Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id

E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari

OPEN ACCESS

https://doi.org/10.35870/jimik.v6i1.1211

Klasifikasi Sentimen Masyarakat di Twitter terhadap Puan Maharani dengan Metode *Modified K-Nearest Neighbor*

Wahyu Eka Putra [1], Muhammad Fikry [2], Yusra [3*], Febi Yanto [4], Eka Pandu Cynthia [5]

[1,2,3*,4,5] Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim, Kota Pekanbaru, Provinsi Riau, Indonesia.

Email: 11950115233@students.uin-suska.ac.id [1], muhammad.fikry@uin-suska.ac.id [2], yusra@uin-suska.ac.id [3*], febiyanto@uin-suska.ac.id [4], eka.pandu.cynthia@uin-suska.ac.id [5]

Histori Artikel:
Dikirim 1 November 2024; *Diterima dalam bentuk revisi* 15 November 2024; *Diterima* 30 Desember 2024; *Diterbitkan* 10 Januari 2025.

Abstrak

Penelitian ini bertujuan untuk mengatasi tantangan dalam klasifikasi sentimen Twitter terhadap Puan Maharani dengan menerapkan metode Modified K-Nearest Neighbor (MK-NN) yang dilengkapi dengan teknik feature weighting dan feature selection. Metode ini dirancang untuk meningkatkan akurasi dengan memberikan bobot lebih pada fitur penting dan mengurangi dimensi data guna menghindari overfitting. Data dikumpulkan menggunakan teknik crawling pada tweet berbahasa Indonesia, yang kemudian dilabeli secara manual dan diproses melalui tahap preprocessing. Hasil pengujian yang dilakukan menggunakan metode modified K-Nearest Neighbor (MK-NN) dengan menggunakan matriks konfusi menunjukkan kinerja model pada tiga nilai K yang berbeda (3, 5, dan 7) dan rasio 90:10, 80:20 dan 70:30. Dengan rasio data 90:10 dan nilai K=3, metode ini mencapai akurasi tertinggi sebesar 89,0%. Hasil ini menunjukkan bahwa kombinasi MK-NN dan teknik terkait sangat efektif dalam mengklasifikasikan data sentimen, memberikan solusi inovatif terhadap keterbatasan metode konvensional. Temuan ini memiliki potensi aplikasi dalam analisis opini publik, khususnya untuk mendukung pengambilan keputusan strategis berbasis data.

Kata Kunci: Puan Maharani; Modified K-Nearest Neighbor; Twitter; Klasifikasi Sentimen.

Abstract

This study aims to address the challenges in classifying sentiment on Twitter regarding Puan Maharani by implementing the Modified K-Nearest Neighbor (MK-NN) method, supplemented with feature weighting and feature selection techniques. This method is designed to improve accuracy by assigning higher weights to important features and reducing data dimensions to avoid overfitting. Data is collected using a crawling technique on Indonesian-language tweets, which are then manually labeled and processed through a preprocessing stage. The testing results using the modified K-Nearest Neighbor (MK-NN) method with confusion matrices show the model's performance at three different values of K (3, 5, and 7) and data ratios of 90:10, 80:20, and 70:30. With a 90:10 data ratio and K=3, the method achieved the highest accuracy of 89.0%. These results indicate that the combination of MK-NN and related techniques is highly effective in sentiment classification, offering an innovative solution to the limitations of conventional methods. These findings have potential applications in public opinion analysis, particularly for supporting data-driven strategic decision-making.

Keyword: Puan Maharani; Modified K-Nearest Neighbor; Twitter; Sentiment Classification.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

