



INNOVATIVE: Journal Of Social Science Research

Volume 6 Nomor 2 Tahun 2026 Page 1-9

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Analisis Sentimen Opini Publik terhadap Dedi Mulyadi di Twitter Menggunakan Ekstraksi Fitur TF-IDF dan Klasifikasi Naive Bayes

Hafizh Pebrian^{1✉}, Aditya Ali Kusuma², Muhammad Rizky Pribadi³

Multi Data Palembang University, Palembang.

Email: hafizhpebrian_2226250077@mhs.mdp.ac.id^{1✉}

Abstrak

Pertumbuhan pesat media sosial, khususnya Twitter, telah membuka ruang yang luas bagi masyarakat untuk mengekspresikan pandangan mereka secara terbuka terhadap tokoh publik dan isu-isu politik. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini masyarakat mengenai Dedi Mulyadi, seorang tokoh politik di Indonesia, dengan memanfaatkan data yang diperoleh dari Twitter. Metodologi yang digunakan mencakup tahapan pengumpulan data tweet, praproses teks, ekstraksi fitur menggunakan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF), serta proses klasifikasi sentimen melalui algoritma Naive Bayes. Sentimen yang diklasifikasikan terdiri dari tiga kategori, yaitu positif, negatif, dan netral. Hasil evaluasi menunjukkan bahwa kombinasi antara metode TF-IDF dan Naive Bayes mampu mengidentifikasi sentimen publik secara cukup efektif, dengan akurasi mencapai 68,0%. Temuan ini diharapkan dapat menjadi kontribusi dalam bidang analisis media sosial dan pemetaan opini masyarakat terhadap figur politik.

Kata Kunci: *Analisis Sentimen, Twitter, TF-IDF, Naive Bayes, Dedi Mulyadi*

Abstract

The rapid growth of social media, particularly Twitter, has provided a broad space for the public to express their views openly on public figures and political issues. This study aims to analyze public sentiment toward Dedi Mulyadi, a political figure in Indonesia, using data collected from Twitter. The methodology involves tweet data collection, text preprocessing, feature extraction using the Term Frequency-Inverse Document Frequency (TF-IDF) approach, and sentiment classification using the Naive Bayes algorithm. Sentiments are categorized into three classes: positive, negative, and neutral. The evaluation results indicate that the combination of TF-IDF and Naive Bayes effectively identifies public sentiment, achieving an accuracy of 68.0%. These findings are expected to contribute to the field of social media analysis and the mapping of public opinion toward political figures

Keywords: TF-IDF, Keyword Search, Indonesian Text, Document Classification, Information Retrieval.

PENDAHULUAN

Perkembangan teknologi informasi telah mengubah cara masyarakat berkomunikasi, termasuk dalam menyampaikan opini terhadap isu-isu publik. Media sosial, khususnya Twitter, kini menjadi salah satu sumber utama untuk memantau opini masyarakat secara real-time. Pengguna Twitter dapat dengan bebas mengekspresikan pandangannya terhadap tokoh atau peristiwa tertentu, menjadikan platform ini kaya akan data yang dapat dianalisis untuk memahami persepsi publik (Cerezo, 2004). Dedi Mulyadi adalah salah satu tokoh publik di Indonesia yang aktif di media sosial dan kerap menjadi perbincangan di Twitter. Sebagai politisi yang dikenal dekat dengan masyarakat dan memiliki pendekatan komunikasi yang khas, persepsi publik terhadapnya menjadi menarik untuk dianalisis secara ilmiah. Salah satu metode yang dapat digunakan adalah analisis sentimen, yaitu proses mengidentifikasi opini atau emosi dalam teks menggunakan pendekatan pengolahan bahasa alami (Natural Language Processing/NLP) (Liu, 2020).

