

Analisis Sentimen Air Mineral dan *Demineral* Menggunakan SVM pada *Dataset* Tidak Seimbang

Dipo Yudhis Rana^{*1}, Nur Fadillah², Asyifa Raudha Syarifah³, Dhea Aulia Afiandri⁴, Harsih Rianto⁵, Yamin Nuryamin⁶, Susi Susilowati⁷

¹ Universitas Bina Sarana Informatika; dipoyudhisrana@gmail.com;

² Universitas Bina Sarana Informatika; nurfadillah010123@gmail.com;

³ Universitas Bina Sarana Informatika; asyifaraudha29@gmail.com;

⁴ Universitas Bina Sarana Informatika; dhea55234@gmail.com;

⁵ Universitas Bina Sarana Informatika; harsih.hhr@bsi.ac.id;

⁶ Universitas Bina Sarana Informatika; yamin.yny@bsi.ac.id;

⁷ Universitas Bina Sarana Informatika; susi.sss@bsi.ac.id;

Abstrak: Penelitian ini menganalisis sentimen opini pengguna platform X terhadap air mineral dan air *demineral* menggunakan pendekatan kuantitatif eksperimental. Data set terdiri dari 140 *tweet* yang dilabeli manual ke dalam tiga kategori: positif (38.57%), negatif (13.57%), dan netral (47.86%). Ketidakseimbangan kelas ekstrem dan ukuran data set yang sangat kecil menjadi tantangan utama yang mendasari *novelty* penelitian ini. Penelitian ini menguji hipotesis bahwa *class weighting* lebih efektif daripada *Synthetic Minority Over-sampling Technique* (SMOTE) pada *dataset* kecil. Klasifikasi menggunakan *Support Vector Machine* (SVM) dengan *pra*-pemrosesan teks (*case folding*, *data cleaning*, *stemming* Sastrawi, *stopword removal*) dan ekstraksi fitur Term *Frequency-Inverse Document Frequency* (TF-IDF). Evaluasi melalui *10-fold stratified cross-validation* menunjukkan SVM dengan *class weighting* mencapai akurasi 57.14% dan F1-score *macro* 42.36%, sedikit mengungguli SMOTE (53.57% akurasi, 40.98% F1 *macro*). Uji statistik (*Cohen's d* = 0.125) mengkonfirmasi tidak ada perbedaan signifikan, membuktikan SMOTE tidak efektif pada *dataset* sangat kecil. Kelas minoritas (negatif) mencapai F1-score hanya 8%, menunjukkan tantangan ketidakseimbangan ekstrem. Kontribusi utama penelitian ini adalah validasi empiris bahwa *class weighting* lebih *robust* daripada SMOTE untuk analisis sentimen berbahasa Indonesia pada kondisi data terbatas.

Keywords: sentimen analisis; air mineral; support vector machine; ketidakseimbangan kelas; dataset kecil

DOI: <https://doi.org/10.47134/jacis.v5i2.138>

*Correspondensi: Dipo Yudhis Rana

Email: dipoyudhisrana@gmail.com

Receive: 20 November 2025

Accepted: 27 November 2025

Published: 28 November 2025



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstrak: This study analyzes sentiment in user opinions on platform X regarding mineral and demineralized water using a quantitative experimental approach. The dataset comprises 140 manually labeled tweets across three categories: positive (38.57%), negative (13.57%), and neutral (47.86%). This research tests the hypothesis that class weighting is more effective than SMOTE for small datasets. Classification employs Support Vector Machine (SVM) with text preprocessing (*case folding*, *data cleaning*, Sastrawi *stemming*, *stopword removal*) and TF-IDF feature extraction. Evaluation via *10-fold stratified cross-validation* shows SVM with class weighting achieves 57.14% accuracy and 42.36% macro F1-score, slightly outperforming SMOTE (53.57% accuracy, 40.98% F1 *macro*). Statistical testing (*Cohen's d* = 0.125) confirms no significant difference, proving SMOTE ineffective for very small datasets. The minority class (negative) achieves only 8% F1-score, demonstrating extreme imbalance challenges.

Keywords: sentiment analysis; mineral water; support

PENDAHULUAN

Air adalah komponen vital: sekitar 70% tubuh tersusun dari air yang menjaga fungsi organ, metabolisme, dan homeostasis elektrolit[1]. Kualitas konsumsi air ditentukan tidak hanya oleh volume tetapi juga kandungan mineralnya. Di Indonesia, masyarakat mengenal dua jenis air minum kemasan yang umum beredar, yaitu air mineral dan air *demineral*. Kedua jenis air ini memiliki karakteristik yang berbeda dan sering menjadi bahan diskusi di ruang publik, terutama terkait dengan aspek kesehatan dan klaim manfaatnya. Berdasarkan Peraturan Menteri Perindustrian Republik Indonesia Nomor 26 Tahun 2019, air mineral adalah air minum kemasan yang mengandung mineral alami tanpa penambahan dari luar, sedangkan air *demineral* adalah air yang telah melalui proses pemurnian seperti *reverse osmosis* atau *deionisasi* hingga sebagian besar kandungan mineralnya dihilangkan[1].

