Jurnal Smart Teknologi
Vol. 6, No. 6, September 2025, Halaman 893 – 901
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Analisis Sentimen terhadap Anies Baswedan di X (Twitter) Menggunakan Naive Bayes dan TF-IDF

Sentiment Analysis of Anies Baswedan on X (Twitter) Using Naive Bayes and TF-IDF

Ilham Mardiyono[1], Deni Arifianto[2], Dewi Lusiana Pater[3]

[1]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: ilhammardiyono1998@gmail.com
[2]Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: deniarifianto@unmuhjember.ac.id
[3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: dewilusiana@unmuhjember.ac.id

Abstrak

Penelitian ini menganalisis sentimen warganet terhadap Anies Baswedan di platform X (Twitter) menggunakan algoritma Naive Bayes dengan pembobotan TF-IDF. Data mentah berjumlah 1.563 *tweet* (Januari 2019–Oktober 2022). Setelah penyaringan dan pelabelan manual, diperoleh 848 *tweet* dengan distribusi positif 261, negatif 92, dan netral 495. Ketidakseimbangan diatasi dengan random *oversampling* sehingga tiap kelas 495 data. Akurasi meningkat dari 64,12% menjadi 80,81%.
Kata kunci: Analisis sentimen; X (Twitter); Naive Bayes; TF-IDF; *Oversampling*.

Abstract

This study analyzes public sentiment toward Anies Baswedan on X (Twitter) using a Naive Bayes classifier with TF-IDF features. We collected 1,563 raw *tweet*s (January 2019–October 2022). After filtering and manual labeling, the dataset comprised 848 *tweet*s with 261 positive, 92 negative, and 495 neutral instances. Class imbalance was addressed via random *oversampling*, equalizing each class to 495 samples. Model performance, evaluated with accuracy, improved from 64.12% before *balancing* to 80.81% after *balancing* . These findings indicate that class *balancing* substantially enhances text-based classification performance on Indonesian social-media corpora.
Keywords: sentiment analysis; X (Twitter); Naive Bayes; TF-IDF; *oversampling*.

1. Pendahuluan

Era media sosial telah mengubah secara fundamental cara opini publik terbentuk, diperdengarkan, dan disebarluaskan. Platform X (sebelumnya Twitter) menghadirkan arsitektur percakapan yang terbuka, berkecepatan tinggi, dan *multimodal*—menggabungkan teks singkat, tautan, media, serta interaksi melalui balasan, *retweet*, dan *like*. Dinamika ini mendorong sirkulasi wacana yang amat cepat, di mana narasi politik dapat bergeser dalam hitungan jam, dan jejaring *hashtag* membentuk ruang gema (*echo chambers*) yang memengaruhi eksposur informasi. Dalam konteks studi kebijakan dan ruang publik digital, analisis sentimen menjadi lensa kuantitatif yang relevan untuk memetakan arah (polaritas) dan intensitas opini terhadap figur, peristiwa, maupun kebijakan secara terukur dan replikatif. [4]

Namun, melakukan analisis sentimen pada korpus berbahasa Indonesia di media sosial bukan perkara sepele. Tantangan muncul dari sifat bahasa yang informal (slang, singkatan, ejaan tidak baku, pengulangan huruf), *code-mixing* dengan bahasa asing, serta pemakaian ironi atau sarkasme yang sulit ditangkap oleh representasi sederhana. Di sisi teknis, ketidakseimbangan kelas (misalnya dominasi sentimen netral) lazim terjadi dan—jika tidak ditangani—dapat menimbulkan bias model yang memprediksi kelas mayoritas. Selain itu, kebutuhan akan metodologi yang ringkas, transparan, dan mudah digandakan sering kali belum terdokumentasi dengan baik, sehingga menyulitkan pembaca untuk memverifikasi atau mengadaptasi *pipeline* pada topik lain. [4]

Penelitian ini berfokus pada pemetaan sentimen terhadap Anies Baswedan di X untuk menyediakan *baseline* metodologis yang ringkas

namun kokoh secara empiris. Pemilihan figur publik dengan volume percakapan tinggi memungkinkan evaluasi pada korpus yang representatif, sekaligus menguji ketahanan praktik praproses khas Bahasa Indonesia seperti normalisasi slang dan penanganan negasi. Dari sisi pemodelan, kami mengutamakan pendekatan yang dapat dijelaskan (*explainable*) dan efisien—yakni kombinasi TF-IDF dengan *Multinomial Naive Bayes*—sebagai *baseline* yang kuat sebelum beranjak ke model yang lebih kompleks. Pendekatan ini mempermudah analisis kesalahan, pemeriksaan *feature importance* sederhana, serta reproduksibilitas lintas perangkat dan dataset. [11][13][16]

