JTIM : Jurnal Teknologi Informasi dan Multimedia

p-ISSN : 2715-2529
e-ISSN : 2684-9151
https://journal.sekawan-org.id/index.php/jtim

Pengaruh Teknik Representasi Teks *Bag-of-Words* dan *TF-IDF* terhadap Akurasi Klasifikasi Sentimen Teks *Multi-Domain*

Angelica Davina Meisya Putri [1,*], Neny Sulistianingsih [2], dan Ria Rismayati [1]

[1] Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Bumigora Mataram, Indonesia
[2] Program Studi Magister Ilmu Komputer, Program Pascasarjana, Universitas Bumigora Mataram, Indonesia

* Korespondensi: angelicameisya2@gmail.com

Sitasi: Putri, A. D. M.; Sulistianingsih, N.; Rismayati, R. (**2025**). Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF terhadap Akurasi Klasifikasi Sentimen Teks Multi-Domain JTIM: Jurnal Teknologi Informasi Dan Multimedia, 7(4), 675-688. https://doi.org/10.35746/jtim.v7i4.756

Diterima: 27-05-2025
Direvisi: 06-07-2025
Disetujui: 14-07-2025

Abstract: Text representation is an essential component in sentiment analysis systems, as it determines how text data is converted into numerical features that can be utilized by classification algorithms. This study aims to analyze the effect of two popular text representation techniques, namely Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), on the performance of short text sentiment classification in a multi-domain context. The dataset used is a combination of original data and synonym-based augmentation data, with a total of 418 text entries. The two machine learning algorithms used in the evaluation were Ridge Classifier and Complement Naïve Bayes. The assessment was conducted using the Stratified K-Fold cross-validation technique as well as four key evaluation metrics: accuracy, precision, recall, and F1-score. Experimental results show that TF-IDF representation consistently performs better than BoW in both models. The best configuration was achieved by Ridge Classifier with TF-IDF, which obtained an accuracy of 0.911 and F1-score of 0.908. These findings underscore the importance of choosing the right feature representation technique in improving the effectiveness of text-based sentiment classification systems.

Keywords: Bag-of-Words; sentiment analysis; text classification; text representation; TF-IDF

Abstrak: Representasi teks merupakan komponen esensial dalam sistem analisis sentimen, karena menentukan bagaimana data teks diubah menjadi fitur numerik yang dapat dimanfaatkan oleh algoritma klasifikasi. Penelitian ini bertujuan untuk menganalisis pengaruh dua teknik representasi teks populer, yaitu Bag-of-Words (BoW) dan Term Frequency–Inverse Document Frequency (TF-IDF), terhadap performa klasifikasi sentimen teks pendek dalam konteks multi-domain. Dataset yang digunakan merupakan hasil kombinasi antara data asli dan data augmentasi berbasis sinonim, dengan total 418 entri teks. Dua algoritma pembelajaran mesin yang digunakan dalam evaluasi adalah *Ridge Classifier* dan *Complement Naïve Bayes*. Penilaian dilakukan menggunakan teknik validasi silang Stratified K-Fold serta empat metrik evaluasi utama: akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa representasi TF-IDF secara konsisten memberikan performa lebih baik dibandingkan BoW pada kedua model. Konfigurasi terbaik dicapai oleh *Ridge Classifier* dengan TF-IDF, yang memperoleh akurasi sebesar 0,911 dan F1-score sebesar 0,908. Temuan ini menggarisbawahi pentingnya pemilihan teknik representasi fitur yang tepat dalam meningkatkan efektivitas sistem klasifikasi sentimen berbasis teks.

Kata kunci: analisis sentimen; Bag-of-Words; klasifikasi teks; representasi teks; TF-IDF

1. Pendahuluan

Pertumbuhan interaksi manusia dalam *platform* digital seperti media sosial, forum daring, dan ulasan produk telah mendorong meningkatnya kebutuhan untuk memahami opini publik secara otomatis melalui analisis sentiment. Analisis sentimen menjadi alat penting dalam berbagai sektor, seperti bisnis, politik, dan kesehatan. Pentingnya teknologi ini tercermin dari pertumbuhan pasarnya yang signifikan, di mana pasar analisis sentimen global diproyeksikan tumbuh dari US5,1 Miliar pada tahun 2024 menjadi US11,4 Miliar pada tahun 2030 [1]–[4]. Hal ini didorong oleh kemampuannya untuk mengungkap persepsi publik terhadap isu atau kebijakan tertentu secara akurat dan terukur. Salah satu tantangan utama dalam analisis sentimen adalah keberagaman gaya penulisan, penggunaan bahasa informal, sarkasme, serta konteks kata, yang dapat mempersulit model dalam mengidentifikasi sentimen secara akurat [5].

Dalam pengolahan teks, representasi fitur memegang peran krusial terhadap keberhasilan model klasifikasi. Dua pendekatan representasi fitur yang umum digunakan adalah *Bag-of-Words* (BoW) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). BoW merepresentasikan teks berdasarkan frekuensi kemunculan kata tanpa mempertimbangkan konteks atau posisi kata, sementara TF-IDF menambahkan bobot berdasarkan pentingnya suatu kata dalam dokumen relatif terhadap keseluruhan korpus [6].

Beberapa penelitian sebelumnya telah mengevaluasi berbagai teknik representasi fitur. Studi dalam domain spesifik seperti ulasan vaksinasi [1], e-commerce [7], dan ulasan makanan [8] telah menguji efektivitas TF-IDF dan Word2Vec, sementara penelitian lain membandingkan performa BoW dan TF-IDF pada data multi-topik [6], [9]. Temuan dari studi-studi ini secara umum mengonfirmasi bahwa pemilihan representasi fitur sangat memengaruhi hasil klasifikasi, dengan TF-IDF seringkali menunjukkan performa yang kompetitif. Namun, seiring kemunculan arsitektur *deep learning* yang lebih kompleks (misalnya BERT), fokus penelitian seringkali bergeser ke model-model tersebut. Akibatnya, kajian yang secara sistematis membandingkan efektivitas metode fundamental seperti BoW dan TF-IDF pada dataset multi-domain dengan algoritma klasifikasi yang lebih baru seperti Ridge Classifier masih jarang dilakukan secara menyeluruh.

