

KOMPARASI KINERJA ALGORITMA MACHINE LEARNING UNTUK ANALISIS SENTIMEN MULTI-KELAS PADA DATA TEKS DIGITAL TERHADAP KEBIJAKAN PEMERINTAH

Sestri Novia Rizki^{1,*}, Sunardi², Herman Yuliansyah³

^{1,3}Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

²Sistem Informasi, Universitas Media Nusantara Citra, Jakarta, Indonesia

Email: ¹2536083029@webmail.uad.ac.id, ²sunardi@uad.ac.id, ³herman.yuliansyah@uad.ac.id

Diajukan: 11-08-2026 Direvisi: 29-08-2026	Disetujui: 01-09-2026 Diterbitkan: 30-09-2026
Abstract. The growth of social media and digital platforms has generated vast amounts of text data, serving as a crucial source for understanding public opinion on government policies. This study aims to compare the performance of machine learning algorithms in multi-class sentiment classification regarding government policies, specifically the Free Nutritious Meal (MBG) program. The dataset comprises 4,969 public opinion entries sourced from Kaggle, classified into three sentiment categories: positive, negative, and neutral. The research methodology involves text preprocessing—including case folding, cleaning, tokenization, stopword removal, and stemming—followed by feature weighting using TF-IDF, data splitting (80% for training and 20% for testing), and classification using Naïve Bayes, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Decision Tree. Performance evaluation metrics include accuracy, precision, recall, and F1-score. The results indicate that Random Forest achieved the best performance, with an accuracy of 0.81 and a Macro F1-score of 0.81, followed by SVM and Decision Tree (0.75 accuracy), Naïve Bayes (0.70), and KNN (0.52). After hyperparameter tuning, Random Forest (with $n_estimators = 100$) achieved an accuracy of 0.839, while LinearSVC (with $C = 1$) achieved 0.794. These findings demonstrate that Random Forest is more robust in classifying positive, negative, and neutral sentiments within digital text data concerning government policies. Keywords: Algoritma, Machine Learning, Sentiment Analysis, Social Media, Teks Digital Abstrak. Perkembangan media sosial dan platform digital menghasilkan volume data teks yang semakin besar serta menjadi sumber penting untuk memahami opini masyarakat terhadap kebijakan pemerintah. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Machine Learning dalam klasifikasi sentimen multi-kelas terhadap kebijakan pemerintah, khususnya program Makan Bergizi Gratis (MBG). Dataset yang digunakan terdiri dari 4.969 data opini publik yang diperoleh dari Kaggle dan diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Tahapan penelitian meliputi preprocessing teks, yang terdiri atas case folding, cleansing, tokenizing, stopword removal, dan stemming, dilanjutkan dengan pembobotan fitur menggunakan TF-IDF, pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian, serta klasifikasi menggunakan Naïve Bayes, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), dan Decision Tree. Evaluasi kinerja dilakukan menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Random Forest memberikan kinerja terbaik dengan akurasi 0,81 dan Macro F1-score 0,81, diikuti SVM dan Decision Tree dengan akurasi 0,75, Naïve Bayes sebesar 0,70, dan KNN sebesar 0,52. Setelah hyperparameter tuning, Random Forest dengan $n_estimators = 100$ memperoleh akurasi sebesar 0,839, sedangkan LinearSVC dengan $C = 1$ memperoleh akurasi sebesar 0,794. Hasil tersebut menunjukkan bahwa Random Forest lebih stabil dalam mengklasifikasikan sentimen positif, negatif, dan netral pada data teks digital terkait kebijakan pemerintah Kata Kunci: Algoritma, analisis Sentimen, Machine Learning, Teks digital, Sosial Media	

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat menyampaikan opini terhadap kebijakan pemerintah termasuk Makan Bergizi Gratis (MBG). Program MBG dirancang untuk memberikan intervensi langsung terhadap persoalan kesehatan dan pendidikan anak-anak Indonesia (Keladu et al., 2026). Media sosial seperti *X*, *Instagram*, *TikTok*, dan *YouTube* menjadi ruang publik digital bagi masyarakat untuk menyampaikan dukungan, kritik, dan respons secara cepat, sehingga data teks digital dapat dimanfaatkan untuk memahami dinamika sentimen publik terhadap kebijakan (Dietterich & Bakiri, 1995; Harwenda et al., 2025). Salah satu metode yang digunakan untuk memahami opini masyarakat adalah analisis sentimen melalui data teks digital dari media sosial (Elinda et al., 2024). Dalam proses analisis sentimen, pendekatan klasifikasi multi-kelas digunakan untuk membagi opini masyarakat ke dalam kategori positif, negatif, dan netral sehingga hasil analisis menjadi lebih detail dibandingkan klasifikasi biner (S. & Yuliansyah, 2024a). Namun, data teks digital berbahasa Indonesia memiliki tantangan tersendiri karena mengandung karakter informal, singkatan, emotikon, dan bahasa campuran. Oleh karena itu, *preprocessing* menjadi tahapan penting sebelum proses klasifikasi dilakukan. Tahapan seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming* digunakan untuk membersihkan serta menyusun data teks agar lebih terstruktur (S. & Yuliansyah, 2024b). Perkembangan penelitian di bidang *Machine Learning* menunjukkan bahwa pendekatan tersebut efektif digunakan dalam analisis sentimen kebijakan publik. Zulfikar et al. (2023) menunjukkan bahwa analisis sentimen berbasis *machine learning* dapat digunakan untuk mengidentifikasi respons masyarakat terhadap kebijakan publik melalui data media sosial. Setyawan et al. (2023) membuktikan bahwa algoritma Naïve Bayes dan SVM mampu mengklasifikasikan opini masyarakat terhadap pemerintah berdasarkan data Twitter. Sementara itu, Zalukhu et al. (2023) menunjukkan bahwa analisis sentimen dapat menggambarkan respons masyarakat terhadap kebijakan kenaikan harga BBM.