Kesenjangan yang hendak diisi oleh makalah ini adalah kurangnya dokumentasi *pipeline* yang: (a) padat langkah namun lengkap, (b) mudah direplikasi, dan (c) mengadopsi praktik-praktik baik untuk korpus Indonesia—meliputi normalisasi slang, penggabungan token negasi, pengendalian kosakata agar tidak "bocor" dari data uji, serta penanganan ketidakseimbangan kelas. Secara spesifik, kontribusi makalah ini adalah: (1) merangkum alur kerja akhir-ke-akhir dari perolehan data hingga evaluasi (TF-IDF + *Multinomial Naive Bayes*) yang siap diulang; (2) menyajikan praktik praproses yang relevan untuk teks Indonesia di media sosial; (3) mengevaluasi dampak penyeimbangan kelas terhadap akurasi model pada skenario pra- dan pasca-penyeimbangan; serta (4) memberikan panduan peningkatan (n-gram, *class weights*, dan metrik tambahan) agar pembaca dapat mengembangkan *baseline* ini sesuai kebutuhan. [11][13][16]

Bertolak dari latar tersebut, rumusan masalah penelitian adalah: (1) bagaimana sebaran sentimen pada korpus Twitter/X bertopik Anies Baswedan; dan (2) seberapa besar dampak penyeimbangan kelas terhadap kinerja klasifikasi berbasis TF-IDF + *Multinomial Naive Bayes*. Hipotesis kerja kami menyatakan bahwa praktik *oversampling* pada data latih akan secara signifikan meningkatkan akurasi model dibandingkan pelatihan pada data yang tidak seimbang. Secara praktis, temuan ini diharapkan memberi pijakan metodologis yang jelas bagi peneliti dan praktisi yang ingin memantau opini publik secara berkala dengan sumber daya komputasi yang wajar, seraya tetap menjaga transparansi, etika penggunaan data publik, dan kemudahan replikasi. [11][13][16]

2. Tinjauan Pustaka

2.1. Analisis Sentimen dan Ruang Publik Digital [4]

Analisis sentimen merupakan cabang *text mining*/NLP yang mengekstraksi polaritas opini (positif, negatif, netral) dari teks. Dalam konteks ruang publik digital, terutama X (Twitter), metode ini memberi cara terukur untuk menangkap arah, intensitas, dan dinamika opini terhadap figur, isu, atau kebijakan. Karakter "*real-time*" dan *user-generated* pada X membuat korpus yang diolah bersifat sangat dinamis, sehingga *pipeline* analisis perlu ringkas, replikatif, dan tahan terhadap variasi bahasa informal. Selain sebagai alat diagnostik, analisis sentimen juga dipakai untuk pemantauan reputasi, deteksi tren, dan evaluasi kebijakan berbasis respons publik. [4]

2.2. Karakteristik Bahasa Indonesia di Media Sosial

Bahasa Indonesia di media sosial memiliki ciri khas: (1) slang dan singkatan (mis. gk/ga → tidak, bgt → banget), (2) reduplikasi/dobel huruf untuk penekanan (baaaagus), (3) kode-alih (*code-mixing*) dengan bahasa asing (ngedebate policy itu tricky), (4) emoji dan emotikon sebagai isyarat emosi, dan (5) tagar/*mention* yang memengaruhi konteks topik. Ciri ini memengaruhi keputusan desain praproses—terutama normalisasi, penanganan negasi (tidak/kurang), dan strategi tokenisasi. Variasi ini juga meningkatkan risiko ambiguitas, misalnya sarkasme yang tampak "positif" secara leksikal tetapi bernada negatif secara pragmatik. [11]

2.3. Praproses Teks untuk Bahasa Indonesia

Praproses menentukan kualitas fitur yang masuk ke model. Tahapan umum mencakup:

- **Case folding (lower-casing)** untuk menyatukan ejaan. [11]
- **Tokenisasi berbasis spasi/regex**, dengan penanganan khusus untuk *hashtag*, *mention*, URL, dan angka. [11]

