

Perbandingan Support Vector Machine dan K-Nearest Neighbor pada Klasifikasi Emosi Tweet

Merry Nehemia Nainggolan ^{1*}, Arif Bijaksana ², Rina septriana ³

^{1,2,3} Program Studi Informatika, Fakultas Teknik, Universitas Tanjungpura, Pontianak, Indonesia

Corresponding Author: merrynehemianainggolan@gmail.com

Info Artikel

Submitted: 28 Agustus 2026

Revised: 09 September 2026

Accepted: 29 September 2026

Published: 12 Oktober 2026

Keywords: emotion classification; Indonesian tweets; support vector machine; k-nearest neighbor; feature extraction

Kata Kunci: klasifikasi emosi; tweet bahasa Indonesia; support vector machine; k-nearest neighbor; ekstraksi fitur

Abstract

Emotion classification of Indonesian tweets is difficult because social-media text often uses nonstandard language. This study compares Support Vector Machine and K-Nearest Neighbor across eight configurations combining term frequency-inverse document frequency or bigram features, synthetic minority oversampling, and grid-search tuning. The experiment used 4,403 manually labelled tweets in five emotion classes, with 90% training and 10% testing data. Text underwent case folding, cleaning, and normalization. Ten-fold cross-validation and confusion-matrix metrics evaluated performance. Support Vector Machine with term frequency-inverse document frequency, oversampling, and tuning produced the best result: 76.04% accuracy and 76.19% F1-score. K-Nearest Neighbor reached 54.08% accuracy and 48.61% F1-score. Term frequency-inverse document frequency consistently outperformed bigram features, while oversampling effects differed by algorithm.

Abstrak

Klasifikasi emosi pada tweet berbahasa Indonesia sulit dilakukan karena teks media sosial sering memakai bahasa tidak baku. Penelitian ini membandingkan Support Vector Machine dan K-Nearest Neighbor melalui delapan konfigurasi yang mengombinasikan term frequency-inverse document frequency atau bigram, synthetic minority oversampling, dan penyetelan grid search. Eksperimen menggunakan 4.403 tweet berlabel manual dalam lima kelas emosi, dengan 90% data latih dan 10% data uji. Teks menjalani case folding, cleaning, dan normalisasi. Kinerja dievaluasi melalui validasi silang sepuluh lipatan dan confusion matrix. Support Vector Machine dengan term frequency-inverse document frequency, oversampling, dan penyetelan menghasilkan akurasi terbaik sebesar 76,04% dan F1-score 76,19%. K-Nearest Neighbor mencapai akurasi 54,08% dan F1-score 48,61%. Term frequency-inverse document frequency secara konsisten unggul bigram, sedangkan pengaruh oversampling berbeda menurut algoritma.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Publisher: Lembaga Penerbit Penelitian Nusantara

Pendahuluan

Media sosial menyediakan teks dalam jumlah besar yang mencerminkan pendapat dan keadaan emosional penggunanya. Tweet menarik untuk dianalisis karena dipublikasikan secara waktu nyata, tetapi bentuknya singkat dan sering menggunakan ejaan tidak baku, singkatan, simbol, serta struktur

kalimat yang tidak teratur. Karakteristik tersebut membuat klasifikasi emosi memerlukan prapemrosesan dan representasi fitur yang sesuai sebelum model pembelajaran mesin diterapkan.

Klasifikasi emosi bertujuan memetakan teks ke kelas emosi tertentu berdasarkan pola yang dipelajari dari data berlabel (Musu et al., 2021). Dataset Indonesian Twitter Emotion yang disusun oleh Saputri et al. (2019) membagi tweet ke dalam lima kelas, yaitu anger, fear, happy, love, dan sadness. Pada dataset tersebut, Logistic Regression menghasilkan F1-score 75,98%, Support Vector Machine 70,84%, dan Random Forest 66,60%. Temuan ini menunjukkan bahwa pilihan algoritma dan representasi fitur dapat menghasilkan perbedaan kinerja yang berarti pada teks media sosial berbahasa Indonesia.

