



Analisis Sentimen Komentar pada Lagu “Takut” Idigitaf Menggunakan SVM Kernel Linear dan Polynomial

Desi Daomara Sitanggang^{1*}, Wahyu S. J. Saputra², Muhammad Nasrudin³
^{1,2,3} Universitas Pembangunan Nasional “Veteran” Jawa Timur
* E-mail: 22083010027@student.upnjatim.ac.id

Abstract

The development of social media, particularly YouTube, has become a platform for people to express their opinions about musical works. The song “Takut” by Idigitaf is one of the pieces that has attracted much attention from listeners because it tells about fear and anxiety in entering adulthood. The great enthusiasm of listeners has generated thousands of diverse comments on the music video. Sentiment analysis of the comments uses a Natural Language Processing (NLP) approach to detect emotional patterns based on two labels, namely positive and negative. This study aims to conduct sentiment analysis on comments about the song “Takut” using the Support Vector Machine (SVM) algorithm by comparing the linear and polynomial kernels. Before the analysis, the data go through preprocessing, labeling, and feature extraction stages. In data labeling, a lexicon-based method is used and then manually updated, where words such as takut, nangis, and sedih are categorized as positive sentiments. The study also applies random undersampling to address data imbalance. Based on testing of 13,037 comments, SVM with a linear kernel showed superior performance with an accuracy of 91%, compared to the polynomial kernel which achieved 87%. This indicates that the linear kernel is more effective in classifying sentiments in this type of text data.

Keywords: Sentiment Analysis, SVM, Idigitaf, Youtube.

Abstrak

Perkembangan media sosial, khususnya Youtube telah menjadi wadah bagi masyarakat untuk mengekspresikan opini terhadap karya musik. Lagu “Takut” karya Idigitaf adalah salah satu musik yang menarik banyak perhatian pendengar karena menceritakan tentang ketakutan kecemasan saat memasuki pendewasaan. Banyaknya antusias pendengar sehingga menghasilkan ribuan komentar yang beragam pada video musik tersebut. Analisis sentimen komentar menggunakan pendekatan NLP untuk mendeteksi pola emosi berdasarkan dua label yaitu positif dan negatif. Penelitian ini bertujuan melakukan analisis sentimen pada komentar lagu “Takut” menggunakan algoritma *Support Vector Machine (SVM)* dengan membandingkan kernel linear dan kernel polynomial. Sebelum dilakukan analisis, data melalui tahap preprocessing, pelabelan, ekstraksi fitur. Pada pelabelan data dilakukan menggunakan lexicon based dan diperbarui manual seperti takut, nangis, sedih dan lainnya menjadi sentimen positif. Penelitian ini juga menerapkan teknik random undersampling untuk mengatasi ketidakseimbangan data. Berdasarkan hasil pengujian terhadap 13.037 komentar, SVM dengan kernel linear menunjukkan kinerja yang lebih unggul dengan akurasi 91%, dibandingkan kernel polynomial yang mencapai 87%. Hal ini menunjukkan bahwa kernel linear lebih efektif dalam mengklasifikasikan sentimen pada data teks seperti ini.

Kata kunci: Analisis Sentimen, SVM, Idigitaf, Youtube.

PENDAHULUAN

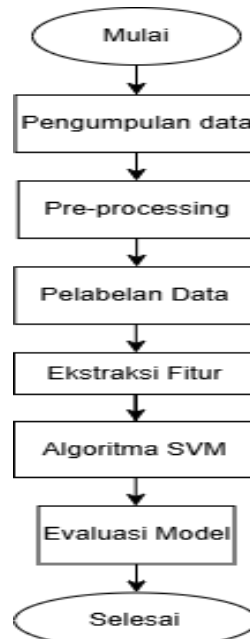
Perkembangan teknologi saat ini telah membawa perubahan yang signifikan dalam kehidupan manusia, khususnya dalam cara masyarakat berkomunikasi atau mengekspresikan pendapat mereka melalui platform digital. Media sosial seperti, *YouTube*, *Twitter*, *Instagram* dan lainnya, menjadi wadah bagi masyarakat berkomunikasi, menyampaikan pandangan, kritik dan apresiasi terhadap suatu postingan, termasuk musik. Dibidang musik, *YouTube* adalah salah satu platform yang terpopuler karena tidak hanya memposting suatu karya musik dalam bentuk video, tetapi platform *Youtube* juga memberikan fitur yang interaktif untuk mengetahui respon publik (Alviyanti et al. 2025). Fitur interaktif yang berupa kolom komentar pada platform *Youtube* memungkinkan masyarakat memberikan pendapat secara bebas terhadap suatu karya musik. Melalui komentar tersebut, beragamnya pendapat masyarakat, mulai dari mendukung hingga mengkritik (Abel Laia and Barus 2025). Oleh karena itu, komentar-komentar publik ini dapat menjadi sumber data yang penting untuk membantu industri musik dalam memahami pendapat publik pada suatu karya tersebut.