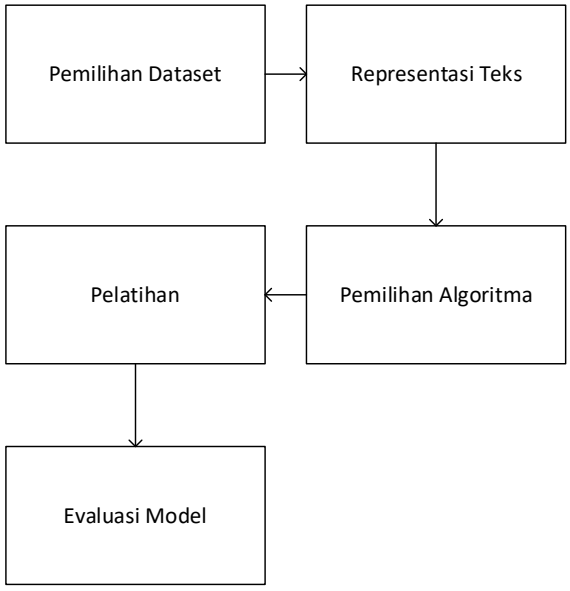
Penelitian ini bertujuan untuk menganalisis pengaruh teknik representasi teks *Bag-of-Words* (BoW) dan *Term Frequency–Inverse Document Frequency* (TF-IDF) terhadap performa klasifikasi sentimen teks dengan pendekatan pembelajaran mesin. Analisis dilakukan melalui serangkaian tahapan, dimulai dari preprocessing data, transformasi teks menggunakan kedua teknik representasi, hingga pelatihan dan evaluasi model klasifikasi. Dua algoritma *machine learning*, yaitu *Complement Naïve Bayes* dan *Ridge Classifier*, digunakan untuk menguji efektivitas masing-masing teknik representasi dalam memetakan fitur teks ke dalam label sentimen. Dataset yang digunakan adalah *Sentiment Analysis Evaluation Dataset*, yang mencakup teks dari empat *genre* berbeda (*Education*, *Finance*, *Politics*, dan *Sports*) dan diklasifikasikan ke dalam dua label sentiment yaitu positif dan negatif. Kontribusi utama penelitian ini adalah mengisi kesenjangan tersebut dengan menyajikan analisis perbandingan langsung dan sistematis antara Bag-of-Words (BoW) dan TF-IDF pada dataset sentimen multi-domain. Secara spesifik, penelitian ini menguji kedua teknik representasi pada dua model klasifikasi yang berbeda—*Ridge Classifier* dan *Complement Naïve Bayes*. Dengan demikian, kontribusi ilmiah yang dihasilkan adalah penyediaan bukti empiris yang kuat mengenai kombinasi representasi fitur dan algoritma mana yang paling optimal, yang hasilnya dapat berfungsi sebagai panduan praktis dan *baseline* yang andal untuk pengembangan sistem serupa di masa depan.

2. Metodologi Penelitian

Penelitian ini menerapkan pendekatan eksperimen kuantitatif. Pendekatan ini dipilih untuk mengukur dan membandingkan secara objektif pengaruh variabel

independen (teknik representasi teks BoW dan TF-IDF) terhadap variabel dependen (akurasi, presisi, *recall*, dan *F1-score*) pada dua model klasifikasi yang berbeda. Sumber data utama dalam eksperimen ini adalah Sentiment Analysis Evaluation Dataset, yang diperoleh dari platform publik Kaggle. Dataset ini tersedia di bawah lisensi CC0 1.0 Universal (Public Domain Dedication), yang memastikannya dapat digunakan secara bebas untuk tujuan akademis tanpa batasan hak cipta. Dataset tersebut dapat diakses melalui tautan berikut https://www.kaggle.com/datasets/prishasawhney/sentiment-analysis-evaluation-dataset. Dataset ini terdiri dari total 209 data teks yang telah diklasifikasikan ke dalam empat kategori domain, yaitu *Education* (52 data), *Finance* (48 data), *Politics* (53 data), dan *Sports* (56 data) [10], [11]. Setiap data dalam dataset memiliki dua atribut utama, yaitu *Text* yang berisi cuplikan kalimat pendek yang mencerminkan isi sentimen dari setiap domain, dan Label yang menunjukkan kategori sentimen berupa *Positive* dan *Negative* [11]. Dataset ini digunakan karena sifatnya yang telah teranotasi dengan baik serta mencakup keragaman domain yang relevan untuk pengujian model klasifikasi sentimen. Selain itu, akses terhadap dataset ini bersifat terbuka dan tidak memerlukan izin khusus untuk penggunaan dalam konteks akademik[11].

Secara garis besar, metodologi penelitian ini mengacu pada alur kerja data mining, dan dapat diselaraskan dengan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahap, yaitu: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment* [12]. Namun, dalam konteks penelitian ini yang berfokus pada eksperimen klasifikasi teks, pendekatan CRISP-DM disesuaikan menjadi tahapan yang lebih praktis, yakni: pemilihan dataset, representasi teks, pemilihan algoritma, pelatihan dan evaluasi model. Gambaran dari tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penelitian ini menerapkan dua teknik representasi fitur teks, yaitu *Bag-of-Words* (BoW) dan *Term Frequency–Inverse Document Frequency* (TF-IDF), untuk mentransformasikan data teks menjadi bentuk numerik yang dapat diproses oleh algoritma klasifikasi berbasis *machine learning*. Representasi BoW menghitung frekuensi kemunculan setiap kata dalam dokumen tanpa mempertimbangkan konteks atau urutan kata. Sebaliknya, TF-IDF mempertimbangkan tingkat kepentingan kata dalam dokumen serta kelangkaannya dalam keseluruhan korpus. Perhitungan bobot representasi TF-IDF dapat dilihat pada Persamaan (1)[13].

$$tfidf_{t,d} = tf_{t,d} \times idf_t \quad (1)$$

dengan,

$tf_{t,d}$: term frekuensi atau banyaknya kata (t) pada dokumen (d) atau pembobotan lokal

idf_t : $\log\left(\frac{N}{df_t}\right)$, di mana N adalah jumlah total dokumen, dan df_t adalah jumlah dokumen yang mengandung term t

Berikut disajikan sebuah ilustrasi perhitungan manual nilai TF-IDF pada salah satu term dalam dokumen tertentu. Asumsikan kita memiliki korpus dengan total $N = 5$ dokumen. Kata "politik" muncul sebanyak 3 kali dalam dokumen D1, dan kata tersebut muncul pada 3 dari 5 dokumen. Maka nilai yang diketahui adalah sebagai berikut:

$$tf(\text{politik}, D1) = 3$$

$$idf(politik) = \log\left(\frac{5}{3}\right) = \log(1{,}67) \approx 0{,}51$$

maka,

$$tfidf(politik, D1) = 3 \times 0{,}51 = 1{,}53$$

Nilai ini menunjukkan bahwa meskipun kata "politik" sering muncul dalam D1, bobotnya tidak terlalu tinggi karena kata tersebut juga muncul cukup sering dalam dokumen lain.