Penelitian lain juga memperlihatkan variasi kinerja antarmodel. Negara et al. (2021) membandingkan Logistic Regression dan K-Nearest Neighbor pada klasifikasi emosi tweet berbahasa Indonesia dengan pembobotan term frequency-inverse document frequency dan penyetelan hiperparameter. Logistic Regression memperoleh F1-score 66%, sedangkan K-Nearest Neighbor memperoleh 55%. Aldawiyati (2018) juga menerapkan K-Nearest Neighbor untuk mengidentifikasi emosi pada tweet. Support Vector Machine telah digunakan pada klasifikasi komentar cyberbullying (Luqyana et al., 2018), sedangkan K-Nearest Neighbor digunakan pada opini tayangan televisi (Nurjanah et al., 2017) dan ulasan tempat wisata (Sari, 2020). Pada analisis sentimen Twitter, Nasution dan Hayaty (2019) melaporkan perbedaan akurasi dan waktu proses antara kedua algoritma. Hasil-hasil tersebut menegaskan bahwa performa model perlu dinilai pada data dan konfigurasi eksperimen yang sama.

Representasi fitur merupakan bagian penting dalam klasifikasi teks. Term frequency-inverse document frequency memberi bobot berdasarkan frekuensi sebuah istilah dalam dokumen dan kelangkaannya dalam kumpulan dokumen. Sebaliknya, bigram merepresentasikan pasangan kata yang muncul berurutan. Nugroho (2018) menggunakan fitur n-gram untuk analisis sentimen Twitter, sedangkan Trianto et al. (2020) memperoleh akurasi 94,06% dengan bigram pada ulasan produk elektronik. Perbedaan domain dan distribusi kata membuat keunggulan suatu fitur tidak dapat langsung digeneralisasi ke klasifikasi emosi tweet.

Ketidakseimbangan jumlah data antarkelas juga dapat memengaruhi kemampuan model mengenali kelas minoritas. Synthetic Minority Over-sampling Technique menghasilkan sampel sintesis dari kelas minoritas untuk memperbaiki representasinya. Siringoringo (2018) melaporkan peningkatan G-Mean dan F-Measure setelah teknik tersebut diterapkan pada data tidak seimbang.

Namun, sampel sintetis tidak selalu meningkatkan kinerja karena hasilnya dipengaruhi oleh karakteristik data, pemilihan fitur, dan algoritma klasifikasi.

Kajian terdahulu umumnya menguji sebagian kombinasi algoritma, fitur, oversampling, atau penyetelan hiperparameter. Penelitian ini menguji Support Vector Machine dan K-Nearest Neighbor pada satu dataset melalui delapan konfigurasi yang mengombinasikan term frequency-inverse document frequency atau bigram, penggunaan oversampling, dan penyetelan hiperparameter. Rancangan yang seragam memungkinkan pengaruh setiap komponen dibandingkan secara langsung. Penelitian bertujuan menentukan algoritma dengan kinerja terbaik, membandingkan pengaruh kedua fitur, serta menilai pengaruh oversampling dan penyetelan hiperparameter pada klasifikasi emosi tweet berbahasa Indonesia.

Metode Penelitian

Penelitian menggunakan pendekatan eksperimen komparatif. Data berasal dari Indonesian Twitter Emotion Dataset yang dikumpulkan melalui Twitter Streaming API pada 1-14 Juni 2018 dan dipublikasikan oleh Saputri et al. (2019). Dataset berisi 4.403 tweet berbahasa Indonesia yang telah diberi label manual ke dalam lima kelas emosi: anger, fear, happy, love, dan sadness. Kualitas anotasi pada sumber data dilaporkan dengan nilai Kappa 0,917.

Prapemrosesan dilakukan melalui case folding, cleaning, dan normalisasi. Case folding mengubah seluruh huruf menjadi huruf kecil. Cleaning menghapus tanda baca, angka, simbol, penanda nama pengguna, alamat uniform resource locator, hashtag, dan penanda nomor sensitif. Normalisasi mengubah singkatan dan kata tidak baku menjadi bentuk baku agar variasi penulisan berkurang.

Teks yang telah diproses diubah menjadi vektor numerik menggunakan dua pendekatan. Pendekatan pertama memakai TfidfVectorizer untuk menghasilkan bobot term frequency-inverse document frequency. Pendekatan kedua memakai CountVectorizer dengan ngram_range=(2,2) untuk membentuk fitur bigram. Dataset kemudian dibagi menjadi 90% data latih dan 10% data uji.

