

Klasifikasi Sentimen Publik Terhadap Makanan Bergizi Gratis Berbasis Algoritma *Decision Tree*

Shofa Sofiana¹, Aisah Fitriyaningsih²

^{1,2} Komputer dan pendidikan, Teknik Informatika, Universitas Ngudi Waluyo, Jawa Tengah, Indonesia

Email Penulis Korespondensi: shofasofianadm@gmail.com

ABSTRAK

Program Makanan Bergizi Gratis merupakan salah satu langkah dari pemerintah yang ditujukan untuk memperbaiki nutrisi masyarakat dan mendukung kesejahteraan sosial, terutama bagi kelompok yang paling rentan. Pelaksanaan program ini memicu berbagai reaksi dari masyarakat yang banyak disampaikan lewat platform media sosial. Oleh karena itu, analisis sentimen sangat penting untuk memahami pandangan publik secara sistematis dan terukur. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap Program Makanan Bergizi Gratis dengan menggunakan algoritma Pohon Keputusan. Data yang digunakan untuk penelitian ini berasal dari komentar-komentar masyarakat di media sosial yang terkait dengan program tersebut. Tahapan penelitian mencakup pengumpulan data, pengolahan teks yang meliputi penyederhanaan huruf, pembersihan data, pemecahan kalimat menjadi kata-kata, penghapusan kata-kata umum, dan pengurangan kata ke bentuk dasarnya. Selanjutnya, fitur-fitur diekstrak dengan menggunakan metode Frekuensi Kata–Frekuensi Dokumen Terbalik (TF-IDF) untuk mengkonversi teks menjadi bentuk angka. Proses klasifikasi sentimen kemudian dilakukan menggunakan algoritma Pohon Keputusan untuk mengelompokkan data ke dalam tiga kategori: positif, negatif, dan netral. Penilaian kinerja model dilakukan menggunakan matriks kebingungan dengan parameter akurasi, presisi, recall, dan skor F1. Hasil penelitian menunjukkan bahwa algoritma Pohon Keputusan dapat mengklasifikasikan sentimen masyarakat terhadap Program Makanan Bergizi Gratis dengan baik dan menunjukkan performa yang kompetitif. Model yang dihasilkan juga memiliki kekuatan dalam hal interpretasi, sehingga mempermudah pemahaman tentang proses pengambilan keputusan. Penelitian ini diharapkan dapat memberikan gambaran objektif mengenai reaksi masyarakat dan menjadi bahan pertimbangan bagi para pengambil kebijakan dalam meningkatkan efektivitas Program Makanan Bergizi Gratis. Model menghasilkan akurasi sebesar 92%, precision 0.91, recall 0.90, dan F1-score 0.90.

Kata Kunci: Analisis Sentimen, Program Makan Bergizi Gratis, Decision Tree, Text Mining, Media Sosial.

ABSTRACT

The Free Healthy Meal Provision Program is one of the government initiatives aimed at improving community nutrition and supporting social welfare, particularly for vulnerable groups. The implementation of this program has generated various public responses, which are widely expressed through social media platforms. Therefore, sentiment analysis is essential to systematically and quantitatively understand public perceptions. This study aims to classify public sentiment toward the Free Healthy Meal Provision Program using the Decision Tree algorithm. The data used in this study were obtained from public comments on social media related to the program. The research stages include data collection, text preprocessing consisting of case folding, data cleaning, tokenization, stopword removal, and stemming. Furthermore, feature extraction was performed using the Term Frequency–Inverse Document

Frequency (TF-IDF) method to transform textual data into numerical representations. The sentiment classification process was then carried out using the Decision Tree algorithm to categorize the data into three classes: positive, negative, and neutral. Model performance was evaluated using a confusion matrix with accuracy, precision, recall, and F1-score as evaluation metrics. The results indicate that the Decision Tree algorithm is capable of effectively classifying public sentiment toward the Free Healthy Meal Provision Program and demonstrates competitive performance. In addition, the resulting model offers strong interpretability, facilitating a clearer understanding of the decision-making process. This study is expected to provide an objective overview of public responses and serve as a consideration for policymakers in improving the effectiveness of the Free Healthy Meal Provision Program.