Air mineral mengandung mineral esensial (Mg, Ca, Na, Se)[1] yang mendukung fungsi saraf dan tulang, sedangkan air *demineral* sangat murni tetapi berisiko[2] jika dikonsumsi jangka panjang karena dapat mengganggu keseimbangan elektrolit. Kajian [3] membandingkan air mineral, air alkali, dan air *demineral* pada populasi sehat, dan menemukan bahwa efek fisiologis jangka pendek (seperti *pH urin* dan keseimbangan cairan) relatif tidak berbeda signifikan. Namun, penelitian tersebut juga menekankan bahwa bukti mengenai efek jangka panjang air *demineral* masih terbatas dan memerlukan penelitian lanjutan. Hal ini menjelaskan mengapa perdebatan publik tentang "air mineral vs air *demineral*" kerap muncul di berbagai platform digital. Selain itu peneliti [4] mengatakan bahwa 24,1% remaja kota Bogor kurang konsumsi air harian; preferensi dan informasi yang rancu memengaruhi pola minum. Seiring maraknya ulasan produk, komentar, dan diskusi publik di media sosial, diperlukan pendekatan otomatis untuk menganalisis sentimen masyarakat terhadap air mineral dan air *demineral* secara objektif dan terukur.

Analisis sentimen telah menjadi area penelitian yang berkembang pesat dalam bidang pemrosesan bahasa alami (*Natural Language Processing/NLP*). Berbagai penelitian menunjukkan efektivitas algoritma *machine learning* untuk klasifikasi sentimen, dengan *Support Vector Machine* (SVM) sebagai salah satu metode yang paling banyak digunakan. Peneliti [5] menggunakan SVM untuk analisis sentimen pada data platform X terkait sengketa Pemilu 2024 dan mencapai akurasi hingga 85% pada *dataset* besar (lebih dari 1.000 sampel)[6]. mengembangkan sistem analisis sentimen berbasis web menggunakan SVM dan TF-IDF dengan fitur statistik, menunjukkan bahwa kombinasi TF-IDF dan SVM menghasilkan performa yang stabil untuk data teks berbahasa Indonesia. Peneliti [7] membandingkan *Naïve Bayes* dan SVM dalam analisis sentimen ulasan produk *marketplace*, dan menemukan bahwa SVM lebih unggul dengan akurasi 82% dibandingkan *Naïve Bayes* yang hanya mencapai 76%.

Namun, tantangan utama dalam analisis sentimen adalah ketidakseimbangan kelas (*class imbalance*), di mana distribusi data antar kategori sentimen tidak merata [8] dalam kajian komprehensif mereka menyatakan bahwa masalah ketidakseimbangan kelas sangat umum terjadi dalam *deep learning* dan dapat menurunkan performa model secara signifikan, terutama pada kelas minoritas. Peneliti [9] lebih lanjut menekankan bahwa pembelajaran dari data tidak seimbang masih menjadi tantangan terbuka dengan berbagai arah penelitian

masa depan. Untuk mengatasi masalah ini, berbagai teknik telah dikembangkan, termasuk *Synthetic Minority Over-sampling Technique* (SMOTE) dan *class weighting*. Peneliti [10] mengevaluasi performa SVM dengan SMOTE dalam klasifikasi sentimen dan menemukan bahwa SMOTE dapat meningkatkan *recall* pada kelas minoritas hingga 15% pada *dataset* dengan ukuran lebih dari 1.000 sampel. Namun, efektivitas SMOTE pada *dataset* yang sangat kecil (kurang dari 200 sampel) masih menjadi pertanyaan terbuka.

Penelitian ini secara khusus memfokuskan pada *Support Vector Machine* (SVM) sebagai metode klasifikasi sentimen utama. Alur kerja mencakup *pra-pemrosesan* teks berbahasa Indonesia (*case folding*, *tokenisasi*, *stemming* Sastrawi, dan *stopword removal*) serta representasi fitur berbasis Term *Frequency-Inverse Document Frequency* (TF-IDF) yang sesuai untuk data teks berdimensi tinggi dan bersifat jarang (*sparse*). Fokus pada SVM didasari pertimbangan teoretis dan empiris bahwa model *margin-based* ini bekerja baik pada ruang fitur besar seperti TF-IDF, relatif tahan terhadap *overfitting* dengan pengaturan parameter yang tepat, serta memberikan kinerja kompetitif pada tugas analisis sentimen[5],[7]. Permasalahan ketidakseimbangan kelas yang lazim pada data opini publik ditangani dengan strategi penyeimbangan yang tepat (*class weighting* dan SMOTE) dan validasi silang terstratifikasi, sehingga evaluasi lebih adil terhadap seluruh kelas. Kinerja model dievaluasi menggunakan metrik yang relevan untuk data tidak seimbang, seperti macro-F1, di samping akurasi.

Beberapa penelitian sebelumnya fokus pada analisis sentimen menggunakan berbagai algoritma *machine learning* pada *dataset* berukuran besar dan relatif seimbang[5],[7]. Penelitian lain fokus pada perbandingan metode klasifikasi tanpa mempertimbangkan tantangan *dataset* kecil dan ketidakseimbangan ekstrem[10]. Hingga saat ini, belum ada penelitian yang secara spesifik menganalisis sentimen terhadap air mineral dan air *demineral* di platform X dengan *dataset* sangat kecil (kurang dari 150 sampel) dan ketidakseimbangan kelas ekstrem (kelas minoritas hanya 13.57% dari total data). Penelitian-penelitian sebelumnya juga belum mengeksplorasi perbandingan langsung antara *class weighting* dan SMOTE pada kondisi data yang sangat terbatas seperti ini.

