 ISSN:2302-8149
e-ISSN:2540-9719

Perbandingan Algoritma *Support Vector Machine* dan *K-Nearest Neighbor* terhadap Efektivitas Program Makan Siang Gratis

Comparison of Support Vector Machine and K-Nearest Neighbor Algorithms on the Effectiveness of a Free Lunch Program

[1]Frenda Farahdinna*, [2]Pipit Prabawati
[1]Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
[2]Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
[1,2]Jl. Sawo Manila No.61, Kelurahan Pejaten Barat, Pasar Minggu, Kota Administrasi Jakarta Selatan, DKI Jakarta 12520, Indonesia
*e-mail: frenda.farahdinna@civitas.unas.ac.id

(***received*:** 24 December 2025, ***revised*:** 22 January 2026, ***accepted*:** 24 January 2026)

Abstrak

Program Makan Gratis adalah sebuah upaya negara yang diberikan atas ide negara dalam memberikan aspek kecukupan atas nutrisi yang harus dipenuhi masyarakat. Tujuan penelitian adalah untuk membahas persepsi publik terhadap program makanan gratis dengan menggunakan analisis sentimen dan membandingkan efektivitas *Support Vector Machine dan K-Nearest Neighbor* model. Sejumlah 6.532 komen publik dikumpulkan dari twitter, YouTube, dan TikTok. Setelah Pra-pemrosesan berupa *Normalization, elimination of stopwords, dan stemming*. Fitur diekstraksi *dengan Term Frequency Inverse Document Frequency* menjadi 5.992 data bersih. Data dipecah menjadi 80% latih, dan 20% uji. Pemodelan eksperimen menggunakan *Hyperparameter Tuning* dengan *GridSearchCV* 3-*fold*. Hasil penelitian menunjukkan sentimen negatif mendominasi sebesar 42,7%. Dalam perbandingan model, SVM dengan *kernel Linear* terbukti jauh lebih unggul dengan akurasi 72%, sedangkan K-NN (k=3) hanya mencapai akurasi 48%. Hal ini menunjukkan bahwa algoritma SVM lebih efektif dalam mengklasifikasikan sentimen opini publik pada data berdimensi tinggi dibandingkan K-NN.

Kata kunci: evaluasi sentimen, support vector machine, k-nearest neighbor, makan siang gratis

Abstract

The Free Lunch Program is a government initiative aimed at ensuring adequate nutrition for the public. This study aims to examine public perceptions of the program through sentiment analysis and to compare the effectiveness of Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) models. A total of 6,532 public comments were collected from Twitter, YouTube, and TikTok. After preprocessing, including normalization, stopword removal, and stemming, features were extracted using Term Frequency–Inverse Document Frequency (TF-IDF), resulting in 5,992 clean data points. The dataset was split into 80% training and 20% testing sets. Model training was conducted with hyperparameter tuning using 3-fold GridSearchCV. The results indicate that negative sentiment dominated at 42.7%. In the model comparison, SVM with a linear kernel significantly outperformed K-NN, achieving an accuracy of 72%, while K-NN (k=3) reached only 48%. These findings suggest that the SVM algorithm is more effective in classifying public opinion sentiment on high-dimensional data compared to K-NN.

Keywords: *sentiment evaluation, support vector machine, k-nearest neighbor, free lunch*

1 Pendahuluan

Pemberian program makanan gratis ialah tindakan sosial dilakukan negara dengan tujuan untuk meningkatkan kecukupan gizi dan membantu masyarakat yang membutuhkan. Dalam era tata kelola digital saat ini, pemantauan dan evaluasi terhadap efektivitas program semacam ini menjadi krusial.

 ISSN:2302-8149
e-ISSN:2540-9719

Permasalahan utama yang muncul adalah belum adanya sistem pengukuran yang terstruktur dan berbasis data untuk menilai secara objektif efektivitas program ini. Tanpa evaluasi yang objektif, mencari tahu sejauh mana keberhasilan, serta apakah manfaat yang diberikan telah sesuai dengan yang diharapkan. Padahal, pembuatan kebijakan di era digital sangat bergantung pada analisis data skala besar untuk memahami dampak dan respons publik [2]. Oleh karena itu, evaluasi yang tepat waktu serta informasi data sangat dibutuhkan guna perbaikan program yang berkelanjutan. Menanggapi kebutuhan ini, analisis sentimen publik di media sosial muncul sebagai salah satu metode evaluasi yang paling efektif [3]. Analisis sentimen merujuk pada serangkaian proses komputasi yang bertujuan guna memperoleh informasi tentang sentimen, diperlukan pemahaman, ekstraksi, serta pengolahan informasi teks. (pendapat, perilaku, dan emosi), sebuah bidang yang telah dikaji secara ekstensif dalam berbagai survei terbaru [4]. Platform-platform media sosial seperti Twitter, YouTube, dan TikTok sudah menjadi sarana bagi masyarakat untuk mengungkapkan pendapat secara massal dan real-time. Hal ini didukung dengan pemanfaatan teknik pengolahan bahasa *Natural Language Processing* (NLP) guna bahasa Indonesia.[5] dan *Machine Learning* (ML) [6] jutaan opini tidak terstruktur ini dapat diolah untuk mengekstraksi dan mengklasifikasikan sentimen. Metode ini pun telah terbukti efektif untuk menganalisis kebijakan publik lain di Indonesia, seperti pemindahan Ibu Kota Negara (IKN) [7].