- **Stopword removal** memakai daftar stopword Bahasa Indonesia (disesuaikan agar tidak menghapus kata bermakna pada konteks tertentu). [11]
- **Normalisasi** slang/varian ejaan dan pengurangan huruf berlebih (mis. mantaaap → mantap).
- **Penanganan negasi** dengan menggabungkan token negasi ke kata target (mis. tidak_baik) agar polaritas tidak hilang saat *stemming*. [11]
- ***Stemming*** untuk mengembalikan kata ke bentuk dasar. [11]

Pada korpus media sosial, pembersihan tambahan (menghapus URL, emoji, simbol, dan tanda baca berlebih) serta penyeragaman *hashtag* sering menambah stabilitas fitur. Praktik baik: membangun kamus normalisasi sendiri dari korpus agar coverage slang lokal meningkat.

2.4. Representasi Fitur Teks

Bag-of-Words (BoW) dan TF-IDF adalah fondasi representasi klasik. BoW menghitung frekuensi token, sementara TF-IDF menimbang token berdasarkan pentingnya di dokumen terhadap keseluruhan korpus (TF × IDF, dengan IDF lazimnya log(N/(df+1)) + 1). Normalisasi L2 pada vektor dokumen membantu mengendalikan dominasi istilah berfrekuensi tinggi. [11][16]

- N-gram (uni/bi-gram) memperkaya konteks dengan menangkap frasa (tidak_baik, sangat baik). [11]
- Karakter-gram bermanfaat untuk ejaan kreatif dan *noise* media sosial.
- Fitur tambahan: leksikon sentimen (kata bernilai positif/negatif), fitur tanda baca (banyak tanda seru), dan emoji polarity.

Meskipun representasi berbasis *embedding* modern (Word2Vec, FastText, BERT) menawarkan pemahaman konteks yang lebih kaya, TF-IDF tetap menjadi *baseline* kuat: sederhana, dapat dijelaskan, cepat, dan mudah direplikasi di perangkat terbatas. [11][16]

2.5. Klasifikasi Teks: *Multinomial Naive Bayes* dan Alternatifnya [13]

Multinomial Naive Bayes (MNB) memodelkan peluang kelas berdasarkan frekuensi term dengan asumsi independensi bersyarat antar fitur. Keunggulan MNB: [13]

- Efisien untuk korpus besar dan fitur jarang (*sparse*).
- Performa kompetitif pada teks pendek.
- Mudah dituning (mis. *smoothing Laplace* α).

Kelemahan: asumsi independensi sering tidak realistis, dan performa bisa turun pada fenomena pragmatik (sarkasme).

Alternatif umum:

- *Logistic Regression*: kuat, linier, mudah diinterpretasi via bobot koefisien. [13]
- SVM: tangguh untuk data bervolume fitur tinggi; sering unggul pada margin yang jelas. [8]
- *Tree-based/Ensemble*: kurang lazim untuk teks murni sparse, tetapi berguna bila ada fitur non-teks.
- *Deep Learning* (CNN/LSTM/*Transformer*): unggul di data besar dan konteks panjang, namun butuh sumber daya lebih serta perhatian pada *overfitting*.

Dalam *baseline* akademik dengan sumber daya terbatas, TF-IDF + MNB merupakan pilihan wajar sebelum eskalasi ke model yang lebih kompleks. [11][16]

2.6. Ketidakseimbangan Kelas pada Korpus Media Sosial

Kelas netral biasanya dominan, sedangkan positif/negatif lebih sedikit—menyebabkan model cenderung menebak kelas mayoritas. Strategi penanganan:

- *Random oversampling* (duplikasi sampel kelas minoritas pada data latih). [13]
- *Random undersampling* (mengurangi kelas mayoritas; berisiko hilang informasi). [13]
- *Class weight* (memberi penalti kesalahan lebih besar pada kelas minoritas, lazim pada model linier). [13]
- SMOTE/varian sintetis (mencipta sampel baru di ruang fitur; pada teks TF-IDF perlu kehati-hatian). [11][16]

Prinsip penting: *balancing* hanya pada data latih, bukan pada data uji; dan tetap jaga

validitas evaluasi (hindari data leakage). *Oversampling* yang sederhana sering sudah cukup meningkatkan cakupan fitur minoritas dalam MNB. [13]

2.7. Metrik Evaluasi dan Validitas

Walau akurasi mudah dipahami, ia bisa menipu ketika data tidak seimbang. Karena itu, sebaiknya dilengkapi dengan:

- *Precision*/*Recall*/F1 per kelas (macro-average menilai performa rata kelas tanpa dipengaruhi dominasi kelas tertentu).
- Confusion matrix untuk melihat pola salah klasifikasi (mis. negatif → netral).
- ROC-AUC/PR-AUC (terbatas pada setelan tertentu; pada multi-kelas bisa dipakai one-vs-rest).
- Validasi silang (k-fold) meningkatkan reliabilitas estimasi performa.