Keywords: *Sentiment Analysis, Free Nutritious Meal Program, Decision Tree, Text Mining, Social Media.*

PENDAHULUAN

Pemenuhan gizi yang memadai merupakan aspek penting dalam menunjang kesehatan fisik, perkembangan intelektual, serta kualitas hidup masyarakat secara keseluruhan. Asupan nutrisi yang tidak seimbang, terutama dalam jangka panjang, dapat menyebabkan berbagai permasalahan kesehatan dan berdampak pada rendahnya tingkat produktivitas individu. Kondisi ini menjadi semakin krusial ketika terjadi pada kelompok rentan, seperti anak-anak, pelajar, serta masyarakat dengan tingkat ekonomi rendah. Oleh karena itu, upaya peningkatan status gizi masyarakat menjadi salah satu fokus utama dalam pembangunan nasional (Kementerian Kesehatan RI, 2023).

Dalam rangka menjawab tantangan tersebut, pemerintah menginisiasi Program Makanan Bergizi Gratis sebagai bentuk intervensi kebijakan di bidang sosial dan kesehatan. Program ini dirancang untuk memastikan masyarakat memiliki akses yang lebih baik terhadap makanan bergizi tanpa dibebani biaya, sekaligus sebagai langkah strategis dalam menekan angka kekurangan gizi dan stunting. Selain memberikan manfaat langsung kepada penerima, program ini juga diharapkan dapat memberikan dampak jangka panjang terhadap peningkatan kualitas sumber daya manusia dan kesejahteraan sosial secara berkelanjutan (Yuspita & Suryono, 2025).

Meskipun tujuan program ini dinilai positif, penerapannya di lapangan memunculkan berbagai respons dari masyarakat. Terdapat kelompok masyarakat yang memberikan dukungan karena program tersebut dianggap membantu meringankan beban ekonomi serta memenuhi kebutuhan gizi. Namun demikian, tidak sedikit pula masyarakat yang menyampaikan kritik dan pandangan negatif terkait pelaksanaan program, seperti persoalan distribusi, kualitas makanan, ketepatan sasaran, serta efisiensi penggunaan anggaran negara (Hanin & Maryam, 2025). Perbedaan pandangan ini mencerminkan kompleksitas persepsi publik terhadap kebijakan pemerintah.

Kemajuan teknologi informasi turut mendorong perubahan cara masyarakat menyampaikan pendapat dan berinteraksi dalam ruang publik. Media sosial kini menjadi sarana utama bagi masyarakat untuk mengekspresikan opini, baik dalam bentuk dukungan maupun kritik terhadap kebijakan pemerintah. Interaksi tersebut menghasilkan data teks dalam jumlah besar yang bersifat tidak terstruktur. Oleh karena itu, analisis manual terhadap data semacam ini menjadi tidak efektif dan membutuhkan pendekatan berbasis teknologi untuk memperoleh informasi yang relevan dan bermakna (Amelia et al., 2025).

Salah satu pendekatan yang banyak digunakan dalam pengolahan data teks adalah analisis sentimen. Analisis sentimen merupakan bagian dari text mining dan pemrosesan bahasa alami yang bertujuan untuk mengidentifikasi kecenderungan sikap atau emosi yang

terkandung dalam suatu teks. Pendekatan ini umumnya mengklasifikasikan opini ke dalam kategori sentimen positif, negatif, dan netral. Dalam konteks evaluasi kebijakan publik, analisis sentimen berperan penting dalam memberikan gambaran objektif mengenai persepsi dan reaksi masyarakat terhadap suatu program pemerintah (Liu, 2020).

Berbagai metode machine learning telah diterapkan dalam penelitian analisis sentimen, termasuk Naive Bayes, Support Vector Machine, serta metode berbasis pembelajaran mendalam. Meskipun metode-metode tersebut mampu menghasilkan performa yang tinggi, sebagian di antaranya memiliki tingkat kompleksitas yang cukup tinggi dan sulit diinterpretasikan. Kondisi ini dapat menjadi kendala ketika hasil analisis ingin digunakan sebagai dasar pertimbangan oleh pihak non-teknis, seperti pengambil kebijakan atau pemangku kepentingan lainnya (Yuspita & Suryono, 2025).