Model klasifikasi yang digunakan dalam penelitian ini adalah *Complement Naïve Bayes* (CNB) dan *Ridge Classifier*. *Complement Naïve Bayes* merupakan varian dari algoritma Naïve Bayes yang dirancang untuk menangani ketidakseimbangan kelas dalam data teks [14]. CNB menggunakan informasi dari kelas pelengkap untuk memperkuat pembedaan antar kelas, sehingga lebih stabil dalam kasus dengan fitur tinggi dan distribusi kelas yang tidak seimbang [15]. Keunggulan CNB telah dibuktikan dalam beberapa studi terkini sebagai model yang unggul dalam klasifikasi teks imbang dan tidak imbang [16], [17]. Perhitungan skor klasifikasi pada CNB dapat dilihat pada Persamaan (2) [18].

$$\hat{y} = \arg\min_y(-\log P(y) + \sum_i f_i \times \log P(w_i \mid \hat{y})) \quad (2)$$

dengan,

$\hat{y}$: kelas hasil prediksi

$P(y)$: probabilitas awal (prior) dari kelas y

f_i : frekuensi kemunculan kata ke-i dalam dokumen

$P(w_i \mid \hat{y})$: probabilitas kata w_i muncul pada kelas pelengkap dari y

Untuk memperjelas bagaimana rumus ini diterapkan dalam praktik klasifikasi teks, berikut disajikan contoh perhitungan manual sederhana. Misalkan sebuah dokumen mengandung tiga kata kunci: data, politik, dan negara, dengan frekuensi berturut-turut yaitu 2, 1, dan 3. Diketahui juga probabilitas awal kelas Positif adalah $P(Positif) = 0{,}6$ dan nilai $\log P(Positif) \approx -0{,}22$. Probabilitas kemunculan masing-masing kata dalam kelas pelengkap dihitung sebagai berikut:

$$\log P(data \mid Negatif) = -0{,}50$$

$$\log P(politik \mid Negatif) = -0{,}70$$

$$\log P(negara \mid Negatif) = -0{,}40$$

maka skor untuk kelas Positif dihitung sebagai:

$$\text{Skor} = -\log(0{,}6) + (2 \times -0{,}50) + (1 \times -0{,}70) + (3 \times -0{,}40)$$

$$Skor = 0{,}22 + (-1{,}00) + (-0{,}70) + (-1{,}20) = 0{,}22 - 2{,}90 = -2{,}68$$

Skor ini akan dibandingkan dengan skor kelas lainnya (misalnya Negatif), dan model akan memilih kelas dengan nilai skor minimum sebagai hasil prediksi akhir.

Sementara itu, *Ridge Classifier* adalah metode klasifikasi berbasis regresi linear dengan penerapan regularisasi L2 untuk mencegah *overfitting* [19]. Model ini bekerja dengan mengestimasi bobot parameter berdasarkan minimisasi fungsi loss yang menggabungkan galat kuadrat dan penalti terhadap bobot besar. *Ridge Classifier* cocok untuk data dengan jumlah fitur tinggi dan korelasi antar fitur, yang merupakan karakteristik umum dalam data teks [20]. Keandalannya dalam kondisi multikolinearitas juga telah dibuktikan dalam penelitian-penelitian terbaru [19], [20]. Perhitungan fungsi *loss Ridge Classifier* dapat dilihat pada Persamaan (3) [20].

$$J(w) = \sum_{i=1}^{n} (y_i - x_i^T w)^2 + \lambda \|w\|_2^2 \tag{3}$$

dengan,

y_i : label target dari dokumen ke-i
x_i : vektor fitur dari dokumen ke-i
w : bobot parameter yang dikalibrasi selama pelatihan
λ : parameter regularisasi yang menentukan seberapa besar penalty diberikan terhadap bobot tinggi

Untuk memberikan ilustrasi penerapan rumus tersebut, berikut disajikan contoh perhitungan sederhana. Misalkan terdapat sebuah dokumen dengan vektor fitur $x_i = [2,\ 1]$, bobot model awal $w = [0{,}5;\ -0{,}3]$, dan target kelas $y_i = 1$. Asumsikan juga parameter regularisasi $\lambda = 0{,}1$.

Langkah 1: Hitung prediksi linier

$$x_i^T w = (2 \times 0{,}5) + (1 \times -0{,}3) = 1{,}0 - 0{,}3 = 0{,}7$$

Langkah 2: Hitung galat kuadrat

$$(y_i - x_i^T w)^2 = (1 - 0{,}7)^2 = 0{,}09$$

Langkah 3: Hitung penalty L2

$$\|w\|_2^2 = (0{,}5)^2 + (-0{,}3)^2 = 0{,}25 + 0{,}09 = 0{,}34$$

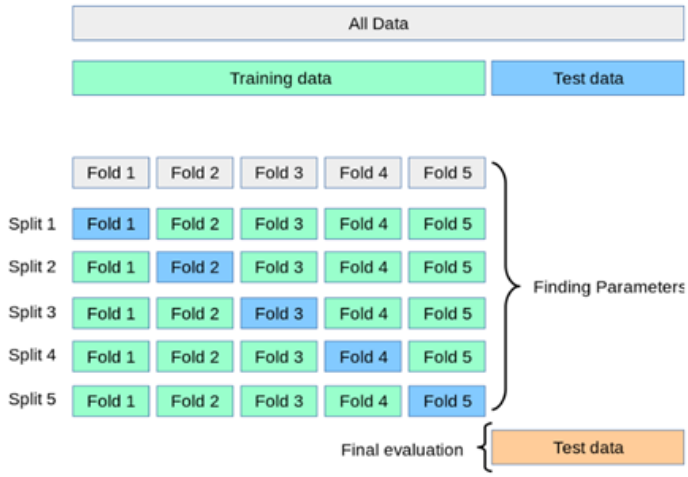
Langkah 4: Hitung total loss

$$J(w) = 0{,}09 + (0{,}1 \times 0{,}34) = 0{,}09 + 0{,}034 = 0{,}124$$

Nilai loss $J(w) = 0{,}124$ ini digunakan dalam proses optimisasi untuk memperbarui bobot w selama pelatihan. *Ridge Classifier* akan meminimalkan nilai ini iteratif, sehingga model memiliki generalisasi yang lebih baik dengan penalti terhadap bobot besar.

