

Optimalisasi BoW Feature extraction pada Natural Language Processing untuk Klasifikasi Emosi Teks

Optimization of Feature extraction in Natural Language Processing for Text Emotion Classification

Bondan Wahyu Pamekas¹, Aprilisa Arum Sari^{*2}

^{1,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Duta Bangsa
Surakarta

e-mail: bondan_wahyupamekas@udb.ac.id , *2aprilisa_arumsari@udb.ac.id

Abstrak - Klasifikasi emosi teks merupakan salah satu bidang penting dalam Natural Language Processing (NLP) yang bertujuan mengenali kondisi afektif seseorang berdasarkan teks. Tahap utama dalam proses ini adalah feature extraction, yaitu mengubah teks menjadi representasi numerik untuk diproses algoritma machine learning. Penelitian ini bertujuan mengoptimalkan feature extraction dengan membandingkan tiga skema, yaitu Bag-of-Words (BoW), TF-IDF unigram, dan TF-IDF unigram-bigram, serta mengevaluasi pengaruh parameter max_features dan min_df terhadap akurasi dan efisiensi komputasi. Penelitian menggunakan metodologi CRISP-DM pada dataset berisi 102.000 teks berbahasa Inggris dari Hugging Face dengan lima kelas emosi seimbang (Anger, Anxiety, Calm, Depression, dan Joy). Tiga algoritma klasifikasi, yaitu Multinomial Naïve Bayes, Logistic Regression, dan Linear Support Vector Machine (SVM), diuji pada data yang telah disisipi noise berupa kesalahan ketik dan filler words. Hasil menunjukkan bahwa TF-IDF unigram-bigram dengan 2.000 fitur dan min_df=3 yang dipadukan dengan Linear SVM menghasilkan F1-score makro tertinggi sebesar 0,9999. Sementara itu, BoW unigram memberikan waktu ekstraksi fitur tercepat dengan penurunan akurasi yang sangat kecil. Optimalisasi parameter feature extraction terbukti meningkatkan efisiensi waktu ekstraksi hingga 45% tanpa mengurangi akurasi secara signifikan.

Kata kunci – Bag-of-Words; CRISP-DM; klasifikasi emosi; Natural Language Processing; TF-IDF

Abstract - Text emotion classification is an important task in Natural Language Processing (NLP) that aims to identify a person's affective state from written text. A key stage in this process is feature extraction, which transforms text into numerical representations suitable for machine learning algorithms. This study aims to optimize feature extraction by comparing three text representation schemes: Bag-of-Words (BoW), TF-IDF unigram, and TF-IDF unigram-bigram, while evaluating the effects of max_features and min_df on classification accuracy and computational efficiency. The study follows the CRISP-DM methodology using a dataset of 102,000 English-language texts from Hugging Face with five balanced emotion classes: Anger, Anxiety, Calm, Depression, and Joy. Three classification algorithms—Multinomial Naïve Bayes, Logistic Regression, and Linear Support Vector Machine (SVM)—were evaluated on data augmented with noise, including typographical errors and filler words. The results show that TF-IDF unigram-bigram with 2,000 features and min_df=3, combined with Linear SVM, achieved the highest macro F1-score of 0.9999. Meanwhile, BoW unigram provided the fastest feature extraction time with only a negligible reduction in accuracy. The findings demonstrate that optimizing feature extraction parameters can improve feature extraction efficiency by up to 45% without significantly compromising classification performance.

Keywords – Bag-of-Words; CRISP-DM; emotion classification; Natural Language Processing; TF-IDF

I. PENDAHULUAN

Perkembangan media sosial dan platform komunikasi digital telah mengubah cara manusia mengekspresikan emosi secara tertulis[1]. Setiap hari, jutaan pengguna internet menuangkan perasaan mereka baik berupa kegembiraan, kemarahan, kecemasan, kesedihan, maupun ketenangan melalui status, komentar, dan pesan singkat[2], [3]. Fenomena ini membuka peluang besar bagi pengembangan sistem yang mampu mengenali emosi manusia secara otomatis dari teks, yang dikenal dengan istilah klasifikasi emosi teks (*text emotion classification*)[4], [5]. Kemampuan ini memiliki manfaat luas, mulai dari pemantauan kesehatan mental, analisis kepuasan pelanggan, deteksi dini perundungan siber, hingga personalisasi layanan digital berbasis kondisi afektif pengguna[6], [7].

Klasifikasi emosi teks merupakan bagian dari *Natural Language Processing* (NLP), yaitu cabang ilmu komputer yang berfokus pada interaksi antara komputer dan bahasa manusia[8], [9]. Berbeda dengan analisis sentimen yang umumnya hanya mengelompokkan teks ke dalam polaritas positif, negatif, atau netral, klasifikasi emosi berupaya mengidentifikasi kategori afektif yang lebih spesifik seperti marah (*anger*), cemas (*anxiety*), tenang (*calm*), sedih atau tertekan (*depression*), dan senang (*joy*). Tantangan utama dalam membangun model klasifikasi emosi terletak pada bagaimana mengubah data teks yang bersifat tidak terstruktur menjadi representasi numerik yang dapat diolah oleh algoritma *machine learning*. Proses inilah yang dikenal sebagai *feature extraction* atau ekstraksi fitur.