2. Tinjauan Pustaka

2.1 Machine Learning

Machine Learning merupakan cabang Artificial Intelligence (AI) yang memungkinkan komputer belajar dari data, mengenali pola, serta membuat prediksi atau keputusan secara otomatis tanpa harus diprogram secara khusus untuk setiap tugas.. Menurut Nichols, Chan, dan Baker (2019), *machine learning* mencakup berbagai algoritma yang mampu menghasilkan prediksi cerdas berdasarkan data yang tersedia. Choi dkk. (2020) *Machine Learning* berfokus pada pengembangan algoritma yang mampu mempelajari pola dari data untuk melakukan prediksi atau klasifikasi secara otomatis. Dalam analisis sentimen, metode ini digunakan untuk mengelompokkan opini pengguna ke dalam sentimen positif, negatif, atau netral. Beberapa algoritma yang umum digunakan antara lain: *Support Vector Machine* (SVM), *Naïve Bayes* (NB), *K-Nearest Neighbor* (KNN), *Decision Tree* (DT), dan *Random Forest* (RF).

2.2 Analisis Sentimen

Analisis sentimen salah satu teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi opini, emosi, dan sikap seseorang terhadap suatu topik (Alfaridzy et al., 2025). Perkembangan media sosial menyebabkan meningkatnya jumlah data teks sehingga analisis sentimen dapat digunakan untuk mengetahui opini publik di media digital (Yuliansyah, H. et al., 2025). Analisis sentimen multi-label digunakan untuk mengelompokkan satu opini ke dalam beberapa kategori sekaligus (Ardiansyah et al., 2025). Metode ini memiliki tantangan seperti struktur bahasa yang kompleks dan ketidakseimbangan data Yuliansyah, H. et al., 2026). Untuk mengatasi ketidakseimbangan data, digunakan teknik SMOTE dan ADASYN untuk menyeimbangkan distribusi data sehingga model lebih stabil dalam mengenali setiap label sentimen (Hetherington, 2025). Teknik *oversampling* juga dapat meningkatkan kinerja klasifikasi, terutama pada data dengan distribusi sentimen yang tidak seimbang. (H. Yuliansyah, H. et al., 2025), sehingga opini publik terhadap kebijakan pemerintah memiliki karakteristik multikelas. Metode klasifikasi memerlukan kemampuan untuk menangani lebih dari dua kategori sentimen (Panjaitan et al., 2025). Pengelompokkan opini publik

dari data teks ke dalam sentimen positif, negatif, atau netral (Fauzi et al., 2026). Penelitian menunjukkan respons masyarakat terhadap kebijakan tersebut cukup beragam. (Leidiyana et al., 2024). Klasifikasi Sentimen Multi-kelas Klasifikasi multi-kelas digunakan untuk mengelompokkan data ke dalam lebih dari dua kategori sentimen, seperti positif, negatif, dan netral. (Dietterich & Bakiri, 1995). Penelitian metode SVM multi-kelas mampu mengklasifikasikan sentimen positif, negatif, dan netral secara efektif (Dasriani et al., 2024). Bentuk opini publik melalui proses pengolahan dan analisis data teks yang berasal dari media digital (Prasetyoningrum & Andono, 2024).

2.3 Data Teks Digital

Data teks digital merupakan data tidak terstruktur yang berasal dari media sosial, forum *online*, dan berita digital sehingga perlu melalui tahap *preprocessing* sebelum dilakukan klasifikasi. Tahapan yang dilakukan meliputi *case folding*, *tokenizing*, *stopword removal*, *stemming*, dan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode TF-IDF digunakan untuk mengukur tingkat kepentingan suatu kata dalam dokumen. Semakin tinggi nilai TF-IDF, semakin besar kontribusi kata tersebut dalam membedakan suatu dokumen dari dokumen lainnya. Maka nilai bobot kata tersebut akan semakin tinggi. Perhitungan TF-IDF ditunjukkan pada Persamaan (1).

$$TF - IDF = tf(t, d) \times \log \frac{N}{df(t)} \quad (1)$$

Keterangan:

$tf(t, d)$ = Jumlah kemunculan term t pada dokumen d

N = Jumlah seluruh dokumen

$df(t)$ = Jumlah dokumen yang mengandung term t

3. Metode Penelitian

Penelitian ini merupakan studi analisis komparatif yang bertujuan untuk membandingkan performa beberapa algoritma *Machine Learning* berbasis *supervised learning* dalam melakukan klasifikasi sentimen multikelas pada data teks digital terkait kebijakan pemerintah, khususnya pada kasus MBG. Tahapan penelitian dilakukan secara terstruktur mulai dari proses pengumpulan data, pelabelan sentimen, *preprocessing* teks, ekstraksi fitur, pembagian *dataset*, implementasi algoritma, hingga evaluasi kinerja model. Data penelitian berupa opini masyarakat pada media digital yang diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Secara umum, penelitian ini difokuskan untuk menganalisis dan membandingkan tingkat akurasi serta performa masing-masing algoritma dalam klasifikasi sentimen multi-kelas. Alur penelitian secara keseluruhan ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

Alur penelitian dimulai dari studi literatur, dilanjutkan pengumpulan 4.969 data Kaggle dan opini publik. Data kemudian melalui *preprocessing*, ekstraksi fitur TF-IDF, serta pembagian data 80% training dan 20% testing. Selanjutnya dilakukan klasifikasi menggunakan SVM, Naive Bayes, Random Forest, KNN, dan Decision Tree, kemudian dievaluasi dengan accuracy, precision, recall, dan F1-score.