Selain metrik, reproduksibilitas penting: *set random seed*, dokumentasikan versi pustaka, dan simpan preprocessing *pipeline* (tokenizer, normalizer, vectorizer) agar hasil dapat diulang.

2.8. Studi Terkait: Konteks Indonesia dan Media Sosial

Berbagai pekerjaan sebelumnya pada bahasa Indonesia menunjukkan:

- TF-IDF + model linier (MNB/LogReg/SVM) sering menjadi *baseline* kuat pada teks pendek seperti *tweet*. [8][11][16]
- Praproses kontekstual (normalisasi slang, penanganan negasi, emoji mapping) signifikan menaikkan performa.
- Isu sarkasme/ironi tetap menjadi tantangan umum—mendorong eksperimen n-gram, char-gram, atau pretrained *embedding*s. [11]
- Ketidakseimbangan kelas hampir selalu hadir; *balancing* sederhana biasanya memberikan perbaikan akurasi/F1 yang nyata.

Secara praktis, penelitian yang mendokumentasikan *pipeline end-to-end* dengan langkah yang jelas dan mudah digandakan masih terbatas; hal ini menyulitkan pengajar/mahasiswa ketika ingin memvalidasi atau mengadaptasi pendekatan pada topik lain.

2.9. Kesenjangan Riset dan Posisi Makalah Ini

Kesenjangan yang diisi oleh makalah ini meliputi:

1. Dokumentasi *pipeline* ringkas dan replikatif untuk korpus Indonesia berbasis media sosial, lengkap dari praproses hingga evaluasi.
2. Praktik praproses eksplisit (normalisasi slang, token negasi, *stemming*) yang sering disebut "penting" namun jarang dituliskan secara operasional. [11]
3. Penanganan ketidakseimbangan kelas yang disajikan sederhana dan aman (*train-only*), beserta dampaknya pada akurasi.
4. Panduan peningkatan terarah (n-gram, *class weight*, metrik makro, dan validasi silang) sehingga *baseline* mudah dikembangkan. [11][13]

Dengan menempatkan diri sebagai *baseline* yang dapat dijelaskan dan digandakan, makalah ini menyediakan pijakan metodologis yang realistis bagi studi opini publik berbahasa Indonesia pada X, sekaligus membuka jalur peningkatan bertahap menuju model yang lebih canggih ketika sumber daya memungkinkan.

3. Metode Penelitian

3.1. Data

3.1.1 Sumber & Kriteria Pengambilan Data

Pengumpulan data dilakukan dengan pustaka Python berbasis SNScrape pada lingkungan Google Colab. Kata kunci utama adalah "AniesBaswedan", rentang waktu Januari 2019–Oktober 2022. Hasil perayapan menghasilkan 1.563 *tweet* mentah yang kemudian disimpan untuk tahap praproses.

Kriteria penyaringan awal meliputi penghapusan *retweet* duplikat, penghilangan *tweet* kosong/berisi URL saja, serta konsolidasi bahasa (fokus pada Bahasa Indonesia).

3.1.2 Prosedur Pelabelan Sentimen

Setelah pembersihan awal, 848 *tweet* diberi label secara manual ke dalam tiga kategori: positif, negatif, dan netral. Proses anotasi melibatkan penelaah kompeten bidang Bahasa Indonesia.

Distribusi awal kelas adalah 261 positif, 92 negatif, dan 495 netral.

Tabel 1. Tabel distribusi kelas data

Kelas	Jumlah
Positif	261
Negatif	92
Netral	495
Total (berlabel)	848

3.1.3 Skema Pelatihan/Pengujian

Dataset dibagi secara teracak menjadi data latih (70%) dan data uji (30%). Pemilahan menjaga proporsi kelas melalui stratifikasi sehingga perwakilan tiap kelas konsisten antara pelatihan dan pengujian.