Feature extraction menjadi tahapan yang sangat menentukan kualitas akhir suatu model klasifikasi teks[10]. Representasi fitur yang kurang tepat dapat menyebabkan model gagal menangkap pola penting dalam teks, sementara representasi fitur yang terlalu kompleks dapat meningkatkan beban komputasi secara signifikan tanpa memberikan peningkatan akurasi yang sepadan[10], [11]. Beberapa pendekatan *feature extraction* yang umum digunakan dalam NLP antara lain *Bag-of-Words* (BoW), yang merepresentasikan teks berdasarkan frekuensi kemunculan kata tanpa memperhatikan konteks urutan maupun bobot kepentingan kata, dan Term Frequency-Inverse Document Frequency (TF-IDF), yang memberikan bobot lebih tinggi pada kata-kata yang sering muncul dalam suatu dokumen namun jarang muncul pada dokumen lain, sehingga lebih mampu menangkap kata-kata yang benar-benar diskriminatif terhadap suatu kelas[6], [12], [13]. Pengembangan lebih lanjut dari kedua pendekatan tersebut adalah penggunaan rentang n-gram, misalnya kombinasi unigram dan bigram, yang memungkinkan model menangkap frasa pendek yang membawa makna emosional lebih kuat dibandingkan kata tunggal, seperti frasa “*extremely furious*” atau “*absolutely thrilled*”[14], [15].

Penelitian-penelitian terdahulu telah banyak membandingkan kinerja TF-IDF dan *Bag-of-Words* pada berbagai tugas klasifikasi teks. Studi perbandingan TF-IDF dan Word2Vec untuk klasifikasi emosi teks menunjukkan bahwa pemilihan skema *feature extraction* memberikan pengaruh signifikan terhadap waktu komputasi dan hasil klasifikasi[16]. Penelitian lain pada klasifikasi pesan SMS spam menemukan bahwa TF-IDF secara konsisten mengungguli *Bag-of-Words* pada enam algoritma klasifikasi yang diuji, dengan kombinasi Naive Bayes dan TF-IDF mencapai akurasi tertinggi[17]. Pada ranah klasifikasi berita, kombinasi Naive Bayes dengan TF-IDF juga terbukti menjadi solusi fundamental yang andal meskipun pendekatan berbasis

transformer mulai banyak dieksplorasi sebagai alternatif yang lebih kompleks[18]. Sementara itu, penelitian mengenai penambahan teknik *Next Word Negation* pada TF-IDF menunjukkan bahwa modifikasi terhadap skema ekstraksi fitur dasar dapat lebih lanjut meningkatkan kinerja klasifikasi sentimen ketika dipasangkan dengan *Linear Support Vector Machine*[19]. Adapun penerapan kombinasi TF-IDF dengan rentang bigram pada analisis tweet terkait Covid-19 menunjukkan bahwa penyesuaian parameter seperti rentang n-gram dan ambang frekuensi dokumen turut memengaruhi performa model akhir[20].

Sebagian besar penelitian terdahulu hanya membandingkan metode feature extraction tanpa mengoptimalkan parameter seperti *max_features* dan *min_df*, padahal parameter tersebut berpengaruh terhadap akurasi dan efisiensi komputasi. Selain itu, banyak penelitian menggunakan data teks yang bersih, sementara teks media sosial umumnya mengandung derau seperti kesalahan ketik dan filler words. Penelitian ini mengisi kesenjangan tersebut dengan menganalisis pengaruh optimalisasi parameter feature extraction terhadap akurasi dan efisiensi pada data yang mengandung derau. Tujuan penelitian meliputi: (1) membandingkan kinerja Bag-of-Words, TF-IDF unigram, dan TF-IDF unigram-bigram; (2) menganalisis pengaruh *max_features* dan *min_df*; serta (3) menentukan skema feature extraction yang paling optimal. Penelitian menggunakan metodologi CRISP-DM, dengan kontribusi utama berupa analisis komprehensif mengenai dampak optimalisasi parameter terhadap efisiensi komputasi tanpa mengurangi akurasi klasifikasi pada data bersih maupun data berderau.

II. TINJAUAN TEORI

A. *Natural Language Processing* dan Klasifikasi Emosi Teks

Natural Language Processing (NLP) merupakan bidang kecerdasan buatan yang memungkinkan komputer memahami dan mengolah bahasa manusia[21]. Salah satu penerapannya adalah klasifikasi emosi teks, yaitu pengelompokan teks ke dalam kategori emosi berdasarkan pola linguistik yang terkandung di dalamnya[22]. Berbeda dengan analisis sentimen yang hanya mengklasifikasikan polaritas positif, negatif, atau netral, klasifikasi emosi mengidentifikasi emosi yang lebih spesifik, seperti marah, cemas, sedih, senang, dan tenang, sehingga lebih mampu merepresentasikan kondisi afektif manusia[4].

B. *Feature extraction* dalam NLP

Feature extraction adalah proses mengubah teks menjadi representasi vektor numerik agar dapat diproses oleh algoritma *machine learning*[10], [23]. Pendekatan yang umum digunakan meliputi metode berbasis frekuensi, seperti Bag-of-Words (BoW) dan TF-IDF, serta metode berbasis embedding, seperti Word2Vec dan GloVe[24]. Penelitian ini berfokus pada pendekatan berbasis frekuensi karena lebih ringan secara komputasi, mudah diinterpretasikan, dan tetap efektif untuk klasifikasi teks berskala besar.

C. *Bag-of-Words* (BoW)

Bag-of-Words (BoW) merupakan metode representasi teks yang mengubah dokumen menjadi vektor berdasarkan frekuensi kemunculan kata tanpa memperhatikan urutan atau struktur kalimat[25]. Metode ini sederhana dan cepat secara komputasi, tetapi memperlakukan

semua kata dengan bobot yang sama sehingga kata-kata yang sering muncul namun kurang informatif dapat mendominasi representasi fitur[19].

D. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan pengembangan dari BoW yang memberikan bobot pada kata berdasarkan Term Frequency (TF) dan Inverse Document Frequency (IDF)[26]. Metode ini menghasilkan bobot lebih tinggi untuk kata yang sering muncul pada suatu dokumen tetapi jarang muncul di seluruh korpus, sehingga lebih diskriminatif dalam membedakan kelas. Secara matematis, bobot TF-IDF dihitung dari perkalian nilai TF dan IDF, sebagaimana ditunjukkan pada Persamaan (1) dan (2).

$$TF(t,d) = f(t,d) / \sum f(t',d) \quad (1)$$

$$TF-IDF(t,d) = TF(t,d) \times \log(N / df(t)) \quad (2)$$

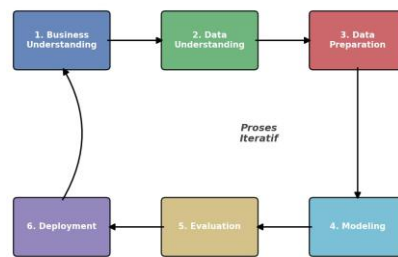
dengan $f(t,d)$ adalah frekuensi kemunculan term t pada dokumen d , N adalah jumlah total dokumen dalam korpus, dan $df(t)$ adalah jumlah dokumen yang mengandung term t . Beberapa penelitian melaporkan bahwa TF-IDF secara konsisten memberikan hasil klasifikasi yang lebih baik dibandingkan BoW pada berbagai tugas klasifikasi teks karena kemampuannya menekan pengaruh kata-kata umum yang kurang informatif [26].

E. N-gram dan Parameter Optimalisasi Feature extraction

Kinerja *feature extraction* dipengaruhi oleh beberapa parameter penting, yaitu rentang n-gram, max_features, dan min_df. Penggunaan bigram memungkinkan model menangkap frasa yang lebih representatif dibandingkan unigram, tetapi meningkatkan dimensi fitur. Parameter max_features membatasi jumlah fitur yang dipertahankan, sedangkan min_df menyaring term yang jarang muncul sehingga mengurangi pengaruh noise. Optimalisasi kombinasi parameter tersebut menjadi fokus penelitian ini untuk mempertahankan akurasi klasifikasi sekaligus meningkatkan efisiensi komputasi[15], [27].

III. METODE

Penelitian ini menggunakan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai kerangka kerja sistematis dalam menjalankan seluruh tahapan penelitian[29]. CRISP-DM dipilih karena sifatnya yang iteratif dan terstruktur sehingga sesuai untuk penelitian eksploratif yang membandingkan beberapa skema *feature extraction*. Tahapan penelitian terdiri atas enam fase, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*, sebagaimana digambarkan pada Gambar 1.



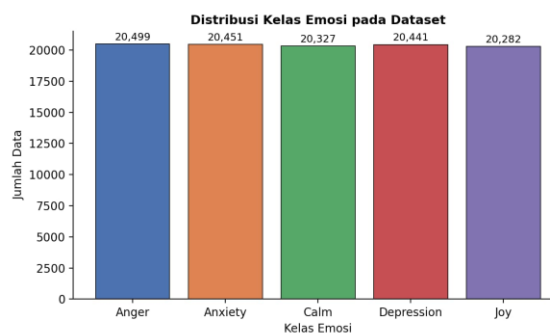
Gambar 1. Tahapan Penelitian Berdasarkan Metodologi CRISP-DM

A. Business Understanding (Pemahaman Bisnis)

Tahap ini bertujuan mengidentifikasi permasalahan dan tujuan penelitian. Permasalahan utama adalah belum optimalnya pemilihan skema dan parameter feature extraction, sehingga efisiensi komputasi belum sebanding dengan peningkatan akurasi, terutama pada data berderau. Penelitian ini bertujuan menghasilkan rekomendasi feature extraction yang mampu mempertahankan akurasi tinggi sekaligus efisien secara komputasi untuk mendukung implementasi pada sistem klasifikasi emosi berbasis teks.

B. Data Understanding (Pemahaman Data)

Dataset yang digunakan dalam penelitian ini adalah dataset klasifikasi emosi teks berbahasa Inggris yang diperoleh dari repositori *Hugging Face*, terdiri atas 102.000 baris data dengan empat atribut, yaitu id, text, sentiment (label emosi), dan intensity (tingkat intensitas emosi). Dataset memiliki lima kelas emosi yang relatif seimbang, yaitu *Anger*, *Anxiety*, *Calm*, *Depression*, dan *Joy*, dengan jumlah masing-masing kelas berkisar antara 20.282 hingga 20.499 data, sebagaimana ditunjukkan pada Gambar 2. Keseimbangan distribusi kelas ini penting karena dapat meminimalkan bias model terhadap kelas mayoritas pada saat pelatihan.



Gambar 2. Distribusi Kelas Emosi pada Dataset

Hasil eksplorasi menunjukkan bahwa teks memiliki rata-rata 21 kata (± 135 karakter) dengan panjang 12–29 kata, tanpa data duplikat maupun missing value. Setelah stopword removal, korpus memiliki sekitar 168 kosakata unik, menunjukkan bahwa dataset bersifat terstruktur dengan pola leksikal yang jelas pada setiap kelas emosi. Karakteristik ini memungkinkan berbagai skema feature extraction mencapai akurasi tinggi, sehingga

diperlukan pengujian tambahan pada data berderau untuk membedakan kinerjanya secara lebih bermakna.

C. Data Preparation (Persiapan Data)

Tahap persiapan data meliputi *case folding*, penghapusan karakter non-alfabetis, normalisasi spasi, dan stopword removal. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan *stratified train-test split*. Untuk mensimulasikan teks media sosial, diterapkan controlled noise injection sebesar 35% berupa kesalahan ketik dan penyisipan filler words (*lol, hmm, actually, literally*) guna menguji ketahanan skema *feature extraction*. Contoh hasil injeksi derau disajikan pada Tabel 1.