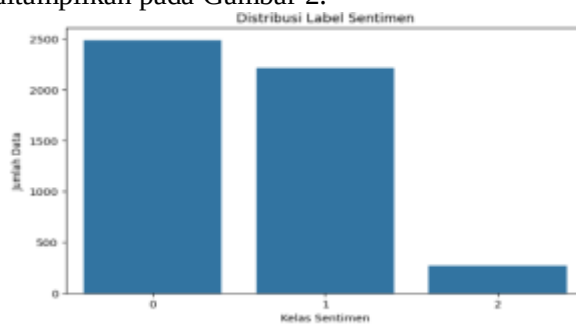
3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa *dataset* opini publik mengenai program MBG yang diperoleh dari platform digital Kaggle melalui tautan berikut. <https://www.kaggle.com/code/zhulfanifaisaladam2/klasifikasi-sentimen-analisis-mbg>. *Dataset* dikumpulkan dari berbagai komentar masyarakat berbahasa Indonesia yang berasal dari media sosial dan sumber digital lain yang membahas kebijakan pemerintah terkait program MBG. Proses pengambilan data dilakukan dengan metode *crawling* dan pengumpulan komentar digital yang tersedia pada *platform* tersebut. Data yang digunakan dalam penelitian ini berjumlah 4.969 data teks yang terdiri dari opini masyarakat berupa dukungan, kritik, maupun tanggapan terhadap kebijakan pemerintah. Seluruh *dataset* kemudian digunakan sebagai bahan analisis dalam proses klasifikasi sentimen *multiclass* untuk membandingkan performa beberapa algoritma *Machine Learning*. Contoh data yang akan diolah berupa teks komentar publik mengenai program Makan Bergizi Gratis (MBG), misalnya:

1. “Program MBG sangat membantu masyarakat dan anak-anak.”
2. “Pelaksanaan MBG masih banyak kekurangan.”
3. “Saya mendukung program MBG karena bermanfaat bagi siswa.”

3.2 Pelabelan Data

Pada tahap pelabelan, data diklasifikasikan menjadi negatif (label 0), positif (label 1), dan netral (label 2) berdasarkan isi komentar MBG. Hasilnya terdiri dari 2.480 data negatif, 2.220 positif, dan 270 netral. Distribusi sentimen ditampilkan pada Gambar 2.



Gambar 2. Distribusi Kelas Sentimen

3.3 Data Preprocessing

Tahap *preprocessing* bertujuan membersihkan dan menyiapkan data teks melalui *case folding*, *cleansing*, *tokenizing*, *filtering*, dan *stemming*. Dari 4.970 data, sebanyak 4.900 data digunakan setelah data kosong dihapus. *Wordcloud* menunjukkan kata dominan seperti “makan”, “gratis”, “anak”, “sekolah”, dan “program”.



Gambar 3. Wordcloud Setelah Preprocessing

3.4 Feature Extraction

Setelah *preprocessing*, dilakukan ekstraksi fitur menggunakan TF-IDF untuk mengubah teks menjadi data numerik. TF-IDF memberikan bobot lebih tinggi pada kata yang penting, sehingga membantu algoritma dalam melakukan klasifikasi sentimen. (Rizkiyanto et al., 2024). Metode *TF-IDF* mampu merepresentasikan kata berdasarkan tingkat frekuensi dan relevansinya dalam dokumen

sehingga efektif digunakan dalam klasifikasi sentimen data teks digital (Abdillah et al., 2024). Perbandingan metode *TF-IDF* dan *GloVe* menunjukkan bahwa *TF-IDF* memberikan performa yang lebih konsisten dan akurat pada analisis sentimen politik di media sosial (Rahmatullah & Annisa, 2026). *Feature extraction* menggunakan *TF-IDF* dan *Word2Vec* mengubah teks menjadi representasi numerik agar model dapat mengenali pola dan karakteristik sentimen positif, negatif, dan netral dengan lebih baik. Perhitungan *TF-IDF* ditunjukkan pada Persamaan (2)-(4).

$$TF = \frac{\text{Jumlah kemunculan term } t \text{ pada dokumen } d}{\text{Total kata dalam dokumen } d} \quad (2)$$

$$IDF_{(t)} = \log \left(\frac{N}{df_{(t)}} \right) \quad (3)$$

$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)} \quad (4)$$

3.5 Split Data

Dataset dibagi menjadi 80% data latih dan 20% data uji. Dari 4.968 data, diperoleh 3.974 data latih dan 994 data uji. Data latih digunakan untuk mempelajari pola sentimen, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam memprediksi data baru.