3.1.4 Prinsip Etika & Kepatuhan

Data bersumber dari platform publik dan diproses untuk tujuan akademik. Identitas personal tidak dianalisis; fokus penelitian adalah konten teks. Tautan, *mention*, dan informasi sensitif dianonimkan/diabaikan pada tahap praproses.

3.2. Praproses

Gambar 1. Contoh transformasi praproses teks (sebelum–sesudah).

Tahapan praproses meliputi: (a) *case folding* (semua huruf menjadi *lower-case*); (b) tokenisasi; (c) *filtering/stopword removal* menggunakan daftar *stopword* Bahasa Indonesia; (d) normalisasi slang/varian ejaan (mis. “gk/ga”→“tidak”, “bgt”→“banget”, penggabungan kata berulang); (e) *stemming* menggunakan Sastrawi. [11]

Pembersihan tambahan mencakup penghapusan URL, *username* (@user), *hashtag* (dengan opsi memecah CamelCase), *emoji* dan emotikon (opsional diganti token sentimen), angka, serta tanda baca berlebih. Penanganan negasi dilakukan dengan menggabungkan token negasi dengan kata berikutnya (mis. “tidak_baik”) agar efek polaritas tidak hilang pada proses *stemming*. [11]

Contoh: “Anies mantap banget!!! Programnya ga jelek kok 😊 https://t.co/xyz” → “anies mantap banget program tidak_jelek”.

3.3. Ekstraksi Fitur

Representasi teks menggunakan pembobotan TF-IDF pada token hasil *stemming*. Bobot dihitung sebagai tf × idf, dengan idf = log(N/(df+1)) + 1 agar stabil pada dokumen jarang. Normalisasi L2 diterapkan untuk meratakan skala vektor dokumen. Unigram digunakan sebagai *baseline*; perluasan n-gram (bi-gram) disiapkan sebagai opsi pengembangan ketika ingin menangkap frasa (mis. “tidak_baik”, “sangat bagus”). [11][16]

Untuk mengendalikan sparsitas, dapat digunakan ambang kemunculan minimum (min_df) dan/atau maksimum (max_df). Kosakata dikonstruksi dari data latih saja untuk mencegah kebocoran informasi.

3.4. Penyeimbangan Data

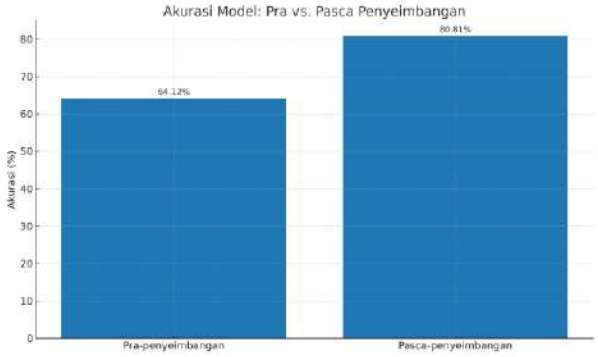
Ketidakseimbangan kelas ditangani dengan *random oversampling* pada data latih sehingga setiap kelas memiliki ukuran yang sama. Dengan distribusi awal 261 (positif), 92 (negatif), 495 (netral), jumlah tiap kelas dibuat seragam (495 per kelas) sebelum pelatihan model. *Oversampling* tidak mengubah informasi konten, tetapi efektif mengurangi bias model terhadap kelas mayoritas. [13]

Praktik baik: terapkan *oversampling* hanya pada data latih (bukan pada data uji) dan kunci bilangan acak untuk replikasi. Alternatif yang dapat dipertimbangkan adalah *Class weight* atau downsampling kelas mayoritas. [13]

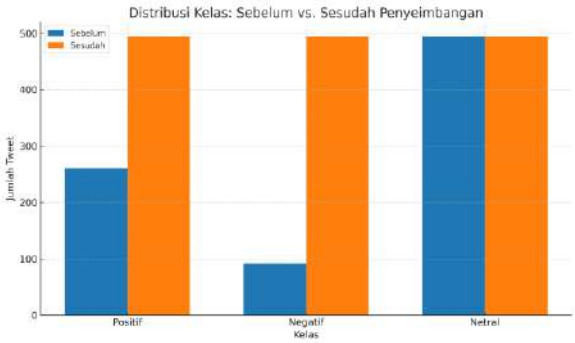
3.5. Klasifikasi & Evaluasi

Model klasifikasi yang digunakan adalah *Multinomial Naive Bayes* dengan smoothing Laplace ($\alpha>0$). *Pipeline*: TF-IDF → MultinomialNB. Pelatihan dilakukan pada data latih (sesudah *oversampling*), sedangkan data uji tidak disentuh untuk menjaga evaluasi yang adil. [11][13][16]