Performa dari kedua model dievaluasi menggunakan teknik Stratified K-Fold Cross-Validation dengan nilai $k = 5$. Pemilihan nilai $k = 5$ ini didasarkan pada praktik umum dalam *machine learning* yang bertujuan untuk mencapai keseimbangan yang baik antara

bias dan varians (*bias-variance trade-off*) dalam estimasi performa model [21], [22]. Nilai ini dianggap mampu memberikan estimasi kinerja yang andal tanpa menimbulkan biaya komputasi yang berlebihan, dan merupakan pilihan yang solid untuk ukuran dataset dalam penelitian ini ($N = 418$). Metode ini membagi dataset menjadi lima subset dengan distribusi label sentimen yang seimbang pada setiap fold [23]. Dalam setiap iterasi, empat subset digunakan sebagai data pelatihan dan satu sebagai data pengujian, sehingga seluruh data dimanfaatkan secara menyeluruh baik untuk pelatihan maupun validasi. Hasil evaluasi dihitung dengan mengambil rata-rata dari lima iterasi untuk mendapatkan estimasi performa yang lebih stabil dan representatif. Teknik ini juga mengurangi risiko *overfitting* karena model diuji terhadap berbagai kombinasi data [23], [24]. Skema dari Teknik ini dapat dilihat pada Gambar 2.



Gambar 2. Skema *Stratified K-Fold cross validation*

Berbeda dengan K-Fold standar yang tidak mempertimbangkan proporsi kelas, Stratified K-Fold memastikan bahwa setiap fold memiliki distribusi kelas yang sama dengan dataset awal [24], [25]. Pendekatan ini dinilai lebih unggul dalam menangani permasalahan ketidakseimbangan kelas, terutama dalam konteks klasifikasi teks seperti pada penelitian ini [24]. Stratified K-Fold juga mempertahankan keunggulan K-Fold dalam memanfaatkan data secara efisien, namun dengan tambahan stabilitas terhadap distribusi kelas[23]–[25].

Penilaian performa dilakukan berdasarkan empat metrik utama yaitu akurasi, *precision*, *recall*, dan *F1-Score*. Keempat metrik ini dipilih karena memberikan evaluasi menyeluruh terhadap performa klasifikasi pada kasus biner, khususnya ketika distribusi label tidak merata. Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data, sebagaimana ditunjukkan pada Persamaan (4) [26].

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{4}$$

Sebagai contoh, misalkan pada satu iterasi validasi diperoleh hasil klasifikasi sebagai berikut:

$$TP = 90, \quad TN = 85, \quad FP = 10, \quad FN = 10$$

maka nilai akurasi dapat dihitung sebagai:

$$Accuracy = \frac{90 + 85}{90 + 85 + 10 + 15} = \frac{175}{200} = 0{,}875$$

Nilai ini menunjukkan bahwa sebanyak 0,875 prediksi yang dilakukan oleh model sudah sesuai dengan label aktual pada data pengujian. Nilai akurasi seperti ini sangat berguna untuk memberikan gambaran umum terhadap tingkat kesalahan total dalam prediksi, namun perlu dikombinasikan dengan metrik lainnya (seperti precision dan recall) untuk interpretasi yang lebih seimbang, terutama pada kasus data tidak seimbang[26].

Precision mengukur tingkat ketepatan dari prediksi model terhadap kelas positif [27]. Dalam konteks klasifikasi sentimen, *precision* menunjukkan seberapa banyak dari seluruh prediksi yang dinyatakan positif oleh model ternyata benar-benar merupakan data dengan sentimen positif [28]. Metrik ini menjadi penting ketika kesalahan dalam memberikan prediksi positif harus diminimalkan, seperti dalam kasus opini publik yang berdampak terhadap keputusan. Perhitungan *precision* dapat dilihat pada Persamaan (5) [26].

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

sebagai ilustrasi, misalkan dalam suatu evaluasi model didapatkan:

$$TP = 90, \qquad FP = 10$$

maka nilai precision dapat dihitung sebagai berikut:

$$Precision = \frac{90}{90 + 10} = \frac{90}{100} = 0{,}90$$

Artinya, dari seluruh prediksi yang dinyatakan positif oleh model, sebesar 0,90 benar-benar merupakan data yang memiliki label sentimen positif. Metrik ini sangat krusial apabila kesalahan memberikan label positif dapat menyebabkan implikasi serius, seperti dalam klasifikasi opini publik, spam filtering, atau sistem rekomendasi [26].

Recall, atau dikenal juga sebagai sensitivitas, mengukur sejauh mana model mampu mengenali seluruh data yang benar-benar positif dari keseluruhan data positif yang ada [1]. Dalam kasus klasifikasi sentimen, *recall* membantu mengetahui apakah model berhasil menangkap semua opini positif yang relevan [5]. Nilai *recall* yang tinggi menunjukkan bahwa sedikit data positif yang luput terdeteksi oleh model. Perhitungan *recall* dapat dilihat pada Persamaan (6) [26].

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

sebagai contoh, misalkan dari data pengujian diketahui:

$$TP = 90, \qquad FN = 15$$

maka recall dihitung sebagai berikut:

$$Recall = \frac{90}{90 + 15} = \frac{90}{105} \approx 0{,}86$$

Artinya, model berhasil mengenali sekitar 0,86 dari seluruh data yang benar-benar positif. Dalam konteks klasifikasi sentimen, nilai recall yang tinggi menunjukkan bahwa opini-opini penting yang bersentimen positif tidak banyak terlewatkan oleh sistem.

F1-Score merupakan metrik gabungan yang menghitung rata-rata harmonis antara *precision* dan *recall* [9]. Metrik ini sangat berguna ketika dibutuhkan keseimbangan antara

ketepatan dan kelengkapan dalam deteksi kelas positif, terutama dalam kondisi di mana data tidak seimbang [2]. *F1-Score* memberikan gambaran menyeluruh mengenai performa model dalam menangani prediksi positif secara akurat dan menyeluruh. Perhitungan *F1-Score* dapat dilihat pada Persamaan (7) [26].

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \tag{7}$$

sebagai contoh, jika diketahui hasil evaluasi model menunjukkan:

$$precision = 0{,}90, \qquad recall = 0{,}86$$

maka nilai F1-Score dapat dihitung sebagai:

$$F1 - Score = 2 \times \frac{0{,}90 \times 0{,}86}{0{,}90 + 0{,}86} = 2 \times \frac{0{,}77}{1{,}76} \approx 2 \times 0{,}44 = 0{,}88$$

Dengan demikian, F1-Score memberikan nilai 0,88 yang mencerminkan keseimbangan antara kemampuan model dalam mengenali data positif (recall) dan menghasilkan prediksi positif yang akurat (precision). Metrik ini sangat relevan ketika kedua aspek tersebut sama-sama penting untuk diperhatikan dalam evaluasi performa klasifikasi [26].