1. Pendahuluan

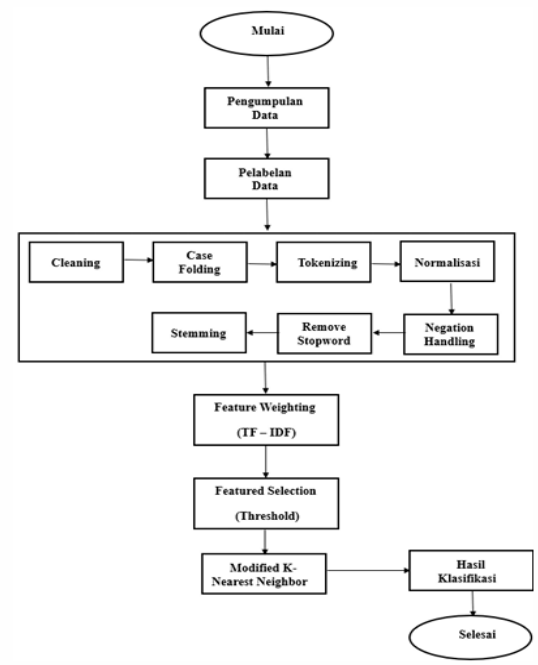
Media sosial, khususnya Twitter atau X, menjadi salah satu platform utama untuk berbagi pendapat dan membahas isu-isu politik, termasuk perhatian terhadap Puan Maharani, seorang politisi sekaligus pejabat pemerintah di Indonesia. Sebagai Menteri Koordinator Bidang Pembangunan Manusia dan Kebudayaan, Puan Maharani memiliki peran strategis dalam perumusan kebijakan pembangunan nasional. Saat ini, ia menjabat sebagai Ketua Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) periode 2019–2024. Diskusi publik mengenai Puan Maharani di Twitter menunjukkan bahwa platform ini telah menjadi ruang signifikan bagi masyarakat untuk menyuarakan pandangan politik mereka. Popularitas dan pengaruh Puan Maharani sebagai tokoh publik menjadikannya salah satu figur yang paling sering dibahas di media sosial (Nainggolan *et al.*, 2023). Twitter adalah platform jejaring sosial berbasis teks yang terus berkembang dengan berbagai fitur tambahan seperti unggahan gambar dan video. Pada Oktober 2021, Indonesia memiliki lebih dari 17 juta pengguna aktif Twitter, menempatkannya di posisi keenam dunia berdasarkan jumlah pengguna. Jakarta, sebagai salah satu kota dengan aktivitas tertinggi, masuk dalam peringkat sepuluh besar secara global, melampaui kota-kota besar seperti New York dan Tokyo (Priyansyah *et al.*, 2022; Utami & Marzuki, 2020). Tingginya tingkat aktivitas ini menjadikan Twitter sumber data potensial untuk penelitian sentimen masyarakat (Salim & Mayary, 2020).

Sentiment analysis digunakan untuk mengidentifikasi opini publik, baik positif maupun negatif, terhadap topik tertentu. Dalam penelitian ini, data opini publik mengenai Puan Maharani dikumpulkan dari Twitter berbahasa Indonesia (Vonega *et al.*, 2022). Studi ini bertujuan untuk memahami pola polaritas opini yang berkembang di masyarakat sekaligus mengevaluasi hubungan antara Puan Maharani dengan opini publik yang muncul. Pendekatan *sentiment analysis* dianggap relevan karena mampu memberikan gambaran opini secara luas dan dalam waktu nyata, meskipun terdapat tantangan teknis seperti keragaman gaya bahasa, keterbatasan panjang teks, dan elemen *noise* seperti *emoji* dan *hashtag* (Singgalen, 2021). Untuk mengatasi tantangan tersebut, penelitian ini menggunakan metode *Modified K-Nearest Neighbor* (MK-NN) yang mengintegrasikan teknik *feature weighting* dan *feature selection*. Pendekatan ini diharapkan dapat meningkatkan akurasi klasifikasi dengan mengurangi dimensi data serta mengoptimalkan bobot pada fitur yang relevan. Hasil penelitian sebelumnya menunjukkan bahwa metode *Naïve Bayes Classifier* mampu mencapai akurasi 88,89% dalam mengklasifikasikan *sentimen* masyarakat terhadap Puan Maharani (Hidayat *et al.*, 2024). Penelitian ini membuktikan bahwa metode MK-NN dapat mencapai akurasi lebih tinggi, yaitu 89,0%. MK-NN merupakan pengembangan dari metode *K-Nearest Neighbor* (K-NN) dengan menambahkan langkah validasi dan pembobotan (*weight voting*) (Kripsiandita *et al.*, 2021). Pengembangan ini bertujuan untuk memperbaiki keterbatasan *K-NN* dalam klasifikasi *sentimen* Twitter dan diharapkan memberikan kontribusi signifikan pada pengolahan data opini publik.