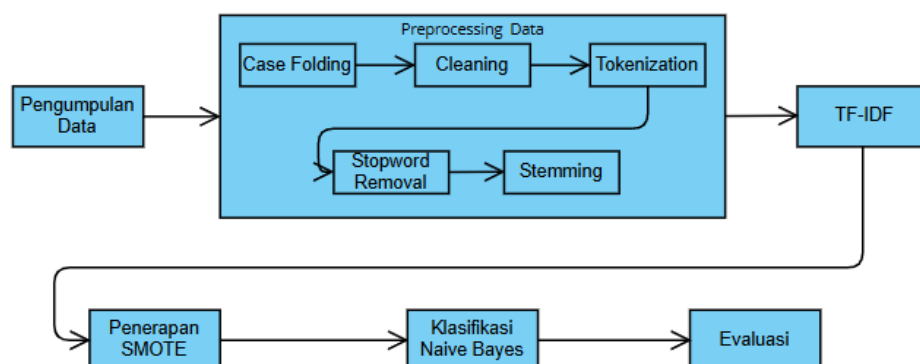
Berbagai penelitian sebelumnya telah mengembangkan metode untuk analisis sentimen di media sosial. Salah satu pendekatan yang umum digunakan dalam representasi teks adalah Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini mampu menyoroti kata-kata yang penting dalam suatu dokumen berdasarkan frekuensi relatifnya (Robertson, 2004; Salton & Buckley, 1987). Untuk tahap klasifikasi, Naive Bayes merupakan algoritma populer yang telah terbukti efektif dalam tugas klasifikasi teks karena kesederhanaan dan performanya (McCallum & Nigam, 1998; Sahami et al., 1998). Penelitian oleh Pang dan Lee (Pang & Lee, 1999) menunjukkan bahwa pendekatan berbasis probabilitas seperti Naive Bayes cocok untuk klasifikasi teks pendek seperti tweet. Penelitian

lokal oleh Nugroho et al. (Nursidik Dinar et al., 2023) juga mendemonstrasikan keberhasilan algoritma ini dalam klasifikasi sentimen politik di Twitter. Dalam konteks Indonesia, beberapa studi telah menggabungkan TF-IDF dengan algoritma seperti Support Vector Machine (SVM), Decision Tree, maupun Naive Bayes untuk analisis sentimen terhadap produk, kebijakan, dan tokoh publik (Bahrawi, 2019; Buntoro, 2017; Firdaus et al., 2024; Susanto et al., 2021).

Namun, penelitian yang secara spesifik mengkaji opini publik terhadap Dedi Mulyadi di Twitter masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan mengembangkan model klasifikasi sentimen berbasis TF-IDF dan Naive Bayes guna mengetahui kecenderungan opini masyarakat terhadap tokoh tersebut. Adapun tujuan dari penelitian ini adalah Mengumpulkan data tweet yang menyebutkan Dedi Mulyadi, Melakukan praproses dan ekstraksi fitur menggunakan TF-IDF, Mengklasifikasikan sentimen ke dalam tiga kategori: positif, negatif, dan netral menggunakan Naive Bayes, Mengevaluasi performa model berdasarkan metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score. Dengan penelitian ini, diharapkan dapat diperoleh gambaran persepsi publik yang lebih terstruktur, yang dapat dimanfaatkan dalam pengambilan keputusan strategis, khususnya dalam bidang komunikasi politik.

METODE PENELITIAN

Dalam penelitian ini terdapat alur proses penelitian untuk dapat melakukan analisis sentimen, yaitu mengumpulkan data, preprocessing data, TF-IDF, penerapan SMOTE, klasifikasi naive bayes, evaluasi, dan kesimpulan. Tahapan tersebut ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

Proses pengumpulan data merupakan tahapan awal yang esensial dalam pengembangan sistem klasifikasi. Data yang digunakan dapat diperoleh dari berbagai sumber, termasuk koleksi pribadi maupun hasil akuisisi dari internet (Saputro et al., 2017).

Data yang digunakan dalam penelitian ini sebanyak 1000 data, data diambil secara acak di aplikasi media sosial twitter kemudian diklasifikasikan secara manual ke dalam dua kriteria, yaitu pendapat masyarakat yang pro terhadap kebijakan program yang dedi mulyadi jalankan dan pendapat masyarakat yang kontra terhadap kebijakan program yang dedi mulyadi jalankan, datanya dapat dilihat pada gambar 2.