3.6 Classification

Tahap klasifikasi menggunakan fitur *TF-IDF* sebagai input dengan pendekatan *Supervised Learning*. Penelitian ini membandingkan empat algoritma klasifikasi untuk mengidentifikasi sentimen positif, negatif, dan netral. *SVM* dengan *Word2Vec* mampu meningkatkan akurasi klasifikasi sentimen pada politik digital Indonesia (Damayanti & Lhaksmana, 2024). *SVM* digunakan untuk memisahkan data berdasarkan *hyperplane* terbaik. *SVM* efektif untuk data teks dan memiliki akurasi serta kemampuan generalisasi yang baik, sehingga cocok digunakan dalam analisis sentimen. (Nur Cahyo et al., 2023). Secara matematis, *SVM* bekerja dengan menentukan *hyperplane* optimal untuk memisahkan data antar kelas. *Hyperplane* pada *SVM* dapat dituliskan sebagai Perhitungan *SVM* ditunjukkan pada Persamaan . *Naïve Bayes* merupakan metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi setiap fitur saling independen. Metode ini cepat dan efisien, sehingga banyak digunakan untuk klasifikasi teks dan analisis sentimen. (Lestari & Hutagalung, 2025). Bagian *Naïve Bayes* sudah cukup baik karena sudah terdapat Persamaan (5-7).

$$f(x) = \omega \times x + b \quad (5)$$

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (6)$$

keterangan:

$P(C|X)$ = probabilitas data X termasuk kelas C

$P(X|C)$ = probabilitas fitur terhadap kelas

$P(C)$ = probabilitas awal kelas

$P(X)$ = probabilitas data

Algoritma *Random Forest* digunakan karena kemampuannya menangani data medis yang kompleks dan non linear. Model ini juga memiliki keunggulan dalam memberikan interpretasi terhadap pengaruh setiap fitur melalui analisis *feature importance* (Ripaldi et al., 2025). *Random forest* bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) secara acak, kemudian hasil klasifikasi ditentukan berdasarkan voting mayoritas dari seluruh pohon, seperti pada Persamaan (7).

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (7)$$

KNN merupakan algoritma *Machine Learning* yang melakukan klasifikasi berdasarkan kedekatan jarak antar data. Metode ini menentukan kelas suatu data baru berdasarkan mayoritas kelas

dari sejumlah tetangga terdekatnya. Karena sederhana dan efektif dalam mengenali pola berdasarkan kemiripan data, KNN banyak digunakan dalam analisis sentimen untuk mengelompokkan data ke dalam kategori positif, negatif, dan netral (Lillah et al., 2023; Muhidin & Wibowo, 2020). Untuk mengukur kedekatan antar data, penelitian ini menggunakan jarak Euclidean sebagaimana ditunjukkan pada Persamaan (8).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (8)$$

Keterangan:

$d(x, y)$ = Jarak antara data x dan y
 x_i = nilai atribut ke-i pada data x
 y_i = nilai atribut ke-i pada data y
 n = jumlah atribut atau fitur

Decision Tree merupakan algoritma Machine Learning yang mengklasifikasikan data melalui struktur pohon keputusan berdasarkan aturan tertentu. Dalam analisis sentimen, algoritma ini digunakan untuk mengelompokkan sentimen positif, negatif, dan netral serta mudah dipahami. (Harun & Fahmi, 2024; Rahman & Lestari, 2024). Untuk mengukur tingkat ketidakpastian (*impurity*) atau keragaman data dalam suatu himpunan ditunjukkan pada Persamaan (8).

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i \dots\dots \quad (9)$$

Keterangan

S = Himpunan Data
 P_i = Probabilitas Tiap Kelas
 n = Jumlah Kelas

3.7 Model Evaluation

Evaluasi performa model dalam penelitian ini menggunakan *confusion matrix*, yang membandingkan label aktual dengan hasil prediksi model. Performa model dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-Score* yang dihitung berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) pada *confusion matrix*. Persamaan yang digunakan untuk menghitung masing-masing metrik ditunjukkan pada Persamaan (10) sampai Persamaan (13).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Nilai *precision* digunakan untuk mengukur ketepatan prediksi positif yang dihasilkan model:

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Nilai *recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data positif:

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

Sementara itu, *F1-score* digunakan untuk mengukur keseimbangan antara nilai *precision* dan *recall* untuk memperoleh evaluasi performa model yang lebih seimbang (Chicco et al., 2021) dengan persamaan sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

4. Hasil dan Pembahasan

4.1 Analisis Karakteristik Data

Berdasarkan hasil pengamatan terhadap 4969 data, ditemukan adanya ketidakseimbangan kelas (*class imbalance*) pada distribusi sentimen. Sentimen negatif memiliki jumlah data terbanyak yaitu 2486

data, diikuti sentimen positif sebanyak 2212 data, sedangkan sentimen netral hanya berjumlah 270 data. Kondisi ini menunjukkan bahwa pembahasan publik terkait kasus MBG di *platform* tersebut lebih banyak berisi kritik, keluhan, maupun kekhawatiran dibandingkan opini yang bersifat netral atau positif.

Machine Learning berdasarkan hasil pengujian yang dilakukan terhadap beberapa algoritma klasifikasi, yaitu *Naive Bayes*, *Random Forest*, *SVM*, *KNN*, *Decision Tree*, diperoleh perbedaan performa yang cukup signifikan. Penilaian kinerja model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* guna memberikan evaluasi yang lebih menyeluruh terhadap kemampuan masing-masing algoritma dalam melakukan klasifikasi data.