2. Metode Penelitian

Metode penelitian merupakan serangkaian langkah dan prosedur yang digunakan oleh peneliti untuk merancang, mengumpulkan, menganalisis, dan menyajikan data dalam suatu studi. Dalam ilmu pengetahuan dan teknologi memainkan peran penting sebagai solusi atas berbagai tantangan yang dihadapi umat manusia. Pola pikir kreatif dan inovatif menjadi aspek yang diperlukan dalam setiap proses penelitian. Pentingnya metodologi penelitian terletak pada kemampuannya memberikan kerangka kerja untuk memahami fenomena, menguji hipotesis, atau menjawab pertanyaan penelitian (Waruwu, 2024). Tahapan metode penelitian yang diterapkan pada tugas akhir ini dapat dilihat pada Gambar 1 di bawah ini.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211



Gambar 1. Desain Penelitian

2.1 Pengumpulan Data

Data dikumpulkan menggunakan teknik *crawling* pada platform Twitter dengan kata kunci "Puan Maharani." Tweet yang dikumpulkan dibatasi pada bahasa Indonesia untuk menjaga konsistensi data. Proses *crawling* dilakukan menggunakan bahasa pemrograman *Python* di lingkungan *Google Colab* selama periode 30 April hingga 30 Oktober 2023, menghasilkan total 9.000 tweet.

Tabel 1. Contoh *Crawling Data*

Tweet
b'@TetapPuan \nPuan Maharani, soso yg sangat berpengalaman dan terampil dlm memimpin dan kelola negara

2.2 Pelabelan Data

Pelabelan data bertujuan untuk memberikan informasi terkait atribut atau karakteristik setiap data, sehingga data tersebut dapat digunakan dalam berbagai tugas, seperti *classification*, *regression*, atau analisis lainnya. Proses pelabelan dilakukan secara manual dengan membaca dan menganalisis setiap *tweet* satu per satu. Pelabelan ini dilakukan oleh seorang ahli Bahasa Indonesia, Gadis Sari Elin, S.Pd. Hasil pelabelan menunjukkan bahwa dari total data yang terkumpul, terdapat 7.800 *tweet* positif dan 1.200 *tweet* negatif.

Tabel 2. Contoh Pelabelan Data

Tweet	Pelabelan Data
b'@TetapPuan \nPuan Maharani, soso yg sangat berpengalaman dan terampil dlm memimpin dan kelola	Positif
b'CNNIndonesia Goblok dia, pikirnya cawapres jd ppp taunya dr PDIP juga, namanya puan, cucu Soekarno putrinya megaloman	Negatif

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

2.3 *Preprocessing*

Meskipun *preprocessing* sangat penting, persiapan data teks seringkali menjadi proses yang kompleks. Konten teks dapat bersifat tidak standar dan memiliki tata bahasa yang tidak akurat akibat perubahan budaya, serta seringkali mengandung makna yang berbeda dari yang dimaksudkan oleh penulis. Selain itu, data mentah biasanya belum melalui tahap pengolahan awal (Hakim, 2021). *Preprocessing* pada data teks melibatkan beberapa langkah, seperti penghapusan kata sambung dan kata penghubung (*stopwords*) seperti "atau," "dan," dan "jika," serta normalisasi kata ke bentuk dasar dan penghilangan tanda baca yang tidak relevan, misalnya "isn't." Proses ini sering kali mengubah struktur kalimat untuk meningkatkan kualitas data yang digunakan dalam analisis. Tahapan *preprocessing* dalam penelitian ini terdiri dari tujuh langkah berikut:

a) *Cleaning*

Proses *cleaning* mencakup penghapusan elemen-elemen yang dapat mengganggu atau memengaruhi validitas dan kualitas data, seperti *hashtags*, *emojis*, *emoticons*, nama pengguna (*usernames*), *URLs*, dan alamat e-mail. Selain itu, langkah ini juga melibatkan penghapusan atribut yang tidak relevan terhadap tujuan analisis yang dilakukan (Gori *et al.*, 2024).