Dua algoritma yang dibandingkan ialah K-Nearest Neighbor dan Support Vector Machine dengan kernel radial basis function. Penyetelan hiperparameter dilakukan menggunakan GridSearchCV. Pencarian K-Nearest Neighbor mencakup nilai n_neighbors 1-45, sedangkan pencarian Support Vector Machine menentukan kombinasi C dan gamma. Evaluasi data latih menggunakan validasi silang sepuluh lipatan, mengikuti prinsip pengujian silang yang juga

digunakan Muhidin dan Burhan (2018). Evaluasi data uji menggunakan confusion matrix dengan ukuran akurasi, presisi, recall, dan F1-score.

Delapan skenario diterapkan pada kedua algoritma. Skenario dibentuk dari kombinasi dua fitur, penggunaan atau tanpa oversampling, serta penggunaan atau tanpa penyetelan hiperparameter. Susunan lengkap skenario ditampilkan pada Tabel 1.

Tabel 1. Skenario pengujian model

Skenario	Fitur	Oversampling	Penyetelan hiperparameter
1	TF-IDF	Tidak	Tidak
2	Bigram	Tidak	Tidak
3	TF-IDF	Tidak	Ya
4	Bigram	Tidak	Ya
5	TF-IDF	Ya	Tidak
6	Bigram	Ya	Tidak
7	TF-IDF	Ya	Ya
8	Bigram	Ya	Ya

Hasil dan Pembahasan

Hasil

Grid search menghasilkan nilai hiperparameter yang berbeda menurut representasi fitur dan penggunaan oversampling. Pada K-Nearest Neighbor, nilai K terbaik untuk TF-IDF, bigram, TF-IDF dengan oversampling, dan bigram dengan oversampling masing-masing adalah 38, 2, 2, dan 8. Pada Support Vector Machine, kombinasi C dan gamma untuk keempat konfigurasi tersebut masing-masing adalah (10; 0,1), (100; 0,005), (10; 0,1), dan (10; 0,005).

Hasil pengujian menunjukkan perbedaan yang konsisten antara algoritma dan fitur. Tabel 2 merangkum akurasi serta F1-score pada data uji untuk seluruh skenario.

Tabel 2. Akurasi dan F1-score data uji pada delapan skenario

Algoritma	Skenario	Konfigurasi	Akurasi (%)	F1-score (%)
KNN	1	TF-IDF	50,11	49,26
KNN	2	Bigram	26,08	17,15
KNN	3	TF-IDF + tuning	53,97	53,76

Algoritma	Skenario	Konfigurasi	Akurasi (%)	F1-score (%)
KNN	4	Bigram + tuning	26,76	15,63
KNN	5	TF-IDF + SMOTE	42,83	33,11
KNN	6	Bigram + SMOTE	22,14	14,37
KNN	7	TF-IDF + SMOTE + tuning	54,08	48,61
KNN	8	Bigram + SMOTE + tuning	23,96	19,47
SVM	1	TF-IDF	61,45	61,60
SVM	2	Bigram	36,05	29,11
SVM	3	TF-IDF + tuning	63,27	64,14
SVM	4	Bigram + tuning	48,53	46,88
SVM	5	TF-IDF + SMOTE	75,68	76,28
SVM	6	Bigram + SMOTE	38,48	37,24
SVM	7	TF-IDF + SMOTE + tuning	76,04	76,19
SVM	8	Bigram + SMOTE + tuning	43,38	42,16

Support Vector Machine menghasilkan akurasi lebih tinggi daripada K-Nearest Neighbor pada seluruh skenario yang sepadan. Kinerja terbaik keseluruhan diperoleh pada skenario 7, yaitu Support Vector Machine dengan TF-IDF, oversampling, dan penyetelan hiperparameter. Model tersebut menghasilkan akurasi data uji 76,04%, presisi 76,37%, recall 76,07%, dan F1-score 76,19%. Akurasi data latih sebesar 76,83% menghasilkan selisih 0,79 poin persentase terhadap data uji.

Tabel 3. Perbandingan model terbaik kedua algoritma

Ukuran	KNN skenario 7 (%)	SVM skenario 7 (%)
Akurasi	54,08	76,04
Presisi	64,92	76,37
Recall	54,12	76,07
F1-score	48,61	76,19

Seluruh model kemudian diintegrasikan ke dalam aplikasi generik. Pengguna dapat memilih algoritma, jenis vektorisasi, penggunaan oversampling, dan penyetelan hiperparameter sebelum memasukkan teks. Aplikasi menampilkan kelas emosi yang diprediksi beserta probabilitas setiap kelas. Contoh implementasi model terbaik ditunjukkan pada Gambar 1.