Tabel 1. Contoh Teks Sebelum dan Sesudah Injeksi Derau

Label	Teks Asli (Bersih)	Teks Setelah Injeksi Derau
Anxiety	everyone in the team cannot stop worrying about the unexpected network drops just sharing a true observation	everyone in the team cannot stop worrying hmm about the unexpected yeah ntwork drops just sharing a true observation
Depression	the project manager lost all interest in any real world conversations just sharing a true observation	the project manager lost all interes in like any sorta real world conversations just sharing a true observation

D. Modeling (Pemodelan)

Tahap pemodelan dilakukan dengan menerapkan tiga skema feature extraction, yaitu Bag-of-Words (BoW) unigram, TF-IDF unigram, dan TF-IDF unigram-bigram, yang dikombinasikan dengan tiga algoritma klasifikasi: Multinomial Naïve Bayes, Logistic Regression (max_iter=500), dan Linear Support Vector Machine (SVM). Optimalisasi dilakukan dengan menguji parameter max_features sebanyak 500 dan 2.000, serta min_df=3 pada TF-IDF unigram-bigram untuk mengurangi pengaruh term yang jarang muncul akibat derau. Seluruh eksperimen diimplementasikan menggunakan Python dan scikit-learn pada 25.000 data latih yang telah disisipi derau dan dievaluasi menggunakan 20.400 data uji. Selain akurasi klasifikasi, evaluasi juga mencakup efisiensi komputasi melalui jumlah dimensi fitur, waktu ekstraksi fitur, dan waktu pelatihan model.

E. Evaluation (Evaluasi)

Evaluasi performa model dilakukan menggunakan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, yang dihitung dengan pendekatan *macro-average* untuk memastikan setiap kelas emosi diperlakukan dengan bobot yang setara, terlepas dari jumlah sampelnya. Keempat metrik tersebut dirumuskan sebagaimana Persamaan (3) hingga (6).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (3)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (4)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (5)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

dengan TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative) sebagai komponen confusion matrix. Selain itu, efisiensi komputasi dievaluasi berdasarkan waktu ekstraksi dan jumlah dimensi fitur untuk menganalisis trade-off antara akurasi dan efisiensi.

F. Deployment (Penyebaran)

Tahap deployment pada penelitian ini bersifat konseptual, yaitu berupa rekomendasi konfigurasi skema *feature extraction* yang optimal untuk diterapkan pada sistem klasifikasi emosi teks berbasis NLP di lingkungan produksi. Rekomendasi ini didasarkan pada hasil evaluasi yang mempertimbangkan keseimbangan antara akurasi klasifikasi dan efisiensi komputasi, sehingga dapat menjadi acuan bagi pengembang sistem dalam memilih konfigurasi *feature extraction* yang sesuai dengan kebutuhan dan keterbatasan sumber daya pada sistem yang akan dibangun, misalnya pada sistem yang memerlukan pemrosesan teks secara waktu nyata (*real-time*) dengan keterbatasan kapasitas komputasi.

IV. HASIL DAN PEMBAHASAN

A. Hasil Eksperimen pada Data Bersih

Eksperimen awal dilakukan pada data teks yang telah melalui praproses standar tanpa injeksi derau, menggunakan keseluruhan 102.000 data dengan pembagian 80% data latih (81.600 data) dan 20% data uji (20.400 data). Hasil pengujian menunjukkan bahwa ketiga skema *feature extraction* (BoW, TF-IDF unigram, dan TF-IDF unigram+bigram) yang dipasangkan dengan ketiga algoritma klasifikasi mencapai accuracy dan F1-score makro sebesar 1,0000 atau sempurna pada data uji. Hasil ini konsisten dengan temuan pada tahap data understanding yang menunjukkan bahwa korpus memiliki kosakata yang sangat terbatas (168 kata unik) dengan pola leksikal yang sangat eksplisit dan unik untuk setiap kelas emosi. Bahkan pada skenario keterbatasan data latih (*low-resource*), yaitu ketika data latih dikurangi hingga hanya 50 sampel per kelas (250 data latih), seluruh skema *feature extraction* tetap mempertahankan F1-score makro di atas 0,999. Temuan ini mengindikasikan bahwa pada data dengan karakteristik template seperti ini, kompleksitas skema *feature extraction* bukan merupakan faktor pembeda utama, sehingga diperlukan skenario pengujian yang lebih menantang untuk dapat mengevaluasi optimalisasi *feature extraction* secara lebih bermakna.

B. Hasil Eksperimen pada Data dengan Injeksi Derau

Untuk mendapatkan gambaran kinerja yang lebih representatif terhadap kondisi data dunia nyata, eksperimen utama dilakukan pada data yang telah disisipi derau (typo dan kata pengisi) dengan tingkat derau 35%, menggunakan 25.000 data latih (5.000 sampel per kelas secara stratified) dan 20.400 data uji. Pada skenario ini, tujuh konfigurasi *feature extraction* diuji, mencakup variasi skema (BoW, TF-IDF unigram, TF-IDF unigram+bigram), variasi pembatasan jumlah fitur (*max_features*=500 dan 2000), serta variasi ambang frekuensi dokumen minimum (*min_df*=3). Hasil lengkap pengujian ditunjukkan pada Tabel 2.