Tabel 3. Contoh *Cleaning*

Sebelum *Cleaning*	Sesudah *Cleaning*
b'@TetapPuan \nPuan Maharani, soso yg sangat berpengalaman dan terampil dlm memimpin dan kelola negara	Puan Maharani soso yg sangat berpengalaman dan terampil dlm memimpin dan kelola negara

b) *Case Folding*

Tahap *case folding* adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil, yaitu dari "A" hingga "Z" menjadi "a" hingga "z." Tidak semua dokumen menggunakan huruf kapital secara konsisten, sehingga langkah ini diperlukan untuk menyamakan format teks. Akibatnya, seluruh isi dokumen diubah menjadi huruf kecil tanpa memedulikan huruf besar atau kecil yang digunakan sebelumnya (Jacob *et al.*, 2019). Tujuan utama dari *case folding* adalah menyederhanakan bentuk kata dalam teks agar tidak terjadi perbedaan dalam identifikasi kata-kata yang secara semantik sama.

Tabel 4. Contoh *Case Folding*

Sebelum Case Folding	Sesudah Case Folding
Puan maharani soso yg sangat berpengalaman dan termpil dlm memimpin dan kelola negara	puan maharani soso yg sangat berpengalaman dan terampil dlm mempin dan kelola negara

c) *Tokenizing*

Tokenizing adalah proses memotong *string input* menjadi setiap kata penyusunnya. Selama proses ini, angka, tanda baca, dan karakter selain huruf alfabet akan dihilangkan (Rachman *et al.*, 2021). Tujuan dari *tokenizing* adalah membagi teks menjadi potongan-potongan kecil yang dapat dikelola untuk keperluan analisis atau pemodelan data pada tahap berikutnya.

Tabel 5. Contoh *Tokenizing*

Sebelum *Tokenizing*	Sesudah *Tokenizing*
puan maharani soso yg sangat berpengalaman dan termpil dlm memimpin dan kelola negara	puan
	maharani
	soso
	yg
	sangat
	berpengalaman

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

	dan
	termpil
	dlm
	memimpin
	dan
	kelola
	negara

d) Normalisasi

Normalisasi, dalam pemrosesan teks, merujuk pada proses mengubah teks menjadi bentuk standar atau normal yang lebih konsisten. Langkah ini bertujuan untuk menyelaraskan variasi dalam teks yang berpotensi mengganggu proses analisis atau pemodelan data (Whendasmoro & Joseph, 2022).

Tabel 6. Contoh Normalisasi

Sebelum Normalisasi	Sesudah Normalsasi
puan	puan
maharani	maharani
soso	sosok
yg	yang
sangat	sangat
berpengalaman	berpengalaman
dan	dan
terampil	terampil
dlm	dalam
memimpin	memimpin
dan	dan
kelola	mengelola
negara	negara

e) *Negation Handling*

Negation handling adalah proses penanganan elemen negasi dalam sebuah teks, yang bertujuan untuk mempertahankan sentimen aslinya meskipun terdapat penambahan kata negasi. Kata-kata negasi meliputi "tidak," "bukan," "belum," "tak," dan istilah serupa. Jika negasi tidak ditangani dengan benar, hasil analisis sentimen dapat kehilangan validitasnya (Tarecha *et al.*, 2022).

Tabel 7. Contoh *Negation Handling*

Sebelum *Negation Handling*	Sesudah *Negation Handling*
untung	untung
saja	saja
tidak	belum
pernah	
pilih	pilih
ibu	ibu
puan	puan

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

f) *Stopword Removal*

Stopword removal adalah proses menghilangkan kata-kata yang dianggap tidak penting dari teks. Kata-kata ini biasanya tidak memberikan kontribusi signifikan terhadap analisis, seperti "dan," "atau," "yang," dan sejenisnya. Tahap *stopword removal* dilakukan untuk meningkatkan efisiensi dan relevansi data dalam proses analisis (Khairunnisa *et al.*, 2021).

Tabel 8. Contoh *Stopword Removal*

Sebelum *Stopword Removal*	Sesudah *Stopword Removal*
puan	puan
maharani	maharani
sosok	sosok
yang	-
sangat	-
berpengalaman	berpengalaman
dan	-
terampil	terampil
dalam	-
memimpin	memimpin
dan	-
kelola	mengelola
negara	negara

g) *Stemming*

Stemming adalah proses mengubah berbagai bentuk kata menjadi representasi dasarnya. Langkah ini mencakup penghilangan semua imbuhan kata, termasuk awalan, sisipan, dan akhiran. Tujuan utama dari *stemming* adalah menyederhanakan kata ke bentuk dasarnya sehingga lebih mudah diproses atau dianalisis. Proses ini merupakan langkah penting dalam mengidentifikasi kata dasar dari sebuah kata (Albab *et al.*, 2023).

