

PERBANDINGAN PERSEPSI PUBLIK DAN PENGURUS TERHADAP PROGRAM KOPERASI DESA MERAH PUTIH DI INDONESIA

Ariel Syahwal^{1*}, Ade Andri Hendriadi², Apriade Voutama³

^{1,2,3}Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur Karawang, Jawa Barat, Indonesia; Telp : (0267) 641177

Keywords:

ADASYN;
Koperasi Desa;
Analisis Sentimen;
Naïve Bayes;
Uji Chi-Square.

Correspondent Email:

2210631250005@student.unsika.ac.id

Abstrak. Program Koperasi Desa Merah Putih (KDMP) merupakan transformasi digital ekonomi pedesaan yang keberhasilannya dipengaruhi oleh penerimaan publik dan pengurus. Penelitian ini bertujuan membandingkan persepsi sentimen antara masyarakat di media sosial X dan pengurus koperasi menggunakan metode Knowledge Discovery in Databases (KDD). Analisis sentimen menggunakan algoritma Multinomial Naïve Bayes, sementara teknik Adaptive Synthetic (ADASYN) diterapkan pada data latih untuk mengatasi masalah ketidakseimbangan kelas (imbalanced data). Komparasi persepsi dievaluasi secara analitik menggunakan uji statistik Chi-Square. Hasil penelitian menunjukkan mayoritas sentimen bernilai positif, baik dari publik (90,9%) maupun pengurus (91,6%). Uji Chi-Square ($p\text{-value} = 0,7289$) mengonfirmasi tidak terdapat perbedaan persepsi yang signifikan antara kedua kelompok. Dari sisi pemodelan, penerapan ADASYN terbukti krusial dalam mengatasi fenomena accuracy paradox, di mana Recall kelas minoritas (negatif) meningkat drastis dari 0,01 menjadi 0,71 dengan akurasi akhir 81,27%. Kesimpulannya, program KDMP diterima dengan sangat baik. Penelitian ini menegaskan pentingnya penanganan imbalanced data dalam klasifikasi teks dan merekomendasikan pengembangan sistem pemantauan sentimen real-time untuk pemerintah.



Copyright © 2026 The Author(s),
Published by Jurnal Informatika dan
Teknik Elektro Terapan. This is an
open-access article distributed under
the terms of the Creative Commons
Attribution-NonCommercial 4.0
International License (CC BY-NC
4.0).

Abstract. The Red and White Village Cooperative (KDMP) program is a digital transformation of the rural economy whose success is influenced by public and management acceptance. This study aims to compare sentiment perceptions between the public on X social media and cooperative management using the Knowledge Discovery in Databases (KDD) method. Sentiment analysis utilized the Multinomial Naïve Bayes algorithm, while the Adaptive Synthetic (ADASYN) technique was applied to the training data to overcome the imbalanced data problem. Perception comparisons were analytically evaluated using the Chi-Square statistical test. The results showed that the majority of sentiments were positive, both from the public (90.9%) and the management (91.6%). The Chi-Square test ($p\text{-value} = 0.7289$) confirmed there was no significant difference in perception between the two groups. In terms of modeling, the application of ADASYN proved crucial in overcoming the accuracy paradox phenomenon, where the Recall of the minority (negative) class increased drastically from 0.01 to 0.71 with a final accuracy of 81.27%. In conclusion, the KDMP program is very well received. This study emphasizes the importance of handling imbalanced data in text classification and recommends developing a real-time sentiment monitoring system for the government.

1. PENDAHULUAN

Pemberdayaan masyarakat melalui penguatan lembaga-lembaga ekonomi lokal, khususnya koperasi desa, merupakan prioritas strategis untuk mendorong pertumbuhan yang inklusif dan berkelanjutan. Program Koperasi Desa Merah Putih, yang diluncurkan pada tahun 2025 oleh Kementerian Koperasi, merupakan inisiatif transformasi digital yang memanfaatkan sistem SIMKOPDES untuk memperkuat kemandirian ekonomi desa secara transparan dan terintegrasi [1]. Namun, keberhasilan implementasi program strategis nasional ini tidak hanya bergantung pada infrastruktur teknologi; melainkan sangat dipengaruhi oleh penerimaan dan persepsi dari dua kelompok kunci: masyarakat umum sebagai penerima manfaat dan pengelola koperasi sebagai pelaksana di lapangan.