Secara spesifik, penelitian ini merumuskan tiga permasalahan utama. Pertama, belum tersedianya metode evaluasi objektif yang memadai untuk menilai efektivitas Program Makan Gratis. Kedua, kurangnya pemahaman mengenai persepsi aktual masyarakat baik positif, netral, maupun negatif terhadap program tersebut. Ketiga, belum diketahuinya algoritma *machine learning* yang paling efektif berdasarkan metrik performa standar [7] untuk mengklasifikasikan *sentimen* pada domain program sosial ini. Untuk menjawab permasalahan ketiga, penelitian ini memfokuskan perbandingan pada dua algoritma, yaitu *Support Vector Machine (SVM)* dan *K-Nearest Neighbor (K-NN)*. SVM dipilih karena tinjauan sistematis terbaru mengonfirmasinya sebagai baseline yang kuat untuk klasifikasi teks [8], khususnya karena kemampuannya menangani data berdimensi tinggi secara efektif [9] saat dikombinasikan dengan pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* [10]. Sebaliknya, K-NN dipilih sebagai pembanding langsung [11] guna menguji secara empiris apakah kelemahan teoritisnya, yakni *curse of dimensionality* [12], terbukti signifikan pada data teks sparse berdimensi tinggi yang sering kali membuat konsep jarak Euklides menjadi tidak relevan [13].

Berdasarkan uraian permasalahan tersebut, penelitian ini difokuskan untuk menjawab dua pertanyaan penelitian esensial. Fokus pertama adalah mengidentifikasi bagaimana kecenderungan sentimen publik terhadap implementasi Program Makan Gratis yang terekam di media sosial. Selanjutnya, fokus kedua bertujuan menentukan algoritma manakah di antara SVM dan K-NN yang terbukti memiliki efektivitas dan stabilitas lebih tinggi dalam mengklasifikasikan sentimen, khususnya pada dataset kebijakan publik yang memiliki karakteristik high-dimensional noise. Adapun kebaruan novelty dari studi ini terletak pada validasi empiris komparatif algoritma pada domain spesifik kebijakan sosial baru di Indonesia yang belum banyak dieksplorasi secara komputasional. Penelitian ini mengisi celah literatur dengan memberikan bukti kuantitatif mengenai batas performa K-NN akibat fenomena *curse of dimensionality* pada data tekstual TF-IDF, sekaligus merekomendasikan SVM *Kernel Linear* sebagai standar baseline yang lebih robust untuk analisis sentimen kebijakan publik di Indonesia.

2 Tinjauan Literatur

Penerapan machine learning untuk analisis sentimen merupakan pendekatan yang valid, seperti ditunjukkan dalam studi komparatif sentimen vaksin COVID-19 [14]. Tinjauan sistematis [5] juga menunjukkan bahwa SVM dan K-NN masih relevan sebagai baseline pembanding, meskipun model baru telah muncul. Studi-studi komparatif di Indonesia secara konsisten menempatkan SVM sebagai salah satu model dengan kinerja terkuat. Sebuah penelitian yang membandingkan SVM dan K-NN untuk analisis sentimen naturalisasi menemukan bahwa SVM memberikan kinerja yang lebih baik [15]. Demikian pula, penelitian lain menyimpulkan SVM secara signifikan mengungguli *Naive Bayes* dan K-NN dalam klasifikasi sentimen ulasan [16]. Keunggulan SVM, yang efektif menangani data sparse berdimensi tinggi [8, 9], semakin meningkat saat dipasangkan dengan fitur TF-IDF [10].

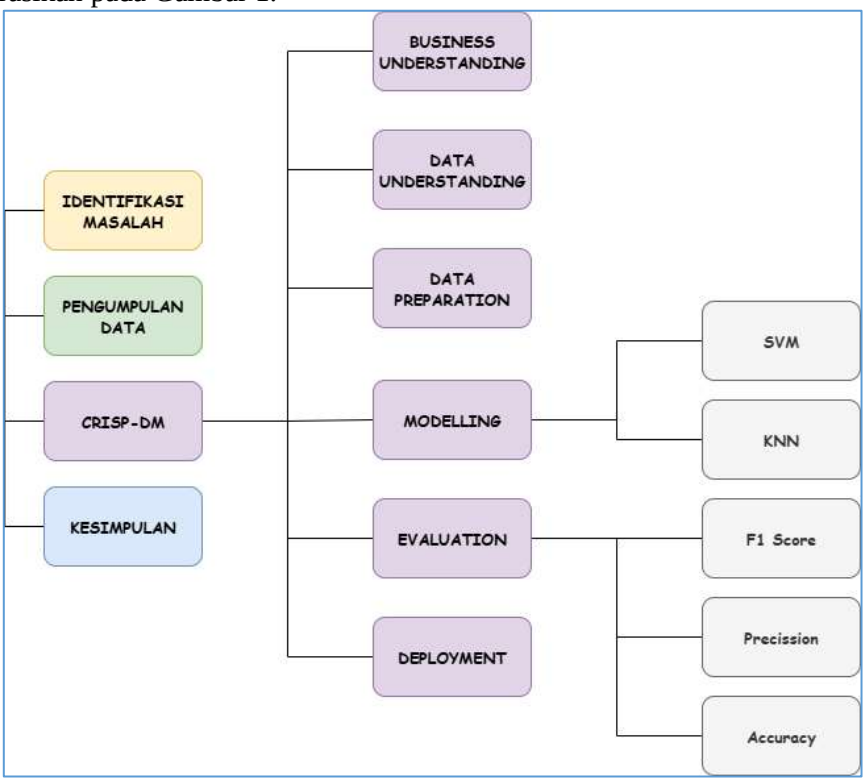
 ISSN:2302-8149
e-ISSN:2540-9719

Sebaliknya, landasan teoritis yang kuat menjelaskan mengapa K-NN sering gagal dalam tugas klasifikasi teks [11]. K-NN rentan terhadap *curse of dimensionality* [12]. Tinjauan sistematis menjelaskan bahwa di ruang dimensi tinggi, jarak Euklides menjadi tidak bermakna [13], sehingga pencarian "tetangga terdekat" menjadi tidak andal [17]. Tinjauan komprehensif lainnya mengkonfirmasi tantangan efisiensi dan akurasi pencarian nearest neighbor di ruang dimensi tinggi [18].

