

PENERAPAN METODE *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN PUBLIK PADA PROGRAM MAKAN BERGIZI GRATIS

Dimas Cahyo Nugroho¹, Amelia Dwi Agustina², Bagas Maulana³, Firmansyah Darussalam⁴, Muhammad Asyraf⁵, Baginda Oloan Lubis⁶

^{1,2,3,4,5,6}Universitas Bina Sarana Informatika, Kota Depok, Indonesia

Responden: 19232509@bsi.ac.id

ABSTRAK

Penelitian ini dilakukan untuk mengevaluasi sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) sebagai salah satu kebijakan sosial pemerintah yang berfokus pada peningkatan asupan gizi peserta didik di Indonesia. Ragam opini yang tersebar di media sosial menggambarkan persepsi publik terhadap implementasi kebijakan tersebut. Untuk mengidentifikasi kecenderungan opini tersebut, penelitian ini menggunakan pendekatan *text mining* dengan dukungan algoritma *machine learning Support Vector Machine (SVM)*. Dataset penelitian diperoleh dari platform Kaggle, yang memuat 3.459 komentar publik mengenai Program MBG. Data tersebut kemudian dibersihkan dan melalui proses *preprocessing* meliputi *tokenizing*, *case folding*, *stopword removal*, serta *stemming* menggunakan pustaka Sastrawi. Selanjutnya, pembobotan kata dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* guna mengubah teks menjadi representasi numerik. Model SVM dioptimasi dengan pendekatan *Grid Search* dan divalidasi menggunakan *5-Fold Cross Validation*, menghasilkan konfigurasi terbaik dengan kernel RBF, $C = 10$, dan $\text{gamma} = 1$. Evaluasi model menunjukkan performa tinggi dengan tingkat akurasi 96%, *precision* 95,44%, *recall* 96,27%, dan *F1-score* 95,85%. Berdasarkan hasil prediksi terhadap data tanpa label, diperoleh distribusi sentimen sebesar 57,43% positif dan 42,57% negatif. Hasil ini mengindikasikan bahwa sebagian besar masyarakat memberikan tanggapan positif terhadap Program MBG, sekaligus menunjukkan efektivitas algoritma SVM dalam mengklasifikasikan sentimen teks berbahasa Indonesia secara akurat dan konsisten.

Kata Kunci: *analisis sentiment, makan bergizi gratis, support vector machine, text mining*

Riwayat Artikel :

Tanggal diterima : 24-10-2025

Tanggal revisi : 19-11-2025

Tanggal terbit : 09-12-2025

DOI :

<https://doi.org/10.31949/infotech.v11i2.16781>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 2025 By Author



1. PENDAHULUAN

1.1. Latar Belakang

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan sosial ini menjadi sorotan publik karena menyangkut hak dasar anak atas pangan bergizi dan akses pendidikan yang inklusif. Menurut Kiftiyah et al. (2025) MBG tidak hanya menyentuh aspek kesehatan, tetapi juga menjadi bagian dari dinamika sosial dan politik yang mencerminkan keadilan sosial dalam kebijakan publik. Di tingkat internasional, program serupa telah diterapkan di berbagai negara seperti India dan Brasil, yang menunjukkan bahwa intervensi gizi di sekolah dapat meningkatkan prestasi belajar dan menurunkan angka putus sekolah (Bundy et al., 2024).

Respon terhadap MBG di Indonesia sangat beragam, terutama di media sosial yang menjadi ruang ekspresi publik. Sebagian masyarakat menyambut positif program ini karena dinilai mampu mengurangi beban ekonomi keluarga dan meningkatkan kualitas hidup anak. Namun, tidak sedikit pula yang meragukan efektivitas dan kesiapan implementasinya. Herdiana (2025) mengidentifikasi sejumlah faktor pendorong dan penghambat dalam pelaksanaan MBG, seperti kesiapan infrastruktur, koordinasi antar lembaga, dan partisipasi masyarakat. Dalam konteks ini, analisis sentimen publik menjadi penting untuk memahami persepsi masyarakat secara luas dan memberikan masukan berbasis data bagi pengambilan kebijakan.