Tabel 9. Contoh *Stemming*

Sebelum *Stemming*	Sesudah *Stemming*
puan	puan
maharani	maharani
sosok	sosok
berpengalaman	pengalaman
terampil	terampil
memimpin	pimpin
mengelola	kelola

2.4 Pembobotan TF-IDF

Pembobotan *TF-IDF* (*Term Frequency-Inverse Document Frequency*) merupakan metode yang digunakan dalam pemrosesan teks untuk memberikan bobot pada setiap kata dalam dokumen berdasarkan frekuensi kemunculannya. Pada tahap ini, dokumen uji dan dataset melalui proses pembobotan untuk menghitung frekuensi kemunculan kata di masing-masing dokumen. Langkah awal dalam pembobotan *TF-IDF* adalah menentukan bobot atau *term frequency* setiap kata dalam dokumen, yang kemudian dikalikan dengan *inverse document frequency* (Sofiah *et al.*, 2023). Metode ini dirancang untuk memberikan bobot lebih tinggi pada fitur yang relevan untuk analisis sentimen.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

2.5 *Feature Selection*

Setelah memperoleh nilai *TF* dan *IDF* dari pembobotan kata, langkah berikutnya adalah melakukan *feature selection* menggunakan nilai ambang (*threshold*). Tujuan dari *feature selection* adalah menyederhanakan proses klasifikasi dengan mengurangi jumlah fitur yang digunakan, sehingga analisis menjadi lebih efisien. Threshold dipilih berdasarkan eksperimen terdahulu yang menunjukkan bahwa nilai 0,001 dapat mempertahankan fitur yang relevan sekaligus mengurangi dimensi data untuk menghindari *overfitting* (Zafitra Fadhlan & Fikry, 2023).

2.6 *Modified K-Nearest Neighbor* (MK-NN)

Metode *K-Nearest Neighbor (K-NN)* dikembangkan menjadi *Modified K-Nearest Neighbor (MK-NN)* untuk meningkatkan performa algoritma dalam proses klasifikasi. Tujuan algoritma *K-NN* adalah memanfaatkan data pelatihan untuk menentukan klasifikasi berdasarkan tetangga terdekat. Variabel *k* menentukan jumlah tetangga yang digunakan dalam klasifikasi, dan nilai prediksi dihitung berdasarkan variabel tersebut. Pengembangan dari *K-NN* menjadi *MK-NN* mencakup penambahan dua langkah komputasi, yaitu validasi data dan *weight voting* (Novitasari *et al.*, 2023). Proses validasi bertujuan meningkatkan akurasi dengan memvalidasi nilai prediksi pada data pelatihan, sedangkan *weight voting* menghitung bobot suara pada data uji berdasarkan hasil validasi (Sofiah *et al.*, 2023). Meskipun prinsip kerja *MK-NN* mirip dengan *K-NN*, metode ini terbukti lebih unggul dalam menangani klasifikasi. Langkah awal dalam algoritma *MK-NN* adalah perhitungan nilai validasi untuk data latih, dilanjutkan dengan proses *weight voting* untuk data uji.

3. Hasil dan Pembahasan

3.1 Hasil

Pengumpulan data dilakukan dengan menggunakan teknik *crawling* pada platform Twitter, hanya mencakup *tweet* berbahasa Indonesia dengan kata kunci "Puan Maharani." Proses ini menghasilkan dataset berjumlah total 9.000 *tweet*. Data dikumpulkan menggunakan bahasa pemrograman *Python* yang dijalankan di lingkungan *Google Colab*, selama periode 30 April 2023 hingga 30 Oktober 2023. Setelah pengumpulan data, dilakukan pelabelan secara manual oleh seorang ahli Bahasa Indonesia, Gadis Sari Elin, S.Pd. Proses ini menghasilkan 7.800 *tweet* positif dan 1.200 *tweet* negatif. Tahapan berikutnya adalah *preprocessing*, yaitu serangkaian langkah untuk membersihkan, memformat, dan mengubah teks mentah menjadi bentuk yang lebih terstruktur. Langkah ini bertujuan untuk menghilangkan elemen *noise* dan meningkatkan kualitas data sehingga dapat diolah secara lebih efektif oleh model atau algoritma. Melalui *preprocessing*, data yang tidak relevan dapat disaring untuk menghasilkan keluaran yang lebih akurat. Tahapan ini mencakup pembersihan data, pembobotan menggunakan *TF-IDF*, pemilihan fitur, serta proses klasifikasi. Pada tahap *TF-IDF* (*Term Frequency-Inverse Document Frequency*), bobot diberikan pada setiap kata dalam dokumen berdasarkan frekuensi kemunculannya. Langkah ini bertujuan untuk menyoroti kata-kata yang signifikan dalam analisis sentimen, sekaligus mengurangi pengaruh kata-kata umum yang kurang relevan. Hasil keluaran *TF-IDF* ditampilkan dalam Gambar 2 untuk menunjukkan distribusi bobot kata pada dataset yang telah diproses.