Dalam kaitannya dengan metodologi, kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) telah diakui. sebagai standar industri, sebagaimana dikonfirmasi oleh tinjauan sistematis [19]. Berbagai penelitian di Indonesia telah berhasil menerapkan CRISP-DM untuk proyek analisis sentimen [20, 21]. Metodologi ini menyediakan alur kerja yang terstruktur mulai dari pemahaman bisnis hingga *deployment*. Dari tinjauan ini, dapat dijelaskan apa yang belum dilakukan *research gap*: Meskipun perbandingan SVM dan K-NN telah umum dilakukan [11, 15, 16], dan kelemahan teoretis K-NN telah diketahui [12, 13, 17], penelitian ini fokus pada bagian yang belum dikerjakan tersebut: yaitu, validasi empiris dari kegagalan teoretis K-NN dibandingkan dengan ketangguhan SVM, diterapkan secara spesifik pada domain baru yang sangat relevan secara sosial di Indonesia: Program Makan Gratis. Penelitian ini mengisi celah dengan menyediakan *benchmark* kinerja pada dataset *multi-platform* (Twitter, YouTube, TikTok) yang *noisy*.

3 Metode Penelitian

Lingkungan metode penelitian deskriptif dengan pendekatan kuantitatif diadopsi. Desain penelitian mengikuti kerangka kerja CRISP-DM yang telah teruji berhasil dalam proyek analisis sentimen. Kerangka kerja ini meliputi enam fase sistematis seperti yang ada pada alur desain penelitian diilustrasikan pada Gambar 1.



Gambar 1 Desain penelitian

Proses penelitian mengikuti alur kerja CRISP-DM yang telah diuraikan dalam desain gambar 1.

1. Fase *Business dan Data Understanding*
   Pada fase ini, difokuskan pada pemahaman tujuan penelitian, yang melibatkan evaluasi persepsi publik terhadap program makanan gratis dan perbandingan algoritma SVM dan K-NN. Di tahap ini, dilaksanakan pengumpulan dan integrasi data. Data mentah diperoleh melalui API resmi Twitter dan *web scraping* YouTube dan TikTok yang menghasilkan total 6.532 komentar. Setelah

 ISSN:2302-8149
e-ISSN:2540-9719

pembersihan data duplikat dan tidak relevan, diperoleh 5.992 komentar yang digunakan untuk di analisis.

2. Fase *Data Preparation*, Fase ini adalah proses krusial untuk mentransformasi data mentah agar siap dimodelkan. Proses ini mencakup beberapa langkah:
   a. Pra-pemrosesan Teks: Setiap komentar mentah dibersihkan melalui serangkaian langkah, meliputi: konversi ke huruf kecil *(lowercasing)*, penghapusan URL, *mention* (@username), angka, serta tanda baca dan karakter non-alfabet.
   b. *Stopword Removal* dan *Stemming*: *Stopwords* (kata-kata umum) untuk Bahasa Indonesia dan Inggris dihilangkan. Selanjutnya, proses *stemming* (pencarian kata dasar) dilakukan.
   c. Ekstraksi Fitur: Teks yang telah difilter dan dibersihkan dikonversi menjadi representasi numerik dengan mengimplementasikan pendekatan *Term Frequency–Inverse Document Frequency* (TF-IDF). [10]. Parameter diatur *max_df*=0.75 (mengabaikan istilah yang muncul di >75% dokumen) dan *min_df*=1 (mengabaikan istilah yang muncul kurang dari 1 kali) untuk mengurangi noise.
   d. Mengingat volume data yang besar (5.992 komentar), proses pelabelan menerapkan pendekatan *Semi-Supervised Learning* untuk menjaga efisiensi tanpa mengorbankan akurasi. Proses ini terdiri dari dua tahap:
      1. *Initial Labeling*: Model *pre-trained* IndoBERT digunakan untuk memprediksi label sentimen awal (Positif, Netral, Negatif) pada seluruh data bersih karena kemampuannya menangkap konteks semantik bahasa Indonesia yang kompleks.
      2. *Validation & Correction*: Hasil prediksi IndoBERT divalidasi menggunakan metode *rule-based* berbasis leksikon *(Lexicon Based)* untuk mengoreksi bias prediksi. Jika terjadi ketidaksesuaian prediksi antara IndoBERT dan Leksikon pada sampel tertentu, dilakukan verifikasi manual oleh peneliti untuk menentukan label *ground truth* yang *final*. Pendekatan hibrida ini memastikan dataset latih memiliki kualitas label yang valid."
   e. Pembagian Dataset: Dataset dibagi secara *stratified* menjaga proporsi kelas sentimen menjadi 80% data latih dan 20% data uji.
3. Fase *Modelling*, Fase ini mencakup implementasi, pelatihan, dan tuning kedua algoritma:
   a. Penanganan data tidak seimbang: Untuk menangani ketidakseimbangan *kelas (imbalanced data)* yang didominasi sentimen negatif, penelitian ini menerapkan teknik *SMOTE (Synthetic Minority Over-sampling Technique).* Teknik ini menyeimbangkan distribusi data latih dengan mensintesis sampel baru untuk kelas minoritas agar model tidak bias
   b. *Support Vector Machine* (SVM)*:* melalui pemetaan kernel, yaitu kernel *linear*. (*Radial Basis Function*), telah dijalankan uji coba. Dilakukan proses optimasi hyperparameter menggunakan *Grid Search* dengan validasi silang 3-fold [22].
   c. *K-Nearest Neighbor* (K-NN)*:* Algoritma dieksperimenkan dengan metrik jarak *Euklides* dan mengeksplorasi nilai k (jumlah tetangga) dari 1 hingga 15, juga dievaluasi menggunakan validasi silang 3-fold.
4. Fase *Evaluation*, Pada tahap ini, performa optimal model dari kedua algoritma dinilai dan dibandingkan dengan menggunakan data uji sebanyak 20%. Evaluasi metrik akurasi, *Precision, Recall, F1 Score* dimanfaatkan untuk pelaksanaan tugas klasifikasi. Metrik-metrik tersebut dianggap standar dalam domain ini [7]. Selain metrik kuantitatif, dilakukan analisis visual menggunakan *Confusion Matrix* dan Kurva ROC.
5. Fase *Deployment*, Fase terakhir ini diwujudkan dalam bentuk prototipe aplikasi dashboard yang fungsional, dikembangkan menggunakan *framework* Streamlit. Prototipe ini mampu mendemonstrasikan prediksi sentimen secara real-time terhadap input teks baru dari pengguna.