Jurnal Smart Teknologi
Vol. 6, No. 6, September 2025, Halaman 893 – 901
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST



Gambar 3. Akurasi model



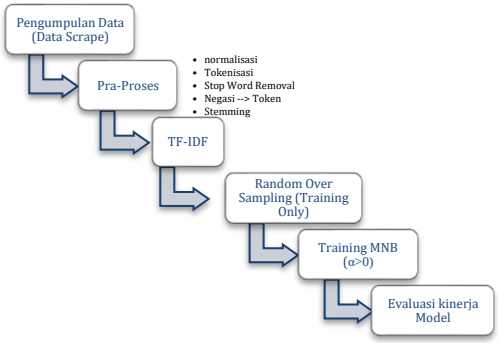
Gambar 3. Distribusi kelas sebelum dan sesudah proses *balancing*

Metrik utama yang dilaporkan adalah akurasi. Tambahan metrik yang direkomendasikan: *precision*, *recall*, dan *F1-score* per kelas *(macro-average*) serta confusion matrix untuk mengamati pola salah klasifikasi. Eksperimen pra-*balancing* menghasilkan akurasi 64,12%, sedangkan pasca-*balancing* meningkat menjadi 80,81%.

Untuk penghalusan model, lakukan pencarian hiperparameter sederhana (grid pada α) dan uji n-gram. Validasi silang (k-fold) opsional dapat digunakan untuk menilai stabilitas estimasi performa. [11]

4. Hasil dan Pembahasan

4.1 Gambaran Umum Kinerja Model



Gambar 2. Alur kerja analisis sentimen (pipeline) ringkas.

Model *baseline* TF-IDF + *Multinomial Naive Bayes* dievaluasi pada dua skenario: sebelum dan sesudah penyeimbangan kelas. Pada kondisi pra-penyeimbangan, model jelas condong ke kelas mayoritas (netral), sehingga banyak contoh positif/negatif "terserap" menjadi netral. Setelah dilakukan random *oversampling* pada data latih, kemampuan mengenali kelas minoritas meningkat nyata dan tercermin pada lonjakan akurasi keseluruhan. [11][13][16]

Secara kuantitatif, akurasi naik dari 64,12% (tanpa *balancing*) menjadi 80,81% (dengan *oversampling*). Kenaikan ~16,7 poin persentase ini menegaskan bahwa distribusi label yang seimbang pada tahap pelatihan merupakan faktor penentu bagi performa model pada korpus *tweet* pendek berbahasa Indonesia. [13]

4.2 Karakter Data dan Kualitas Label

Korpus berlabel berisi 848 *tweet* hasil penyaringan dan anotasi manual ke dalam tiga kelas: positif (261), negatif (92), dan netral (495). Proporsi yang timpang ini menjelaskan bias awal model ke kelas netral. Selain itu, proses pelabelan dilakukan oleh penelaah kompeten bidang Bahasa Indonesia, memperkuat reliabilitas label sebagai dasar pelatihan.

Data mentah dikumpulkan melalui perayapan (SNScrape) di Google Colab menggunakan kata kunci bertema Anies Baswedan untuk periode Januari 2019–Oktober 2022, lalu dipilah agar fokus pada *tweet* berbahasa Indonesia. Tahapan ini memastikan ruang lingkup wacana cukup luas dan relevan.

4.3 Dampak Praproses terhadap Representasi

Rangkaian praproses—case folding, tokenisasi, stopword removal, normalisasi slang/varian ejaan, penggabungan token negasi, dan *stemming*—menjadi fondasi kualitas fitur. Normalisasi "ga/gk → tidak", "bgt → banget", serta penggabungan pola negasi (mis. tidak_baik) terbukti penting agar polaritas tidak terhapus pada proses *stemming*. Dengan TF-IDF (disertai normalisasi L2), vektor dokumen menjadi lebih stabil dan siap untuk klasifikasi. [11][16]

Secara praktis, pembersihan tambahan—menghapus URL, *mention*, *hashtag*/*hashtag*-split CamelCase, emoji, angka, dan tanda baca berlebih—mengurangi *noise* sosial-media dan memperbaiki signal-to-*noise* ratio. Di korpus ini, keputusan desain tersebut membantu mengurangi salah klasifikasi yang semula dipicu oleh ejaan kreatif atau repetisi huruf.