Oleh karena itu, *novelty* penelitian ini terletak pada tiga aspek utama: (1) fokus pada domain air mineral dan *demineral* yang belum pernah diteliti sebelumnya dalam konteks analisis sentimen berbahasa Indonesia, (2) evaluasi empiris efektivitas *class weighting* dibandingkan SMOTE pada *dataset* sangat kecil dengan ketidakseimbangan ekstrem, dan (3) penyediaan *baseline* metodologis untuk analisis sentimen pada kondisi data tidak ideal yang sering ditemui dalam penelitian awal atau domain spesifik. Tujuan penelitian ini adalah untuk memberikan pemahaman kuantitatif terhadap persepsi masyarakat terhadap air mineral dan *demineral*, sekaligus menyediakan landasan metodologis untuk analisis sentimen berbahasa Indonesia pada *dataset* kecil dengan ketidakseimbangan kelas yang signifikan. Manfaat praktis penelitian ini adalah memberikan rekomendasi berbasis bukti kepada peneliti pemula yang bekerja dengan data terbatas untuk memilih metode penanganan ketidakseimbangan kelas yang tepat, serta menyediakan wawasan bagi produsen air minum kemasan tentang persepsi publik terhadap produk mereka di media sosial.

METODE

Bagian ini menyajikan desain penelitian, populasi, teknik sampel dan sampling, serta metode analisis yang digunakan dalam penelitian ini.

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan desain *comparative study*. Tujuan eksperimen ini adalah untuk membandingkan dua metode penanganan ketidakseimbangan kelas, yaitu *class weighting* dan SMOTE, dalam konteks klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM). Dengan pendekatan ini, peneliti dapat mengukur secara objektif perbedaan performa antara kedua metode tersebut melalui pengujian yang sistematis dan terkontrol.

Variabel yang digunakan dalam penelitian ini terdiri dari tiga jenis. Variabel independen adalah metode penanganan ketidakseimbangan kelas (*class weighting* dan SMOTE). Variabel dependen meliputi metrik evaluasi performa model yaitu akurasi, presisi, *recall*, dan *F1-score macro*. Sementara itu, variabel kontrol mencakup penggunaan algoritma SVM linear, representasi teks menggunakan TF-IDF, pembagian data menggunakan 10-fold *stratified cross-validation*, pengaturan *random_state=42*, serta penerapan *preprocessing pipeline* untuk menjaga konsistensi proses eksperimen.

Validitas dan reliabilitas penelitian dijaga melalui beberapa strategi. Validitas internal diperkuat dengan penerapan *stratified cross-validation*, *randomization*, dan penggunaan *pipeline* terintegrasi guna mencegah data *leakage*. Validitas eksternal dibatasi pada domain air mineral, platform X, penggunaan bahasa Indonesia informal, serta ukuran *dataset* yang relatif kecil (<200 sampel). Sementara itu, reliabilitas penelitian diukur berdasarkan nilai *standard deviation* dari hasil 10-fold *cross-validation*, dan *reproducibility* dijamin melalui penggunaan nilai *random state* yang tetap.

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif di bidang Ilmu Komputer untuk analisis sentimen teks, dengan fokus pada klasifikasi menggunakan *Support Vector Machine* (SVM). Penelitian ini diawali dengan tahap pengumpulan data ulasan publik yang diambil dari platform media sosial X (sebelumnya dikenal sebagai Twitter) menggunakan metode *web scraping*. Data yang diperoleh berisi opini dan komentar pengguna terkait topik air mineral dan air *demineral*. Tahapan utama meliputi: pengumpulan data, *pra*-pemrosesan, ekstraksi fitur TF-IDF, pemodelan SVM, dan evaluasi.



Gambar 1. Alur penelitian

Tahap Pra-pemrosesan Data

Tahap *pra*-pemrosesan data merupakan langkah awal yang krusial dalam analisis teks. Pada tahap ini, data mentah yang berasal dari ulasan publik umumnya masih mengandung berbagai *noise* seperti tanda baca, URL, angka, huruf kapital yang tidak seragam, serta kata-kata umum yang tidak memiliki makna signifikan terhadap konteks analisis, seperti *stopwords*. Oleh karena itu, *pra*-pemrosesan bertujuan untuk membersihkan dan menormalkan teks agar dapat diproses dengan lebih efektif oleh algoritma pembelajaran mesin. Proses ini diharapkan dapat meningkatkan efisiensi serta akurasi model klasifikasi sentimen yang dibangun.

Seluruh tahapan *pra*-pemrosesan diimplementasikan menggunakan bahasa pemrograman *Python* dengan mengikuti urutan proses yang sistematis. Tahap pertama adalah *case folding*, yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk memastikan konsistensi dalam representasi kata. Tahap berikutnya adalah data *cleaning*, yang bertujuan untuk menghapus elemen-elemen tidak relevan seperti URL, tagar, *mention*, angka, serta tanda baca, sekaligus melakukan normalisasi pada kata-kata yang disingkat, misalnya "yg" diubah menjadi "yang". Setelah teks dibersihkan, dilakukan proses *stemming* menggunakan pustaka Sastrawi. Proses ini berfungsi untuk mengubah setiap kata berimbuhan menjadi bentuk kata dasarnya sehingga mengurangi variasi kata dan menyederhanakan korpus yang akan dianalisis.