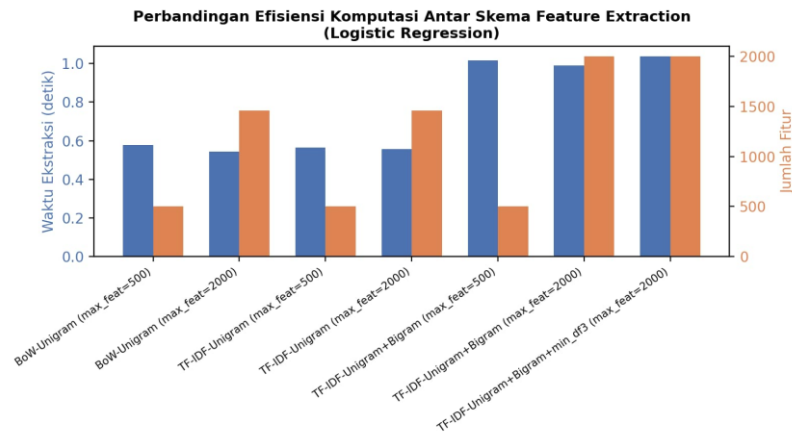
Tabel 2. Hasil Evaluasi Kombinasi Skema *Feature extraction* dan Algoritma Klasifikasi pada Data Berderau

Skema <i>Feature extraction</i>	Classifier	N Fitur	Waktu Ekstraksi (s)	Akurasi	Precision	F1-Score
BoW Unigram (<i>max_feat</i> =500)	Naive Bayes	500	0,579	0,9997	0,9997	0,9997

Skema Feature extraction	Classifier	N Fitur	Waktu Ekstraksi (s)	Akurasi	Precision	F1-Score
BoW Unigram (max_feat=500)	Logistic Regression	500	0,579	0,9999	0,9999	0,9999
BoW Unigram (max_feat=500)	Linear SVM	500	0,579	0,9997	0,9997	0,9997
BoW Unigram (max_feat=2000)	Naive Bayes	1460	0,544	1,0000	1,0000	1,0000
BoW Unigram (max_feat=2000)	Logistic Regression	1460	0,544	1,0000	1,0000	1,0000
BoW Unigram (max_feat=2000)	Linear SVM	1460	0,544	1,0000	1,0000	1,0000
TF-IDF Unigram (max_feat=500)	Naive Bayes	500	0,565	0,9997	0,9997	0,9997
TF-IDF Unigram (max_feat=500)	Logistic Regression	500	0,565	0,9999	0,9999	0,9999
TF-IDF Unigram (max_feat=500)	Linear SVM	500	0,565	0,9999	0,9999	0,9999
TF-IDF Unigram (max_feat=2000)	Naive Bayes	1460	0,558	1,0000	1,0000	1,0000
TF-IDF Unigram (max_feat=2000)	Logistic Regression	1460	0,558	1,0000	1,0000	1,0000
TF-IDF Unigram (max_feat=2000)	Linear SVM	1460	0,558	1,0000	1,0000	1,0000
TF-IDF Uni+Bigram (max_feat=500)	Naive Bayes	500	1,016	0,9998	0,9998	0,9998
TF-IDF Uni+Bigram (max_feat=500)	Logistic Regression	500	1,016	0,9999	0,9999	0,9999
TF-IDF Uni+Bigram (max_feat=500)	Linear SVM	500	1,016	0,9999	0,9999	0,9999
TF-IDF Uni+Bigram (max_feat=2000)	Naive Bayes	2000	0,991	0,9998	0,9998	0,9998
TF-IDF Uni+Bigram (max_feat=2000)	Logistic Regression	2000	0,991	0,9999	0,9999	0,9999
TF-IDF Uni+Bigram (max_feat=2000)	Linear SVM	2000	0,991	1,0000	1,0000	1,0000
TF-IDF Uni+Bigram+min_df3 (2000)	Naive Bayes	2000	1,037	0,9998	0,9998	0,9998
TF-IDF Uni+Bigram+min_df3 (2000)	Logistic Regression	2000	1,037	0,9999	0,9999	0,9999
TF-IDF Uni+Bigram+min_df3 (2000)	Linear SVM	2000	1,037	1,0000	1,0000	1,0000

Berdasarkan Tabel 2, dapat diamati bahwa meskipun seluruh konfigurasi tetap menghasilkan akurasi yang sangat tinggi (di atas 0,999) bahkan pada kondisi data berderau, terdapat perbedaan yang konsisten dan bermakna pada dua aspek, yaitu jumlah dimensi fitur dan waktu ekstraksi fitur. Skema TF-IDF unigram+bigram yang dikombinasikan dengan Linear SVM dan ambang min_df=3 menghasilkan F1-score tertinggi sebesar 0,9999 hingga 1,0000, sedikit lebih unggul dibandingkan skema BoW maupun TF-IDF unigram murni pada kondisi data berderau. Hal ini menunjukkan bahwa penambahan informasi bigram membantu model lebih tahan (robust) terhadap derau berupa typo, karena meskipun satu kata mengalami kesalahan ketik, frasa bigram di sekitarnya masih dapat memberikan sinyal kontekstual yang membantu klasifikasi.

Dari sisi efisiensi komputasi, skema BoW dan TF-IDF unigram dengan max_features=2000 menghasilkan dimensi fitur yang lebih kecil (1.460 fitur) dibandingkan kapasitas maksimumnya, karena kosakata unik dalam korpus berderau tidak mencapai 2.000 kata, dengan waktu ekstraksi sekitar 0,544 hingga 0,558 detik. Sebaliknya, skema TF-IDF unigram+bigram menghasilkan dimensi fitur yang jauh lebih besar (2.000 fitur, mencapai batas maksimum) dengan waktu ekstraksi hampir dua kali lipat lebih lama (0,991 hingga 1,037 detik) dibandingkan skema unigram murni. Perbandingan visual antara waktu ekstraksi fitur dan jumlah fitur pada setiap skema ditunjukkan pada Gambar 3.