Brata & Wijayanto (2025) melakukan analisis sentimen pengguna YouTube terhadap MBG menggunakan algoritma SVM, dan menemukan bahwa mayoritas komentar bersifat positif. Sementara itu, Pratama (2025) menggabungkan metode *stacking* antara *Naive Bayes*, SVM, dan *Random Forest* untuk menganalisis sentimen di platform X (sebelumnya Twitter), menunjukkan bahwa kombinasi algoritma dapat meningkatkan akurasi klasifikasi. Penelitian lain oleh Ramadhani et al. (2024) juga menggunakan *Naive Bayes* dan SVM untuk menganalisis tanggapan publik terhadap program makan siang gratis Prabowo-Gibran, dengan hasil yang menunjukkan dominasi sentimen positif namun disertai kritik terhadap aspek teknis pelaksanaan.

Meski demikian, terdapat celah penelitian yang belum banyak dieksplorasi, yaitu analisis sentimen publik terhadap MBG dengan pendekatan SVM yang terstandarisasi dan dioptimalkan menggunakan teknik *GridSearch* pada data terbuka. Sebagian besar penelitian sebelumnya menggunakan data terbatas dari satu platform atau tidak menjelaskan secara rinci proses preprocessing dan parameterisasi model. Dalam konteks ini, Fairuzabadi (2025) dalam bukunya “Mahir Data Science dengan Python” menekankan bahwa efektivitas algoritma SVM sangat dipengaruhi oleh pemilihan kernel, parameterisasi model, dan teknik ekstraksi fitur yang

digunakan. Ia menjelaskan bahwa kernel RBF (*Radial Basis Function*) sering digunakan untuk menangani data non-linear, dan pemilihan parameter seperti C dan γ harus dilakukan secara sistematis, misalnya melalui *Grid Search* dan *Cross Validation*. Buku tersebut juga menyoroti pentingnya *preprocessing* data teks, termasuk penghapusan karakter khusus dan penggunaan teknik seperti TF-IDF untuk meningkatkan performa klasifikasi.

Oleh karena itu, penelitian ini memusatkan perhatian pada analisis sentimen publik terhadap Program Makan Bergizi Gratis dengan menerapkan algoritma *Support Vector Machine (SVM)* yang terstandarisasi dan dioptimalkan menggunakan teknik *Grid Search*. Melalui pendekatan ini, penelitian bertujuan memperoleh klasifikasi sentimen yang lebih akurat serta mengidentifikasi sentimen dominan positif atau negatif sehingga dapat memberikan kontribusi dalam pengembangan sistem analisis opini publik berbasis kecerdasan buatan.

1.2. Tinjauan Pustaka

a. Text Mining

Text mining merupakan pengembangan dari *data mining* yang berfokus pada penggalian informasi dari data teks tidak terstruktur. Berbeda dengan *data mining*, *text mining* memerlukan tahap *preprocessing* untuk membersihkan teks dan mengubahnya menjadi bentuk numerik agar dapat diolah secara komputasional (Faisal et al., 2023). Metode yang sering digunakan adalah TF-IDF, yang berfungsi memberikan bobot pada setiap kata sesuai dengan frekuensi kemunculannya dalam suatu dokumen. Dalam penelitian ini, *text mining* digunakan untuk mengekstraksi dan mengklasifikasikan opini publik terhadap Program MBG.

b. Term Frequency–Inverse Document Frequency (TF-IDF)

Term Frequency–Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk menentukan bobot suatu kata berdasarkan seberapa sering kata tersebut muncul serta penyebarannya di antara sejumlah dokumen (R. R. Putra et al., 2024). *Term Frequency (TF)* digunakan untuk mengukur frekuensi kemunculan suatu kata dalam sebuah dokumen, sementara *Inverse Document Frequency (IDF)* memberikan bobot lebih tinggi pada kata yang jarang muncul di keseluruhan koleksi dokumen. Kombinasi keduanya menghasilkan nilai TF-IDF yang digunakan untuk mengonversi teks menjadi bentuk vektor numerik sebelum diproses oleh algoritma SVM.