4 Hasil dan Pembahasan

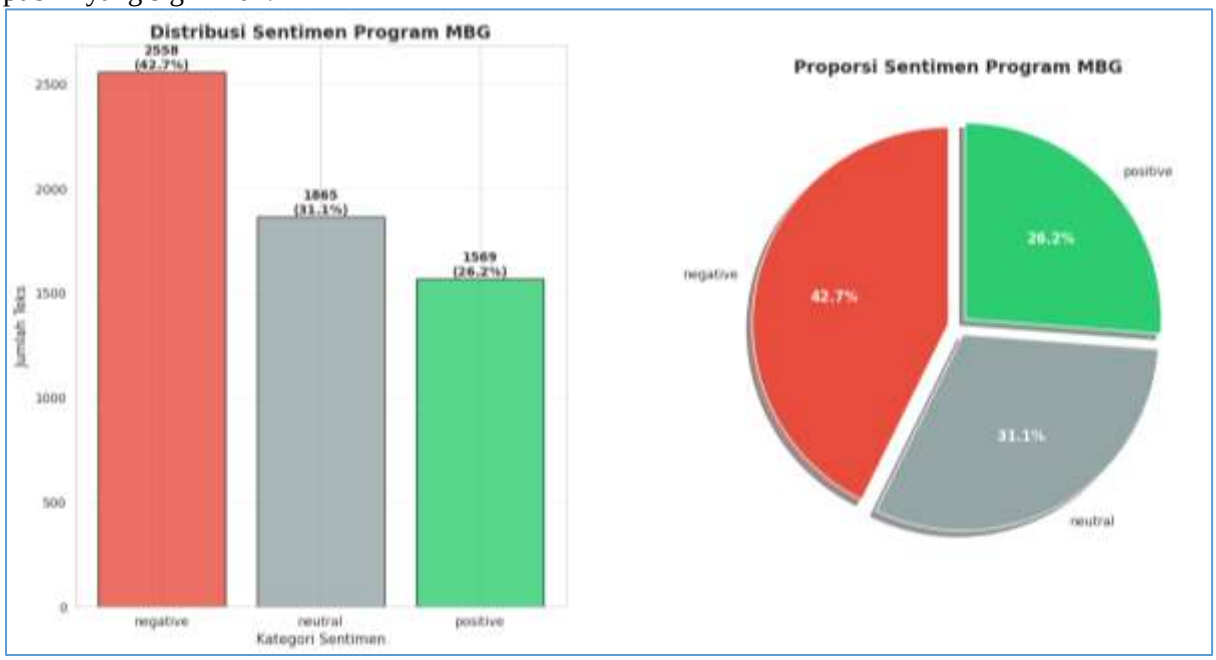
Bagian ini menguraikan hasil eksperimen komputasi dan analisis mendalam yang dilakukan untuk menjawab pertanyaan penelitian. Pembahasan diawali dengan eksplorasi karakteristik dataset hasil crawling dari berbagai platform media sosial guna memetakan distribusi sentimen masyarakat secara deskriptif. Selanjutnya, disajikan evaluasi komparatif kinerja algoritma *Support Vector Machine (SVM)* dan *K-Nearest Neighbor (K-NN)* yang diukur menggunakan metrik standar validasi

 ISSN:2302-8149
e-ISSN:2540-9719

model. Analisis kemudian diperdalam melalui bedah kesalahan *(error analysis)* menggunakan *Confusion Matrix* untuk mengidentifikasi kelemahan spesifik masing-masing algoritma, dan diakhiri dengan demonstrasi implementasi model terbaik ke dalam prototipe aplikasi dasbor interaktif sebagai bukti konsep.

4.1 Eksplorasi Dataset yang Dianalisis.

Analisis pertama terhadap 5.992 data komentar mengindikasikan adanya kecenderungan sentimen publik yang kritis. terhadap program makanan gratis. Seperti ditunjukkan pada Gambar 2, sentimen negatif menjadi kategori terbesar (42,7%), diikuti oleh sentimen netral (31,6%), dan sentimen positif (26,2%). Tingginya proporsi sentimen negatif mengindikasikan adanya kekhawatiran publik yang signifikan.



Gambar 2 Distribusi sentimen dataset

Untuk memahami konteks di balik data mentah dan data yang telah dibersihkan, Tabel 1 menyajikan beberapa contoh komentar dari dataset.

Tabel 1 Contoh dataset

Teks Asli	Teks Bersih (Cleaned_text)
"ini MBG ga cocok namanya Makan bergizi gratis lah parameternya suksesnya aja statistik yg keracunan atau engga. lebih cocok jadi MSG (makan siang gratis) biar 2045 Indonesia Generasi Micin"	"ini mbg ga cocok namanya makan gizi gratis lah parameternya suksesnya aja statistic yg keracunan atau engga lebih cocok jadi msg makan siang gratis biar 2045 indonesia generasi micin"
"kerjasama BRICS bikin ekspor naik, industri lokal berkembang, angka stunting turun berkat makan siang gratis di sekolah"	"kerjasama brics bikin ekspor naik industri lokal kembang angka stunting turun berkat makan siang gratis sekolah"
"lah kan mbg stands for makanan basi dan gaenak... itu program tujuan utamanya buat korupsi"	"lah kan mbg stands for makan basi dan gaenak itu program tujuan utamanya buat korupsi"

Analisis word cloud dapat dilihat pada Gambar 3 yang memberikan wawasan lebih dalam mengenai topik utama per kategori sentimen. Pada sentimen positif berjumlah 1.569 berfokus pada kata "program", "sukses", "sehat", dan "dukung". Pada sentimen netral berjumlah 1.865 berfokus pada istilah faktual seperti "program", "presiden", dan "anggar" (anggaran). Sedangkan sentimen negative berjumlah 2.558 pada Gambar 3 secara jelas menyoroti kekhawatiran publik, dengan dominasi kata-

 ISSN:2302-8149
e-ISSN:2540-9719

kata seperti "korupsi", "proyek", "basi", dan "masalah", yang menyoroti area spesifik untuk evaluasi program.