Seiring dengan tingginya tingkat penetrasi internet, platform media sosial X telah menjadi wadah utama bagi masyarakat Indonesia yang menempati peringkat keempat secara global dalam hal jumlah pengguna untuk menyuarakan pendapat dan sentimen mereka terkait berbagai kebijakan publik secara *real-time*. Meskipun banyak cuitan dan pendapat yang beredar di platform X mengenai program Koperasi Desa Merah Putih, data ini masih berupa teks tak terstruktur yang belum dipetakan secara sistematis. Oleh karena itu, pendekatan *text mining* dan analisis sentimen diperlukan untuk mengekstrak dan mengorganisasikan informasi ini menjadi gambaran objektif mengenai persepsi public [2], [3].

Beberapa penelitian sebelumnya telah menerapkan algoritma klasifikasi pembelajaran mesin untuk analisis sentiment [4]. Penelitian oleh Nuria dkk. (2024) [5] dan Noviyanti & Putra (2025) [6] menunjukkan bahwa algoritma Naïve Bayes sangat efektif dan kompetitif dalam mengklasifikasikan data teks berskala besar. Selain itu, penelitian oleh Fauzi dkk. (2025) [7] secara khusus mengeksplorasi sentimen publik terhadap program Koperasi Desa Merah Putih. Namun, salah satu tantangan utama dalam klasifikasi sentimen berdasarkan data dunia nyata adalah ketidakseimbangan kelas (data yang tidak seimbang), yang dapat menyebabkan model menjadi bias terhadap kelas mayoritas. Penelitian oleh Zuhanda dkk. (2025) [8] dan Nurhopipah & Magnolia (2023) [9] menunjukkan bahwa penerapan teknik

oversampling ADASYN (*Adaptive Synthetic Sampling*) dapat secara signifikan meningkatkan kinerja pengenalan model pada kelas minoritas dengan secara adaptif menghasilkan data sintesis di sekitar batas keputusan.

Berdasarkan tinjauan *state of the art* tersebut, terdapat kesenjangan penelitian di mana evaluasi program hingga saat ini hanya berfokus pada perspektif eksternal (masyarakat) tanpa membandingkannya dengan perspektif internal (manajemen). Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan membandingkan persepsi masyarakat terhadap Platform X dengan persepsi para manajer koperasi. Pendekatan *Knowledge Discovery in Databases* (KDD), dikombinasikan dengan algoritma Naïve Bayes dan metode *oversampling* ADASYN, digunakan untuk memodelkan sentimen publik, dan hasilnya kemudian diuji perbedaannya menggunakan statistik *Chi-Square* terhadap hasil survei dari para manajer. Temuan penelitian ini diharapkan dapat memberikan evaluasi komprehensif yang akan berguna bagi pembuat kebijakan dalam mengembangkan ekosistem koperasi yang lebih inklusif dan adaptif di masa depan.

2. TINJAUAN PUSTAKA

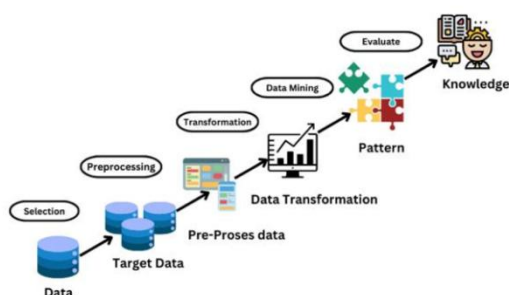
2.1 Koperasi Desa Merah Putih (KDMP)

Program KDMP merupakan inisiatif strategis nasional yang diluncurkan secara resmi oleh Kementerian Koperasi pada tahun 2025 untuk mendorong kemandirian ekonomi di tingkat desa [1]. Program ini diintegrasikan melalui ekosistem digital yang dikenal sebagai Sistem Informasi Koperasi Desa/Kecamatan Merah Putih (SIMKOPDES), yang memungkinkan pengelolaan transaksi, keanggotaan, dan pelaporan yang lebih transparan dan terukur.