Langkah terakhir adalah *stopword removal*, yaitu penghapusan kata-kata umum dalam bahasa Indonesia yang tidak memiliki makna sentimen signifikan, seperti "yang", "dan", atau "di". Daftar *stopword* yang digunakan berasal dari pustaka Sastrawi dan diperluas dengan beberapa kata tambahan yang diidentifikasi tidak relevan selama proses analisis awal. Seluruh tahapan ini merupakan praktik umum dalam pemrosesan bahasa alami (*Natural Language Processing*) berbahasa Indonesia, dan didukung oleh temuan sebelumnya yang menunjukkan bahwa kualitas tahap *pra*-pemrosesan memiliki pengaruh signifikan terhadap performa klasifikasi sentimen[11], [12].

Ekstraksi Fitur

representasi numerik agar bisa diproses oleh algoritma *machine learning*. Penelitian ini menggunakan metode Term *Frequency-Inverse Document Frequency* (TF-IDF), yang menghitung bobot tiap kata berdasarkan frekuensinya dalam satu dokumen dibandingkan seluruh korpus.

Implementasi TF-IDF dalam penelitian ini memanfaatkan fungsi *vectorization* dari pustaka *Scikit-learn*. Proses ini mengubah kumpulan data teks yang telah dibersihkan menjadi representasi matriks numerik, di mana setiap kolom merepresentasikan sebuah kata unik (term) yang diberi bobot TF-IDF. Menurut Sammut and Webb[13], TF-IDF adalah statistik numerik yang mencerminkan seberapa penting sebuah kata dalam dokumen di dalam koleksi atau korpus. Metode ini juga digunakan dalam berbagai penelitian di Indonesia. Misalnya[6] menerapkan TF-IDF untuk pembobotan kata dalam analisis sentimen publik di Twitter menggunakan algoritma SVM. Selain itu, penelitian tentang penerapan algoritma Term *Frequency Inverse-Documents Frequency* untuk *text mining* menjelaskan bahwa TF-IDF efektif dalam *text mining* karena dapat membantu pengguna menemukan informasi penting pada kumpulan dokumen teks[14].

Pemodelan SVM

Klasifikasi sentimen dilakukan menggunakan SVM linear (*LinearSVC*) pada fitur TF-IDF. Untuk mengurangi bias akibat ketidakseimbangan kelas, digunakan pembobotan kelas (*class_weight=balanced*). Penalaan *hiperparameter* dilakukan terhadap koefisien regularisasi C pada himpunan nilai representatif (*mis.* {0.01, 0.1, 1, 10}) dan rentang n-gram pada TF-IDF (*mis.* (1, 1) hingga (1, 2)) dalam *pipeline* terintegrasi *Scikit-learn*.

Penanganan Ketidakseimbangan Kelas

Ketidakeimbangan kelas ditangani terutama melalui pembobotan kelas pada SVM dan validasi silang terstratifikasi. Jika diperlukan, teknik *oversampling* (*mis.* SMOTE) diterapkan

hanya pada lipatan data pelatihan dalam *pipeline* untuk mencegah data *leakage*. Fokus penelitian tetap pada SVM sebagai model tunggal.

Evaluasi Model

Kinerja model SVM dievaluasi menggunakan *K-Fold Cross Validation* terstratifikasi untuk memastikan hasil yang stabil dan dapat digeneralisasi. Implementasi evaluasi menggunakan pustaka *Scikit-learn* menyatakan bahwa *ten-fold stratified cross validation* memberikan *trade-off* terbaik antara bias dan *variance* untuk data dunia nyata, sehingga penelitian ini menggunakan *10-fold stratified cross validation*[15].

Metrik evaluasi yang dilaporkan meliputi akurasi, presisi, *recall*, dan F1-Score; F1-Score disajikan terutama sebagai *macro-average* untuk merefleksikan performa seimbang pada seluruh kelas.

HASIL DAN PEMBAHASAN

Karakteristik Data set

Penelitian ini menggunakan total 140 data ulasan (*tweet*) yang dikumpulkan dari platform X terkait air mineral dan air *demineral*. Setelah proses pelabelan manual, *dataset* ini terdistribusi ke dalam tiga kategori sentimen: positif, negatif, dan netral. Distribusi sentimen ini disajikan pada Tabel 1.

Tabel 1. Distribusi sentimen pada *dataset* (sumber peneliti)

Kategori Sentimen	Jumlah Data	Persentase (%)
Positif	54	38.57
Negatif	19	13.57
Netral	67	47.86
Total	140	100.00

Berdasarkan Tabel 1 diketahui bahwa terdapat ketidakseimbangan kelas yang signifikan, dengan kelas negatif hanya mencakup 13.57% dari total data. Ketidakseimbangan ini menjadi tantangan utama dalam proses klasifikasi.

Contoh Pra-pemrosesan Teks

Tabel 2 menunjukkan transformasi teks melalui empat tahapan: *case folding*, *data cleaning*, *stemming* Sastrawi, dan *stopword removal*.