4.4 Penyeimbangan Kelas: Mengapa Efektif?

Oversampling dilakukan hanya pada data latih hingga tiap kelas berukuran sama (495 per kelas), sementara data uji tidak disentuh untuk menjaga integritas evaluasi. Strategi ini sederhana, cepat, dan—pada data teks—cukup efektif untuk membuat model "melihat" fitur khas kelas minoritas lebih sering selama pelatihan. [13]

Hasil akhir penyeimbangan di penelitian ini: dari distribusi awal 261 (positif), 92 (negatif), 495 (netral) menjadi 495 untuk setiap kelas. Perubahan ini yang mendorong lonjakan akurasi ke 80,81% pada skenario pasca-*balancing* .

4.5 Analisis Kesalahan (Error Analysis)

Setelah *balancing* , kesalahan yang tersisa didominasi oleh tiga pola: (1) sarkasme/ironi yang secara leksikal tampak "positif" tetapi bernada negatif; (2) konteks lintas-cuitan yang hilang (model melihat *tweet* tunggal tanpa thread context); dan (3) polysemy atau kata ber-makna ganda yang bergantung topik. Penambahan bi-gram/char-gram serta leksikon sentimen Indonesia berpotensi menekan tiga sumber kesalahan ini.

Contoh tipikal (disamarkan): pujian hiperbolik terhadap tokoh yang sebenarnya satir. Secara token unigram, kata "mantap", "bagus", "luar biasa" memberi sinyal positif, padahal keseluruhan pragmatik mengarah ke kritik. Ini menjelaskan beberapa false positive yang masih bertahan pada pasca-*balancing* .

4.6 Uji Kepekaan (Sensitivity) dan Opsi Peningkatan

Dua kenop mudah yang layak diuji adalah smoothing Laplace (α) pada MultinomialNB dan rentang n-gram pada TF-IDF. α yang terlalu kecil membuat model sensitif pada kata langka; terlalu besar meratakan distribusi secara berlebihan. Grid sederhana pada α serta perbandingan unigram vs. uni+bi-gram biasanya sudah memberi perbaikan. [11][16]

Selain itu, pelaporan metrik makro (macro-*precision*/*recall*/F1) dan confusion matrix akan membantu melihat trade-off antarkelas secara lebih adil—terutama pada korpus yang semula tidak seimbang. Ini juga selaras dengan praktik evaluasi yang direkomendasikan pada rancangan penelitian.

4.7 Implikasi Praktis

Pipeline ini ringan, dapat dijelaskan, dan mudah direplikasi pada perangkat terbatas—cocok sebagai *baseline* pemantauan opini berkala. Untuk operasional jangka panjang, disarankan (i) refresh kosakata/normalisasi secara periodik guna mengatasi concept drift, (ii) pemeriksaan manual sampel acak sebagai quality gate, dan (iii) eksplorasi *Class weight* atau ensembling linier (mis. Logistic Regression/SVM) bila ingin menekan kesalahan minoritas tanpa menggandakan data. [8][13]

4.8 Ringkasan Temuan Utama

Data berlabel menunjukkan ketimpangan kelas (261/92/495), yang setelah di-oversample menjadi seimbang (495/495/495). 2) Akurasi meningkat tajam dari 64,12% menjadi 80,81% setelah *balancing* . 3) Sisa kesalahan terutama berasal dari sarkasme, hilangnya konteks thread, dan polysemy; mitigasinya: n-gram/char-gram dan leksikon. 4) Tuning α dan pelaporan metrik makro direkomendasikan untuk evaluasi yang lebih robust. [11]

5. Rekomendasi Peningkatan

Arah lanjutan yang direkomendasikan meliputi: (1) perluasan fitur ke n-gram dan karakter-gram; (2) eksplorasi model linear lain (Logistic Regression, SVM) dan pembelajaran dalam ringan; (3) pemodelan leksikon sentimen khusus