Gambar 3 *Word cloud* analisis sentimen

4.2 Perbandingan Kinerja Model

Perbandingan kinerja model divalidasi menggunakan dua metode. Hasil pengujian pada split data 80% dan 20% yang merupakan metrik evaluasi utama pada setiap kelas sentiment dan dan hasil rata-rata dari validasi silang 3 fold yang menunjukkan stabilitas setiap model dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2 Pengujian kinerja model support vector machine (SVM)

Kelas Sentimen	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	0.77	0.79	0.78	512
Netral	0.64	0.65	0.65	373
Positif	0.76	0.69	0.72	314
Rata – Rata *Macro*	0.72	0.71	0.72	1199
Rata – Rata *Weighted*	0.72	0.72	0.72	1199
Akurasi			0.72	1199

Tabel 3 Pengujian kinerja model k-nearest neighbor (K-NN)

Kelas Sentimen	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	0.79	0.30	0.44	512
Netral	0.41	0.59	0.49	373
Positif	0.43	0.64	0.51	314
Rata – Rata *Macro*	0.54	0.51	0.48	1199
Rata – Rata *Weighted*	0.58	0.48	0.47	1199
Akurasi			0.48	1199

Berdasarkan Tabel 2 dan Tabel 3, terlihat perbedaan performa yang signifikan antara kedua algoritma. SVM (*Kernel Linear*) mencapai akurasi 72%, jauh mengungguli K-NN yang hanya mencapai 48%.

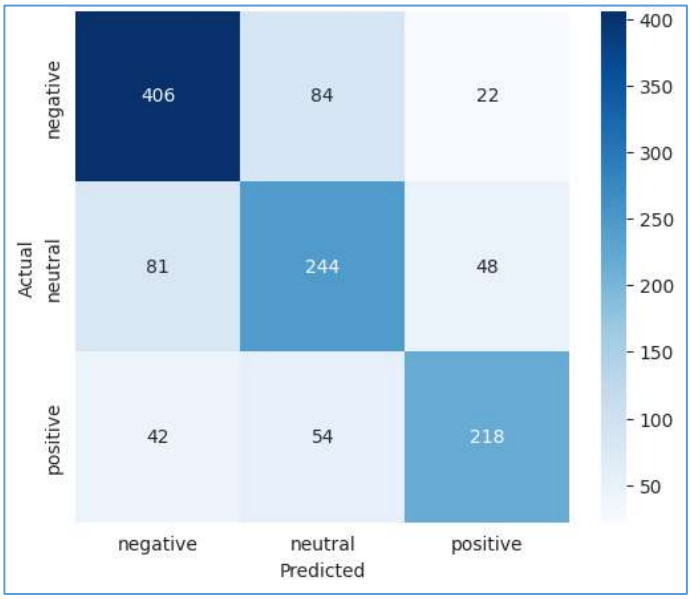
Kelemahan K-NN terlihat sangat jelas pada nilai *Recall* untuk kelas Negatif yang hanya 0.30. Ini berarti K-NN kurang tepat dalam mendeteksi 70% dari total komentar negatif yang ada, dan banyak kesalahan dalam klasifikasi ke kelas lain. Sebaliknya, SVM mampu menyeimbangkan nilai *Precision*

 ISSN:2302-8149
e-ISSN:2540-9719

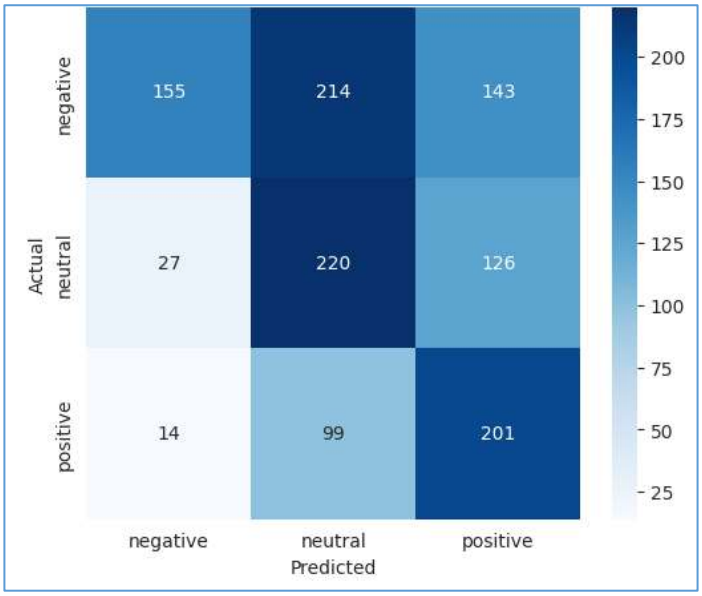
dan *Recall* dengan baik (*F1-Score* rata-rata 0.72), membuktikan bahwa pemisah *linear (hyperplane)* bekerja lebih efektif pada representasi vektor teks TF-*IDF* yang berdimensi tinggi dibandingkan metode berbasis jarak (*distance-based*) seperti K-NN. Berdasarkan temuan tersebut validasi empiris dari kegagalan K-NN yaitu kinerjanya yang sangat rendah menghasilkan manifestasi praktis dari *curse of dimensionality* yang telah dibahas dalam tinjauan literatur [12, 13]. Data *TF-IDF* yang digunakan memiliki dimensi yang sangat tinggi (ribuan fitur/kata unik). Dalam ruang dimensi tinggi ini, metrik jarak *Euklides* K-NN gagal membedakan 'tetangga' yang relevan [17, 18], sehingga kinerjanya tidak jauh lebih baik.