Tabel 2. Transformasi teks melalui *pra*-pemrosesan (sumber peneliti)

Tahap	Hasil
Mentah	Kalian ada yang <i>ngerti</i> tentang air mineral dan <i>demineral</i> <i>nggak</i> sih? Aku jadi bingung.
<i>Case Folding</i>	kalian ada yang <i>mengerti</i> tentang air mineral dan <i>demineral</i> <i>enggak</i> sih? aku jadi bingung.
<i>Cleaning</i>	kalian ada yang <i>mengerti</i> tentang air mineral dan <i>dimineral</i> <i>tidak</i> sih aku jadi bingung
<i>Stemming</i>	kalian ada yang <i>mengerti</i> tentang air mineral dan <i>demineral</i> <i>tidak</i> sih aku jadi bingung
<i>Stopword</i>	kalian <i>ngerti</i> air mineral <i>demineral</i> sih aku jadi bingung

Klasifikasi SVM

Klasifikasi sentimen dilakukan menggunakan SVM linear (*LinearSVC*) dengan fitur TF-IDF. Model dilatih dengan parameter *class_weight='balanced'* untuk menangani ketidakseimbangan kelas. Evaluasi dilakukan menggunakan validasi silang 10-fold terstratifikasi. Parameter optimal yang diperoleh melalui *Grid Search* adalah: $C = 0.1$, *max_features* = 500, dan *ngram_range* = (1, 1). Hasil kinerja model *baseline* disajikan pada Tabel 3.

Tabel 3. Kinerja model SVM *baseline* dengan validasi 10-fold (sumber peneliti)

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1 Macro (%)
SVM <i>Baseline</i>	57.14	44.55	43.98	42.36

Hasil detail per kelas ditunjukkan pada Tabel 4. Model *baseline* menunjukkan performa yang mencerminkan tantangan klasifikasi pada *dataset* kecil dengan ketidakseimbangan kelas ekstrem. Kelas negatif (13.57% dari *dataset*) memiliki F1-score sangat rendah (8.00%) dengan *recall* hanya 5.00%, menunjukkan model kesulitan mendeteksi sentimen negatif.

Tabel 4. Metrik per kelas untuk model *baseline* (sumber peneliti)

Kelas	Presisi (%)	Recall (%)	F1-Score (%)	Support
Negatif	17.00	5.00	8.00	19
Netral	56.00	79.00	65.84	67
Positif	65.00	48.00	55.32	54

Mengingat performa kelas negatif yang rendah, dilakukan eksperimen dengan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk meningkatkan representasi kelas minoritas. SMOTE diterapkan pada data pelatihan dalam setiap *fold* dengan parameter optimal: $C = 1.0$, *max_features* = 500, dan *ngram_range* = (1, 1). Hasil kinerja model dengan SMOTE disajikan pada Tabel 5.

Tabel 5. Kinerja model SVM dengan SMOTE (sumber peneliti)

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1 Macro (%)
SVM + SMOTE	53.57	42.07	42.02	40.98

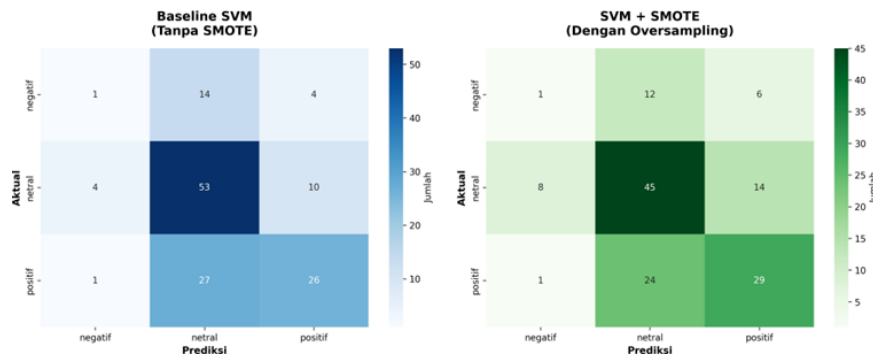
Perbandingan kinerja antara model *baseline* dan model dengan SMOTE disajikan pada Tabel 6. Hasil menunjukkan SMOTE justru menurunkan performa keseluruhan model, dengan penurunan akurasi sebesar 3.57%, F1-score kelas negatif turun 1.10%, dan F1 *macro* turun 1.37%.

Tabel 6. Perbandingan kinerja *baseline* vs SMOTE (sumber peneliti)

Metode	Akurasi (%)	F1 Negatif (%)	F1 Macro (%)
<i>Baseline</i>	57	8.00	42.36
+ SMOTE	53	6.90	40.98
Selisih	-3.57	-1.10	-1.37