Dalam konteks evaluasi ini, TP (True Positive) merujuk pada data positif yang diklasifikasikan benar, TN (True Negative) adalah data negatif yang diklasifikasikan benar, FP (False Positive) merupakan data negatif yang salah diklasifikasikan sebagai positif, dan FN (False Negative) adalah data positif yang diklasifikasikan salah sebagai negatif [26]. Penjabaran metrik ini penting untuk memastikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga sensitif dan tepat dalam mengidentifikasi sentimen yang relevan.

Seluruh proses eksperimen dilakukan menggunakan bahasa pemrograman Python (versi 3.10) di platform Google Colaboratory. Kerangka kerja eksperimen ini didukung oleh beberapa pustaka kunci, termasuk scikit-learn (versi 1.6.1) untuk implementasi model *Complement Naïve Bayes* dan *Ridge Classifier*. Untuk pemrosesan teks, augmentasi data memanfaatkan NLTK (versi 3.9.1) dengan korpus WordNet, sementara proses *lemmatization* dilakukan menggunakan spaCy (versi 3.8.7). Visualisasi hasil evaluasi, seperti *heatmap confusion matrix* dan grafik performa, dibuat menggunakan Seaborn (versi 0.13.2) dan Matplotlib (versi 3.10.0). Penggunaan versi pustaka yang spesifik ini bertujuan untuk menjamin reprodusibilitas dan keandalan hasil penelitian.

3. Hasil

Evaluasi dilakukan untuk mengukur kinerja dari empat konfigurasi eksperimen yang menggabungkan dua metode representasi fitur teks, *Bag-of-Words* (BoW) dan *Term Frequency–Inverse Document Frequency* (TF-IDF), dengan dua algoritma klasifikasi, yaitu *Ridge Classifier* dan *Complement Naïve Bayes* (CNB). Prosedur validasi menggunakan teknik *Stratified K-Fold Cross-Validation* sebanyak lima lipatan ($k = 5$) untuk menjamin distribusi kelas tetap seimbang pada setiap iterasi. Eksperimen diawali dengan dataset asli yang terdiri dari 209 teks unik, dengan distribusi label sebanyak 115 sentimen positif dan 94 sentimen negatif. Teks-teks ini berasal dari empat domain berbeda, *Education* (52 data), *Finance* (48 data), *Politics* (53 data), dan *Sports* (56 data). Untuk memperkaya data pelatihan dan meningkatkan generalisasi model, dilakukan augmentasi data berbasis sinonim, sehingga total dataset yang digunakan dalam evaluasi menjadi 418 entri teks. Setiap teks kemudian melalui tahap pra-pemrosesan untuk standardisasi. Proses ini meliputi *case folding*, pembersihan karakter non-alfanumerik, dan *lemmatization* untuk mengubah setiap

kata ke bentuk dasarnya. Contoh transformasi yang dilakukan pada data teks dapat dilihat pada Tabel 1.

Tabel 1. Contoh Hasil Pra-pemrosesan Teks

Teks Asli	Hasil Case Folding & Cleaning	Hasil Lemmatization
healthy scientist health irrationallyare choose find scientist rationally behave people help environment design spend effort life rational live effort despite live irrationality study nt do scientist behavioral	healthy scientist health irrationallyare choose find scientist rationally behave people help environment design spend effort life rational live effort despite live irrationality study not scientist behavioral	healthy scientist health irrationallyare choose find scientist rationally behave people help environment design spend effort life rational live effort despite live irrationality study not scientist behavioral
cross good nt would save good trafford old goal play nt wo decide ve i glove want character bubbly dubai goal add week east middle break midseason club session training stick impress read goal play chance supersub tell mcdermott start dream likely season goal 12 topscorer royal fondre le adam fan united mindset different forever think nt do young re you day single appreciate look appreciate aways trafford old walk opportunity appreciate great attitude player case year happen thing think young re you united manchester manage m i old m i player mean game think funny company good s he time lot guy s he incredible take club figurehead s he right ferguson alex sir stein jock shankly bill lot read ve i man know need tell opposition want say stoke game come sum enthusiastic bos legendary facetoface come opportunity relish ferguson alex sir height reach unlikely know 51 mcdermott man polite museum say ask 12 say blase medal ask quiet humble hand shake come impressed mcdermott say manager ingredient get prolicence burton recently meet tie round fifth cup fa tonight trafford old struggler league premier take eve winger 39yearold speak manager reading dugout room dressing gap bridge player united manchester giggs ryan tip mcdermott brian	cross good not would save good trafford old goal play not will decide glove want character bubbly dubai goal add week east middle break midseason club session training stick impress read goal play chance supersub tell mcdermott start dream likely season goal topscorer royal fondre le adam fan united mindset different forever think not young day single appreciate look appreciate away trafford old walk opportunity appreciate great attitude player case year happen thing think young united manchester manage old player mean game think funny company good time lot guy incredible take club figurehead right ferguson alex sir stein jock shankly bill lot read man know need tell opposition want say stoke game come sum enthusiastic bos legendary facetoface come opportunity relish ferguson alex sir height reach unlikely know mcdermott man polite museum say ask say blase medal ask quiet humble hand shake come impressed mcdermott say manager ingredient get prolicence burton recently meet tie round fifth cup fa tonight trafford old struggler league premier take eve winger yearold speak manager read dugout room dress gap bridge player united manchester giggs ryan tip mcdermott brian	cross good not would save good trafford old goal play not will decide glove want character bubbly dubai goal add week east middle break midseason club session training stick impress read goal play chance supersub tell mcdermott start dream likely season goal topscorer royal fondre le adam fan united mindset different forever think not young day single appreciate look appreciate away trafford old walk opportunity appreciate great attitude player case year happen thing think young united manchester manage old player mean game think funny company good time lot guy incredible take club figurehead right ferguson alex sir stein jock shankly bill lot read man know need tell opposition want say stoke game come sum enthusiastic bos legendary facetoface come opportunity relish ferguson alex sir height reach unlikely know mcdermott man polite museum say ask say blase medal ask quiet humble hand shake come impressed mcdermott say manager ingredient get prolicence burton recently meet tie round fifth cup fa tonight trafford old struggler league premier take eve winger yearold speak manager read dugout room dress gap bridge player united manchester giggs ryan tip mcdermott brian

Setelah teks bersih, tahap selanjutnya adalah transformasi menjadi fitur numerik menggunakan BoW dan TF-IDF. Kedua metode ini menghasilkan ruang fitur dengan dimensi yang sama, yaitu sebanyak 4.890 fitur unik yang menjadi representasi kosakata dari keseluruhan korpus, seperti dirangkum pada Tabel 2.