	007	008	009	01	0403	06	07	100	1000	1000000	...	yuuu	yuuuu	yuuuuu	zaitun	zayed	zaytun	zionis	zul	zulhas	Kelas
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif

5 rows × 5849 columns

Gambar 2. Pembobotan TF-IDF

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

Setelah pembobotan menggunakan *TF-IDF*, tahap selanjutnya adalah pemilihan fitur (*feature selection*). Proses ini dilakukan dengan menerapkan *threshold* sebesar 0,001, yang dipilih berdasarkan hasil eksperimen sebelumnya. Nilai *threshold* ini terbukti mampu mempertahankan fitur yang relevan sekaligus mengurangi dimensi data, sehingga risiko *overfitting* dapat diminimalkan. Pemilihan fitur ini bertujuan untuk menyederhanakan proses klasifikasi tanpa mengorbankan kualitas hasil analisis. Pada tahap *feature selection*, proses klasifikasi dilanjutkan menggunakan metode *Modified K-Nearest Neighbor (MK-NN)*. Dalam penelitian ini, pengujian dilakukan untuk mengevaluasi pengaruh berbagai nilai *K* (3, 5, dan 7) dalam menentukan tetangga terdekat. Pengujian juga dilakukan dengan membagi data latih dan data uji menggunakan tiga rasio, yaitu 70:30, 80:20, dan 90:10, untuk menentukan kombinasi nilai *K* dan rasio data yang memberikan akurasi terbaik. Evaluasi hasil klasifikasi dilakukan menggunakan matriks kebingungan (*confusion matrix*) yang disajikan dalam Tabel 10. Matriks ini digunakan untuk menganalisis kinerja model berdasarkan jumlah prediksi yang benar dan salah untuk masing-masing rasio pembagian data. Hasil evaluasi menunjukkan tingkat akurasi yang dicapai oleh model pada berbagai konfigurasi nilai *K* dan rasio data.

Tabel 10. Hasil *Confusion Matrix* untuk Nilai *K* = 3,5,7

Rasio data	Nilai *K*	TP	TN	FP	FN
70:30	*K* = 3	2329	50	297	24
	K = 5	2348	25	322	5
	K = 7	2352	19	328	1
80:20	*K* = 3	1549	42	185	24
	K = 5	1569	20	207	4
	K = 7	1572	15	212	1
90:10	*K* = 3	780	21	86	13
	K = 5	785	14	93	8
	K = 7	792	8	99	1

Tabel 11. Nilai *Accuracy* untuk Nilai *K* = 3,5,7

Rasio data	Nilai *K*	Akurasi
70:30	*K* = 3	88.1%
	K = 5	87.9%
	K = 7	87.8%
80:20	*K* = 3	88.4%
	K = 5	88.3%
	K = 7	88.2%
90:10	*K* = 3	89.0%
	K = 5	88.8%
	K = 7	88.9%

Dari tabel 11, menggunakan perbandingan dari 90:10 dan *K* = 3 mengahasilkan skor tertinggi yaitu dengan akurasi 89.0%.

3.2 Pembahasan

Hasil penelitian menunjukkan bahwa metode *Modified K-Nearest Neighbor (MK-NN)* memberikan hasil yang lebih baik dalam klasifikasi sentimen dibandingkan metode seperti *Naïve Bayes Classifier* (Hidayat *et al.*, 2024). Penggunaan *feature selection* berbasis *TF-IDF* dengan *threshold* sebesar 0,001 membantu mengurangi dimensi data tanpa menghilangkan fitur penting. Hal ini selaras dengan penelitian sebelumnya yang menyatakan bahwa prapemrosesan data, seperti *stemming* dan *feature selection*, dapat meningkatkan akurasi analisis teks (Albab *et al.*, 2023; Gori *et al.*, 2024).