2.2 Knowledge Discovery in Databases (KDD) dan Analisis Sentimen

Analisis sentimen adalah cabang dari *text mining* yang secara khusus digunakan untuk memahami polaritas opini (positif atau negatif) yang terkandung dalam dokumen teks tidak terstruktur [10]. Dalam penerapannya, penelitian ini menggunakan metodologi *Knowledge Discovery in Databases* (KDD) [11]. KDD adalah proses sistematis untuk

mengidentifikasi tren dan informasi penting dalam data.



Gambar 1 Tahapan metode KDD

Tahapan KDD ini meliputi pengumpulan data (seleksi), prapemrosesan, transformasi, penambahan data, serta evaluasi dan interpretasi hasil (presentasi pengetahuan).

2.3 Algoritma Naïve Bayes dan TF-IDF

Naïve Bayes *Classifier* adalah metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dan beroperasi dengan asumsi kemandirian fitur [12]. Algoritma ini dipilih karena efisiensi memori dan kecepatan komputasinya dalam menangani data teks berskala besar. Persamaan matematika dasar algoritma Naïve Bayes ditunjukkan pada Persamaan di bawah ini:

$$P(H|X) = \frac{p(X|H) \cdot p(H)}{p(X)} \quad (1)$$

Agar model dapat memproses data teks, transformasi *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan untuk mengubah teks menjadi representasi vektor numerik. Metode TF-IDF memberikan bobot tinggi pada kata-kata yang sering muncul dalam suatu dokumen tetapi jarang muncul dalam dokumen lain di seluruh kumpulan data.

2.4 Adaptive Synthetic Sampling (ADASYN)

Ketidakseimbangan data terjadi ketika satu kategori secara signifikan mendominasi jumlah data dibandingkan dengan kategori lainnya [13]. ADASYN adalah teknik *oversampling* yang bekerja dengan menyeimbangkan distribusi di seluruh kategori [14]. Pendekatan ini memanfaatkan bobot distribusi pada data minoritas dan secara adaptif menghasilkan data sintetis di area dengan tingkat kesulitan pembelajaran tertinggi.

2.5 Uji Chi-Square

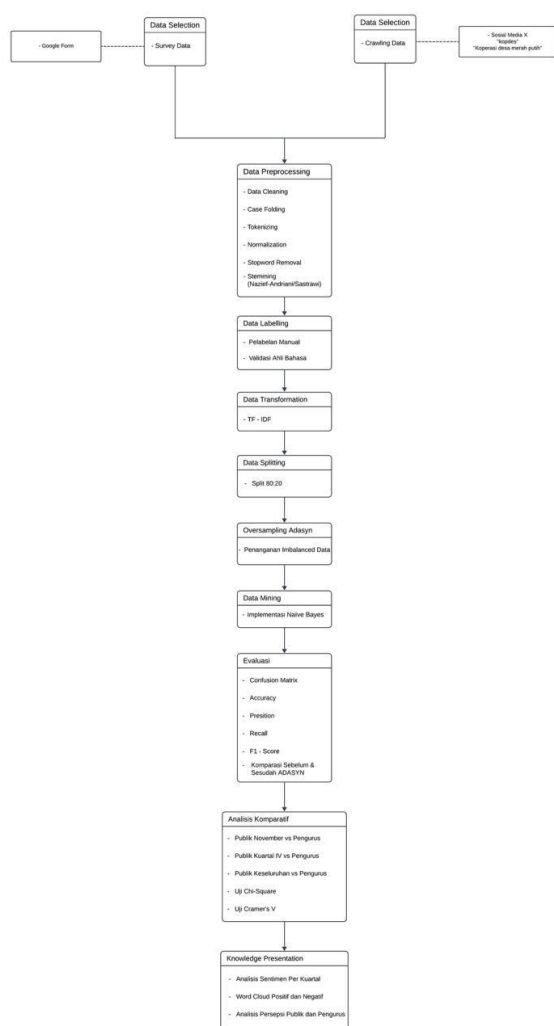
Uji *Chi-Square* adalah metode statistik non-parametrik yang dirancang untuk menguji hubungan antara dua variabel kategorikal dalam bentuk tabel kontingensi [15]. Dalam studi ini, uji *Chi-Square* digunakan untuk mengevaluasi secara analitis apakah terdapat perbedaan yang signifikan dalam persepsi sentimen antara sumber data eksternal (masyarakat di media sosial) dan data internal (pengelola koperasi).