Tabel 1. Perbandingan Algoritma Machine Learning

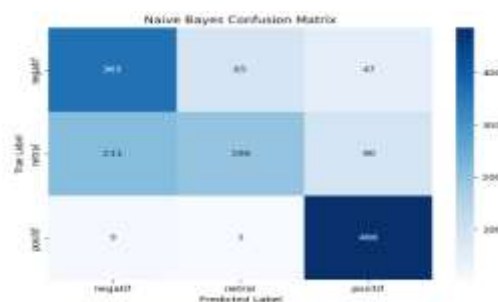
No	Algoritma	Akurasi	Presisi Positif	Presisi Negatif	Presisi Netral
1	<i>Naive Bayes</i>	0,70	0,77	0,63	0,68
2	<i>Random Forest</i>	0,81	0,92	0,78	0,73
3	<i>Support Vector Machine (SVM)</i>	0,75	0,90	0,71	0,64
4	<i>K-Nearest Neighbors (KNN)</i>	0,52	0,49	0,60	0,58
5	<i>Decision Tree</i>	0,75	0,89	0,69	0,65

Berdasarkan Tabel 1, *Random Forest* memiliki akurasi terbaik sebesar 0,81, diikuti *SVM* dan *Decision Tree* sebesar 0,75, sedangkan *KNN* terendah sebesar 0,52. Untuk mengetahui kemampuan model dalam mengenali setiap kelas secara lebih menyeluruh, evaluasi dilanjutkan menggunakan *recall* dan *F1-score* pada Tabel 2.

Tabel 2. Recall dan F1-Score Algoritma Machine Learning

No	Algoritma	Recall Positif	Recall Negatif	Recall Netral	F1-Score	F1 Positif	F1 Negatif	F1 Netral
1	<i>Naive Bayes</i>	0,98	0,72	0,40	0,68	0,86	0,67	0,50
2	<i>Random Forest</i>	0,98	0,72	0,74	0,81	0,95	0,75	0,74
3	<i>SVM</i>	0,95	0,66	0,66	0,75	0,92	0,68	0,65
4	<i>KNN</i>	0,98	0,02	0,56	0,42	0,65	0,04	0,57
5	<i>Decision Tree</i>	0,94	0,64	0,66	0,75	0,91	0,67	0,65

Hasil pengujian menunjukkan *Random Forest* memiliki performa terbaik dengan akurasi 0,81 dan presisi kelas positif 0,92. Keunggulan ini dipengaruhi oleh penggunaan banyak pohon keputusan (*ensemble learning*) yang meningkatkan stabilitas dan mengurangi kesalahan klasifikasi. Sementara itu, *Naive Bayes* juga diuji menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk mengetahui kemampuan klasifikasi pada setiap kelas sentimen. Hasilnya ditampilkan pada Gambar 4.



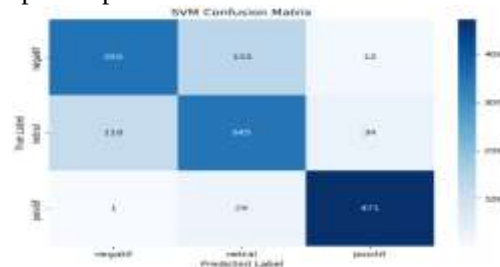
Gambar 4. Confusion Matrix Naive Bayes

Berdasarkan *confusion matrix*, *Naive Bayes* cukup baik dalam mengklasifikasikan sentimen positif dan negatif, tetapi masih kurang optimal pada kelas netral. Sementara itu, *Random Forest* menghasilkan prediksi yang lebih stabil dan akurat pada setiap kelas. Evaluasi *confusion matrix* *Random Forest* ditampilkan pada Gambar 5.



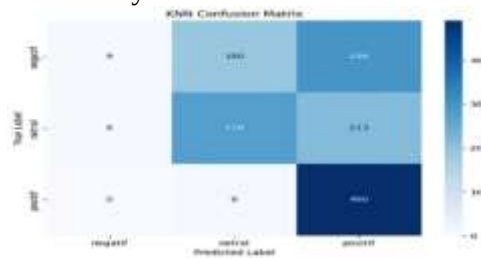
Gambar 5. Hasil Evaluasi Random Forest

Hasil pengujian SVM menunjukkan kemampuan yang baik dalam mengklasifikasikan sentimen positif, negatif, dan netral. Evaluasi dilakukan menggunakan akurasi, presisi, recall, dan F1-score, dengan hasil performa ditampilkan pada Gambar 6.



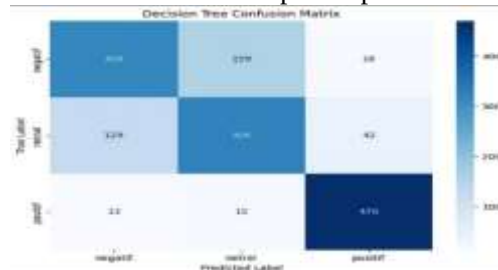
Gambar 6. Confusion Matrix SVM

Berdasarkan *confusion matrix*, SVM mampu mengklasifikasikan sebagian besar data dengan benar pada ketiga kelas sentimen. Sementara itu, evaluasi *confusion matrix* juga dilakukan pada KNN untuk melihat tingkat ketepatan klasifikasinya. Hasil evaluasi KNN ditampilkan pada Gambar 7.



Gambar 7. Hasil Evaluasi KNN

Berdasarkan *confusion matrix*, KNN menunjukkan hasil klasifikasi pada ketiga kelas sentimen. Selanjutnya, *Decision Tree* dievaluasi menggunakan *confusion matrix* untuk melihat jumlah prediksi yang benar dan salah. Hasil evaluasi *Decision Tree* ditampilkan pada Gambar 8.