Penelitian terdahulu menunjukkan bahwa teknik TF-IDF memberikan hasil yang lebih baik dibandingkan *Bag of Words (BoW)* saat digunakan bersama algoritma SVM. Kombinasi TF-IDF dan SVM menghasilkan akurasi 86%, sedangkan BoW dan SVM hanya unggul pada nilai recall sebesar 89%. Hal ini membuktikan bahwa TF-IDF lebih efektif

dalam merepresentasikan kata berdasarkan tingkat kepentingannya dalam dokumen, sehingga cocok diterapkan pada analisis sentimen berbasis teks (K. T. Putra et al., 2023).

c. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah algoritma *machine learning* yang digunakan untuk tugas klasifikasi dan regresi dengan cara mencari garis pemisah (*hyperplane*) yang mampu membedakan data antar kelas secara optimal (R. F. Putra et al., 2024). SVM menggunakan parameter utama C (penalty parameter) dan jenis kernel seperti *polynomial*, *linear*, atau *radial basis function (RBF)*. Algoritma ini mampu bekerja secara efisien pada data yang memiliki banyak fitur serta umum digunakan dalam pemrosesan teks, khususnya untuk analisis sentimen.

d. *Grid Search*

Grid Search merupakan salah satu teknik optimasi *hyperparameter* yang digunakan untuk meningkatkan performa model *machine learning*. Nugraha & Sasongko (2022), metode ini bekerja dengan menguji seluruh kombinasi nilai parameter yang telah ditentukan, kemudian memilih kombinasi terbaik berdasarkan hasil evaluasi model. *Grid Search* sering dipadukan dengan *Cross Validation* untuk memastikan bahwa parameter yang diperoleh tidak hanya optimal pada satu subset data, melainkan konsisten pada keseluruhan dataset.

Fajri & Primajaya (2023) menunjukkan bahwa *Grid Search* mampu menghasilkan akurasi yang sebanding dengan *Random Search*, yaitu rata-rata sebesar 84%. Namun, penggunaan *Grid Search* memerlukan waktu komputasi yang jauh lebih lama karena metode ini harus menguji seluruh kombinasi parameter yang telah ditentukan. Pada beberapa dataset, proses validasi *Grid Search* bahkan memerlukan jumlah percobaan yang jauh lebih besar dibandingkan *Random Search*. Meskipun demikian, hasil penelitian tersebut menegaskan bahwa dari sisi akurasi, *Grid Search* tetap kompetitif dan dapat digunakan secara efektif untuk pemilihan parameter pada model seperti *Support Vector Machine (SVM)*.

2. METODE

2.1. Desain Penelitian

Penelitian ini menggunakan desain kuantitatif eksperimental dengan metode *text mining* berbasis *machine learning*. Pendekatan ini digunakan karena mampu mengonversi teks ke dalam bentuk vektor numerik untuk dianalisis secara statistik menggunakan algoritma SVM.

Seluruh proses diimplementasikan menggunakan RapidMiner 2026.0.1 dengan bantuan Python Sastrawi untuk tahap *stemming* Bahasa Indonesia.

2.2. Tahapan Penelitian



Gambar 1. Tahapan penelitian

Proses penelitian terdiri dari delapan tahapan utama sebagaimana ditunjukkan pada Gambar 1, meliputi:

a. Pengumpulan Data

Dataset diperoleh dari sumber terbuka Kaggle (<https://www.kaggle.com/datasets/hanif281103/sentimen-publik-terhadap-makan-bergizi-gratis>) yang berisi 3.459 komentar publik terkait Program Makan Bergizi Gratis. Data diunduh pada 28 September 2025 dan hanya menggunakan kolom *full_text* sebagai objek analisis.

b. Pembersihan Data

Data dibersihkan menggunakan ekspresi reguler (*regex*) untuk menghapus URL, *mention*, *hashtag*, karakter khusus, dan karakter non-ASCII, serta menghapus nilai kosong dan data duplikat. Hasil pembersihan menghasilkan 2.983 data bersih.

c. Pelabelan dan Pembagian Data

Sebanyak 1.500 data diberi label manual (positif dan negatif) dan dibagi menjadi 80% data *training* serta 20% data *testing* menggunakan *stratified sampling* (*seed* = 42). Sisanya (1.483 data) digunakan untuk prediksi sentimen.