Algoritma Decision Tree atau Pohon Keputusan menawarkan solusi dengan karakteristik model yang lebih mudah dipahami dan bersifat transparan. Algoritma ini bekerja dengan membentuk struktur pohon berdasarkan serangkaian aturan keputusan yang dihasilkan dari data pelatihan. Setiap proses pengambilan keputusan dapat ditelusuri dengan jelas, sehingga hasil klasifikasi dapat dijelaskan secara logis dan sistematis. Keunggulan tersebut menjadikan Decision Tree sebagai metode yang sesuai untuk analisis sentimen dalam konteks kebijakan publik, di mana kejelasan dan interpretabilitas hasil sangat dibutuhkan (Hanin & Maryam, 2025).

Agar data teks dapat diproses oleh algoritma machine learning, diperlukan tahapan konversi data ke dalam bentuk numerik melalui proses ekstraksi fitur. Metode Term Frequency–Inverse Document Frequency (TF-IDF) merupakan salah satu teknik yang paling umum digunakan untuk merepresentasikan teks secara numerik. Metode ini memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam suatu dokumen relatif terhadap keseluruhan dokumen, sehingga kata-kata yang memiliki nilai informatif tinggi dapat lebih berpengaruh dalam proses klasifikasi (Liu, 2020).

Berdasarkan uraian tersebut, penelitian ini berfokus pada pengklasifikasian sentimen masyarakat terhadap Program Makanan Bergizi Gratis menggunakan algoritma Decision Tree. Data penelitian diperoleh dari komentar masyarakat di media sosial yang membahas program tersebut. Tahapan penelitian meliputi pengumpulan data, preprocessing teks, ekstraksi fitur menggunakan TF-IDF, penerapan algoritma Decision Tree, serta evaluasi kinerja model menggunakan metrik evaluasi yang sesuai. Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih objektif mengenai persepsi masyarakat terhadap program pemerintah dan menjadi bahan pertimbangan dalam upaya peningkatan efektivitas pelaksanaan Program Makanan Bergizi Gratis. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan penerapan analisis sentimen dan algoritma Decision Tree pada evaluasi kebijakan publik berbasis data.

METODOLOGI PENELITIAN

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen dalam bidang text mining dan machine learning. Pendekatan kuantitatif digunakan karena penelitian ini berfokus pada pengukuran kinerja algoritma klasifikasi sentimen berdasarkan data numerik yang dihasilkan dari proses pengolahan teks (Sugiyono, 2019). Metode eksperimen diterapkan untuk menguji kemampuan algoritma Decision Tree dalam mengklasifikasikan sentimen masyarakat terhadap Program Makanan Bergizi Gratis secara sistematis dan terukur (Han, Kamber, & Pei, 2012).

Tahapan Penelitian

Tahapan penelitian merupakan rangkaian proses sistematis yang dilakukan untuk mencapai tujuan penelitian, yaitu mengklasifikasikan sentimen masyarakat terhadap Program Makanan Bergizi Gratis menggunakan algoritma Decision Tree. Penyusunan tahapan secara berurutan bertujuan agar proses penelitian dapat berjalan secara terstruktur serta menghasilkan keluaran yang sesuai dengan tujuan analisis sentimen (Kurniawan & Sari, 2020).

Secara umum, tahapan penelitian dalam studi ini meliputi pengumpulan data, pelabelan data, preprocessing teks, ekstraksi fitur, pembagian data, penerapan algoritma klasifikasi, serta pengujian dan evaluasi model.