Gambar 8. Hasil Evaluasi Decision Tree

Berdasarkan *confusion matrix*, *Decision Tree* mampu mengklasifikasikan sebagian besar data dengan benar pada ketiga kelas sentimen. Secara keseluruhan, hasil *confusion matrix* dan *classification report* menunjukkan bahwa setiap algoritma memiliki performa yang berbeda dalam mengklasifikasikan sentimen negatif, netral, dan positif.

4.2 Best $n_estimators$ for Random Forest: 100

Best accuracy for Random Forest with tuned $n_estimators$: 0.8390963887939256
Hyperparameter Tuning untuk SVM (LinearSVC)

```
from sklearn.model_selection import GridSearchCV

# Define the parameter grid for Random Forest
param_grid_rf = {
    'n_estimators': [50, 100, 200, 300]
}

# Initialize GridSearchCV
grid_search_rf = GridSearchCV(
    RandomForestClassifier(random_state=42),
    param_grid_rf,
    cv=5, # Using the same kfold as before
    scoring='accuracy',
    n_jobs=-1 # Use all available cores
)

# Fit GridSearchCV to the SMML-processed data
grid_search_rf.fit(X_smote, y_smote)

print("Best n_estimators for Random Forest: (grid_search_rf.best_params_['n_estimators'])")
print("Best accuracy for Random Forest with tuned n_estimators: (grid_search_rf.best_score_)"

Best n_estimators for Random Forest: 100
Best accuracy for Random Forest with tuned n_estimators: 0.8390963887939256
```

Gambar 9. Hyperparameter Tuning untuk SVM (LinearSVC)

Untuk *LinearSVC*, parameter yang relevan adalah C . Parameter γ umumnya digunakan untuk kernel SVM seperti RBF atau polinomial, yang tidak digunakan oleh *LinearSVC*. Best C for *LinearSVC*: 1 Best accuracy for *LinearSVC* with tuned C : 0.7940439390395306

```
param_grid_svm = {
    'C': [0.1, 1, 10, 100]
}

grid_search_svm = GridSearchCV(
    LinearSVC(random_state=42, dual=False), # dual=False is recommended for n_samples > n
    param_grid_svm,
    cv=5, # Using the same kfold as before
    scoring='accuracy',
    n_jobs=-1 # Use all available cores
)

grid_search_svm.fit(X_smote, y_smote)

print("Best C for LinearSVC: (grid_search_svm.best_params_['C'])")
print("Best accuracy for LinearSVC with tuned C: (grid_search_svm.best_score_)"

*** Best C for LinearSVC: 1
Best accuracy for LinearSVC with tuned C: 0.7940439390395306
```

Gambar 10. Best accuracy for LinearSVC with tuned

Setelah melakukan *tuning*, kita mendapatkan nilai optimal untuk $n_estimators$ pada Random Forest dan C pada *LinearSVC*. Nilai-nilai ini dapat digunakan untuk melatih model akhir guna mendapatkan performa terbaik. Analisis Mendalam Mengapa KNN Berkinerja Buruk, KNN mengklasifikasikan data baru berdasarkan kelas mayoritas dari k tetangga terdekat. Pada data teks berbasis TF-IDF, KNN cenderung berkinerja rendah karena memiliki banyak fitur, data yang sparse, dan fitur yang kurang relevan, sehingga perhitungan jarak menjadi kurang akurat.

1. *Curse of Dimensionality*: TF-IDF memiliki 5.000 fitur, sehingga jarak antar data menjadi sulit dibedakan.
2. Data Sangat *Sparse*: Sebagian besar nilai TF-IDF adalah 0, sehingga perhitungan jarak kurang efektif.
3. Fitur Tidak Relevan: KNN menganggap semua fitur sama pentingnya, sehingga fitur yang kurang relevan dapat mengganggu hasil klasifikasi.
4. Komputasi Berat: KNN harus menghitung jarak ke banyak data, sehingga kurang efisien pada data berdimensi tinggi.

Perbandingan dengan Model Lain:

1. *Random Forest*: Menggunakan banyak pohon keputusan sehingga lebih kuat dalam menangani data teks yang memiliki banyak fitur dan mengurangi kesalahan prediksi.
2. *SVM (LinearSVC)*: Mencari batas pemisah terbaik antar kelas dan efektif untuk data teks dengan jumlah fitur yang tinggi.
3. *Naïve Bayes (MultinomialNB)*: Cocok untuk klasifikasi teks karena mampu bekerja dengan baik pada data TF-IDF yang memiliki banyak fitur.
4. *Decision Tree*: Membagi data berdasarkan fitur yang paling penting sehingga dapat mengabaikan sebagian fitur yang kurang relevan.