d. *Text Preprocessing*

Proses *stemming* dilakukan menggunakan Sastrawi sebelum tahap *Process Documents from Data*. Di RapidMiner, *text preprocessing* meliputi *tokenization*, *case folding*, *stopword removal*, *filter token by length* (4–25 karakter), dan pembobotan kata menggunakan metode TF-IDF.

e. *Modeling*

Model dikembangkan menggunakan algoritma SVM, dengan parameter optimal yang diperoleh melalui *Optimize Parameters (Grid)*. Proses optimasi melibatkan *5-Fold Cross Validation* dan menghasilkan parameter terbaik C = 10, gamma = 1, dengan kernel RBF.

f. Evaluasi Model

Kinerja model dievaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk menilai ketepatan dan keseimbangan hasil klasifikasi.

g. Prediksi dan Penggabungan Hasil

Model terbaik digunakan untuk memprediksi 1.483 data tanpa label menggunakan *Apply Model*. Hasil prediksi digabungkan dengan data berlabel

menggunakan *Join* untuk membentuk dataset akhir sebanyak 2.983 data yang siap dianalisis.

3. PEMBAHASAN

3.1. Pembersihan Data

Dataset dibersihkan dari berbagai macam noise seperti tautan (URL), *mention*, *hashtag*, karakter khusus, serta karakter non-ASCII. Data yang duplikat juga dihapus pada tahapan ini. Hasil akhir dari tahapan ini adalah sebuah file CSV yang sudah dibersihkan.



Gambar 2. Proses pembersihan dataset

Tabel 1. Jumlah data sebelum dan sesudah pembersihan

Tahapan	Nilai Akhir
Dataset Awal	3.459
Setelah Pembersihan	2.983

3.2. Pelabelan Data

Dataset diberi label berjumlah 1.500 data. Sementara itu, sebanyak 1.483 data lainnya tetap tidak berlabel dan dimanfaatkan sebagai data uji untuk prediksi sentimen. Sentimen akan dibagi menjadi dua kategori, yaitu positif dan negatif. Adapun sentimen yang bersifat informatif akan diberi kategori positif.

Tabel 2. Contoh komentar publik berdasarkan kategori sentimen

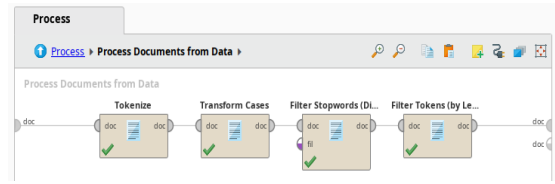
Kategori Sentimen	Contoh Komentar
positif	Sukseskan program makan bergizi gratis
negatif	di dunia ini gada yg gratis ya gusy makan bergizi gratis aja mengorbkan banyak hal

3.3. Pre-Processing Data

Dataset berlabel dibagi menjadi 80% data *training* dan 20% data *testing* untuk proses pelatihan dan pengujian model. Setelah pembagian dilakukan, kedua bagian data tersebut melalui tahap *stemming* menggunakan library Python Sastrawi guna menyeragamkan bentuk dasar kata.

Selanjutnya, data hasil *stemming* diproses dalam tahap *text preprocessing*, yang mencakup

tokenization, *case folding*, *stopword removal*, *filter token by length*, serta pembobotan menggunakan metode TF-IDF.



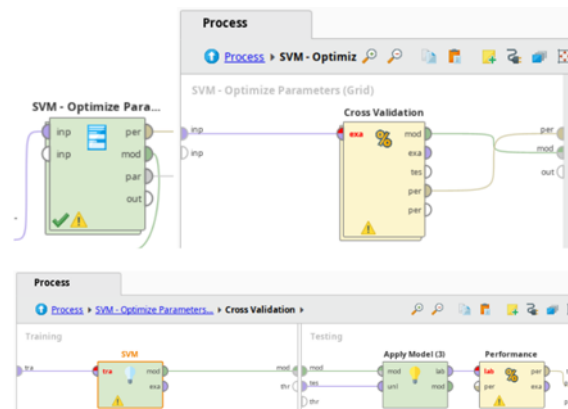
Gambar 3. Proses *text preprocessing*

3.4. Modeling

Setelah tahap *text preprocessing* selesai, dilakukan proses *modeling* menggunakan algoritma SVM yang dikonfigurasi dalam skema *Optimize Parameters (Grid)*. Proses ini menggunakan *5-Fold Cross Validation* untuk menilai performa setiap kombinasi parameter secara otomatis. Hasil proses optimasi menunjukkan bahwa penggunaan kernel RBF dengan parameter $C = 10$ dan $\gamma = 1$ memberikan kinerja paling optimal, dengan keseimbangan yang baik antara tingkat akurasi dan kemampuan model dalam melakukan generalisasi.