Salah satu karya musik dari seorang musisi Idgitaf, lagu yang berjudul "Takut" telah menarik banyak perhatian pendengar. Lagu ini menceritakan tentang ketakutan dan kecemasan saat memasuki usia dewasa, terutama di masa *quarter life crisis* (Kurniawan, Y. W., & Rizki 2023). Tema yang sangat relevan dengan pendengar ini memicu ribuan interaksi pada kolom komentar musik *video officialnya*. Komentar tersebut tidak hanya berisi pujian, tetapi juga curahan hati mengenai pengalaman pribadi mereka yang mendalam. Dengan besarnya jumlah komentar, menjadi tantangan tersendiri jika dianalisis secara manual. Oleh karena itu, diperlukan pelabelan pada data opini publik secara otomatis dan efisien. Penelitian ini menerapkan pelabelan otomatis menggunakan *Lexicon based* serta dilakukannya penyesuaian pada kamus leksikon dengan manual untuk menangani kata-kata yang memiliki sentimen khusus dalam konteks lagu ini. Setelah pelabelan, data komentar seringkali memiliki hasil distribusi data yang tidak seimbang.

Analisis sentimen merupakan bagian dari penambangan teks, menjadi semakin relevan dalam memahami opini publik terhadap berbagai entitas, termasuk karya seni seperti lagu (Natasuwarna et al. 2020). Studi ini berfokus pada analisis sentimen komentar publik terhadap lagu "Takut" dari Idgitaf, sebuah karya yang seringkali memicu beragam interpretasi emosional. Dengan memanfaatkan algoritma *Support Vector Machine* dengan kernel *Linear* dan *Polynomial*, penelitian ini bertujuan untuk mengklasifikasikan sentimen positif, negatif dari komentar-komentar tersebut (Rabbani, Safitri, and Rahmadhani 2023). Pendekatan ini dipilih karena *SVM* telah terbukti efektif dalam tugas klasifikasi teks, termasuk analisis sentimen, dengan kemampuan generalisasi yang baik pada data yang kompleks (Ananda and Pristyanto 2021; Rofiqi and Akbar 2024; Wahyuni, N. A., Ayu, D. P., & Irsyad 2024). Untuk mengatasi masalah ketidakseimbangan data yang telah disebutkan sebelumnya, penelitian ini menerapkan teknik *undersampling* sebelum penerapan *SVM* dilakukan. Penelitian ini bertujuan untuk menganalisis sentimen komentar pada lagu "Takut" dengan membandingkan kinerja *SVM* kernel *linear* dan kernel *polynomial*. Perbandingan ini dilakukan untuk mengetahui kernel mana yang paling optimal dalam data opini publik seperti ini dan yang telah melalui proses pelabelan *lexicon* serta *undersampling* data untuk mengatasi ketimpangan kelas.

METODE PENELITIAN

Pada penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen yang bersifat komparatif. Metode yang digunakan pada penelitian ini yaitu *Support Machine Vector* dengan kernel *Linear* dan *Polynomial*. Dimana pada beberapa penelitian yang membandingkan metode pada penerapan analisis sentimen dan hasilnya menunjukkan bahwa *SVM* menghasilkan akurasi tertinggi dan terbaik. Dengan tahapan seperti pada gambar 1.



Gambar 1. Rancangan Penelitian

Pada penelitian ini, tahapan-tahapan sebagai berikut:

1. Pengumpulan data

Pengumpulan data merupakan tahapan awal dalam penelitian. Pengumpulan data dilakukan dengan cara *crawling* data menggunakan *API* pada youtube menggunakan *google colab*. Waktu pengumpulan data mulai dari rilisnya musik video lagu “Takut” di Youtube hingga 18 September 2025.

