explanation Part of the Computer Sciences Commons*. [https://scholarworks.utep.edu/cs\\_techrephttps://scholarworks.utep.edu/cs\\_techrep/1209](https://scholarworks.utep.edu/cs_techrephttps://scholarworks.utep.edu/cs_techrep/1209)
- Jimly Hanif, A., Nur Farid, M., & Hasanah, B. (2023). Penerapan Natural Language Processing untuk Klasifikasi Bidang Minat berdasarkan Judul Tugas Akhir. *Jurnal Sistim Informasi Dan Teknologi*, 5(1), 41–49. <https://doi.org/10.37034/jsisfotek.v5i1.196>
- Kambey, W. M., Santa, K., & Togas, P. V. (2021). Penerapan Model Pembelajaran Problem Based Learning Pada Mata Pelajaran Multimedia Di Smk. *Edutik : Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, 1(2), 195–208. <https://doi.org/10.53682/edutik.v1i2.2258>
- Kaparang, S. G., Kaparang, D. R., & Rantung, V. P. (2021). Analisis Sentimen New Normal Pada Masa Covid-19 Menggunakan Algoritma Naïve Bayes. In *JOINTER-JOURNAL OF INFORMATICS ENGINEERING* (Vol. 02, Issue 01). <https://t.co/CxptQpYGU9>
- Kurniawan, I., Lia Hananto, A., Shofia Hilabi, S., Hananto, A., Priyatna, B., Yuniar Rahman, A., Studi Sistem Informasi, P., Ilmu Komputer, F., & Buana Perjuangan, U. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 10(1), 731–740.
- Muhammad Furqan, Sriani, M. N. S. (2023). Chatbot Telegram Menggunakan Natural Language Processing. *WJIT: Walisongo Journal of Information Technology Chatbot*, 5(1), 15–26. <https://doi.org/10.1002/9781405164535.ch14>
- Natasa Kumalalah Putri. (2023). *Timnas Indonesia Bantai Brunei 6-0, Ini Kata Pelatih Brunei*. Liputan6. <https://www.liputan6.com/regional/read/5422584/timnas-indonesia-bantai-brunei-6-0-ini-kata-pelatih-brunei?page=3>
- Nurkholis, A., Alita, D., & Munandar, A. (2022). Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(2), 227–233. <https://doi.org/10.29207/resti.v6i2.3906>
- Salsabila, S. N., Sari, B. N., Mayasari, R., Komputer, I., & Karawang, U. S. (2023). *KLASIFIKASI ULASAN PENGGUNA APLIKASI DISCORD MENGGUNAKAN*. 9(2), 383–392.
- Solecha, K., & Irnawati, O. (2023). *Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Berbasis Particle Swarm Optimization Pada Analisis Sentimen Ulasan Aplikasi Flip*. 07, 10–15.
- Undap, M. G., Rantung, V. P., & Rompas, P. T. D. (n.d.). Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based. In *JOINTER-JOURNAL OF INFORMATICS ENGINEERING* (Vol. 02, Issue 02).