Tabel 3. Parameter hasil optimasi model SVM melalui *Grid Search*

Parameter	Nilai Optimal
Kernel	RBF
C	10
Gamma	1



Gambar 4. Struktur operator *Optimize Parameters (Grid)* di RapidMiner

3.5. Evaluasi

Evaluasi kinerja model dilakukan menggunakan empat metrik utama, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, untuk mengukur sejauh mana model SVM mampu mengklasifikasikan data secara akurat dan seimbang. Gambar 5 memperlihatkan

hasil pengujian model pada data uji yang divisualisasikan dalam bentuk *Confusion Matrix*.

accuracy: 96.00%

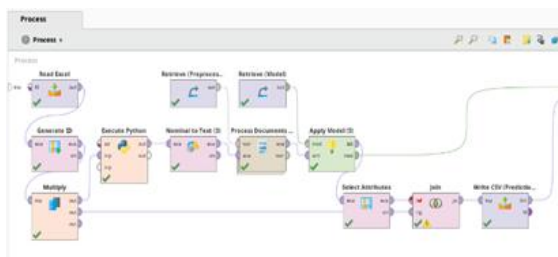
	true negatif	true positif	class precision
pred. negatif	112	9	92.56%
pred. positif	3	176	98.32%
class recall	97.39%	95.14%	

Gambar 5. *Confusion Matrix* hasil pengujian model SVM di RapidMiner

Model SVM menunjukkan performa yang sangat baik dengan accuracy 96%. Nilai recall mencapai 97,39% untuk sentimen negatif dan 95,14% untuk sentimen positif, sedangkan precision masing-masing sebesar 92,56% dan 98,32%. Selain itu, nilai *F1-Score* rata-rata sebesar 95,85% menunjukkan hasil yang seimbang antara kemampuan model dalam mendeteksi dan memprediksi dengan tepat. Secara umum, hasil penelitian ini menunjukkan bahwa model memiliki kinerja yang konsisten dan cukup handal dalam melakukan klasifikasi sentimen publik terhadap Program Makan Bergizi Gratis.

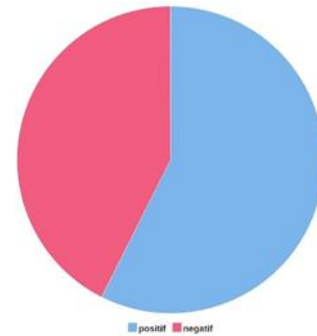
3.6. Prediksi dan Penggabungan Hasil

Model yang telah dilatih digunakan untuk melakukan klasifikasi sentimen pada 1.483 data tanpa label. Data tanpa label akan diklasifikasikan menjadi dua kategori sentimen, yaitu positif dan negatif.



Gambar 6. Proses Prediksi data tanpa label

Setelah itu, data yang dilabeli secara manual dan data hasil prediksi digabungkan untuk melihat distribusi sentimen terhadap Program Makan Bergizi Gratis. Hasil analisis menunjukkan bahwa opini publik terhadap Program MBG didominasi oleh sentimen positif sebesar 57,43%, sedangkan 42,57% komentar tergolong negatif. Persentase tersebut mengindikasikan bahwa mayoritas masyarakat memberikan tanggapan baik terhadap kebijakan pemerintah ini.



Gambar 7. Distribusi sentimen publik terhadap Program Makan Bergizi Gratis

4. KESIMPULAN

Penelitian ini berhasil melakukan analisis sentimen publik terhadap Program MBG dengan memanfaatkan algoritma SVM dan metode TF-IDF. Model dengan parameter optimal (kernel RBF, C = 10, gamma = 1) menunjukkan performa tinggi dengan accuracy 96% serta nilai *precision*, *recall*, dan *F1-score* yang seimbang.