2. Pre-processing

Pre-processing dilakukan beberapa tahap, sebagai berikut:

a. *cleaning data*

Cleaning data merupakan tahap awal dari *preprocessing*, dimana seluruh atribut pada dataset *diselection* sedemikian untuk mendapatkan atribut-atribut yang berisi nilai-nilai yang relevan tidak terdapat *missing value* dan *duplicate data*. Dikatakan *missing value* adalah jika atribut-atribut dalam dataset tidak berisi nilai atau nilai kosong (Purnomo et al. 2021).

b. *case folding*

case folding ini merupakan tahapan untuk menyamaratakan huruf yang sebelumnya terdapat kapital dan ditahap ini akan diubah menjadi huruf kecil semua (Alam, S., & Sulisty 2023).

c. normalisasi data

normalisasi data merupakan tahapan untuk pembentukan struktur basis data, sehingga sebagian besar ambiguitas bisa dihilangkan (Suryadi 2019).

d. *tokenizing*

tokenizing merupakan tahapan untuk memisahkan kalimat menjadi kata per kata (Alam, S., & Sulisty 2023).

e. *stopword removal*

Stopwords merupakan tahapan untuk menghilangkan kata yang nilai informasinya rendah (Shevira et al. 2022).

f. *stemming*

stemming merupakan tahapan untuk mengubah semua kata menjadi kata dasarnya dan mengembalikannya ke dalam bentuk kalimat (Permata Aulia, Arifin, and Mayasari 2021).

3. Pelabelan data

Pelabelan data merupakan proses memberikan label pada tiap data, seperti positif atau negatif. Pada tahapan ini dilakukan pelabelan data menggunakan *lexicon based*, dimana

lexicon based merupakan kamus yang berisi kata berbahasa Indonesia yang biasanya digunakan untuk pelabelan data dengan polaritas berdasarkan score (Pratama, Ramadhan, and Komara 2023).

4. Ekstraksi Fitur

Ekstraksi fitur merupakan proses mengubah kata menjadi data numerik. Pada penelitian ini menggunakan TF-IDF, *frekuensi dokumen frekuensi inverse* (TF-IDF) adalah proses pembobotan dimana kata akan diekstraksi ke dalam file bentuk nilai.

5. Algoritma SVM

Support Vector Machine merupakan metode pembelajaran supervised yang menganalisis data dan mengenali pola yang digunakan untuk klasifikasi (Rahman, O. H., Abdillah, G., & Komarudin 2021). Metode SVM mengandalkan vektor pendukung untuk memisahkan kelas data yang memiliki ciri-ciri yang berbeda. Konsep dari algoritma SVM ini yaitu mencari hyperplane yang optimal (Rabbani, Safitri, and Rahmadhani 2023). SVM mencoba menemukan *hyperplane* yang paling baik memisahkan data komentar dengan sentimen positif dan negatif. Penelitian ini mencoba untuk menerapkan kernel linear dan kernel polynomial dari SVM, pemilihan kernel tersebut karena mengambil jenis kernel yang sangat sederhana (*linear*) dan jenis kernel yang sangat kompleks (*polynomial*). Berikut penjelasan tentang fungsi kernel dan polynomial. :

a. Kernel *Linear*

Kernel linear untuk menghitung (dot product) antara dua vektor input, yang secara efektif mengukur seberapa mirip kedua vektor tersebut dalam ruang aslinya.

$$K(x, x^l) = x^T x^l \quad (1)$$

b. Kernel *Polynomial*

Kernel *polynomial* untuk menghitung (dot product) dari vektor data dalam ruang asli dan menaikkannya ke suatu pangkat, secara efektif mensimulasikan perhitungan dalam ruang dimensi yang lebih tinggi.

$$K(x, x^l) = (x^T x^l + r)^d \quad (2)$$

Keterangan:

x, x^l, x, x^l : Vektor

$x^T x^l, x^T x^l$: Dot product

rr : Suku bias

dd : Degree polinomial

6. Evaluasi model

Evaluasi model dilakukan untuk memeriksa hasil dari model yang telah dibangun. Hasil penelitian ini dilakukan pengujian untuk mengetahui ketepatan dari hasil penelitian yang telah dilakukan. Pada penelitian ini akan menghitung nilai akurasi, *precision*, *recall*, *f-measure* serta *confusion matrix*.