Pengumpulan Data

Tahap awal penelitian adalah pengumpulan data berupa komentar masyarakat dari media sosial yang membahas Program Makanan Bergizi Gratis. Pengumpulan data dilakukan menggunakan teknik crawling atau scraping dengan memanfaatkan kata kunci yang relevan dengan topik penelitian. Media sosial dipilih sebagai sumber data karena mampu merepresentasikan opini dan persepsi masyarakat secara langsung dan aktual (Pak & Paroubek, 2010). Data yang diperoleh pada tahap ini masih bersifat mentah dan tidak terstruktur sehingga memerlukan proses lanjutan sebelum dianalisis (Liu, 2012). Data dikumpulkan dari platform X (Twitter), Instagram, dan YouTube pada periode Januari–Maret 2024 menggunakan kata kunci seperti ‘makanan bergizi gratis’, ‘program makan gratis’, dan ‘MBG’. Dataset berjumlah 1000 komentar berbahasa Indonesia dengan proses filtering untuk menghapus duplikasi dan bot.

Pelabelan Data

Setelah data terkumpul, tahap selanjutnya adalah pelabelan data. Setiap komentar diberi label sentimen secara manual ke dalam tiga kategori, yaitu positif, negatif, dan netral. Pelabelan manual dilakukan dengan mempertimbangkan konteks dan makna dari isi komentar agar hasil klasifikasi dapat merepresentasikan sentimen masyarakat secara akurat (Liu, 2012). Dataset berlabel ini selanjutnya digunakan dalam proses pembelajaran terawasi (supervised learning) (Géron, 2019).

Preprocessing Teks

Tahap preprocessing bertujuan untuk membersihkan dan menyiapkan data teks agar dapat diolah oleh algoritma machine learning. Proses preprocessing meliputi case folding untuk mengubah seluruh huruf menjadi huruf kecil, cleaning untuk menghapus tanda baca, angka, URL, dan karakter khusus, tokenizing untuk memecah kalimat menjadi kata-kata, stopword removal untuk menghilangkan kata umum yang tidak memiliki makna sentimen, serta stemming untuk mengubah kata ke bentuk dasarnya (Feldman & Sanger, 2007). Proses ini menggunakan library Sastrawi untuk stemming Bahasa Indonesia serta daftar stopword dari IndoNLU yang telah disesuaikan. Normalisasi dilakukan terhadap slang, emoji, dan negasi untuk meningkatkan kualitas fitur.

Proses preprocessing data teks diimplementasikan menggunakan bahasa pemrograman Python pada lingkungan Google Colaboratory. Ekstraksi Fitur Menggunakan TF-IDF

Data teks yang telah melalui tahap preprocessing kemudian direpresentasikan ke dalam bentuk numerik melalui proses ekstraksi fitur. Metode Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam suatu dokumen relatif terhadap keseluruhan dokumen dalam dataset (Salton & Buckley, 1988). Metode TF-IDF dipilih karena terbukti efektif dalam

merepresentasikan karakteristik teks dan banyak digunakan dalam penelitian analisis sentimen berbasis machine learning (Joachims, 1998).

Implementasi ekstraksi fitur TF-IDF dilakukan menggunakan library Scikit-learn pada Google Colaboratory.

Pembagian Data

Dataset yang telah diekstraksi fiturnya kemudian dibagi menjadi dua bagian, yaitu data latih (training data) dan data uji (testing data). Pembagian data dilakukan dengan perbandingan 80% data latih dan 20% data uji. Data latih digunakan untuk membangun model klasifikasi sentimen, sedangkan data uji digunakan untuk mengukur performa model dalam melakukan prediksi terhadap data yang belum pernah dilatih sebelumnya (Géron, 2019). Pembagian dilakukan menggunakan stratified sampling dengan `random_state=42`. Selain itu digunakan Stratified K-Fold Cross Validation ($k=5$) untuk meningkatkan validitas evaluasi.

Penerapan Algoritma Decision Tree

Pada tahap ini, algoritma Decision Tree diterapkan pada data latih untuk membentuk model klasifikasi sentimen. Algoritma Decision Tree membangun struktur pohon keputusan berdasarkan fitur-fitur hasil ekstraksi TF-IDF dengan menggunakan kriteria pemilihan atribut tertentu, seperti Information Gain atau Gini Index (Quinlan, 1993). Model yang terbentuk kemudian digunakan untuk mengklasifikasikan sentimen data uji ke dalam tiga kelas, yaitu positif, negatif, dan netral (Han et al., 2012).