4.3 Analisis Kesalahan Model

Analisis *confusion matrix* menunjukkan detail kinerja model. Pada Gambar 4 model SVM sangat baik dalam mengidentifikasi kelas dimana terdapat kelas 'negatif' dengan jumlah 406 prediksi benar dan 'netral' dengan jumlah 244 prediksi benar. Kelemahan utamanya adalah pada recall kelas 'positif' yang hanya berjumlah 42 prediksi benar, atau di bawah 40%, maka dari itu menunjukkan bahwa model ini kesulitan mengidentifikasi pujian, karena jumlahnya yang lebih sedikit dan data tidak seimbang. Pada Gambar 5 model K-NN menunjukkan ketidakstabilan total dan kesalahan klasifikasi yang masif di semua kelas.



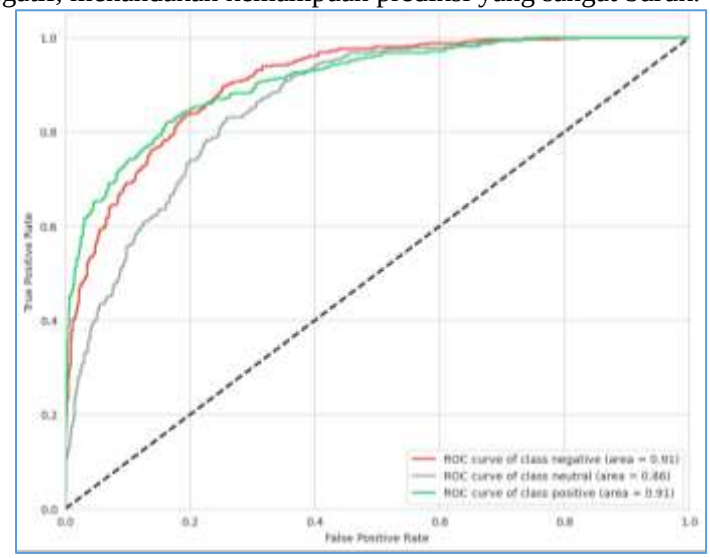
Gambar 4 *Confusion matriks* SVM



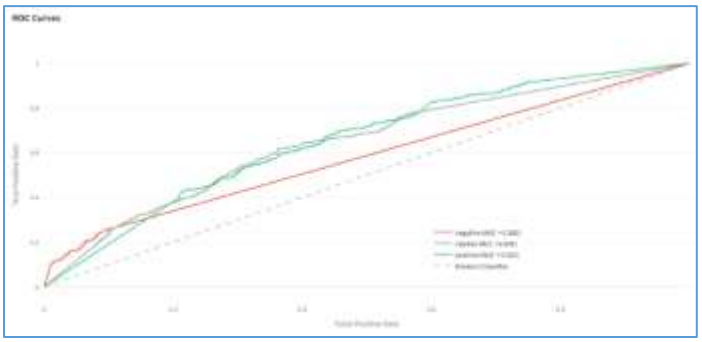
 ISSN:2302-8149
e-ISSN:2540-9719

Gambar 5 *Confusion matriks* K-NN

Hasil dari kurva ROC pada gambar 6 dan 7 mengkonfirmasi hal ini. Pada gambar 6 hasil model SVM menunjukkan kemampuan diskriminatif yang kuat, dengan nilai area untuk semua kelas di atas 0.86 (Negatif: 0.91, Positif: 0.91, Netral: 0.86). Sebaliknya, K-NN pada hasil kurva ROC di gambar 7 menunjukkan kurva yang dekat dengan garis acak *(Random Classifier)*, dengan nilai area terendah 0.582 untuk kelas negatif, menandakan kemampuan prediksi yang sangat buruk.