HASIL DAN PEMBAHASAN

Langkah pertama yang dilakukan pada penelitian ini adalah mengumpulkan data dari komentar youtube. Berdasarkan proses pengumpulan data, diperoleh sebanyak 13.037 komentar dari video musik lagu "Takut" karya Idris. Data tersebut diperoleh dengan cara crawling menggunakan google colab dengan rentang waktu mulai dari rilis video hingga 18 september 2025. Pada gambar 2 merupakan sampel data yang diperoleh masih berupa teks tidak terstruktur dan diperlukan pembersihan serta pelabelan data sebelum pemodelan data.

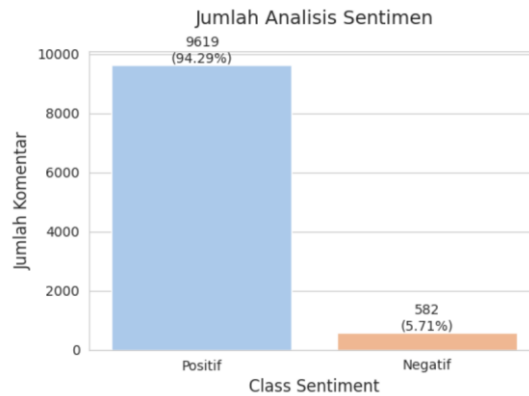
Gambar 2. Sampel Data

Langkah dalam *preprocessing data* yang dilakukan yaitu penghapusan *link*, emoji, simbol, angka, *username*, dan komentar yang bahasa inggris, mengubah ke huruf kecil, melakukan normalisasi data, melakukan *tokenisasi*, melakukan *stopwords* dan *stemming*.

Tabel 1. Hasil *preprocessing*

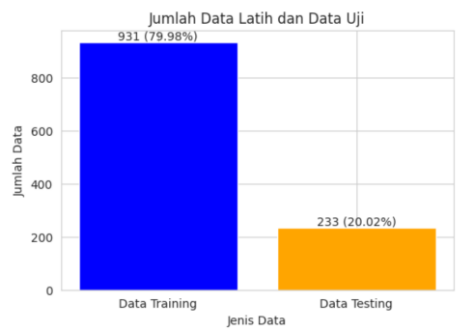
Komentar	Cleaning	Case Folding	Normalisasi	Tokenisasi	Stopword	Stemming
Terharu banget sama semua komentar disiniii, b...	Terharu banget sama semua komentar disiniii bi...	terharu banget sama semua komentar disiniii bi...	terharu banget sama semua komentar disini biki...	['terharu', 'banget', 'sama', 'semua', 'koment...']	['terharu', 'banget', 'komentar', 'bikin', 'be...']	haru banget komentar bikin berat juang thankyo...
Terharu, senang, baca cerita-cerita teman-tema..	Terharu senang baca ceritacerita temanteman Se...	terharu senang baca ceritacerita temanteman se...	terharu senang baca ceritacerita temanteman se...	['terharu', 'senang', 'baca', 'ceritacerita', '...']	['terharu', 'senang', 'baca', 'ceritacerita', '...']	haru senang baca ceritacerita temanteman semangat
Semangat!!	Semangat	semangat	semangat	['semangat']	['semangat']	semangat
Emang realite bgt sih ini lagu apalagi di masa...	Emang realite bgt sih ini lagu apalagi di masa...	emang realite bgt sih ini lagu apalagi di masa...	memang relevan banget sih ini lagu apalagi di ...	['memang', 'relevan', 'banget', 'sih', 'ini', '...']	['relevan', 'banget', 'sih', 'lagu', 'timbang'...']	relevan banget sih lagu bimbang tentu kuiahker...
sial aku teringat benda menggantung itu	sial aku teringat benda menggantung itu.	sial aku teringat benda menggantung itu	sial aku teringat benda menggantung itu	['sial', 'aku', 'teringat', 'benda', 'menggant...']	['sial', 'benda', 'menggantung']	sial benda gantung

Setelah dilakukan tahap *preprocessing*, selanjutnya pelabelan data menggunakan kamus *lexicon* dan dilakukan pembaruan manual dalam kamus karena konteks lagu yang menceritakan ketakutan dan kecemasan saat memasuki usia 20an. Pembaruan dilakukan seperti kata “takut”, “nangis”, “cemas”, “relate” dan lain sebagainya, dimana kata-kata tersebut terdapat pada sentimen negatif pada kamus *lexicon*. Oleh karena itu, dilakukan pembaruan ini untuk mengubah kata-kata tersebut menjadi sentimen positif. Berikut distribusi data setelah dilakukannya tahap pelabelan.