4. Simpulan

Berdasarkan hasil pengujian, Random Forest merupakan algoritma dengan performa terbaik, dengan akurasi 0,81 dan Macro *F1-score* 0,81, dibandingkan Naive Bayes (akurasi 0,70; Macro *F1-score* 0,68), SVM (0,75; 0,75), KNN (0,52; 0,42), dan Decision Tree (0,75; 0,75). Macro *F1-score* digunakan sebagai metrik agregat karena memberikan gambaran performa model pada seluruh kelas sentimen, termasuk kelas netral yang memiliki jumlah data lebih sedikit. Hasil evaluasi menunjukkan bahwa Random Forest lebih stabil dalam mengenali sentimen positif, negatif, dan netral. Keunggulan ini berkaitan dengan mekanisme *ensemble learning* yang menggabungkan banyak pohon keputusan sehingga mampu menangani data teks berdimensi tinggi dengan lebih baik. Sementara itu, KNN memiliki performa terendah karena sensitif terhadap dimensi tinggi dan sparsitas fitur TF-IDF. Setelah dilakukan *hyperparameter tuning*, Random Forest dengan $n_estimators = 100$ menghasilkan akurasi sebesar 0,839, sedangkan LinearSVC dengan $C = 1$ menghasilkan akurasi 0,794. Dengan demikian, Random Forest dipilih sebagai model terbaik dan paling optimal untuk klasifikasi sentimen pada data kebijakan MBG dalam penelitian ini.

Ucapan Terimakasih

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Profesor pembimbing atas arahan, masukan, dan bimbingan yang diberikan selama proses penyusunan disertasi. Penulis juga mengucapkan terima kasih kepada pengelola jurnal atas kesempatan dan proses penelaahan yang konstruktif, serta kepada Universitas MNC atas dukungan sistem, fasilitas, dan lingkungan akademik yang mendukung terlaksananya penelitian ini.

Referensi

- Abdillah, T. F., Hasmawati, H., & Bunyamin, B. (2024). Comparison of TF-IDF and GloVe Word Embedding for Sentiment Analysis of 2024 Presidential Candidates. *Building of Informatics, Technology and Science (BITS)*, 6(2), 961–969. <https://doi.org/10.47065/bits.v6i2.5668>
- Alfaridzy, M. A., Haerani, E., Jasril, & Oktavia, L. (2025). keuangan negara . Secara umum , efisiensi anggaran merujuk pada kemampuan pemerintah dalam pemerintahan (Alfandi Safira & Hasan , 2023). yang dominan dalam masyarakat (Doloksaribu & Yusran Timur Samuel , 2022). Salah satu metode yang efektif untuk mel. 6, 169–182.
- Ardiansyah, R., Yuliansyah, H., & Yudhana, A. (2025). Multi-Label Opinion Mining Based on Random Forest with SMOTE and ADASYN. *Compiler*, 14(2), 65–75. <https://doi.org/10.28989/compiler.v14i2.3185>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Damayanti, L., & Lhaksana, K. M. (2024). Sentiment Analysis of the 2024 Indonesia Presidential Election on Twitter. *Sinkron*, 8(2), 938–946. <https://doi.org/10.33395/sinkron.v8i2.13379>
- Dasriani, N. G. A., Pariandi, L. A. G., & Dharma, I. M. Y. (2024). Analisis Sentimen Program Jaminan Kesehatan Nasional Menggunakan Multiclass Support Vector Machine. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 6(1), 20–30. <https://doi.org/10.37148/bios.v6i1.136>
- Dietterich, T. G., & Bakiri, G. (1995). Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research*, 2, 263–286. <https://doi.org/10.1613/jair.105>
- Elinda, E., Yuliansyah, H., & Latiffi, M. I. A. (2024). Sentiment Analysis of the Sheikh Zayed Grand Mosque's Visitor Reviews on Google Maps Using the VADER Method. *International Journal of Advances in Data and Information Systems*, 5(1), 71–84. <https://doi.org/10.59395/ijadis.v5i1.1320>
- Fauzi, R. H., Kusumah, K., & Rahmatulloh, A. (2026). Analisis Sentimen Media Sosial Terhadap DPR RI: Perbandingan Akurasi Extra Trees Dan Random Forest Dengan Pendekatan Komputasi Hijau Sentiment Analysis of Social Media Towards DPR RI: Accuracy Comparison of Extra