PERBANDINGAN KINERJA MODEL ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN) TERHADAP KLASIFIKASI EMOSI TWEET BERBAHASA INDONESIA

Model : SVM-RBF + SMOTE + GRID SEARCH + TFIDF

Masukkan teks berbahasa Indonesia

gimana gak telat kalau gini terus

Klasifikasi

Hasil kelas emosinya adalah : anger

Nilai probabilitas setiap kelas :

	anger	fear	happy	love	sadness
0	0.7259	0.2611	0.0091	0.0004	0.0034

Gambar 1. Implementasi model klasifikasi emosi pada aplikasi generik

Pembahasan

Keunggulan Support Vector Machine terlihat pada semua pasangan skenario. Pada konfigurasi dasar TF-IDF, akurasi Support Vector Machine mencapai 61,45%, sedangkan K-Nearest Neighbor 50,11%. Setelah oversampling dan penyetelan hiperparameter diterapkan, selisih akurasi kedua algoritma meningkat menjadi 21,96 poin persentase. Hasil ini sejalan dengan perbandingan pada analisis sentimen Twitter oleh Nasution dan Hayaty (2019), yang menunjukkan bahwa karakteristik pemisahan kelas pada ruang fitur teks dapat lebih sesuai bagi Support Vector Machine daripada pendekatan berbasis kedekatan.

TF-IDF menghasilkan akurasi lebih tinggi daripada bigram pada seluruh pasangan skenario. Pada model terbaik Support Vector Machine, TF-IDF menghasilkan akurasi 76,04%, sedangkan hasil terbaik berbasis bigram adalah 48,53% pada skenario 4. Pola serupa muncul pada K-Nearest Neighbor, yaitu 54,08% untuk konfigurasi terbaik TF-IDF dan 26,76% untuk konfigurasi terbaik bigram berdasarkan tabel evaluasi. TF-IDF mempertimbangkan pentingnya istilah dalam keseluruhan dokumen, sedangkan bigram pada penelitian ini hanya menghitung kemunculan pasangan kata. Temuan ini berbeda dari hasil Trianto et al. (2020) pada ulasan produk elektronik, sehingga efektivitas bigram tampak bergantung pada domain, panjang teks, dan pola bahasa dalam data.

Pengaruh oversampling tidak seragam. Pada Support Vector Machine dengan TF-IDF, oversampling meningkatkan akurasi dari 61,45% menjadi 75,68%, atau 14,23 poin persentase. Pada K-Nearest Neighbor dengan fitur yang sama, akurasi justru turun dari 50,11% menjadi 42,83%. Perbedaan tersebut menunjukkan bahwa penambahan sampel sintetis dapat memperbaiki batas

keputusan Support Vector Machine, tetapi belum tentu memperbaiki struktur kedekatan lokal yang digunakan K-Nearest Neighbor. Hasil ini melengkapi temuan Siringoringo (2018): oversampling bermanfaat untuk data tidak seimbang, tetapi dampaknya tetap dipengaruhi model dan representasi fitur.

Penyetelan hiperparameter memperbaiki beberapa konfigurasi. Pada TF-IDF tanpa oversampling, akurasi K-Nearest Neighbor naik 3,86 poin persentase dan Support Vector Machine naik 1,82 poin persentase. Pada bigram, kenaikan Support Vector Machine mencapai 12,48 poin persentase, sedangkan K-Nearest Neighbor hanya meningkat 0,68 poin persentase. Ketika oversampling telah diterapkan pada TF-IDF, penyetelan menambah akurasi Support Vector Machine sebesar 0,36 poin persentase dan K-Nearest Neighbor sebesar 11,25 poin persentase dibanding konfigurasi oversampling tanpa tuning. Dengan demikian, manfaat tuning bergantung pada kondisi awal model dan tidak dapat dinilai terpisah dari fitur serta strategi penyeimbangan data.