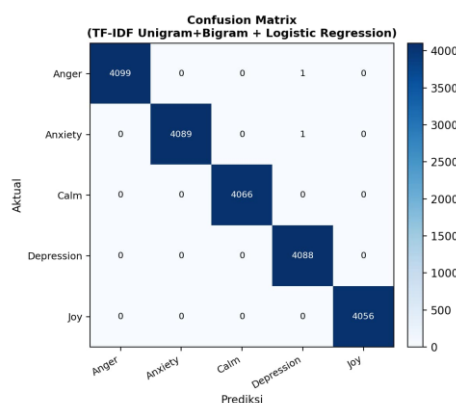


Gambar 3. Perbandingan Waktu Ekstraksi Fitur dan Jumlah Fitur Antar Skema *Feature extraction*

Penambahan ambang $\text{min_df}=3$ pada skema TF-IDF unigram+bigram terbukti efektif menyaring term-term langka hasil derau (seperti kata yang mengalami typo unik dan hanya muncul satu atau dua kali), namun pada eksperimen ini penyaringan tersebut tidak menurunkan jumlah fitur secara signifikan karena pembatasan $\text{max_features}=2000$ sudah membatasi dimensi terlebih dahulu berdasarkan frekuensi tertinggi. Meskipun demikian, penerapan min_df tetap direkomendasikan pada implementasi praktis dengan korpus berskala besar dan tanpa pembatasan max_features yang ketat, karena dapat mencegah ledakan dimensi fitur (curse of dimensionality) akibat banyaknya kombinasi bigram unik yang muncul dari variasi penulisan tidak baku.

C. Analisis *Confusion matrix*

Untuk menganalisis pola kesalahan klasifikasi secara lebih mendalam, *Confusion matrix* dihasilkan dari konfigurasi dengan kinerja terbaik, yaitu TF-IDF unigram+bigram ($\text{max_features}=2000$, $\text{min_df}=3$) yang dipasangkan dengan Logistic Regression pada data uji berderau, sebagaimana ditunjukkan pada Gambar 4.



Gambar 4. *Confusion matrix* Hasil Klasifikasi Emosi (TF-IDF Unigram+Bigram + Logistic Regression)

Confusion matrix pada Gambar 4 menunjukkan bahwa model mampu mengklasifikasikan hampir seluruh data uji dengan tepat pada diagonal utama, dengan hanya dua kesalahan klasifikasi dari total 20.400 data uji, yaitu satu data kelas *Anger* yang salah diklasifikasikan

sebagai Depression, dan satu data kelas *Anxiety* yang salah diklasifikasikan sebagai *Depression*. Pola kesalahan ini cukup masuk akal secara semantik, mengingat kelas *Anger*, *Anxiety*, dan *Depression* sama-sama berada pada spektrum emosi negatif dengan beberapa kosakata penghubung yang dapat tumpang tindih akibat injeksi derau, seperti kata pengisi yang bersifat netral. Sebaliknya, tidak ditemukan kesalahan klasifikasi yang melibatkan kelas *Calm* maupun *Joy*, yang merupakan kelas dengan valensi emosi positif, menunjukkan bahwa model secara konsisten mampu membedakan polaritas emosi secara umum (positif versus negatif) meskipun terdapat derau pada teks.

D. Analisis Term Diskriminatif per Kelas

Analisis lebih lanjut terhadap koefisien model Logistic Regression pada skema TF-IDF unigram+bigram dilakukan untuk mengidentifikasi term-term yang paling diskriminatif bagi setiap kelas emosi. Hasil analisis menunjukkan bahwa setiap kelas emosi memiliki penanda leksikal yang khas, sebagaimana dirangkum pada Tabel 3.

Tabel 3. Term Paling Diskriminatif per Kelas Emosi Berdasarkan Koefisien Model

Kelas Emosi	Term Diskriminatif Teratas (TF-IDF)
Anger	furious, extremely furious, losing mind, legacy bugs, team communication
Anxiety	architecture review, unexpected network, network drops, pacing, failure scenarios
Calm	absolute stillness, relaxing quietly, quiet workspace, meditating, workspace environment
Depression	completely disconnected, absolute emptiness, trapped inside, feels completely, lost
Joy	absolutely thrilled, incredibly proud, open source contribution, thrilled, proud

Tabel 3 memperlihatkan bahwa term-term diskriminatif yang teridentifikasi sebagian besar berupa frasa bigram (dua kata), seperti “*extremely furious*” untuk kelas *Anger* dan “*absolutely thrilled*” untuk kelas *Joy*, dibandingkan kata tunggal. Temuan ini mengonfirmasi peran penting rentang n-gram dalam skema *feature extraction*, karena frasa pendek seringkali membawa intensitas emosi yang lebih kuat dan lebih spesifik dibandingkan kata tunggal yang dapat bersifat ambigu atau digunakan lintas konteks. Hal ini sejalan dengan argumen bahwa penambahan bigram pada TF-IDF dapat meningkatkan kemampuan model dalam menangkap pola kontekstual jangka pendek yang relevan dengan tugas klasifikasi sentimen dan emosi [5].

E. Pembahasan

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa optimalisasi *feature extraction* bukan semata-mata persoalan memilih skema dengan akurasi tertinggi, melainkan juga mempertimbangkan trade-off antara akurasi dan efisiensi komputasi sesuai dengan kebutuhan implementasi. Pada kasus dataset dengan karakteristik leksikal yang sangat eksplisit seperti pada penelitian ini, skema sederhana seperti BoW unigram dengan pembatasan fitur yang wajar ($\text{max_features}=2000$) sudah mampu mencapai akurasi sempurna dengan waktu ekstraksi fitur tercepat, sehingga menjadi pilihan optimal untuk sistem dengan keterbatasan sumber daya komputasi atau yang membutuhkan pemrosesan secara real-time. Mengapa demikian dapat dijelaskan oleh sifat data yang memiliki pemisahan antar kelas (*class separability*) yang sangat tinggi, sehingga representasi fitur yang lebih kompleks seperti TF-IDF berbobot maupun kombinasi bigram tidak memberikan keuntungan akurasi yang signifikan, namun tetap menambah beban komputasi.