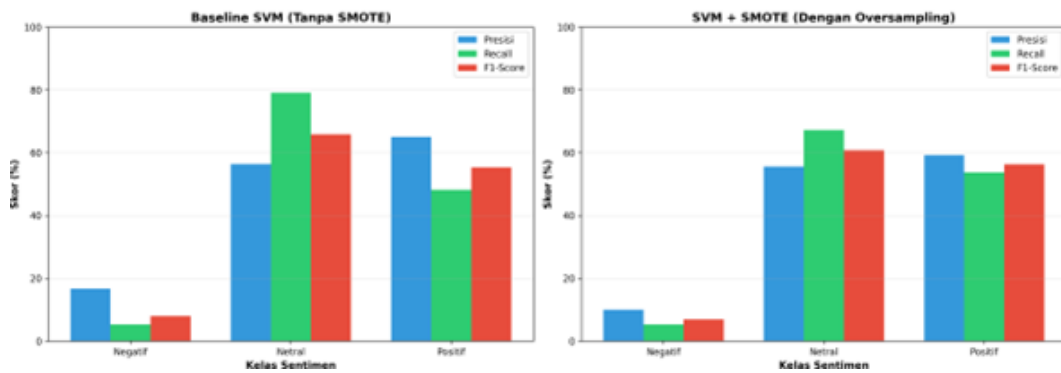
Visualisasi perbandingan kinerja kedua metode ditampilkan pada Gambar 2, Gambar 3, dan Gambar 4. Terlihat bahwa SMOTE tidak memberikan peningkatan konsisten pada semua metrik, bahkan cenderung menurunkan performa secara keseluruhan. Peningkatan pada kelas minoritas tidak signifikan dan tidak mengimbangi penurunan pada kelas lainnya, sehingga performa makro justru menurun.

Gambar 3: Confusion Matrix - Perbandingan Baseline vs SMOTE



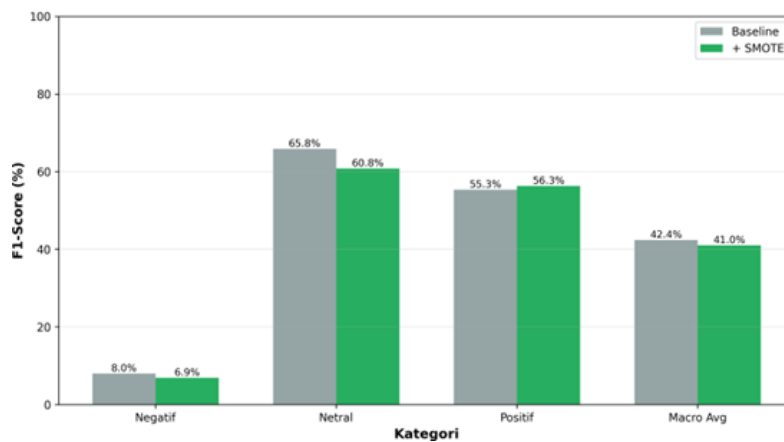
Gambar 2. Perbandingan lengkap *confusion matrix* dan metrik kinerja *baseline vs SMOTE* (sumber peneliti)

Gambar 4: Metrik Per Kelas - Perbandingan Baseline vs SMOTE



Gambar 3: Perbandingan metrik per kelas antara *baseline* dan SMOTE (sumber peneliti)

Gambar 5: Peningkatan F1-Score dengan SMOTE



Gambar 4. Perubahan F1-score per kelas dengan SMOTE (sumber peneliti)

Uji Hipotesis Statistik

Berdasarkan hasil perbandingan di atas, perlu dilakukan uji statistik formal untuk menentukan apakah perbedaan kinerja antara *baseline* dan SMOTE signifikan secara statistik. Analisis dilakukan terhadap nilai F1-score *macro* dari 10-fold *cross-validation*. Ringkasan statistik deskriptif tersaji pada Tabel 7.

Tabel 7. Statistik deskriptif F1-Score *macro* dari 10 fold

Metode	Mean(%)	Std Dev(%)	Min(%)	Max(%)
<i>Baseline</i>	42.36	10.86	25.00	58.33
SMOTE	40.98	11.25	23.08	56.25
Selisih	1.38	-0.39	1.92	2.08

Berdasarkan ambang 2% untuk signifikansi praktis, selisih *mean* F1 *macro* sebesar 1.38% (<2%), sehingga gagal menolak H₀. *Cohen's d* = 0.125 (<0.2, trivial) memperkuat bahwa perbedaan tidak bermakna; SMOTE tidak memberi keuntungan pada *dataset* sangat kecil.

Pembahasan

Pembahasan penelitian ini berfokus pada analisis kinerja model SVM yang digunakan serta dampak penggunaan teknik penanganan ketidakseimbangan data. Hasil eksperimen awal menunjukkan bahwa model SVM *baseline* menghasilkan akurasi 57.14% dan F1-macro 42.36%. Nilai ini menggambarkan tantangan yang muncul ketika bekerja dengan *dataset* yang sangat kecil—hanya 140 sampel—dan memiliki ketidakseimbangan kelas yang tinggi. Kelas negatif yang hanya mencakup sekitar 13.57% dari keseluruhan data menghasilkan F1-score yang sangat rendah, yaitu 8.00%. Dalam kondisi seperti ini, penerapan pembobotan kelas (*class_weight='balanced'*) terbukti lebih efektif karena mampu memberikan perhatian lebih pada kelas minoritas tanpa menambahkan *noise* yang biasanya muncul dari data sintetis.

Namun, ketika SMOTE digunakan sebagai alternatif penanganan ketidakseimbangan, kinerja model justru menurun. Akurasi berkurang 3.57% (dari 57.14% menjadi 53.57%) dan F1-macro turun 1.37% (dari 42.36% menjadi 40.98%). Temuan ini konsisten dengan penelitian Blagus dan Lusa yang menyatakan bahwa SMOTE kurang efektif pada *dataset* dengan ukuran kurang dari 200 sampel. Pada kasus ini, hanya terdapat 19 sampel kelas negatif sehingga proses interpolasi dari SMOTE cenderung menghasilkan *overfitting*[16]. Selain itu, karakteristik sentimen Twitter yang sering tumpang tindih antar kelas serta kemunculan *noise* akibat sintesis data pada korpus berukuran kecil turut memperburuk performa model.