Gambar 3. Distribusi Sentimen

Dapat dilihat pada gambar 3, hasil pelabelan data menunjukkan adanya ketidakseimbangan kelas yang signifikan, dimana sentimen positif mendominasi data. Untuk mengatasi bias pada model *SVM*, dilakukan penyeimbangan data terlebih dahulu yaitu *random undersampling*. Selanjutnya datanya dipisahkan menjadi 2 yaitu data training dan testing dengan perbandingan 8:2. Berikut distribusi data setelah melakukan data *splitting*.



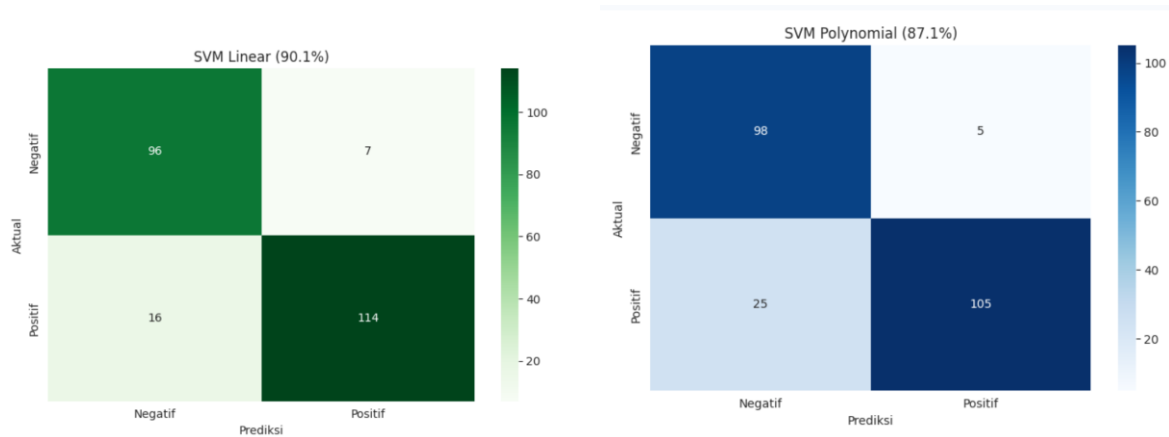
Gambar 4. Data Splitting

Selanjutnya penelitian ini akan mencoba membandingkan hasil dari kernel *linear* dan *polynomial SVM*. Kemudian pengujiannya menghitung nilai akurasi dari klasifikasi sentimen tersebut. Hasilnya seperti pada Tabel 2.

Tabel 2. Hasil Akurasi

Kernel	Akurasi	Precision	Recall	F1-score
Linear	0,90	0,90	0,90	0,90
Polynomial	0,87	0,88	0,87	0,87

Tabel 2 menunjukkan bahwa kernel *linear* lebih optimal daripada kernel *polynomial* dengan hasil yang didapatkan 0,90 dan *polynomial* 0,87. Oleh karena itu pada penelitian ini kernel *linear* dianggap lebih optimal daripada kernel *polynomial*. Berikut adalah gambar dari *confusion matrix*.



Gambar 5. Confusion Matrix

Gambar 5, menunjukkan bahwa kernel linear lebih optimal dari mengklasifikasikan sentimen pada komentar Youtube pada lagu “Takut”. Keunggulan kernel *linear* terletak pada penyeimbangan dalam mendeteksi kedua kelas sentimen. Sebaliknya, meskipun kernel *polynomial* memiliki *False Positive* lebih sedikit yaitu 5 data dibandingkan dengan kernel *linear* sebanyak 7 data, namun kernel ini memiliki kelemahan yang signifikan bahwa kernel *polynomial* cenderung lebih sering gagal mengenali sentimen positif dibandingkan kernel linear.