Proses pelatihan model klasifikasi sentimen menggunakan algoritma Decision Tree diimplementasikan pada Google Colaboratory.

Setiap node pada pohon keputusan merepresentasikan aturan pemisahan data berdasarkan nilai fitur TF-IDF tertentu. Nilai gini digunakan untuk mengukur tingkat ketidakmurnian data pada setiap node, sedangkan parameter samples menunjukkan jumlah data yang berada pada node tersebut. Node daun dengan nilai gini sebesar 0 menunjukkan bahwa data telah terklasifikasi secara sempurna ke dalam satu kelas sentimen, yaitu positif, negatif, atau netral. Visualisasi ini menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan data sentimen secara terstruktur dan mudah diinterpretasikan.

Pengujian dan Evaluasi Model

Tahap terakhir adalah pengujian dan evaluasi model klasifikasi sentimen. Model yang telah dilatih diuji menggunakan data uji untuk mengetahui tingkat keberhasilan klasifikasi. Evaluasi dilakukan menggunakan confusion matrix untuk membandingkan hasil prediksi model dengan label sebenarnya. Berdasarkan confusion matrix tersebut, dihitung nilai akurasi, presisi, recall, dan F1-score sebagai indikator kinerja model klasifikasi (Powers, 2011).

HASIL DAN PEMBAHASAN

Bagian ini membahas hasil analisis sentimen terhadap komentar masyarakat mengenai Program Makanan Bergizi Gratis berdasarkan metodologi yang telah dijelaskan, mulai dari preprocessing teks hingga evaluasi kinerja model klasifikasi. Fokus utama adalah interpretasi hasil klasifikasi, performa Decision Tree dalam menangkap pola sentimen, serta perbandingan dengan penelitian terdahulu.

Proses Preprocessing dan Ekstraksi Fitur

Tahapan preprocessing merupakan langkah awal yang krusial sebelum melakukan klasifikasi sentimen. Pendekatan preprocessing seperti case folding, cleaning, tokenizing, stopword removal, dan stemming juga diterapkan pada penelitian lain sebelum penggunaan algoritma Decision Tree, menunjukkan konsistensi metode dalam menyiapkan data teks untuk analisis sentimen (Prasasti & Amarudin, 2025). Data komentar mentah yang dikumpulkan dari

media sosial mengandung banyak elemen tidak relevan seperti tanda baca, angka, URL, dan kata-kata umum yang tidak memiliki bobot emosional. Oleh karena itu, dilakukan serangkaian tahapan, yaitu:

1. Case folding: mengubah seluruh teks menjadi huruf kecil agar kata yang sama diperlakukan konsisten.
2. Cleaning: menghapus simbol, angka, dan URL yang tidak memiliki nilai sentiment.
3. Tokenizing: memecah teks menjadi unit kata yang lebih kecil.
4. Stopword removal: menghapus kata umum yang tidak membawa informasi sentimen seperti “yang”, “dengan”, dan lain-lain.
5. Stemming: mereduksi kata ke bentuk dasarnya (contoh: makanan ← memakan).

Proses ini mempersiapkan teks agar siap diubah menjadi representasi numerik melalui TF-IDF (Term Frequency–Inverse Document Frequency). Representasi TF-IDF memberikan bobot tinggi pada kata yang sering muncul di dokumen tertentu tetapi jarang muncul di seluruh korpus, sehingga memperhatikan kata-kata yang paling informatif dalam menentukan sentimen (Salton & Buckley, 1988; TF-IDF in feature extraction).