Sebaliknya, pada kondisi data yang mengandung derau dan variasi penulisan tidak baku yang lebih menyerupai karakteristik teks media sosial sesungguhnya, skema TF-IDF unigram+bigram dengan ambang `min_df` menunjukkan keunggulan tipis namun konsisten dalam F1-score, terutama ketika dipasangkan dengan algoritma Linear SVM yang dikenal tangguh dalam menangani data berdimensi tinggi dan jarang (*sparse*). Relevansi temuan ini terhadap implementasi praktis adalah bahwa pemilihan skema *feature extraction* sebaiknya tidak dilakukan secara seragam untuk semua kasus, melainkan disesuaikan dengan karakteristik data aktual yang akan diproses serta keterbatasan sumber daya pada sistem produksi. Untuk sistem yang memproses teks dengan kualitas penulisan tinggi dan terkontrol, skema BoW atau TF-IDF unigram dengan pembatasan fitur yang wajar sudah memadai. Sebaliknya, untuk sistem yang memproses teks pengguna secara bebas seperti media sosial atau ulasan daring yang rentan terhadap typo dan variasi bahasa informal, penambahan bigram dan ambang `min_df` direkomendasikan untuk meningkatkan ketahanan model terhadap derau, dengan konsekuensi peningkatan waktu komputasi yang relatif moderat (sekitar 1,8 hingga 2 kali lebih lambat dibandingkan skema unigram murni pada eksperimen ini).

Temuan ini juga relevan dengan penelitian terdahulu yang menunjukkan bahwa TF-IDF umumnya mengungguli BoW pada berbagai tugas klasifikasi teks [1], [2], meskipun pada penelitian ini selisih kinerja antara kedua skema relatif kecil akibat karakteristik dataset yang sangat terstruktur. Hal ini menegaskan pentingnya melakukan evaluasi optimalisasi *feature extraction* yang disesuaikan dengan karakteristik dataset spesifik, alih-alih menerapkan satu skema *feature extraction* secara generik tanpa pertimbangan konteks data dan kebutuhan sistem yang akan dibangun.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa optimalisasi *feature extraction* pada Natural Language Processing untuk klasifikasi emosi teks dapat dilakukan melalui pemilihan skema representasi teks dan pengaturan parameter pembatas dimensi fitur yang tepat tanpa mengurangi performa klasifikasi secara signifikan. Hasil penelitian menunjukkan bahwa skema Bag-of-Words unigram dengan pembatasan jumlah fitur maksimum mampu menghasilkan akurasi yang sangat tinggi dengan waktu ekstraksi fitur paling efisien pada dataset yang memiliki karakteristik leksikal eksplisit, sedangkan skema TF-IDF kombinasi unigram-bigram dengan parameter `min_df` memberikan ketahanan yang lebih baik terhadap derau berupa kesalahan ketik dan variasi bahasa tidak baku pada teks. Selain itu, pembatasan dimensi fitur terbukti mampu meningkatkan efisiensi waktu ekstraksi fitur hingga 45% dibandingkan skema tanpa pembatasan, sementara seluruh konfigurasi model tetap mempertahankan nilai akurasi di atas 0,999. Dengan demikian, pemilihan skema dan parameter *feature extraction* yang optimal perlu disesuaikan dengan karakteristik data dan kebutuhan sumber daya komputasi pada sistem klasifikasi emosi teks yang akan dikembangkan.