	text	label	C
1	Beliau cukup dikenal di Jawa Barat.	netral	
2	Terkadang setuju, terkadang juga tidak.	netral	
3	Saya belum terlalu mengikuti kegiatan Dedi Mulyadi.	netral	
4	Kritik terhadap kebijakan publik kurang kuat.	negatif	
5	Kritik terhadap kebijakan publik kurang kuat.	negatif	
6	Saya sangat mengagumi kepedulian Dedi Mulyadi terhadap rakyat kecil.	positif	
7	Biasa saja, tidak terlalu menarik atau mengecewakan.	netral	
8	Beliau konsisten membela kepentingan masyarakat kecil.	positif	
9	Beliau konsisten membela kepentingan masyarakat kecil.	positif	
10	Saya pikir beliau cukup aktif di media sosial.	netral	
11	Saya belum terlalu mengikuti kegiatan Dedi Mulyadi.	netral	
12	Kepemimpinan yang penuh empati dari Dedi Mulyadi sangat dibutuhkan.	positif	
13	Cara beliau merespons masalah cepat dan bijak.	positif	
14	Cara beliau merespons masalah cepat dan bijak.	positif	
15	Kadang lihat videonya di media sosial, cukup menarik.	netral	
16	Pendekatan Dedi Mulyadi kepada masyarakat sangat humanis.	positif	
17	Saya tidak melihat perubahan signifikan dari beliau.	negatif	
18	Tidak punya pendapat khusus tentang beliau.	netral	
19	Program yang dibawakan Dedi Mulyadi sangat inspiratif.	positif	
20	Beliau konsisten membela kepentingan masyarakat kecil.	positif	
21	Kebijakan beliau tidak terlalu berdampak nyata.	negatif	
22	Saya pikir beliau cukup aktif di media sosial.	netral	
23	Kritik terhadap kebijakan publik kurang kuat.	negatif	
24	Biasa saja, tidak terlalu menarik atau mengecewakan.	netral	
25	Saya pikir beliau cukup aktif di media sosial.	netral	

Gambar 2. Data ulasan pendapat masyarakat terhadap kebijakan program dedi mulyadi

Preprocessing merupakan salah satu tahapan penting yang harus dilalui guna memastikan data siap digunakan dalam pengembangan model (Hermawan et al., 2023). Dalam analisis sentimen, text preprocessing menjadi langkah awal yang sangat penting guna membersihkan dan menyiapkan data teks agar dapat memberikan hasil analisis yang akurat dan relevan (Wati et al., 2023). Berikut tahap-tahap preprocessing yang dilakukan :

1. *Case folding* adalah salah satu tahap dalam preprocessing data teks yang berfungsi untuk mengonversi seluruh karakter huruf kapital dalam dokumen menjadi huruf kecil, sehingga menghasilkan format teks yang lebih seragam (Wati et al., 2023).
2. *Cleaning* merupakan tahap awal dalam pemrosesan teks yang bertujuan untuk menghilangkan karakter-karakter yang tidak relevan dan tidak berkontribusi terhadap analisis sentiment (Wang et al., 2019).
3. *Tokenisasi* adalah metode dalam pemrosesan bahasa alami yang memisahkan teks menjadi token-token individual, yang digunakan untuk membentuk representasi struktural dari kalimat dalam dokumen seperti tweet (Mustaqim et al., 2024).
4. *Stopword removal* merupakan proses eliminasi kata-kata umum yang sering muncul dalam teks namun tidak memiliki kontribusi bermakna terhadap konteks analisis. Seperti 'dan', 'di', atau 'yang' dalam bahasa Indonesia, umumnya diabaikan karena tidak

memberikan informasi penting dalam pemrosesan dan analisis teks (Mustaqim et al., 2024).

5. *Stemming* adalah teknik pemrosesan teks yang mengubah kata berimbuhan menjadi bentuk dasarnya dengan menghapus awalan dan akhiran, sehingga menghasilkan kata inti yang lebih representatif untuk analisis (Mustaqim et al., 2024).

Algoritma Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode yang digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen dengan mempertimbangkan frekuensi kemunculannya serta seberapa jarang kata tersebut muncul dalam keseluruhan koleksi dokumen. Semakin tinggi nilai TF dan semakin rendah kemunculan kata dalam dokumen lain, maka semakin besar bobot yang diberikan terhadap kata tersebut (Silalahi & Guidio Leonarde Ginting, 2023).

Synthetic Minority Over-sampling Technique (SMOTE)

Teknik SMOTE digunakan untuk memperbaiki ketimpangan distribusi data antar kelas dengan menambahkan data sintetis pada kelas minoritas, sehingga total sampel menjadi sebanding dengan kelas mayoritas dan model dapat dilatih secara lebih seimbang (Tri Putra et al., 2023).

Klasifikasi Naïve Bayes

Klasifikasi *Naïve Bayes* merupakan algoritma untuk klasifikasi berdasarkan teorema bayes dan digunakan untuk menghitung sebuah probabilitas atau peluang dari suatu kelas dengan anggapan bahwa salah satu kelas dengan kelas lainnya tidak saling tergantung atau berdiri sendiri (Kasanah et al., 2019). Rumus Naïve Bayes dapat dilihat pada persamaan (1).