Tabel 2. Hasil Ekstraksi Fitur

Vectorizer	Jumlah Fitur Unik
BoW	4890
TF-IDF	4890

Dengan data yang telah siap dalam bentuk fitur numerik, evaluasi sistematis dilakukan terhadap keempat konfigurasi model berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-Score*. Seluruh nilai yang dilaporkan pada Tabel 3 merupakan rata-rata dari lima kali pelaksanaan validasi silang untuk memperoleh estimasi kinerja yang stabil dan representatif. Nilai tertinggi pada masing-masing metrik diberi penekanan untuk menunjukkan konfigurasi yang paling optimal.

Tabel 3. Hasil Evaluasi Perbandingan Model Klasifikasi Menggunakan BoW dan TF-IDF

Model	*Vectorizer*	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Ridge Classifier	BoW	0.906	0.917	0.899	0.904
Complement Naïve Bayes	BoW	0.861	0.881	0.850	0.855
Ridge Classifier	TF-IDF	**0.911**	**0.923**	**0.905**	**0.908**
Complement Naïve Bayes	TF-IDF	0.907	0.909	0.903	0.905

Sebagai pelengkap dari hasil evaluasi tersebut, berikut disajikan contoh perhitungan manual pada salah satu konfigurasi model, yaitu *Ridge Classifier dengan TF-IDF*, guna memperjelas proses evaluasi metrik. Dalam salah satu fold evaluasi, diperoleh confusion matrix sebagai berikut: $TP = 226$, $TN = 155$, $FP = 4$, $FN = 33$. Berdasrkan nilai-nilai ini, metrik evaluasi dihitung sebagai berikut:

- Accuracy

$$\frac{TP\ + TN}{(TP + TN + FP + FN)} = \frac{226 + \ 155}{418} = \frac{381}{418} = \ 0{,}911$$

- Precision

$$\frac{TP}{(TP + FP)} = \frac{226}{230} \approx 0{,}983$$

- Recall

$$\frac{TP}{(TP + FN)} = \frac{226}{259} \approx 0{,}872$$

- F1-Score

$$2 \times \frac{0{,}923 \times 0{,}872}{0{,}923 + 0{,}872} \approx 0{,}897$$

Sebagai catatan nilai metrik yang diperoleh dari perhitungan manual ini tidak identik dengan nilai rata-rata dalam Tabel 3, karena hanya diambil dari satu fold evaluasi, bukan dari keseluruhan hasil Stratified K-Fold Cross-Validation. Dengan demikian, contoh ini digunakan semata-mata untuk ilustrasi cara kerja evaluasi kuantitatif model, dan bukan sebagai representasi mutlak dari performa keseluruhan. Penekanan pada evaluasi manual ini juga mempertegas pentingnya proses validasi dalam memastikan bahwa setiap metrik yang dihasilkan dapat ditelusuri, dijelaskan, dan digunakan sebagai dasar pengambilan keputusan berbasis data.

4. Pembahasan

Penelitian ini difokuskan untuk mengevaluasi efektivitas teknik representasi teks *Bag-of-Words* (BoW) dan *Term Frequency–Inverse Document Frequency* (TF-IDF) dalam meningkatkan performa klasifikasi sentimen, khususnya ketika diterapkan pada data teks pendek dari berbagai domain. Berdasarkan hasil yang diperoleh dari eksperimen, ditemukan bahwa pemilihan metode representasi teks memiliki pengaruh yang signifikan terhadap performa algoritma klasifikasi yang digunakan. Temuan ini juga diperkuat oleh bukti empiris dari studi-studi terdahulu [2], [29]–[32] yang menunjukkan bahwa pemilihan representasi fitur dapat menentukan efisiensi model secara keseluruhan dalam konteks klasifikasi teks.

4.1. Pengaruh Representasi Teks terhadap Performa Klasifikasi

Dari hasil evaluasi, terlihat bahwa teknik TF-IDF secara konsisten unggul dibandingkan BoW pada seluruh metrik yang diuji. Sebagai contoh, pada model *Ridge Classifier*, nilai *F1-Score* meningkat dari 0,903 menjadi 0,908 saat beralih dari BoW ke TF-IDF. Peningkatan yang lebih substansial terlihat pada model *Complement Naïve Bayes*, dari 0,855 menjadi 0,905. Keunggulan performa ini dapat dijelaskan dari cara kedua teknik merepresentasikan teks. Metode BoW memiliki keterbatasan fundamental, yaitu hanya mengandalkan frekuensi kata tanpa mempertimbangkan relevansi atau distribusinya di seluruh dokumen. Akibatnya, kata-kata umum yang sering muncul namun tidak memiliki muatan sentimen (misalnya, "yang", "pada", "dengan") bisa mendapatkan bobot yang sama tingginya dengan kata kunci sentimen, sehingga berpotensi menjadi *noise* yang mengganggu performa model klasifikasi. Sebaliknya, TF-IDF dirancang untuk mengatasi masalah ini. Dengan mengalikan *Term Frequency* (TF) dengan *Inverse Document Frequency* (IDF), TF-IDF mampu secara efektif menurunkan bobot dari kata-kata umum yang kurang informatif dan pada saat yang sama meningkatkan bobot dari kata-kata unik yang lebih relevan terhadap sentimen suatu teks. Pengaruh ini sejalan dengan temuan oleh [31], yang menjelaskan bahwa TF-IDF merupakan salah satu teknik vektorisasi yang efektif dalam mengatasi ambiguitas semantik pada representasi teks pendek dan mendukung konteks klasifikasi berbasis domain [31].

4.2. Perbandingan Dua Model Terhadap Representasi

Secara umum, *Ridge Classifier* menghasilkan performa tertinggi di antara kedua algoritma yang diuji, dengan kombinasi TF-IDF menghasilkan skor terbaik pada seluruh metrik. Model ini secara teoritis lebih cocok untuk data berdimensi tinggi seperti representasi teks karena menggunakan regularisasi L2 yang efektif untuk mencegah *overfitting*, sehingga membuatnya lebih kokoh (*robust*) terhadap variasi fitur. Hasil ini juga didukung oleh studi oleh [33] yang menunjukkan bahwa *Ridge Classifier* menjadi salah satu model terbaik dalam deteksi ujaran kebencian multibahasa dengan skor F1 mencapai 0,890.