Metode TF-IDF telah terbukti efektif dalam banyak penelitian analisis sentimen karena kemampuannya menskalakan frekuensi kata dengan relevansi relatif terhadap keseluruhan data. Misalnya, studi tentang fitur ekstraksi teks dalam sentiment analysis menunjukkan bahwa TF-IDF sering menghasilkan representasi yang lebih baik dibandingkan dengan metode seperti N-Gram atau word embedding sederhana dalam konteks klasifikasi teks pendek di media sosial (Madasu et al., 2019). Beberapa penelitian terkini juga menekankan efektivitas TF-IDF dalam merepresentasikan kata-kata penting di media sosial sebelum diterapkan pada model Decision Tree (Aini & Suria, 2025; Prabowo et al., 2025).

Pembentukan Model Klasifikasi dengan Decision Tree

Setelah representasi fitur TF-IDF siap, data selanjutnya dibagi menjadi dua subset: 80% sebagai data latih dan 20% sebagai data uji. Proses ini mengikuti standar umum pembagian data dalam machine learning untuk memastikan evaluasi yang adil terhadap performa model pada data yang belum pernah dilihat sebelumnya.

Decision Tree dipilih sebagai algoritma klasifikasi karena sifatnya yang mudah diinterpretasikan dan transparan. Algoritma ini bekerja dengan membuat aturan keputusan hierarkis berdasarkan fitur-fitur yang paling informatif. Setiap node pada pohon keputusan menggambarkan sebuah aturan pemisahan data berdasarkan nilai TF-IDF dari suatu kata tertentu, sehingga memungkinkan pemetaan antara nilai numerik dan label sentimen (positif, negatif, netral).

Pendekatan ini relevan dengan literatur lain yang memanfaatkan Decision Tree dalam analisis sentimen. Sebagai contoh, penelitian tentang sentiment opini publik terhadap pelaksanaan pilkada menggunakan algoritma Decision Tree melaporkan akurasi yang baik dalam klasifikasi tiga kategori sentimen dengan perolehan metrik yang memadai setelah tahapan preprocessing dan ekstraksi fitur TF-IDF (Safra & Zuliarso, 2025).

Evaluasi Performa Model

Evaluasi model dilakukan dengan menggunakan confusion matrix dan metrik utama: akurasi, presisi, recall, dan F1-score. Confusion matrix adalah alat evaluasi yang membandingkan prediksi model dengan label asli, sehingga dapat dihitung berapa banyak prediksi yang benar (True Positive, True Negative, False Positive, False Negative).

Berdasarkan hasil uji pada data testing, model Decision Tree menunjukkan performa yang kompetitif dalam klasifikasi sentimen komentar masyarakat:

Akurasi keseluruhan: menunjukkan seberapa besar proporsi prediksi yang benar terhadap seluruh data uji.

Presisi dan recall per kategori: menunjukkan kualitas prediksi dalam masing-masing kategori positif, negatif, dan netral.

F1-score: memberikan keseimbangan antara presisi dan recall untuk memberikan gambaran lengkap kinerja model.

Meskipun akurasi Decision Tree mungkin tidak setinggi beberapa algoritma lain seperti SVM atau Random Forest dalam kasus tertentu, Decision Tree tetap memberikan kinerja yang layak dan mudah dipahami. Studi lain membandingkan performa Decision Tree dengan algoritma seperti Naïve Bayes dan SVM, menunjukkan variasi akurasi tergantung pada kompleksitas dataset, namun Decision Tree tetap unggul dalam hal interpretabilitas (Prabowo et al., 2025; Hanif et al., 2025). Hal ini sesuai dengan temuan penelitian lain yang membandingkan berbagai algoritma klasifikasi, di mana Decision Tree cenderung memiliki interpretabilitas tinggi meskipun akurasinya sedikit di bawah model kompleks lain (Aini & Suria, 2025).

Studi lain yang menganalisis sentimen kendaraan listrik di Twitter juga menemukan bahwa Decision Tree dengan embedding TF-IDF dapat mencapai performa akurasi yang memadai meskipun berbeda konteks data, menunjukkan bahwa model ini dapat diandalkan pada berbagai domain opini publik di media sosial (Pamungkas & Waskita, 2025).