Indonesia; (4) teknik penanganan imbangan alternatif (*class weight*, SMOTE); (5) validasi silang dan analisis ablation untuk menilai kontribusi setiap komponen *pipeline*. [8][11][13]

6. Kesimpulan

Penelitian menyajikan *baseline* terstruktur untuk analisis sentimen berbahasa Indonesia di X menggunakan TF-IDF dan *Multinomial Naive Bayes*. Dengan penyeimbangan kelas sederhana pada data latih, akurasi meningkat signifikan dibandingkan skenario tidak seimbang. Praktik praproses yang diperinci—khususnya normalisasi slang dan penanganan negasi—berperan penting dalam meningkatkan kualitas representasi teks. [4][11][13][16]

Daftar Pustaka (IEEE)

[1] APJII, "Laporan Survei Internet Indonesia 2023," Asosiasi Penyelenggara Jasa Internet Indonesia, 2023.

[2] A. Baiyere, V. Grover, K. J. Lyytinen, S. Woerner, and A. Gupta, "Digital ‘x’—Charting a Path for Digital-Themed Research," Information Systems Research, 2023.

[3] K. T. Baldwin and C. Eroglu, "Astrocytes 'Coordinate' Synapse Maturation and Plasticity," Neuron, vol. 100, no. 5, pp. 1010–1012, 2018.

[4] C. Bhadane, H. Dalal, and H. Doshi, "Sentiment Analysis: Measuring Opinions," Procedia Computer Science, 2015.

[5] R. Feldman and J. Sanger, *Text mining* Handbook: Advanced Methods for Unstructured Data Analysis. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[6] R. Greenlaw et al., "PERFORM: Building and Mining Electronic Records of Neurological Patients Being Monitored in the Home," in World Congress on Medical Physics and Biomedical Engineering, 2009, pp. 533–535.

[7] A. R. Isnain, N. S. Marga, and D. Alita, "Sentiment Analysis of Government Policy on Corona Case Using Naive Bayes Algorithm," IJCCS, vol. 15, no. 1, p. 55, 2021.

[8] T. Joachims, "Text categorization with Support Vector Machines: Learning with many relevant features," in Proc. ECML, 1998.

[9] A. J. Kim and H. Gil, "The future of social media in marketing," Journal of the Academy of Marketing Science, vol. 47, pp. 237–254, 2019.

[10] K. N. Manasa and M. C. Padma, "A Study on Sentiment Analysis on Social Media Data," in Emerging Research in Electronics, Computer Science and Technology, 2019, pp. 661–667.

[11] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[12] I. Mardiyono, "Analisis Sentimen terhadap Anies Baswedan di X (Twitter) Menggunakan Naive Bayes dan TF-IDF," Undergraduate Thesis, Univ. Muhammadiyah Jember, 2022.

[13] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.

[14] S. Rahate, V. Dehanka, T. Teppalwar, and V. R. Surjuse, "Review Sentimental Analysis," International Journal of Computer Science and Mobile Computing, vol. 11, no. 3, pp. 37–41, 2022.

[15] N. C. R. Riego and D. B. Villarba, "Utilization of Multinomial Naive Bayes Algorithm and TF-IDF Vectorizer in Checking the Credibility of News *Tweet* in the Philippines," arXiv preprint, 2023.

[16] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval,"

Information Processing & Management, vol. 24, no. 5, 1988.

[17] Y. Sandria, "Penerapan Algoritma Selection Sort untuk Melakukan Pengurutan Data dalam Bahasa Pemrograman PHP," Article (in Indonesian), 2022.

[18] G. A. Sofa, "Yang Tersembunyi dari Pidato Politik Pertama Anies Baswedan sebagai Gubernur DKI Jakarta: Sebuah Analisis Wacana Kritis," Jurnal Kata, vol. 2, no. 2, pp. 384–402, 2018.

[19] D. Sumardani, I. Midaraeni, and N. I. Sumardani, "Virtual reality sebagai media pembelajaran relativitas khusus berbasis Google Cardboard pada smartphone Android," Prosiding Seminar Nasional Pendidikan KALUNI, 2019.

[20] H. Thorpe and B. Wheaton, "The X Games: Re-imagining youth and sport," in Sport, Media and Mega-events, London, U.K.: Routledge, 2017, pp. 247–261.

[21] B. Wheaton and H. Thorpe, "Action sport media consumption trends across generations," Communication & Sport, vol. 7, no. 4, pp. 415–445, 2019.