Perbandingan dengan Penelitian Sejenis Hasil penelitian ini menunjukkan perbedaan signifikan dengan studi-studi sebelumnya yang menggunakan *dataset* lebih besar. [5] mencapai akurasi 85% pada analisis sentimen platform X dengan lebih dari 1.000 sampel, sementara [7] melaporkan akurasi 82% untuk SVM pada *dataset* ulasan produk *marketplace*. Perbedaan performa ini dapat dijelaskan oleh tiga faktor utama: (1) ukuran *dataset*—penelitian ini hanya menggunakan 140 sampel dibandingkan ribuan sampel pada studi lain, (2) tingkat ketidakseimbangan—kelas minoritas pada penelitian ini hanya 13.57% dibandingkan distribusi yang lebih merata pada studi lain, dan (3) kompleksitas domain—

sentimen terhadap air minum cenderung lebih ambigu dan subjektif dibandingkan sentimen terhadap produk konsumen atau isu politik yang lebih jelas polaritasnya.

Temuan tentang *ketidakefektifan* SMOTE pada *dataset* kecil konsisten dengan peringatan [16] yang menyatakan bahwa SMOTE memerlukan minimal 200 sampel per kelas untuk menghasilkan data sintetis yang berkualitas. [10] melaporkan peningkatan *recall* hingga 15% dengan SMOTE, namun penelitian mereka menggunakan *dataset* dengan lebih dari 1.000 sampel—jauh lebih besar dari penelitian ini. Hal ini mengkonfirmasi bahwa efektivitas teknik *oversampling* sangat bergantung pada ukuran *dataset* awal. Penelitian ini memberikan bukti empiris tambahan bahwa untuk *dataset* sangat kecil (< 150 sampel), *class weighting* merupakan pilihan yang lebih aman dan stabil dibandingkan *oversampling* sintetis.

Kontribusi utama penelitian ini terhadap literatur yang ada adalah penyediaan *baseline* metodologis untuk analisis sentimen pada kondisi data tidak ideal—sebuah skenario yang sering dihadapi oleh peneliti pemula atau saat mengeksplorasi domain baru. Meskipun performa absolut lebih rendah dari studi pada *dataset* besar, temuan ini tetap memberikan nilai praktis: (1) memvalidasi bahwa SVM dengan *class weighting* tetap dapat menghasilkan klasifikasi yang bermakna bahkan pada data sangat terbatas, (2) memberikan peringatan eksplisit tentang risiko menggunakan SMOTE pada *dataset* kecil, dan (3) menyediakan protokol eksperimental yang dapat *direplikasi* untuk domain-domain spesifik lain yang memiliki tantangan serupa.

Implikasi Praktis dan Keterbatasan Meskipun F1-macro sebesar 42.36% tergolong rendah dibandingkan penelitian pada *dataset* besar[5], [7], hasil ini tetap dapat diterima mengingat berbagai keterbatasan yang ada, seperti ukuran *dataset* yang sangat kecil, ketidakseimbangan kelas yang ekstrem, sifat teks Twitter yang informal, dan tidak adanya *inter-annotator agreement* pada proses pelabelan. Untuk konteks praktis, temuan ini memberikan tiga implikasi penting: (1) peneliti yang bekerja dengan data terbatas sebaiknya memprioritaskan *class weighting* dibandingkan SMOTE, (2) ekspektasi performa harus disesuaikan dengan ukuran dan kualitas *dataset*, dan (3) pengumpulan data tambahan menjadi prioritas utama untuk meningkatkan kinerja model secara signifikan.

Penelitian ini menyediakan *baseline* awal untuk penelitian lanjutan di domain air mineral dan memberikan pemahaman mengenai tantangan analisis sentimen pada media sosial Indonesia dengan topik yang relatif baru. Walaupun demikian, beberapa keterbatasan perlu diperhatikan. *Dataset* yang hanya terdiri dari 140 sampel belum cukup untuk menghasilkan model yang benar-benar *robust*. Ketidakseimbangan kelas yang sangat ekstrem juga menyulitkan proses pembelajaran terutama pada kelas minoritas. Selain itu, penggunaan TF-IDF tidak mampu menangkap aspek semantik yang lebih dalam, dan tidak adanya *inter-annotator agreement* dapat menurunkan konsistensi label. Penelitian ini juga hanya mengevaluasi satu jenis model, yaitu SVM linear, sehingga ruang eksplorasi model lain masih terbuka untuk pengembangan berikutnya.

SIMPULAN

Penelitian ini memberikan kontribusi empiris terhadap literatur analisis sentimen berbahasa Indonesia dengan fokus pada kondisi data tidak ideal—*dataset* sangat kecil dengan ketidakseimbangan kelas ekstrem. Hasil menunjukkan bahwa model SVM dengan

representasi TF-IDF masih mampu melakukan klasifikasi sentimen pada opini terkait air mineral meskipun bekerja pada *dataset* yang hanya terdiri dari 140 sampel. Dengan *class_weight='balanced'*, model *baseline* menghasilkan akurasi 57.14% dan F1-macro 42.36% melalui 10-fold *stratified cross-validation*. Nilai ini memang lebih rendah dibanding penelitian lain yang menggunakan ribuan data, namun tetap mencerminkan kemampuan SVM yang stabil ketika berhadapan dengan fitur *sparse* seperti TF-IDF. Rendahnya performa pada kelas minoritas—ditandai dengan F1-score hanya 8%—menunjukkan bahwa ukuran *dataset* yang kecil dan ketidakseimbangan ekstrem (kelas negatif hanya 13.57%) sangat membatasi proses pembelajaran model.