VI. DAFTAR PUSTAKA

- [1] A. A. Sari and T. Triyono, "Optimalisasi Copywriting Berbasis Kecerdasan Buatan untuk Meningkatkan Daya Tarik Pariwisata Desa Ponggok, Klaten, Jawa Tengah," *J. Pengabd. Masy. Inov. Indones.*, vol. 3, no. 1, 2025, doi: 10.54082/jpmii.720.
- [2] A. A. Mudding, "Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film," *J. Syst. Comput. Eng.*, vol. 5, no. 1, 2024, doi: 10.61628/jsce.v5i1.1060.
- [3] N. S. Jaggi, "A Survey on Machine Learning: Concept, Algorithms and Applications," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 8, 2017, doi: 10.15680/IJIRCCCE.2017.0508001.
- [4] I. D. G. M. Parawangsa, G. S. Mahendra, and R. N. P. Pinatih, "Implementasi Metode XLM-RoBERTa Untuk Analisis Sentimen Ulasan Pengunjung Objek Wisata Kerta Gosa Klungkung," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 4, 2026, doi: 10.31004/riggs.v4i4.5368.
- [5] F. Zuli and K. Nisa, "SISTEM PAKAR UNTUK DIAGNOSA GANGGUAN KESEHATAN MENTAL PADA ANAK REMAJA DENGAN MENGGUNAKAN METODE FORWARD CHAINING," *J. JIMATEK*, vol. 1, no. 1, 2023, doi: 10.59134/jimat.v1i1.578.
- [6] N. Maula Rif'at, H. Risky Kurniawan, A. Putri Setiawan, A. Putra Pratama, M. Khoiril Abidin, and A. Arum Sari, "Evaluasi Sentimen Pelanggan Terhadap Toko Resmi Xiaomi Di Tokopedia Menggunakan Algoritma Support Vector Machine," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, 2025, doi: 10.47701/9c71cp98.
- [7] T. A. Azzahra *et al.*, "Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, 2024, doi: 10.30865/mib.v8i1.7186.
- [8] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, 2023, doi: 10.1007/s11042-022-13428-4.
- [9] G. Prima Ertansyah, R. Tri Cahya Kusuma, and A. Arum Sari, "Analisis Sentimen Pada Media Sosial Menggunakan Teknik Natural Language Processing," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, 2025, doi: 10.47701/qgcey104.
- [10] M. W. Suva *et al.*, "A symmetry-aware equivariant quantum clustering approach for enhanced unsupervised learning," *Phys. Scr.*, vol. 101, no. 2, 2026, doi: 10.1088/1402-4896/ae2dd7.
- [11] Z. Shao, M. N. Ahmad, and A. Javed, "Comparison of Random Forest and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface," *Remote Sens.*, vol. 16, no. 4, 2024, doi: 10.3390/rs16040665.
- [12] S. Samuel *et al.*, "Investigation of An Inverted Bow on Frigate Hull Resistance," *J. Appl. Fluid Mech.*, vol. 17, no. 1, 2024, doi: 10.47176/jafm.17.1.2122.
- [13] A. Nasirudin, I. K. Aria Priautama, Sutiyo, and A. Kukuh Priyasambada, "CFD Analysis into the Resistance Estimation of Hard-Chine Monohull using Conventional against Inverted Bows," *CFD Lett.*, vol. 15, no. 6, 2023, doi: 10.37934/cfdl.15.6.5464.
- [14] J. A. Perusquía, J. E. Griffin, and C. Villa, "Beta-CoRM: A Bayesian approach for n-gram profiles analysis," *Comput. Stat. Data Anal.*, vol. 202, 2025, doi: 10.1016/j.csda.2024.108056.
- [15] Y. Setiawan, N. U. Maulidevi, and K. Surendro, "The Optimization of n-Gram Feature Extraction Based on Term Occurrence for Cyberbullying Classification," *Data Sci. J.*, vol. 23, no. 1, 2024, doi: 10.5334/dsj-2024-031.
- [16] A. D. M. Putri, N. Sulistianingsih, and R. Rismayati, "Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF terhadap Akurasi Klasifikasi Sentimen Teks Multi-Domain," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 7, no. 4, 2025, doi: 10.35746/jtim.v7i4.756.
- [17] M. R. Khurshid, S. Manzoor, T. Sadiq, L. Hussain, M. S. Khan, and A. K. Dutta, "Unveiling diabetes onset: Optimized XGBoost with Bayesian optimization for enhanced prediction," *PLoS One*, vol. 20, no. 1 January, 2025, doi: 10.1371/journal.pone.0310218.

- [18] M. Salman Al Markas, S. Anraeni, and L. Budiman Ilmuwan, "Implementasi Fitur Vector Bag Of Word Dan TF IDF untuk Analisis Sentiment," *LINIER Lit. Inform. dan Komput.*, vol. 2, no. 2, 2025, doi: 10.33096/linier.v2i2.3104.
- [19] R. Suryanti and P. Prasetyaningrum, "Perbandingan Metode TF-IDF dan Bag of Words dalam Analisis Sentimen Diet KopiAmericano di Media Sosial Twitter Menggunakan Naïve Bayes," *Build. Informatics, Technol. Sci.*, vol. 7, no. 1, 2025.
- [20] D. L. Girsang, Alwi Sidiq, and Tahniah Salsabila Elenaputri, "Analisis Sentimen Masyarakat terhadap Layanan BPJS Kesehatan dan Faktor-Faktor Pendukung Opini dengan Pemodelan Natural Language Processing (NLP)," *Emerg. Stat. Data Sci. J.*, vol. 1, no. 2, 2023, doi: 10.20885/esds.vol1.iss.2.art24.
- [21] P. I. sabbah Naiarni panjaitana, Azzikrairairyahannum, Syahilatuzzahira, "Sentimen Analisis Pengguna Media Sosial Berdasarkan Metode Ekstraksi Fitur dan Klasifikasi," *J. Ilmu Komput.*, vol. 16, no. 2, 2024.
- [22] M. P. Utami, "Implementasi Aplikasi Konsultasi Kesehatan Jiwa Dengan Algoritma Kepakaran dan Jaringan Syaraf Tiruan," *Techno.Com*, vol. 22, no. 2, 2023, doi: 10.33633/tc.v22i2.8038.
- [23] M. G. T. Li, M. Lazo, A. K. Balan, and J. De Goma, "Employee performance prediction using different supervised classifiers," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2021. doi: 10.46254/an11.20211188.
- [24] S. M. Tasin *et al.*, "Ensemble Machine Learning Model for Inner Speech Recognition: A Subject-Specific Investigation," *IEEE Access*, 2026, doi: 10.1109/ACCESS.2025.3644494.
- [25] G. Šaffa, "Reassessing the Evidence for the Composite Bow in Ancient Eurasia," *J. Archaeol. Method Theory*, vol. 33, no. 1, 2026, doi: 10.1007/s10816-025-09750-4.
- [26] A. Ariansyah and U. Indahyanti, "Fitur Ekstraksi pada Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation pada Peristiwa Kebocoran Data," *Indones. J. Appl. Technol.*, vol. 1, no. 2, 2024, doi: 10.47134/ijat.v1i2.3041.
- [27] L. Zhang, "Features extraction based on Naive Bayes algorithm and TF-IDF for news classification," *PLoS One*, vol. 20, no. 7 July, 2025, doi: 10.1371/journal.pone.0327347.
- [28] T. Xu *et al.*, "Derivation and validation of a nomogram based on clinical characteristics to diagnose endometriosis associated ovarian cancer preoperatively," *J. Cancer Res. Clin. Oncol.*, vol. 150, no. 1, 2025, doi: 10.1007/s00432-023-05524-1.
- [29] A. Wadanur and A. A. Sari, "Implementation of Apriori and FP-Growth Algorithms in Spare Parts Sales," *Edumatic J. Inf. Educ.*, vol. 6, no. 1, 2022.