- Trees and Random Forest with a Green Com. 14(1), 36–45. <https://doi.org/10.26418/justin.v14i1.92556>
- Harun, R. M., & Fahmi, F. (2024). Analisis Sentimen Menggunakan Algoritma Naïve Bayes, KNN, dan Decision Tree Terhadap Ulasan Aplikasi KitaLulus. *Building of Informatics, Technology and Science (BITS)*, 6(3), 2033–2042. <https://doi.org/10.47065/bits.v6i3.6367>
- Harwenda, R. W., Angelo, M. D., Budi, I., Santoso, A. B., & Putra, P. K. (2025). Sentiment Analysis on Government Public Policies: A Systematic Literature Review. *Dinasti International Journal of Education Management And Social Science*, 6(5), 4192–4211. <https://doi.org/10.38035/dijemss.v6i5.4699>
- Hetherington, M. J. (2025). The Presidential Election: The Elections of 2024, 9(1), 68–93. <https://doi.org/10.2307/jj.28833768.6>
- Keladu, Y., Darling, D., Sugar, F., & Armansi, A. (2026). Program Makan Bergizi Gratis (MBG) di Era Kepemimpinan Prabowo-Gibran dalam Perspektif Amartya Sen. *Politeia: Jurnal Ilmu Politik*, 18(1), 1–8. <https://doi.org/10.32734/politeia.v18i1.20855>
- Leidiyana, H., Misriati, T., & Aryanti, R. (2024). Klasifikasi Sentimen Terhadap Kebijakan Tapera Menggunakan Komparasi Machine Learning dan SMOTE. *Jurnal Komtika (Komputasi Dan Informatika)*, 8(2), 125–135. <https://doi.org/10.31603/komtika.v8i2.12595>
- Lestari, V. B., & Hutagalung, C. A. (2025). Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes. *Journal of Computer Science and Applications*, 8(1), 22–2025.
- Lillah, M. R. R., Maylawati, D. S., Zulfikar, W. B., Uriawan, W., & Wahana, A. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) untuk analisis sentimen pengguna aplikasi Tokopedia. *Intellect : Indonesian Journal of Learning and Technological Innovation*, 02(02), 171–184.
- Muhidin, D., & Wibowo, A. (2020). Perbandingan Kinerja Algoritma Support Vector Machine dan K-Nearest Neighbor Terhadap Analisis Sentimen Kebijakan New Normal. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 5(2), 153. <https://doi.org/10.30998/string.v5i2.6715>
- Nur Cahyo, D. D., Farasalsabila, F., Lestari, V. B., Hanafi, Lestari, T., Al Islami, F. R., & Maulana, M. A. (2023). Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 8(2), 90–95. <https://doi.org/10.25139/inform.v8i2.5700>
- Panjaitan, F., Ce, W., Oktafiandy, H., Kanugrahan, G., Ramdhani, Y., Hafizh Cahaya Putra, V., & Permai, A. (2025). Studi Komparatif Algoritma Machine Learning Pada Analisis Sentimen Media Sosial. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 3145–3152. <https://doi.org/10.36040/jati.v9i2.13277>
- Prasetyoningrum, D., & Andono, P. N. (2024). Perbandingan Model Machine Learning dalam Analisis Sentimen Pada Kasus Monkeypox di Media Sosial X. *Building of Informatics, Technology and Science (BITS)*, 6(3), 2005–2014. <https://doi.org/10.47065/bits.v6i3.6447>
- Pratama, M. L., Via, Y. V., & Mandyartha, E. P. (2023). Analisis performansi Naive Bayes dan Random Forest terhadap sentimen kenaikan harga BBM di Indonesia. *SCAN: Jurnal Teknologi Informasi dan Komunikasi*, 18(1). <https://doi.org/10.33005/scan.v18i1.3837>
- Rahman, F. N., & Lestari, S. (2024). Analisis Sentimen Masyarakat Terhadap Pemerintah di Era Kabinet Joko Widodo Berdasarkan Sosial Media X Menggunakan Naïve bayes dan K-Nearest Neighbor (KNN). *INTECOMS: Journal of Information Technology and Computer Science*, 7(5), 1537–1544. <https://doi.org/10.31539/intecom.v7i5.11823>
- Rahmatullah, A., & Annisa, Q. (2026). Application Of TF-IDF And Word2vec For Feature Extraction In Sentiment Analysis Of Free Nutritious Food Policies. *Journal of Computer Electronic and Telecommunication*, 6(2). <https://doi.org/10.52435/complete.v6i2.741>
- Ripaldi, R., Tambunan, L. S., Edowardo, S., Rahkmah, S. N., & Sutoyo, I. (2025). Penerapan Algoritma Machine Learning Random Forest untuk Prediksi Risiko Konversi Sindrom Terisolasi Klinis Menjadi Multiple Sclerosis. 5(1), 23–32. <https://doi.org/10.47065/jogtc.v5i1.8847>
- Rizkiyanto, M. D., Purbolaksono, M. D., & Astuti, W. (2024). Sentiment Analysis Classification on PLN Mobile Application Reviews using Random Forest Method and TF-IDF Feature Extraction. *INTEK: Jurnal Penelitian*, 11(1), 37–43. <https://doi.org/10.31963/intek.v11i1.4774>

- Setyawan, H., Azizah, L. M., & Pradani, A. Y. (2023). Sentiment analysis of public responses on Indonesia government using Naïve Bayes and Support Vector Machine. *Emerging Information Science and Technology*, 4(1), 1–7. <https://doi.org/10.18196/eist.v4i1.18681>
- Yuliansyah, H., Ardiansyah, R., & Yudhana, A. (2026). View of A Machine Learning Approach for Multi-Label Classification in Candidate Election Social Media Analysis.pdf.
- Yuliansyah, H., Raihan, H. A., & Murinto. (2025). Sentiment Analysis Model for VTuber Live Stream Chat using Decision Tree and Support Vector Machine. *Journal of Innovation Information Technology and Application (JINITA)*, 7(2), 258–268. <https://doi.org/10.35970/jinita.v7i2.2872>
- Yuliansyah, H. S. &. (2024a). Sistemasi: Jurnal Sistem Informasi Analisis Sentimen Berbasis Aspek Terhadap Ulasan Pengguna pada Game Honkai: Star Rail Menggunakan Naïve Bayes Classifier Aspect-Based Sentiment Analysis of User Reviews on the Game Honkai: Star Rail Using Naïve Bayes Cla. 13(5), 2540–9719. <http://sistemasi.ftik.unisi.ac.id>
- Yuliansyah, H. S. &. (2024). Analisis Sentimen Berbasis Aspek Terhadap Ulasan Pengguna pada Game Honkai: Star Rail Menggunakan Naïve Bayes Classifier Aspect-Based Sentiment Analysis of User Reviews on the Game Honkai: Star Rail Using Naïve Bayes Cla. 13(5), 2540–9719.
- Zalukhu, P. S., Handhayani, T., & Sitorus, M. (2023). Analisis sentimen terhadap kenaikan BBM di Indonesia pada media sosial Twitter menggunakan metode Naïve Bayes. *Simtek: Jurnal Sistem Informasi dan Teknik Komputer*, 8(1), 65–69. <https://doi.org/10.51876/simtek.v8i1.177>