Model terbaik memiliki akurasi latih 76,83% dan akurasi uji 76,04%. Selisih 0,79 poin persentase menunjukkan performa uji sedikit lebih rendah, tetapi relatif dekat dengan hasil pelatihan. Generalisasi model tetap perlu diuji pada data dari periode dan sumber berbeda karena penelitian hanya menggunakan satu dataset yang dikumpulkan pada 2018 dan membatasi keluaran pada lima kelas emosi. Penelitian selanjutnya juga perlu menerapkan oversampling hanya pada data latih di setiap lipatan validasi, membandingkan representasi fitur kontekstual, dan melaporkan hasil per kelas agar kemampuan model terhadap kelas minoritas dapat dinilai lebih rinci.

Simpulan

Support Vector Machine memberikan kinerja terbaik untuk klasifikasi lima emosi pada tweet berbahasa Indonesia. Konfigurasi TF-IDF, oversampling, dan penyetelan hiperparameter menghasilkan akurasi 76,04% serta F1-score 76,19%, lebih tinggi daripada konfigurasi terbaik K-Nearest Neighbor yang menghasilkan akurasi 54,08% dan F1-score 48,61%. TF-IDF secara konsisten mengungguli bigram pada kedua algoritma. Oversampling meningkatkan kinerja Support Vector Machine berbasis TF-IDF, tetapi menurunkan akurasi beberapa skenario K-Nearest Neighbor, sehingga pengaruhnya bergantung pada algoritma dan representasi fitur. Penyetelan hiperparameter memberi peningkatan pada sebagian besar perbandingan, terutama pada Support Vector Machine berbasis bigram dan K-Nearest Neighbor berbasis TF-IDF. Hasil tersebut mendukung penggunaan

Support Vector Machine dengan TF-IDF, oversampling, dan grid search sebagai model utama pada aplikasi klasifikasi emosi yang dikembangkan.

Daftar Pustaka

- Aldawiyati, O. N. (2018). *Identifikasi emosi pada tweet dengan algoritma K-Nearest Neighbor untuk rekomendasi judul lagu*.
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis sentimen cyberbullying pada komentar Instagram dengan metode klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4704-4713.
- Muhidin, A., & Burhan, A. (2018). Klasifikasi tingkat produksi tanaman padi Kabupaten Karawang menggunakan metode Naive Bayes dan K-Fold Cross Validation. *Jurnal Teknologi Pelita Bangsa SIGMA*, 8(2), 225-232.
- Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh komposisi data training dan testing terhadap akurasi algoritma C4.5. *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, 10(1), 186-195.
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan akurasi dan waktu proses algoritma K-NN dan SVM dalam analisis sentimen Twitter. *Jurnal Informatika*, 6(2), 226-235. <https://doi.org/10.31311/ji.v6i2.5129>
- Negara, A. B. P., Muhandi, H., & Sajid, F. (2021). Perbandingan algoritma klasifikasi terhadap emosi tweet berbahasa Indonesia. *Jurnal Edukasi dan Penelitian Informatika*, 7(2), 242-249. <https://doi.org/10.26418/jp.v7i2.48198>
- Nugroho, A. (2018). Analisis sentimen pada media sosial Twitter menggunakan Naive Bayes Classifier dengan ekstraksi fitur N-Gram. *J-SAKTI Jurnal Sains Komputer dan Informatika*, 2(2), 200. <https://doi.org/10.30645/j-sakti.v2i2.83>
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial Twitter menggunakan metode K-Nearest Neighbor dan pembobotan jumlah retweet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1750-1757.
- Saputri, M. S., Mahendra, R., & Adriani, M. (2019). Emotion classification on Indonesian Twitter dataset. *Proceedings of the 2018 International Conference on Asian Language Processing*, 90-95. <https://doi.org/10.1109/IALP.2018.8629262>

- Sari, R. (2020). Analisis sentimen pada review objek wisata Dunia Fantasi menggunakan algoritma K-Nearest Neighbor. *Evolusi Jurnal Sains dan Manajemen*, 8(1), 10-17. <https://doi.org/10.31294/evolusi.v8i1.7371>
- Siringoringo, R. (2018). Klasifikasi data tidak seimbang menggunakan algoritma SMOTE dan K-Nearest Neighbor. *Jurnal ISD*, 3(1), 44-49.
- Trianto, R. B., Triyono, A., & Arum, D. M. P. (2020). Klasifikasi rating otomatis pada dokumen teks ulasan produk elektronik menggunakan metode N-Gram dan Naive Bayes. *Jurnal Informatika Universitas Pamulang*, 5(3), 295. <https://doi.org/10.32493/informatika.v5i3.6110>