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif dengan mengadopsi kerangka kerja *Knowledge Discovery in Databases* (KDD) sebagai metodologi utama. Kerangka kerja ini dipilih karena sifatnya yang komprehensif dan sistematis, mencakup seluruh rangkaian integrasi data mulai dari seleksi data mentah hingga penyajian pengetahuan yang bernilai.

3.1. Rancangan Penelitian

Rancangan penelitian ini disusun secara terstruktur untuk menjawab masalah komparasi persepsi sentimen menggunakan data riil berdimensi tinggi mengikuti alur KDD. Alur tahapan penelitian secara visual direpresentasikan pada Gambar 2.



Gambar 2 Tahapan Rancangan Penelitian

Metode pengolahan data untuk pemodelan dikembangkan dalam dua skenario eksperimen paralel. Skenario A menetapkan kondisi *baseline* di mana klasifikasi dilakukan langsung pada distribusi data latih asli yang mengalami ketidakseimbangan kelas (*imbalanced data*). Skenario B mengimplementasikan penyeimbangan distribusi data menggunakan algoritma *Adaptive Synthetic Sampling* (ADASYN) sebelum fase pelatihan model dilakukan. Desain berpasangan ini dirancang untuk mengevaluasi dampak penanganan data tidak seimbang secara objektif terhadap performa pengenalan sentimen.

3.2. Sumber dan Partisipan Data

Objek penelitian ini mencakup data opini dan persepsi yang diperoleh dari dua kelompok subjek berbeda namun saling melengkapi:

1. **Data Publik (Eksternal):** Berupa dokumen teks cuitan berbahasa Indonesia di platform media sosial X mengenai kebijakan pembentukan Koperasi Desa Merah Putih. Data ini merepresentasikan opini masyarakat umum dari luar sistem.
2. **Data Pengurus (Internal):** Partisipan data adalah para pengurus aktif Koperasi Desa/Kelurahan Merah Putih yang tersebar di berbagai wilayah geografis Indonesia. Data ini merepresentasikan pandangan pelaksana program yang berinteraksi langsung dengan sistem di lapangan.

3.3. Teknik Pengumpulan Data

Proses pengumpulan data dilakukan secara terpisah untuk kedua kelompok sumber data:

1. **Data Publik:** Pengumpulan dilakukan secara otomatis menggunakan teknik *crawling* melalui alat *tweet-harvest* dengan kata kunci pencarian "kopdes" dan "koperasi desa merah putih". Pengambilan data dibatasi pada rentang waktu temporal dari April hingga Desember 2025. Setelah proses dedublikasi berdasarkan kesamaan isi teks komentar, diperoleh 4.176 data publik yang valid.
2. **Data Pengurus:** Pengumpulan dilakukan menggunakan kuesioner daring (*Google Form*) yang didistribusikan kepada pengurus saat kegiatan pelatihan resmi serta melalui grup komunitas. Dari total 512 respons, proses seleksi menghasilkan 439 data pengurus yang valid.

3.4. Teknik Analisis Data

Teknik analisis data teks dan pemodelan statistik inferensial dijalankan melalui tahapan *pipeline* terperinci sebagai berikut:

3.4.1. Data Preprocessing

Seluruh data teks publik dan pengurus melewati enam tahapan prapemrosesan guna mengonversi teks tidak terstruktur menjadi bentuk baku:

1. **Cleaning:** Penghapusan tautan URL, *mention*, *hashtag*, angka, tanda baca, karakter khusus/emoji, serta normalisasi spasi.
2. **Case Folding:** Transformasi seluruh karakter teks menjadi huruf kecil secara menyeluruh.

3. *Tokenizing*: Pemecahan kalimat menjadi daftar *token* (kata) individual menggunakan fungsi `word_tokenize` NLTK.
4. *Normalization*: Rekonstruksi kata tidak baku, slang, dan singkatan (misalnya "kopdes" menjadi "koperasi desa") menggunakan kamus normalisasi terkurasi.
5. *Stopword Removal*: Eliminasi kata-kata umum (*functional words*) menggunakan daftar *stopword* NLTK, beserta penghapusan *token* dengan panjang kurang dari dua karakter.
6. *Stemming*: Reduksi kata berimbuhan menjadi kata dasar menggunakan pustaka Sastrawi yang mengimplementasikan algoritma Nazief-Andriani.