Temuan utama penelitian ini adalah konfirmasi bahwa SMOTE tidak memberikan keuntungan signifikan pada *dataset* sangat kecil. Hasil uji statistik menunjukkan *Cohen's d* = 0.125 dengan selisih F1-macro hanya 1.38%, bahkan cenderung menurunkan performa dibanding pembobotan kelas. Temuan ini konsisten dengan literatur yang menyatakan bahwa *oversampling* sintesis kurang efektif pada korpus berukuran di bawah 200 sampel[16]. *Novelty* penelitian ini terletak pada validasi empiris efektivitas *class weighting* versus SMOTE pada *dataset* sangat kecil dalam konteks analisis sentimen berbahasa Indonesia, khususnya pada domain air mineral yang belum pernah diteliti sebelumnya. Penelitian ini menyediakan *baseline* metodologis dan rekomendasi praktis untuk peneliti yang menghadapi keterbatasan data serupa.

Perbaikan selanjutnya disarankan untuk menggunakan *dataset* lebih besar (≥ 500 per kelas), menerapkan fitur lanjutan seperti *lexicon* berbasis sentimen atau *embeddings* (Word2Vec, FastText, IndoBERT), mengeksplorasi model *ensemble* atau *deep learning*, menerapkan pendekatan *semi-supervised learning*, melakukan analisis multi-aspek atau multi-sumber, serta memvalidasi label dengan *inter-annotator agreement* (IAA) untuk meningkatkan konsistensi dan kualitas data pelatihan. Penelitian lanjutan juga dapat mengeksplorasi teknik *active learning* untuk optimalisasi pengumpulan data pada domain spesifik dengan sumber daya terbatas.

DAFTAR PUSTAKA

- [1] R. Salim and T. Taslim, "Edukasi manfaat air mineral pada tubuh bagi anak sekolah dasar secara online," *J. Pengabd. Kpd. Masy.*, vol. 27, no. 2, pp. 126–135, 2021.
- [2] F. Kozisek, "Health risks from drinking demineralised water," in *Nutrients in Drinking Water*, vol. 1, 2005, pp. 148–163.
- [3] D. Sunardi *et al.*, "Health effects of alkaline, oxygenated, and demineralized water compared to mineral water among healthy population: A systematic review," *Rev. Environ. Health*, vol. 39, no. 2, pp. 339–349, 2024.
- [4] D. Briawan, T. R. Sedayu, and I. Ekayanti, "Kebiasaan minum dan asupan cairan remaja di perkotaan," *J. Gizi Klin. Indones.*, vol. 8, no. 1, pp. 36–41, 2011.
- [5] Z. Maharani, A. Luthfiarta, and N. Farsya, "Sentiment Analysis of the 2024 Indonesian Presidential Dispute Trial Election using SVM and Naïve Bayes on Platform X," *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 440–449, 2024, doi: 10.47065/bits.v6i1.5380.
- [6] M. Q. H. Octava, D. G. P. Putri, F. M. Hilmy, U. Farooq, R. A. Nurhaliza, and G. Alfian, "Web-based sentiment analysis system using SVM and TF-IDF with statistical

- feature,” in *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2023, pp. 466–471.
- [7] N. Z. B. Jannah and K. Kusnawi, “Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces,” *Sinkron*, vol. 8, no. 2, pp. 727–733, 2024, doi: 10.33395/sinkron.v8i2.13559.
- [8] J. M. Johnson and T. M. Khoshgoftaar, “The class imbalance problem in deep learning,” *Mach. Learn.*, vol. 111, pp. 4845–4888, 2022, doi: 10.1007/s10994-022-06268-8.
- [9] B. Krawczyk, “Learning from imbalanced data: open challenges and future directions,” *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016, doi: 10.1007/s13748-016-0094-0.
- [10] J. Mantik and Y. A. Singgalen, “Performance evaluation of SVM with synthetic minority over-sampling technique in sentiment classification,” *Mantik J.*, vol. 8, no. 1, pp. 1–10, 2024.
- [11] A. W. Pradana and M. Hayaty, “The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on Indonesian-language texts,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 4, pp. 375–380, 2019.
- [12] S. Shevira, I. Suarjaya, and P. Buana, “Pengaruh kombinasi dan urutan pre-processing pada tweets Bahasa Indonesia,” *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 2, pp. 78–87, 2022.
- [13] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning and Data Mining*, 2nd ed. Springer Publishing Company, 2017.
- [14] H. Sari, G. L. Ginting, T. Zebua, and Mesran, “Penerapan algoritma term frequency inverse-document frequency untuk text mining,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 4, no. 1, pp. 100–107, 2021.
- [15] R. Kohavi and others, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1995, vol. 14, no. 2, pp. 1137–1145.
- [16] R. Blagus and L. Lusa, “SMOTE for high-dimensional class-imbalanced data,” *BMC Bioinformatics*, vol. 14, no. 1, pp. 1–16, 2013.