PENUTUP

Berdasarkan hasil dan pembahasan maka dapat disimpulkan, bahwa sentimen masyarakat pada lagu “Takut” karya Idgitaf lebih condong ke sentimen positif. Hasil evaluasi model yang dilakukan pada kernel linear mendapatkan akurasi 90% sedangkan pada kernel polynomial mendapatkan akurasi 87%. Dimana dapat ditarik kesimpulan bahwa metode SVM dengan kernel linear lebih baik dari SVM kernel polynomial pada dataset ini. Pada penelitian ini, analisis sentimen terbatas hanya menggunakan Bahasa Indonesia dan pelabelan datanya secara otomatis. Oleh karena itu diharapkan, untuk kedepannya dapat dikembangkan lagi menggunakan Bahasa Inggris atau pelabelan datanya secara manual.

DAFTAR PUSTAKA

- Abel Laia, Nico, and Simon Prananta Barus. 2025. “Analisis Sentimen YouTube: ‘Di Balik Ambisi Jokowi Dalam IKN.’” *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)* 5(1): 07–12. doi:10.55382/jurnalpustakaai.v5i1.891.
- Alam, S., & Sulistyono, M. I. 2023. “Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes 1,2,3.” 2(3): 100–108.
- Alviyanti, Risma, Nur Hidayat, Muh Said, Universitas Islam, and Negeri Alauddin. 2025. “Respon Publik Terhadap Strategi Dakwah Melalui Stand Up Comedy Pada Platform YouTube.” 4: 21–31.
- Ananda, Fadhilah Dwi, and Yoga Pristyanto. 2021. “Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Twitter Users on Internet Service Providers Using Support Vector Machine Algorithm.” 20(2): 407–16. doi:10.30812/matrik.v20i2.1130.
- Kurniawan, Y. W., & Rizki, R. C. 2023. “REPRESENTASI QUARTER LIFE CRISIS MELALUI ANALISIS WACANA PADA LIRIK LAGU ‘ TAKUT ’ KARYA IDGITAF TERHADAP REALITA MENTAL HEALTH.” 5(2): 486–510.
- Natasuwarna, Amar P, Jurusan Sistem Informasi, Analisis Sentimen, and Pembelajaran Daring. 2020. “Seleksi Fitur Support Vector Machine Pada Analisis Sentimen Keberlanjutan Pembelajaran Daring.” 19(4): 437–48.

- Permata Aulia, Thalita Meisya, Nur Arifin, and Rini Mayasari. 2021. "Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19." *SINTECH (Science and Information Technology) Journal* 4(2): 139–45. doi:10.31598/sintechjournal.v4i2.762.
- Pratama, Muhamad Rizky, Yudhi Raymond Ramadhan, and Mutiara Andayani Komara. 2023. "Analisis Sentimen BRImo Dan BCA Mobile Menggunakan Support Vector Machine Dan Lexicon Based."
- Purnomo, Bambang Setio, Putri Taqwa Prasetyaningrum, Universitas Mercu, Buana Yogyakarta, Data Mining, and Algoritma K-means. 2021. "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN KUNJUNGAN WISATAWAN DI KOTA YOGYAKARTA MENGGUNAKAN METODE K-MEANS." 1(1): 27–32.
- Rabbani, Salsabila, Dea Safitri, and Nadila Rahmadhani. 2023. "Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis Perbandingan Evaluasi Kernel SVM Untuk Klasifikasi Sentimen Dalam Analisis Kenaikan Harga BBM." 3(October): 153–60.
- Rahman, O. H., Abdillah, G., & Komarudin, A. 2021. "Klasifikasi Ujaran Kebencian Pada Media Sosial Twitter Menggunakan." 1(10): 17–23.
- Rofiqi, Lutfi, and Mutaqin Akbar. 2024. "Analisis Sentimen Terkait RUU Perampasan Aset Dengan Support Vector Machine." 4(3).
- Shevira, Sheila, I Made Agus, Dwi Suarjaya, and Putu Wira. 2022. "Pengaruh Kombinasi Dan Urutan Pre-Processing Pada Tweets Bahasa Indonesia." 3(2).
- Suryadi, Sudi. 2019. "Implementasi Normalisasi Dalam Perancangan Database Relational." 3(2): 1–5.
- Wahyuni, N. A., Ayu, D. P., & Irsyad, H. 2024. "Analisis Sentimen Di Youtube Terhadap Kenaikan UKT Menggunakan Metode Support Vector Machine Salah Satu Teknik Atau Metode Yang Paling Efektif Untuk Analisis Sentimen Ini Adalah Support Vector Machine (SVM). SVM Merupakan Teknik Machine Learning Yang Ba." 4(1): 57–71.