3.4.2. Data Labelling

Anotasi kelas sentimen dilakukan secara biner ke dalam kategori Positif dan Negatif dengan mengacu pada panduan anotasi kontekstual. *Ground truth* label kemudian divalidasi secara independen oleh ahli bahasa (*linguist*) Bahasa Indonesia guna menjamin objektivitas nilai interpretasi semantik.

3.4.3. Data Transformation (TF-IDF)

Vektorisasi teks menjadi representasi numerik dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) lewat pustaka *scikit-learn*. Langkah ini menghasilkan matriks berdimensi 4.165 dokumen dan 7.982 fitur.

3.4.4. Data Splitting

Pembagian subset dilakukan menggunakan metode *stratified sampling* dengan proporsi 80% data latih (3.332 dokumen) dan 20% data uji (833 dokumen). Parameter `random_state=42` digunakan untuk menjaga hasil pembagian tetap deterministik.

3.4.5. Penanganan Imbalanced Data (ADASYN)

Ketidakeimbangan kelas pada data latih ditangani menggunakan algoritma *Adaptive Synthetic Sampling* (ADASYN). Pembangkitan sampel sintetis kelas minoritas difokuskan secara adaptif pada area batas keputusan yang memiliki tingkat kesulitan pembelajaran tertinggi. Proses ini diaplikasikan secara eksklusif hanya pada data latih guna mencegah kebocoran data (*data leakage*).

3.4.6. Klasifikasi (Data Mining)

Klasifikasi sentimen dimodelkan menggunakan algoritma *Multinomial Naïve Bayes* (MultinomialNB). Algoritma ini

dimanfaatkan karena efektivitas dan kompleksitas komputasinya yang rendah dalam memproses matriks *sparse* berdimensi tinggi seperti fitur TF-IDF.

3.4.7. Evaluasi Model

Pengukuran performa model klasifikasi dievaluasi pada data uji menggunakan instrumen *Confusion Matrix*. Parameter performa diekstraksi berdasarkan nilai rata-rata tertimbang (*weighted average*) dari metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

3.4.8. Analisis Komparatif

Pengujian hipotesis perbedaan persepsi dilakukan dengan membandingkan distribusi sentimen data publik terhadap data pengurus menggunakan pengujian statistik:

1. Uji *Chi-Square*: Untuk mengevaluasi signifikansi perbedaan sebaran frekuensi sentimen antar kelompok dengan ambang batas signifikansi 0,05 ($p\text{-value} < 0,05$).
2. *Cramer's V*: Untuk mengukur kekuatan asosiasi antar kedua kelompok variabel tersebut.

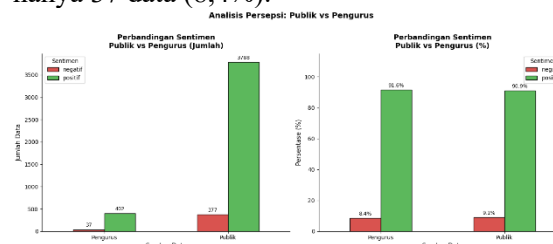
4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Bagian ini menyoroti tiga temuan utama dari ekstraksi data: distribusi sentimen, evaluasi performa model klasifikasi dengan skenario penanganan data tidak seimbang, dan hasil uji hipotesis komparatif.

4.1.1 Distribusi Sentimen Publik dan Pengurus

Berdasarkan hasil pelabelan pada 4.165 data publik yang valid, sentimen masyarakat didominasi oleh opini positif sebanyak 3.788 data (90,9%), sedangkan opini negatif hanya berjumlah 377 data (9,1%). Pola yang sangat identik juga ditemukan pada data pengurus internal (439 responden), di mana sentimen positif mencapai 402 data (91,6%) dan negatif hanya 37 data (8,4%).



Gambar 3 Perbandingan Persentase Sentimen Publik vs Pengurus

4.1.2 Performa Klasifikasi Naïve Bayes dan Dampak ADASYN

Ketidakseimbangan kelas (*imbalanced data*) dengan rasio 10:1 pada data latih memberikan dampak signifikan terhadap performa model. Pemodelan dievaluasi melalui dua skenario: Skenario A (*Baseline* tanpa ADASYN) dan Skenario B (*Proposed* dengan ADASYN). Perbandingan performa kedua skenario dievaluasi pada data uji yang identik menggunakan metrik rata-rata tertimbang (*weighted average*), seperti disajikan pada Tabel 1.