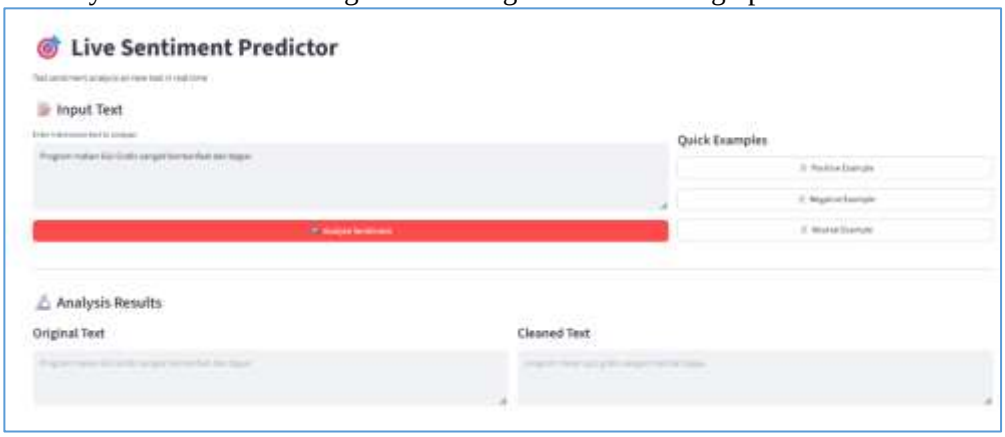
Gambar 6 Kurva ROC SVM



Gambar 7 Kurva ROC K-NN

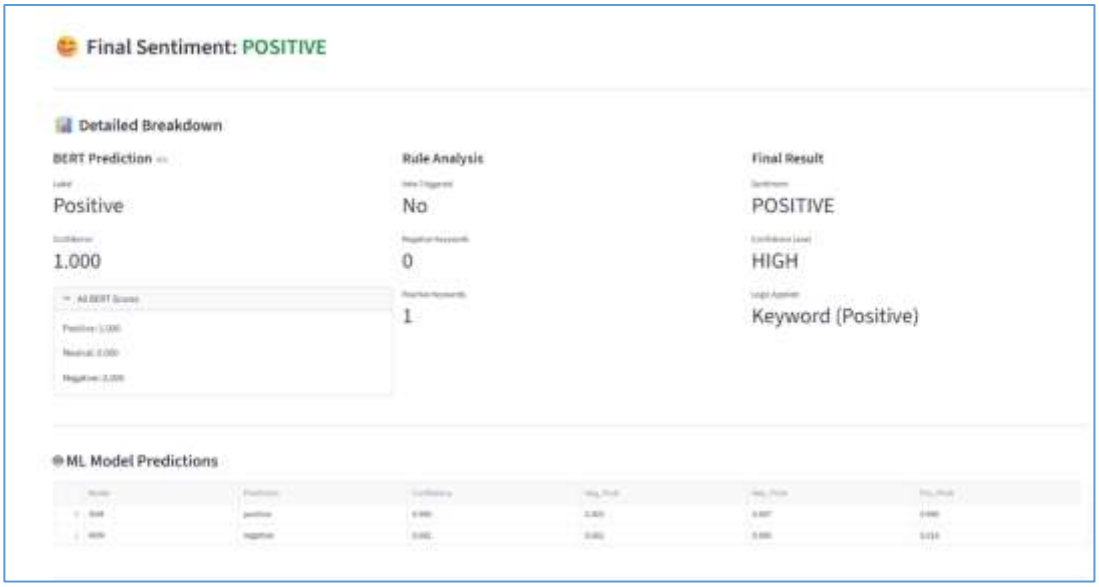
4.4 Implementasi Prototipe

Sebagai bukti konsep *(proof-of-concept)* untuk fase *deployment*, model SVM terbaik diimplementasikan ke dalam *dashboard* prototipe menggunakan Streamlit dapat dilihat pada gambar 8. Aplikasi ini berhasil memprediksi sentimen *real-time* dari input teks baru, pada gambar 9 menunjukkan kelayakan model untuk digunakan sebagai alat monitoring operasional.



Gambar 8 Prototipe aplikasi

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 9 Hasil sentimen pada aplikasi

5 Kesimpulan

Berdasarkan hasil pengujian dan analisis terhadap 5.992 data ulasan masyarakat, dapat disimpulkan bahwa sentimen publik terhadap Program Makan Gratis cenderung kritis. Hal ini dibuktikan dengan dominasi sentimen negatif sebesar 42,7%, diikuti sentimen netral 31,1%, dan positif 26,2%. Tingginya angka sentimen negatif mengindikasikan adanya urgensi bagi negara untuk mengevaluasi aspek transparansi anggaran dan teknis distribusi program. Secara komparatif teknis, penelitian ini menjawab pertanyaan penelitian dengan membuktikan bahwa *Support Vector Machine (SVM)* berbasis *Kernel Linear* secara signifikan mengungguli *K-Nearest Neighbor (K-NN)*. SVM mencatatkan akurasi sebesar 72% dan menunjukkan stabilitas yang baik pada metrik Precision-Recall. Sebaliknya, algoritma K-NN mengalami penurunan performa drastis dengan akurasi hanya 48% dan nilai Recall kelas negatif yang sangat rendah (0.30). Kegagalan K-NN ini mengonfirmasi fenomena *curse of dimensionality*, di mana perhitungan jarak *Euklides* menjadi tidak efektif pada representasi fitur teks TF-IDF yang memiliki ribuan dimensi.[24][25] Saran Penelitian selanjutnya disarankan untuk tidak hanya berhenti pada klasifikasi sentimen umum, tetapi menerapkan metode *Aspect-Based Sentiment Analysis (ABSA).* Pendekatan ABSA diperlukan untuk memetakan kritik masyarakat ke aspek spesifik, seperti "kualitas makanan", "anggaran", atau "distribusi". Selain itu, untuk mengatasi ketidakseimbangan data yang ekstrem, penelitian mendatang dapat bereksperimen dengan teknik augmentasi data teks seperti *back-translation* atau menggunakan arsitektur *Deep Learning* mutakhir seperti *Fine-Tuned* IndoBERT untuk meningkatkan akurasi klasifikasi di atas performa SVM."

Ucapan Terima Kasih

Peneliti mengungkapkan apresiasi serta terimakasih ke semua partisipan serta pengguna media sosial yang telah berbagi ulasan dan umpan balik, yang merupakan data utama dalam kajian ini juga, terima kasih kepada keluarga dan rekan-rekan atas dukungan, semangat, dan masukan konstruktif selama proses penelitian hingga penyelesaian manuskrip ini.

Referensi (Reference)

[1] F. P. Wahyu, S. Alia, E. Susanti, F. H. Abdillah, and R. Febrianti, "*AAPA-EROPA-AGPA-IAPA International Conference 2024 Towards World Class Bureaucracy A Systematic Review of the Data-Driven Public Policy Making in Indonesia*", DOI: 10.30589/proceedings.2024.1111.

[2] Q. A. Xu, V. Chang, and C. Jayne, "*A Systematic Review of Social Media-based Sentiment Analysis: Emerging Trends and Challenges*," *Decision Analytics Journal*, Vol. 3, p. 100073, Jun. 2022, DOI: 10.1016/j.dajour.2022.100073.

 ISSN:2302-8149
e-ISSN:2540-9719

[3] B. Liu, J. Zhao, K. Liu, and L. Xu, "*Book Review Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*," *Press*, 2015, DOI: 10.1162/COLI.

[4] M. Amien, "Sejarah dan Perkembangan Teknik *Natural Language Processing (NLP)* Bahasa Indonesia: Tinjauan tentang Sejarah, Perkembangan Teknologi, dan Aplikasi *NLP* dalam Bahasa Indonesia", DOI: 10.48550/arXiv.2304.02746.

[5] Y. Song, X. Liu, and Z. Zhou, "*A Comprehensive Review of Text Classification Algorithms*," J. Electron. Inf. SCI., Vol. 9, No. 2, pp. 34–42, 2024, DOI: 10.23977/jeis.2024.090205.