Hasil analisis menunjukkan bahwa 57,43% komentar publik bersentimen positif, sedangkan 42,57% bersentimen negatif, menandakan bahwa mayoritas masyarakat memberikan tanggapan baik terhadap kebijakan tersebut. Temuan ini juga membuktikan bahwa kombinasi *text mining*, *preprocessing*, TF-IDF, dan SVM efektif dalam mengklasifikasikan opini publik berbahasa Indonesia.

PUSTAKA

Brata, D., & Wijayanto, S. (2025). Analisis Sentimen Pengguna YouTube Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Support Vector Machine. *E-Proceeding of Engineering*, 12(4), 5784–5789.

Bundy, D. A., Gentilini, U., Schultz, L., Bedasso, B., Singh, S., Okamura, Y., TMM Iyengar, H., & Monique Blakstad, M. (2024). *School Meals, Social Protection and Human Development: Revisiting Trends, Evidence, and Practices in South Asia and Beyond*. www.worldbank.org

Fairuzabadi, M. (2025). *Mahir Data Science dengan Python : Analisis Data & Machine Learning Untuk Pemula*. Yashmedia.

Faisal, M. R., Kartini, D., Arrahimi, A. R., & Saragih, T. H. (2023). *Belajar Data Science: Text Mining Untuk Pemula I*. M Reza Faisal. <https://books.google.co.id/books?id=5CKIEA AAQBAJ>

Fajri, M., & Primajaya, A. (2023). Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random

- Search. *Journal of Applied Informatics and Computing (JAIC)*, 7(1), 10–15.
<https://doi.org/10.30871/jaic.v7i1.5004>
- Herdiana, D. (2025). Implementasi Kebijakan Makan Bergizi Gratis (MBG): Faktor-faktor Pendorong dan Penghambat. *Jurnal Ilmiah Multidisipliner*, 3(2), 470–478.
<https://doi.org/10.5281/zenodo.15063556>
- Kiftiyah, A., Palestina, F. A., Abshar, F. U., & Rofiah, K. (2025). Program Makan Bergizi Gratis (MBG) dalam Perspektif Keadilan Sosial dan Dinamika Sosial – Politik. *Pancasila: Jurnal Keindonesiaan*, 5(1), 101–112. <https://doi.org/10.52738/pjk.v5i1.726>
- Nugraha, W., & Sasongko, A. (2022). Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid. *Jurnal Sistem Informasi*, 11(2), 391–401.
<https://doi.org/10.32520/stmsi.v11i2.1750>
- Pratama, A. W. (2025). *ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X DENGAN STACKING NAIVE BAYES SVM DAN RF* [Skripsi S1]. Universitas Multimedia Nusantara.
- Putra, K. T., Hariyadi, M. A., & Crysdiyan, C. (2023). PERBANDINGAN FEATURE EXTRACTION TF-IDF DAN BOW UNTUK ANALISIS SENTIMEN BERBASIS SVM. *Jurnal Cahaya Mandalika*, 3, 1449–1463.
<https://www.ojs.cahayamandalika.com/index.php/jcm/article/view/2292>
- Putra, R. F., Mukhlis, I. R., Datya, A. I., Pipin, S. J., Reba, F., Al-Husaini, M., Mandowen, S. A., Zain, N. N. L. E., Judijanto, L., & Sepriano, S. (2024). *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*. PT. Sonpedia Publishing Indonesia.
<https://books.google.co.id/books?id=MX7-EAAAQBAJ>
- Putra, R. R., Putri, N. A., & Putra, A. D. (2024). *Teknik Cosine Similarity Dan TF-IDF Dalam Analisis Data*. Serasi Media Teknologi.
<https://books.google.co.id/books?id=gZU0EQAAQBAJ>
- Ramadhani, A., Permana, I., Afdal, M., & Fronita, M. (2024). Analisis Sentimen Tanggapan Publik di Twitter Terkait Program Kerja Makan Siang Gratis Prabowo–Gibran Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1509–1516.
<https://doi.org/10.47065/bits.v6i3.6188>