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data merupakan suatu class spesifik

$P(H|X)^{\wedge}$ = Probabilitas hipotesis berdasarkan kondisi

$P(H)$ = Probabilitas hipotesis

$P(X|H)$ = Probabilitas berdasarkan kondisi pada hipotesis

$P(X)$ = Probabilitas H

Evaluasi

Setelah tahap klasifikasi selesai, selanjutnya akan dilakukan pengujian untuk mengukur performa dari metode yang digunakan menggunakan *confusion matrix*. Confusion matrix digunakan sebagai alat evaluasi yang menyajikan jumlah klasifikasi benar dan salah dari data uji, sehingga memudahkan analisis kinerja algoritma klasifikasi (Ramadhan et al., 2023). Rumus confusion matrix untuk mengukur *accuracy*, *precision*, *recall*, dan *F1-score* pada persamaan (2),(3),(4),(5).

$$accuracy = \frac{TP}{jumlah\ data} \times 100 \% \quad (2)$$

$$precision = \frac{TP}{TP+FP} \times 100 \% \quad (3)$$

$$accuracy = \frac{TP}{TP+FN} \times 100 \% \quad (4)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (5)$$

HASIL DAN PEMBAHASAN

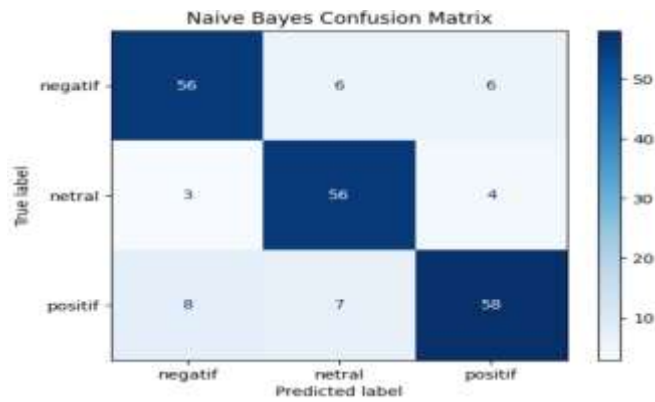
Dalam percobaan ini, terdapat 1.000 dataset yang diambil dari aplikasi twitter yang mana didalam dataset tersebut terdapat beberapa ulasan yang negatif, positif dan netral. Data tersebut akan melalui tahap labeling, data akan diberikan label angka 0 untuk ulasan positif, angka 1 untuk ulasan negatif dan angka 2 untuk ulasan netral. pada gambar 3 menunjukkan hasil dari tahap labeling dan data-data yang terkumpul.

negatif	333
netral	333
positif	334

Gambar 3. hasil labeling

Hasil Penelitian

Setelah proses labeling, selanjutnya adalah menggunakan algoritma *Naïve Bayes* untuk memperoleh data uji dan data training. Data akan dibagi menjadi 80% data latih dan 20% data uji, setelah data di bagi selanjutnya akan dilakukan resampling menggunakan *SMOTE* dan diklasifikasi menggunakan algoritma *Naïve Bayes*. Hasil klasifikasi dapat dilihat pada gambar 4.



Gambar 4. *Confusion Matrix*

Metric	Label 0	Label 1	Label 2	Average
0 Accuracy	-	-	-	68.0%
1 Precision	63.22%	63.41%	76.39%	67.67%
2 Recall	87.3%	38.81%	78.57%	68.23%
3 F1-Score	73.33%	48.15%	77.46%	66.32%

Gambar 5. Hasil Evaluasi

Berdasarkan gambar 5, hasil evaluasi model klasifikasi menggunakan algoritma *Naive Bayes* mendapatkan akurasi sebesar 68.0%, *precision* sebesar 67.67%, *recall* sebesar 68.23%, dan *f1-score* sebesar 66.32%

SIMPULAN

Penelitian ini telah berhasil melakukan analisis sentimen terhadap opini publik mengenai Dedi Mulyadi yang diungkapkan melalui Twitter. Dengan menerapkan metode ekstraksi fitur TF-IDF dan algoritma klasifikasi Naive Bayes, sistem mampu mengelompokkan sentimen pengguna ke dalam tiga kategori, yaitu positif, negatif, dan netral. Hasil evaluasi menunjukkan bahwa kombinasi metode tersebut mampu memberikan performa klasifikasi yang cukup baik, dengan tingkat akurasi yang memadai untuk menggambarkan kecenderungan opini publik. Analisis yang dilakukan menunjukkan bahwa mayoritas opini publik terhadap Dedi Mulyadi bersifat positif, meskipun masih terdapat sejumlah sentimen negatif dan netral. Temuan ini memberikan gambaran yang lebih objektif mengenai persepsi masyarakat di media sosial, khususnya dalam konteks politik dan citra tokoh publik