Hal ini berbeda dengan pendekatan *Complement Naïve Bayes* (CNB), yang beroperasi di bawah asumsi independensi fitur (*feature independence*) yang kuat. CNB menganggap setiap kata dalam dokumen bersifat independen satu sama lain, sebuah asumsi yang jarang terpenuhi dalam bahasa alami di mana konteks sangat penting. Keterbatasan teoretis ini menjelaskan mengapa CNB menunjukkan sensitivitas yang lebih tinggi terhadap jenis representasi yang digunakan. Seperti yang ditunjukkan oleh hasil eksperimen, model ini mengalami peningkatan performa yang lebih drastis ketika beralih dari BoW ke TF-IDF, membuktikan bahwa metode probabilistik ini sangat bergantung pada kualitas fitur yang diskriminatif. Tren ini juga diamati oleh [34], yang mencatat bahwa *Complement Naïve Bayes* dapat mengungguli model lain pada data sentimen media sosial ketika dipasangkan dengan TF-IDF, dengan akurasi mencapai 0,820 dan F1-Score sebesar 0,810.

4.3. Analisis terhadap Data Multi-Domain

Dataset yang digunakan mencakup teks dari empat domain berbeda—Education, Finance, Politics, dan Sports—yang masing-masing memiliki karakteristik linguistik dan semantik yang unik. Dalam konteks ini, pemilihan teknik representasi menjadi sangat penting untuk memastikan model dapat menyesuaikan diri terhadap variasi konteks. Metode BoW, yang hanya menghitung frekuensi kata tanpa mempertimbangkan relevansi atau distribusi globalnya, gagal membedakan makna kontekstual antar domain. Misalnya, kata "score" dapat bermakna nilai akademik dalam domain Education atau hasil pertandingan dalam domain Sports.

Sebaliknya, TF-IDF memberikan bobot lebih tinggi pada kata-kata yang memiliki signifikansi kontekstual tinggi tetapi distribusi global rendah. Ini membuat representasi TF-IDF lebih adaptif dalam menangkap perbedaan terminologi antar domain. Efektivitas pendekatan ini juga tercermin dalam penelitian oleh [34], yang menunjukkan bahwa dalam pengklasifikasian sentimen politik di media sosial, TF-IDF menghasilkan performa yang lebih baik daripada BoW untuk model berbasis Naïve Bayes dan SVM [34].

5. Kesimpulan

Berdasarkan evaluasi sistematis yang telah dilakukan, penelitian ini menegaskan bahwa pemilihan teknik representasi fitur secara signifikan memengaruhi performa klasifikasi sentimen pada data multi-domain. Secara khusus, representasi TF-IDF terbukti secara konsisten lebih efektif dibandingkan BoW dalam menangkap karakteristik semantik yang relevan dari teks pendek. Konfigurasi terbaik dalam penelitian ini diraih oleh model Ridge Classifier yang dipadukan dengan TF-IDF, yang menunjukkan performa paling unggul dengan akurasi 0,911 dan F1-Score tertinggi sebesar 0,908. Dengan temuan ini, kombinasi tersebut dapat direkomendasikan sebagai pendekatan *baseline* yang andal, efisien, dan adaptif untuk tugas klasifikasi sentimen lintas domain.

Meskipun penelitian ini telah berhasil membandingkan efektivitas BoW dan TF-IDF secara sistematis pada konteks multi-domain, terdapat keterbatasan utama yang perlu diakui. Penelitian ini masih terbatas pada penerapan model klasifikasi *machine learning* konvensional. Oleh karena itu, sebagai arah penelitian lanjutan, terdapat peluang besar untuk meningkatkan performa dengan mengeksplorasi arsitektur *deep learning* yang lebih canggih. Studi di masa depan dapat diarahkan pada penerapan model seperti *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM), atau model berbasis *transformer* seperti BERT, yang berpotensi menangkap nuansa semantik yang lebih dalam dan kompleks pada data teks.

Referensi

[1] C. . Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Comput. Sci.*, vol. 234, pp. 156–163, 2024, https://doi.org/10.1016/j.procs.2024.02.162.

[2] A. Y. Sain, S. A. S. Mola, A. Y. Huan, and I. R. Nomleni, "Analisis Sentimen Sunscreen Azarine dengan Naïve Bayes di Toko Aneka Kosmetik Kupang pada Marketplace Shopee," *J. SAINTEKOM*, vol. 15, no. 1, pp. 94–105, Mar. 2025, https://doi.org/10.33020/saintekom.v15i1.783.

[3] F. Greco, "Sentiment analysis and opinion mining," in *Elgar Encyclopedia of Technology and Politics*, vol. 4, no. 27, Edward Elgar Publishing, 2022, pp. 105–108. https://hdl.handle.net/11573/1616253

[4] A. Iyengar, S. Divyashree, N. Huda, S. Katira, and K. Jitendra Vasudevshahapur, "Insight into Sentimental Analysis," *JETIR2006559 J. Emerg. Technol. Innov. Res.*, vol. 7, no. 6, pp. 1561–1566, 2020, https://www.jetir.org/papers/JETIR2006559.pdf.

[5] A. Ananta Firdaus, A. Id Hadiana, and A. Kania Ningsih, "Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest," *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 6, no. 5, pp. 1678–1683, Jul. 2024, https://doi.org/10.38035/rrj.v6i5.994.

[6] J. Yuan, Y. Zhao, and B. Qin, "Learning to share by masking the non-shared for multi-domain sentiment classification," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 9, pp. 2711–2724, 2022, https://doi.org/10.1007/s13042-022-01556-0.

[7] J. Lu, "Text vectorization in sentiment analysis: A comparative study of TF-IDF and Word2Vec from Amazon Fine Food Reviews," *ITM Web Conf.*, vol. 70, no. 03001, p. 03001, Jan. 2025, https://doi.org/10.1051/itmconf/20257003001.

[8] Z. Zhan, "Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis : A Case of Food Reviews," *ITM Web Conf.*, vol. 70, no. 02013, pp. 1–7, 2025, https://doi.org/10.1051/itmconf/20257002013.

[9] M. S. Kobari, N. Karimi, B. Pourhosseini, and R. Mousa, "weighted CapsuleNet networks for Persian multi-domain sentiment analysis," *arXiv.org*, vol. 2306, no. 17068, pp. 1–20, Jul. 2023, https://doi.org/10.48550/arXiv.2306.17068.