E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

Tahapan prapemrosesan, termasuk *stopword removal*, *normalization*, dan *stemming*, berhasil mengurangi elemen *noise* dalam teks, membuat data lebih siap untuk dianalisis. Proses *cleaning*, yang mencakup penghapusan elemen seperti *hashtags* dan simbol lain, memastikan konsistensi data. Teknik pembobotan *TF-IDF* berperan dalam memberikan bobot pada kata-kata yang relevan untuk analisis sentimen (Khairunnisa *et al.*, 2021; Whendasmoro & Joseph, 2022). Algoritma *MK-NN* berhasil mengatasi keterbatasan algoritma *K-NN* konvensional dalam menghadapi data yang tidak terstruktur. Penambahan langkah validasi dan *weight voting* meningkatkan akurasi model, dengan akurasi tertinggi sebesar 89,0% pada rasio data latih dan uji 90:10. Hal ini sesuai dengan penelitian lain yang menunjukkan efektivitas *MK-NN* dalam klasifikasi data dengan pola distribusi yang tidak merata (Novitasari *et al.*, 2023; Kripsiandita *et al.*, 2021). Penelitian ini juga menangani tantangan dalam analisis sentimen di media sosial, seperti keragaman gaya bahasa, panjang teks yang terbatas, serta elemen non-teks seperti *emoji* dan *hashtags*. Langkah prapemrosesan yang dilakukan terbukti meningkatkan kualitas data input sehingga model dapat bekerja dengan lebih optimal (Hakim, 2021; Tarecha *et al.*, 2022). Metode yang diterapkan membuktikan bahwa kombinasi teknik prapemrosesan data dan algoritma *MK-NN* dapat menghasilkan klasifikasi sentimen yang akurat. Pendekatan ini memungkinkan penggunaan analisis sentimen untuk memahami opini publik di media sosial secara lebih luas dan terarah (Hidayat *et al.*, 2024; Priyansyah *et al.*, 2022).

4. Kesimpulan

Berdasarkan hasil penelitian dan uji coba, metode *Modified K-Nearest Neighbors* (*Modified* KNN) menunjukkan kinerja terbaik dalam klasifikasi data. Metode ini memberikan nilai tertinggi pada metrik evaluasi utama, yaitu akurasi. Akurasi tertinggi sebesar 89,0% dicapai pada pengaturan rasio data 90:10 dengan nilai $K = 3$. Hasil ini juga hampir konsisten untuk rasio data lainnya, seperti 80:20 dan 70:30, yang mengindikasikan bahwa metode Modified KNN mampu menangani klasifikasi data Twitter secara efektif pada berbagai pengaturan data.

5. Daftar Pustaka

Albab, M. U., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, *20*(2), 1-12. https://doi.org/10.26623/transformatika.v20i2.5374.

FADHLAN, Y. Z. (2023). Klasifikasi Sentimen Masyarakat Di Twitter Terhadap Ganjar Pranowo Dengan Metode Modified K-Nearest Neighbor. *Jurnal Informatika Universitas Pamulang*, *8*(2), 191-198. https://doi.org/10.32493/informatika.v7i2.30686.

Fikry, M., & Oktavia, L. (2023). Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga Bahan Bakar Minyak dengan Metode Modified K-Nearest Neighbor. *SATIN-Sains dan Teknologi Informasi*, *9*(1), 137-148. https://doi.org/10.33372/stn.v9i1.988.

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(1), 215-224. https://doi.org/10.25126/jtiik.20241118074.

Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *Jbase-Journal of business and audit information systems*, *4*(2). https://doi.org/10.30813/jbase.v4i2.3000.

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

Hidayat, R., Fikry, M., Yusra, Y., Yanto, F., & Cynthia, E. P. (2024). Penerapan Naïve Bayes Classifier dalam Klasifikasi Sentimen Publik di Twitter terhadap Puan Maharani. *JUKI: Jurnal Komputer dan Informatika*, *6*(1), 93-101. https://doi.org/10.53842/juki.v6i1.479.