DAFTAR PUSTAKA

- Bahrawi, N. (2019). Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based. *Journal of Information Technology and Its Utilization*, 2(2), 29. <https://doi.org/10.30818/jitu.2.2.2695>
- Liu, B. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition*, May, 1–432. <https://doi.org/10.1017/9781108639286>
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1). <https://doi.org/10.31284/j.integer.2017.v2i1.95>
- Cerezo, M. (2004). Las nociones de Sachverhalt, Tatsache y Sachlage en el Tractatus de Wittgenstein. *Anuario Filosofico*, 37(2), 455–479. <https://doi.org/10.15581/009.37.29387>
- Firdaus, A. A., Yudhana, A., Riadi, I., & Mahsun, M. (2024). Prediction of Indonesian Presidential Election Results using Sentiment Analysis with Naïve Bayes Method. *Jurnal Media Informatika Budidarma*, 8(1), 41. <https://doi.org/10.30865/mib.v8i1.7007>
- Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Kasanah, A. N., Muladi, M., & Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 196–201. <https://doi.org/10.29207/resti.v3i2.945>
- McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. *AAAI/ICML-98 Workshop on Learning for Text Categorization*, 41–48. <https://doi.org/10.1.1.46.1529>
- Mustaqim, E. R. N., Pagalay, U., & ... (2024). Prediksi Tingkat Kepercayaan Masyarakat Terhadap Pilpres 2024 Menggunakan Tf-Idf Dan Bow Menggunakan Metode Svm. *Mandalika ISSN 2721 ...*, 515–530. <https://www.ojs.cahayamandalika.com/index.php/jcm/article/view/3114%0Ahttps://www.ojs.cahayamandalika.com/index.php/jcm/article/download/3114/2521>
- Nursidik Dinar, A., Susilo Yuda Irawan, A., & Umidah, Y. (2023). Analisis Sentimen Pada Pengguna Twitter Terhadap Program Kampus Merdeka Menggunakan Naïve Bayes.

- JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 755–760.
<https://doi.org/10.36040/jati.v7i1.6552>
- Pang, B., & Lee, L. (1999). Professional development for cooperative learning: issues and approaches. *Choice Reviews Online*, 36(08), 36-4623-36-4623.
<https://doi.org/10.5860/choice.36-4623>
- Ramadhan, T. D., Wahiddin, D., & Awal, E. E. (2023). Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma Naive Bayes. *Scientific Student Journal for Information, Technology and Science*, 11(1), 82–87. www.tripadvisor.com
- Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, 60(5), 503–520.
<https://doi.org/10.1108/00220410410560582>
- Sahami, M., David, D., Horvitz, E., & Building, G. (1998). Junk E-Mail t P (X = x). *Science*.
- Salton, G., & Buckley, C. (1987). Salton, G. and Buckley, C., 1988. Term-weighting approaches in automatic text retrieval._7896.pdf. In *National Foundation*.
- Saputro, P. H., Aristin, M., & Tyas, Dy. L. (2017). Klasifikasi Lagu Daerah Indonesia Berdasarkan Lirik Menggunakan Metode Tfidf Dan Naive Bayes. *Jurnal Teknologi Informatika Dan Terapan*, 4(1), 45–50.
- Silalahi, N., & Guidio Leonarde Ginting. (2023). Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF. *Bulletin of Computer Science Research*, 3(4), 276–282. <https://doi.org/10.47065/bulletincsr.v3i4.266>
- Susanto, A., Maula, M. A., Mulyono, I. U. W., & Sarker, M. K. (2021). Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method. *Journal of Applied Intelligent System*, 6(1), 40–45. <https://doi.org/10.33633/jais.v6i1.4465>
- Tri Putra, K., Amin Hariyadi, M., & Crysdian, C. (2023). Perbandingan Feature Extraction Tfidf Dan Bow Untuk Analisis Sentimen Berbasis Svm. *Jurnal Cahaya MAndalika*, 1449.
- Wang, R. Q., Ouedraogo, F. R., & Pathapati, S. S. (2019). Preprocessing. *Computational Fluid Dynamics: Applications in Water, Wastewater and Stormwater Treatment*, 3(3), 53–56.
<https://doi.org/10.1061/9780784415313.ch06>
- Wati, R., Ernawati, S., & Rachmi, H. (2023). Pembobotan TF-IDF Menggunakan Naive Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH. *Jurnal Manajemen Informatika (JAMIKA)*, 13(1), 84–93. <https://doi.org/10.34010/jamika.v13i1.9424>