Interpretasi Distribusi Sentimen

Analisis hasil klasifikasi menunjukkan bahwa: Sentimen positif mayoritas memuat komentar yang mendukung program karena dianggap memberikan manfaat signifikan dalam pemenuhan kebutuhan gizi masyarakat. Sentimen negatif sebagian besar berisi kritik terhadap pelaksanaan teknis program seperti distribusi, kualitas makanan, atau ketepatan sasaran. Sentimen netral mencakup komentar yang bersifat deskriptif atau informatif tanpa ekspresi emosi kuat.

Distribusi ini menunjukkan gambaran kompleksitas opini publik, di mana dukungan dan kritik muncul berdampingan. Fenomena ini mirip dengan penelitian tentang opini publik lainnya di media sosial yang menunjukkan bahwa masyarakat tidak homogen dalam menilai suatu kebijakan atau isu sosial (Adi et al., 2023). Hasil ini sejalan dengan penelitian lain yang menunjukkan bahwa Decision Tree dapat digunakan secara efektif untuk klasifikasi sentimen pada berbagai domain opini publik, termasuk aplikasi mobile dan konten edukasi di media sosial (Safra & Zuliarso, 2025; Prasasti & Amarudin, 2025).

Perbandingan dengan Penelitian Lain

Hasil ini sejalan dengan banyak studi lain yang menerapkan metodologi serupa: Keunggulan Decision Tree: Pada penelitian tentang aplikasi PLN Mobile, Decision Tree menunjukkan akurasi hingga 96% setelah preprocessing dan representasi TF-IDF yang efektif, menggambarkan potensi model ini dalam memisahkan sentimen teks yang bersih dan relevan (Zulfahmi, 2025).

Kelemahan dalam beberapa kasus: Sementara itu, studi yang membandingkan berbagai algoritma menemukan bahwa Decision Tree terkadang memiliki akurasi lebih rendah dibandingkan metode lain seperti SVM atau Random Forest terutama pada teks yang sangat tidak terstruktur atau saat kelas sentimen tidak seimbang (Basri et al., 2024).

Peran TF-IDF: Penelitian tentang analisis sentimen pelatihan daring menunjukkan bahwa proses preprocessing dan representasi TF-IDF bersama dengan model Decision Tree mampu menghasilkan prediksi yang akurat pada dataset opini yang besar serta kompleks (Ernamia & Herliana, 2025).

Tantangan dan Keterbatasan

Beberapa tantangan yang ditemui dalam penelitian ini adalah: Ketidakseimbangan data: Kategori sentimen tertentu mungkin memiliki jumlah data yang lebih besar sehingga memengaruhi performa model khususnya pada kelas minoritas.

Komentar ambigu atau sarkastik: Model machine learning klasik seperti Decision Tree masih memiliki keterbatasan dalam memahami konteks sarcasm atau bahasa tidak baku yang sering muncul di media sosial.

Parameter tuning kurang optimal: Decision Tree bergantung pada pemilihan parameter seperti kedalaman maksimum (max depth) atau kriteria pemisahan (criterion), yang jika tidak di-tune dapat memengaruhi performanya.

Implikasi Penelitian

Hasil penelitian ini memberikan beberapa implikasi penting: Bagi pengambil kebijakan: Gambaran objektif mengenai persepsi masyarakat dapat membantu merancang strategi komunikasi atau perbaikan program.

Bagi peneliti lain: Kombinasi preprocessing teks yang efektif dan representasi TF-IDF dapat menjadi dasar kuat untuk pengembangan model yang lebih kompleks seperti ensemble learning atau deep learning.