[10] Allanatrix, "How to Use Sentiment Analysis Evaluation Dataset," *Kaggle*, 2025. https://www.kaggle.com/code/allanwandia/how-to-use-sentiment-analysis-evaluation-dataset/.

[11] Prishasawhney, "sentiment-analysis-evaluation-dataset," *Kaggle*, 2025. https://www.kaggle.com/datasets/prishasawhney/sentiment-analysis-evaluation-dataset.

[12] A. Muzaki and S. Agustin, "Sistem Informasi Agenda Rapat Berbasis Web untuk Optimalisasi Kinerja Dinas Kominfo Lamongan," *remik*, vol. 9, no. 1, pp. 161–174, Jan. 2025, https://doi.org/10.33395/remik.v9i1.14366.

[13] V. Meida Hersianty, E. Larasati Amalia, D. Puspitasari, and D. Wahyu Wibowo, "Penerapan Algoritma Tf-Idf Dan Cosine Similarity Dalam Sistem Rekomendasi Lowongan Pekerjaan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1619–1625, Jan. 2025, https://doi.org/10.36040/jati.v9i1.12406.

[14] T. I. Simanjuntak, M. Muhathir, F. Fadlisyah, and I. Safira, "Performance Analysis of Naive Bayes Variation Method in Spice Image Classification Using Histogram of Gradient Oriented (HOG) Feature Extraction," *J. Informatics Telecommun. Eng.*, vol. 7, no. 1, pp. 282–291, 2023, https://doi.org/10.31289/jite.v7i1.7957.

[15] H. D. Cahyono, A. Mahadewa, A. Wijayanto, D. W. Wardani, and H. Setiadi, "Fast Naïve Bayes classifiers for COVID-19 news in social networks," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 34, no. 2, pp. 1033–1041, 2024, https://doi.org/10.11591/ijeecs.v34.i2.pp1033-1041.

[16] A. Budiman, J. C. Young, and A. Suryadibrata, "Implementasi Algoritma Naïve Bayes untuk Klasifikasi Konten Twitter dengan Indikasi Depresi," *J. Inform. J. Pengemb. IT*, vol. 6, no. 2, pp. 133–138, 2021, https://doi.org/10.30591/jpit.v6i2.2419.

[17] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. Alzami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia," *J. Masy. Inform.*, vol. 13, no. 2, pp. 85–98, 2022, https://doi.org/10.14710/jmasif.13.2.47983.

[18] B. Wicaksono and N. Cahyono, "Analisis Sentimen Komentar Instagram Pada Program Kampus Merdeka Dengan Algoritma Naive Bayes Dan Decision Tree," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2372–2381, 2024, https://doi.org/10.36040/jati.v8i2.9473.

[19] S. Sasidharan Nair and M. Subaji, "Automated Identification of Breast Cancer Type Using Novel Multipath Transfer Learning and Ensemble of Classifier," *IEEE Access*, vol. 12, pp. 87560–87578, 2024, https://doi.org/10.1109/ACCESS.2024.3415482.

[20] T. M. Umar, "Klasifikasi teks multilabel mengenai bencana alam pada media sosial twitter menggunakan ridge classifier," Universitas Islam Negeri Syarif Hidayatullah Jakarta, 2023. https://repository.uinjkt.ac.id/dspace/handle/123456789/66805

[21] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," Nov. 2020, http://arxiv.org/abs/1811.12808.

[22] M. B. - and D. B. B. -, "A Comprehensive Review of Cross-Validation Techniques in Machine Learning," *Int. J. Sci. Technol.*, vol. 16, no. 1, pp. 1–4, Jan. 2025, https://doi.org/10.71097/IJSAT.v16.i1.1305.

[23] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *J. Algoritm.*, vol. 21, no. 1, pp. 239–248, 2024, https://doi.org/10.33364/algoritma/v.21-1.1618.

[24] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Front. Nanotechnol.*, vol. 4, no. August, pp. 1–12, 2022, https://doi.org/10.3389/fnano.2022.972421.

[25] M. K. Mayangsari, I. Syarif, and A. Barakbah, "Evaluation of Stratified K-Fold Cross Validation for Predicting Bug Severity in Game Review Classification," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 277–288, Jul. 2023, https://doi.org/10.22219/kinetik.v8i3.1740.

[26] G. H. Martono and N. Sulistianingsih, "Perbandingan Matriks Jarak pada Algoritma K-NN untuk Prediksi Penyakit Diabetes," *JoMI J. Millenn. Informatics*, vol. 2, no. 1, pp. 1–6, 2024, https://journal.mudaberkarya.id/index.php/JoMI/article/view/110.

[27] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, https://doi.org/10.35134/komtekinfo.v10i1.330.

[28] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562.

[29] N. N. Hidayati, "Improving Aspect-Based Sentiment Analysis for Hotel Reviews with Latent Dirichlet Allocation and Machine Learning Algorithms," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 9, no. 2, pp. 144–157, 2023, https://doi.org/10.26594/register.v9i2.3441.

[30] A. Pratama, R. I. Alhaqq, and Y. Ruldeviyani, "Sentiment Analysis of the Covid-19 Booster Vaccination Program As a Requirement for Homecoming During Eid Fitr in Indonesia," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 1, pp. 248–261, 2023, https://api.semanticscholar.org/CorpusID:264346213.

[31] S. Akhtar, "Machine Learning in Business Analytics: Advancing Statistical Methods for Data- Driven Innovation," *J. Comput. Sci. Technol. Stud.*, pp. 104–111, 2025, https://doi.org/10.32996/jcsts.2023.5.3.8.

[32] A. Zamsuri, S. Defit, and G. W. Nurcahyo, "Classification Of Multiple Emotions In Indonesian Text Using The K-Nearest Neighbor Method," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 2, pp. 1012–1021, 2023, https://doi.org/10.37385/jaets.v4i2.1964.

[33] K. Shanmugavadivel, M. Subramanian, S. Shahidkhan, S. S. S, and S. Yashica, "KEC _ AI _ GRYFFINDOR @ Dravidian-LangTech 2025 : Multimodal Hate Speech Detection in Dravidian languages," *Proc. ofthe Fifth Work. Speech, Vision, Lang. Technol. Dravidian Lang.*, pp. 269–273, 2025, https://aclanthology.org/2025.dravidianlangtech-1.31/.

[34] A. Zamsuri, S. D, and E. Asril, "Deteksi Hate Speech Pada Pemilu 2024 Menggunakan Algoritma Machine Learning," *Zo. J. Sist. Inf.*, vol. 7, no. 1, pp. 228–241, 2024, https://journal.unilak.ac.id/index.php/zn/article/view/22049.