Jacob, V. E., Lumenta, A. S., & Jacobus, A. (2019). Rancang Bangun Aplikasi Kemiripan Dokumen Dengan Sumber–Sumber Internet. *Jurnal Teknik Informatika*, *14*(2), 159-164. https://doi.org/10.35793/jti.v14i2.23990.

Khairunnisa, S., Adiwijaya, A., & Al Faraby, S. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, *5*(2), 406-414.

Kripsiandita, Y., Arifianto, D., & A'yun, Q. (2021). Deteksi Gangguan Autis Pada Anak Menggunakan Metode Modified K-Nearst Neighbor. *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, *6*(1), 31-37. https://doi.org/10.32528/justindo.v6i1.4357.

Nainggolan, I. D. P., Widyawan, P. A., Akbar, N., & Sholihatin, E. (2023). ANALISIS FRAMING PEMBERITAAN PUAN MAHARANI TERDAHAP ISU PERATURAN PERUNDANG-UNDANGAN CIPTA KERJA DI PORTAL BERITA KOMPAS. COM DAN DETIK. COM PADA EDISI OKTOBER 2020. *Sabda: Jurnal Sastra dan Bahasa*, *2*(1), 1-10. https://doi.org/10.572349/sabda.v2i1.434.

Novitasari, F. (2023). Sistem Klasifikasi Penyakit Jantung Menggunakan Teknik Pendekatan SMOTE Pada Algoritma Modified K-Nearest Neighbor. *Jurnal Building of Informatics, Technology and Science (BITS)*, *5*(1), 274-284. https://doi.org/10.47065/bits.v5i1.3610.

Priyansyah, R. N., Wibowo, K. A., & Fuady, I. (2022). TWITTER SEBAGAI MEDIA KOMUNIKASI KRISIS PEMIMPIN PEMERINTAHAN DI INDONESIA (STUDI GELOMBANG COVID-19 VARIAN DELTA DAN OMICRON). *Jurnal Studi Komunikasi dan Media*, *26*(1), 31-52. https://doi.org/10.17933/jskm.2022.4788.

Rachman, D. A. C., Goejantoro, R., & Amijaya, F. D. T. (2021). Implementasi Text Mining Pengelompokkan Dokumen Skripsi Menggunakan Metode K-Means Clustering. *EKSPONENSIAL*, *11*(2), 167-174. https://doi.org/10.30872/eksponensial.v11i2.660.

Salim, S. S., & Mayary, J. (2020). Analisis Sentimen pengguna Twitter terhadap dompet elektronik dengan metode lexicon based dan k–nearest neighbor. *Jurnal Ilmiah Informatika Komputer*, *25*(1), 1-17. https://doi.org/10.35760/ik.2020.v25i1.2411.

Sihombing, D. Y., & Nataliani, Y. (2021). Analisis Interaksi Pengguna Twitter pada Strategi Pengadaan Barang Menggunakan Social Network Analysis. *Sistemasi: Jurnal Sistem Informasi*, *10*(2), 434-444.

Singgalen, Y. A. (2021). Pemilihan metode dan algoritma dalam analisis sentimen di media sosial: sistematic literature review. *Journal of Information Systems and Informatics*, *3*(2), 278-302. https://doi.org/10.33557/journalisi.v3i2.125.

Tarecha, R. I., Wahyudi, F., & Jannah, U. M. (2022). Penanganan Negasi dalam Analisa Sentimen Bahasa Indonesia. *Jurnal Sistem Informasi dan Informatika (JUSIFOR)*, *1*(1), 51-58. https://doi.org/10.33379/jusifor.v1i1.1276.

E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 6 No. 1 (2025) | Januari
OPEN ACCESS
https://doi.org/10.35870/jimik.v6i1.1211

Utami, I., & Marzuki, M. (2020). Analisis sistem informasi banjir berbasis media twitter. *Jurnal Fisika Unand*, *9*(1), 67-72. https://doi.org/10.25077/jfu.9.1.67-72.2020.

Vonega, D. A., Fadila, A., & Kurniawan, D. E. (2022). Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024. *Journal of Applied Informatics and Computing*, *6*(2), 129-135.

Waruwu, M. (2024). Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan. *Jurnal Ilmiah Profesi Pendidikan*, *9*(2), 1220-1230. https://doi.org/10.29303/jipp.v9i2.2141.

Whendasmoro, R. G., & Joseph, J. (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, *9*(4), 872-876.