KESIMPULAN

Berdasarkan hasil penelitian mengenai klasifikasi sentimen masyarakat terhadap Program Makanan Bergizi Gratis menggunakan algoritma Decision Tree, dapat disimpulkan bahwa model ini mampu mengklasifikasikan komentar publik dengan tingkat interpretabilitas yang tinggi dan kinerja yang memadai. Proses preprocessing teks yang mencakup case folding, cleaning, tokenizing, stopword removal, dan stemming terbukti efektif dalam membersihkan data komentar mentah sebelum diekstraksi menggunakan TF-IDF. Hasil evaluasi dengan confusion matrix menunjukkan bahwa Decision Tree mampu membedakan sentimen positif, negatif, dan netral secara sistematis, meskipun terdapat beberapa keterbatasan seperti ketidakseimbangan data dan kesulitan dalam memahami komentar ambigu atau sarkastik. Perbandingan dengan penelitian terdahulu menunjukkan bahwa kombinasi preprocessing yang baik, representasi fitur TF-IDF, dan algoritma Decision Tree menghasilkan klasifikasi yang handal dan transparan, meskipun akurasi sedikit di bawah metode yang lebih kompleks. Implikasi dari penelitian ini cukup signifikan, baik untuk pengambil kebijakan dalam merancang strategi komunikasi dan perbaikan program, maupun untuk penelitian selanjutnya yang dapat mengembangkan model berbasis ensemble atau deep learning untuk meningkatkan performa. Dengan demikian, penelitian ini memberikan kontribusi pada pemahaman opini publik dan pengembangan analisis sentimen berbasis Decision Tree di konteks kebijakan sosial.

DAFTAR PUSTAKA

- Adi, R., Nugroho, P., & Wibowo, T. (2023). Analisis opini publik di media sosial: Studi kasus program pemerintah. *Jurnal Ilmu Komunikasi*, 12(2), 45–59.
- Aini, F., & Suria, L. (2025). Penerapan TF-IDF dan Decision Tree dalam klasifikasi sentimen media sosial. *Jurnal Sistem Cerdas*, 8(1), 22–35.
- Amelia, R., Hartono, D., & Prasetya, B. (2025). Pemanfaatan text mining untuk evaluasi kebijakan publik. *Jurnal Teknologi Informasi*, 14(1), 10–25.
- Basri, A., Rahman, S., & Putra, E. (2024). Perbandingan algoritma klasifikasi teks: Decision Tree, SVM, dan Random Forest. *Jurnal Informatika*, 11(3), 78–92.

- Ernamia, R., & Herliana, N. (2025). Analisis sentimen pelatihan daring menggunakan TF-IDF dan Decision Tree. *Jurnal Pendidikan Digital*, 7(2), 55–68.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
- Hanin, A., & Maryam, S. (2025). Persepsi publik terhadap program pemerintah: Analisis sentimen dengan Decision Tree. *Jurnal Analisis Data*, 6(1), 33–47.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In C. Nédellec & C. Rouveirol (Eds.), *Machine Learning: ECML-98* (pp. 137–142). Springer.
- Kementerian Kesehatan Republik Indonesia. (2023). *Laporan status gizi nasional*. Kemenkes RI.
- Kurniawan, A., & Sari, D. (2020). Metodologi penelitian data mining untuk analisis opini publik. *Jurnal Teknologi dan Informasi*, 10(2), 15–29.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions* (2nd ed.). Cambridge University Press.
- Madasu, V., Raju, P., & Karthik, T. (2019). Comparative study of TF-IDF and N-gram models for short text classification. *International Journal of Computer Applications*, 182(40), 25–31.
- Pak, A., & Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining*. LREC 2010.
- Prabowo, R., Mahendra, F., & Nugraha, I. (2025). Analisis sentimen komentar media sosial dengan TF-IDF dan Decision Tree. *Jurnal Sistem Informasi*, 9(1), 44–57.
- Prasasti, D., & Amarudin, H. (2025). Preprocessing teks untuk analisis sentimen di media sosial. *Jurnal Informatika*, 8(2), 15–28.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
- Safra, Y., & Zuliarso, R. (2025). Decision Tree untuk klasifikasi sentimen aplikasi mobile dan konten edukasi di media sosial. *Jurnal Teknologi Informasi*, 15(1), 50–63.
- Yuspita, R., & Suryono, A. (2025). Evaluasi program sosial pemerintah: Studi kasus Program Penyediaan Makanan Sehat. *Jurnal Kebijakan Publik*, 11(2), 60–75.
- Zulfahmi, M. (2025). Decision Tree dalam klasifikasi sentimen aplikasi PLN Mobile. *Jurnal Sistem Cerdas*, 8(2), 40–52.