[6] A. K. Hidayah, Y. Erwadi, and S. Handayani, "Analisis Sentimen Publik terhadap Pemindahan Ibu Kota Negara di *Twitter* menggunakan Metode Klasifikasi *Random Forest* dan *Smote*," 2025.

[7] A. de la Cruz Huayanay, J. L. Bazán, and C. M. Russo, "*Performance of Evaluation Metrics for Classification in Imbalanced Data*," *Comput Stat*, Vol. 40, No. 3, pp. 1447–1473, 2025, DOI: 10.1007/s00180-024-01539-5.

[8] N. Arifin, U. Enri, and N. Sulistiyowati, "Satuan Tulisan Riset dan Inovasi Teknologi Penerapan Algoritma *Support Vector Machine (SVM)* dengan *TF-IDF N-Gram* untuk *Tect Classification*."

[9] A. Salhi, R. Alshamrani, A. Althbiti, A. Ismail, M. Abd-ElRahman, and B. M. Hassan, "*Optimizing High Dimensional Data Classification with a Hybrid AI Driven Feature Selection Framework and Machine Learning Schema*," *SCI Rep*, Vol. 15, No. 1, p. 35038, Dec. 2025, DOI: 10.1038/s41598-025-08699-4.

[10] I. Rifky Hendrawan, E. Utami, and A. D. Hartanto, "Analisis Perbandingan Metode *TF-IDF* dan *Word2vec* pada Klasifikasi Teks Sentimen Masyarakat terhadap Produk Lokal di Indonesia."

[11] M. Bansal, A. Goyal, and A. Choudhary, "*A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning*," *Decision Analytics Journal*, Vol. 3, p. 100071, Jun. 2022, DOI: 10.1016/j.dajour.2022.100071.

[12] V. Pestov, "Is the K-NN *Classifier in High Dimensions Affected by the Curse of Dimensionality?*," *Computers and Mathematics with Applications*, Vol. 65, No. 10, pp. 1427–1437, 2013, DOI: 10.1016/j.camwa.2012.09.011.

[13] R. Agrawal, M. Majumder, I. Yadav, N. Taneja, S. Hamdare, and P. Hemnani, "*Evaluating Sentiment Analysis Models: A Comparative Analysis of Vaccination Tweets During the COVID-19 Phase Leveraging DistilBERT for Enhanced Insights*," *MethodsX*, Vol. 14, Jun. 2025, DOI: 10.1016/j.mex.2025.103407.

[14] S. K. Adharani, S. Kacung, and A. V. Vitianingsih, "*Sentiment Analysis on Indonesian National Football Team Naturalization using KNN and SVM*," *Edumatic: Jurnal Pendidikan Informatika*, Vol. 9, No. 1, pp. 189–197, Apr. 2025, DOI: 10.29408/edumatic.v9i1.29653.

[15] L. Widiastuti, D. Nurlaela, A. Vol, L. D. Utami, A. Surniandari, and P. Studi Sistem Informasi Akuntansi Kampus Kota Bogor, "Jusikom : Jurnal Sistem Komputer Musi Rawas."

[16] A. Kataria, M. Singh, W. : Www, and M. D. Singh, "*A Review of Data Classification using K-Nearest Neighbour Algorithm International Journal of Emerging Technology and Advanced Engineering A Review of Data Classification using K-Nearest Neighbour Algorithm*," 2008. [Online]. Available: https://www.researchgate.net/publication/353306410

[17] M. Wang, X. Xu, Q. Yue, and Y. Wang, "*A Comprehensive Survey and Experimental Comparison of Graph-based Approximate Nearest Neighbor Search*," May 2021, [Online]. Available: http://arxiv.org/abs/2101.12631

[18] C. Schröer, F. Kruse, and J. M. Gómez, "*A Systematic Literature Review on Applying CRISP-DM Process Model*," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. DOI: 10.1016/j.procs.2021.01.199.

[19] T. Oswari, Murniyati, T. Yusnitasari, Nurasiah, and S. Wijaya, "*Sentiment Analysis of Indonesian YouTube Reviews about Lesbian, Gay, Bisexual, and Transgender (LGBT) using IndoBERT Fine Tuning*," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, Vol. 15, No. 01, pp. 26–37, Oct. 2025, DOI: 10.24843/LKJITI.2024.v15.i01.p03.

[20] Y. A. Singgalen, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel berbasis Algoritma *Decision Tree (DT)*," *Jurnal Sistem*

ISSN:2302-8149
e-ISSN:2540-9719

Komputer dan Informatika (JSON), Vol. 5, No. 2, p. 237, Dec. 2023, DOI: 10.30865/json.v5i2.7081.

[21] W. Nugraha and A. Sasongko, "*Hyperparameter Tuning* pada Algoritma Klasifikasi dengan *Grid Search*," *Sistemasi: Jurnal Sistem Informasi.*, Vol. 11, No. 2, pp. 391–401, May 2022, DOI: 10.32520/stmsi.v11i2.1766.

[22] M. Hasan, M. F. Rabbi, M. N. Sultan, A. M. Nitu, and M. P. Uddin, "*A Novel Data Balancing Technique Via Resampling Majority and Minority Classes Toward Effective Classification*," *Telkomnika (Telecommunication Computing Electronics and Control)*, Vol. 21, No. 6, pp. 1308–1316, 2023, Doi: 10.12928/TELKOMNIKA.V21I6.25211.

[23] L. A. Pekandi, R. G. Widjaja, A. Ananta, J. Harefa, and K. Jingga, "*Evaluating IndoBERT for Indonesian Hoax News Detection: A Comparative Study with Ensemble and CNN-LSTM Models*," *Procedia Comput SCI*, Vol. 269, pp. 1625–1633, 2025, DOI: 10.1016/j.procs.2025.09.105.

[24] F. Koto, J. H. Lau, and T. Baldwin, "*IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization*," Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.04607