Tabel 1 Komparasi Hasil Evaluasi Model Naïve Bayes

Metrik Evaluasi	Skenario A (Tanpa ADASYN)	Skenario B (Dengan ADASYN)
Accuracy	0,9112	0,8127
Precision	0,9191	0,9045
Recall	0,9112	0,8127
F1-Score	0,8700	0,8453
Recall Kelas Negatif	0,01	0,71

Meskipun Skenario A menghasilkan *Accuracy* tinggi (91,12%), model ini gagal mengidentifikasi sentimen negatif (Recall kelas negatif hanya 0,01 atau 1 dari 75 data). Penerapan ADASYN pada Skenario B mengoreksi kegagalan tersebut dengan meningkatkan Recall kelas negatif secara drastis menjadi 0,71, meskipun mengorbankan sedikit akurasi keseluruhan menjadi 81,27%.

4.1.3 Hasil Uji Hipotesis (Analisis Komparatif)

Pengujian statistik inferensial menggunakan *Chi-Square* dilakukan untuk membandingkan distribusi sentimen publik secara keseluruhan (April–Desember 2025) dengan sentimen pengurus. Hasil uji menghasilkan nilai *Chi-Square* sebesar 0,1201 dengan *p-value* sebesar 0,7289 ($p > 0,05$). Nilai *Cramer's V* yang dihasilkan adalah 0,0051.

4.2 Pembahasan

Pembahasan penelitian ini berfokus pada interpretasi temuan pemodelan *machine learning*, keselarasan persepsi, serta implikasinya secara teoritis maupun praktis guna menjawab rumusan masalah secara komprehensif.

4.2.1 Mengatasi Accuracy Paradox pada Klasifikasi Teks

Temuan evaluasi model secara eksplisit menunjukkan terjadinya fenomena *accuracy paradox* pada Skenario A. Model terlihat memiliki akurasi tinggi (91,12%), namun pada kenyataannya model tersebut hanya memprediksi mayoritas data sebagai kelas positif (mengabaikan kelas negatif). Kondisi ini sejalan dengan penelitian Zuhanda et al. (2025) [8] dan Nurhopipah & Magnolia (2023) [9] yang menegaskan bahwa klasifikasi pada data yang sangat timpang tanpa *resampling* akan menghasilkan evaluasi yang menyesatkan. Mengapa Skenario B lebih baik? Algoritma ADASYN memecahkan masalah ini dengan tidak sekadar menduplikasi data minoritas, melainkan membangkitkan data sintesis secara adaptif di area batas keputusan (*decision boundary*) yang sulit dikenali oleh Naïve Bayes. Penurunan akurasi pada Skenario B (menjadi 81,27%) bukanlah sebuah kemunduran, melainkan representasi model yang lebih rasional, jujur, dan memiliki sensitivitas tinggi dalam mendeteksi opini kritis masyarakat.

4.2.2 Keselarasan Persepsi dan Potensi Social Desirability Bias

Apa makna dari hasil uji *Chi-Square*? Nilai *p-value* sebesar 0,7289 dan *Cramer's V* sebesar 0,0051 secara statistik menegaskan bahwa tidak ada perbedaan yang signifikan antara persepsi publik (eksternal) dan pengurus (internal) terkait Program Koperasi Desa Merah Putih. Kedua entitas ini secara mayoritas (>90%) menunjukkan penerimaan dan optimisme yang tinggi terhadap digitalisasi koperasi.

Namun, temuan ini harus diinterpretasikan dengan hati-hati. Mengapa sentimen pengurus mencapai angka positif yang sangat tinggi (91,6%)? Tingginya angka ini kemungkinan besar dipengaruhi oleh *social desirability bias*, mengingat kuesioner disebarkan pada saat kegiatan pelatihan institusional resmi Kemenkop. Responden internal cenderung memberikan jawaban yang dipersepsikan aman dan sesuai dengan ekspektasi program, berbeda dengan data publik di media sosial X yang bersifat organik dan spontan. Selain itu, analisis temporal publik mencatat adanya lonjakan sentimen negatif pada Kuartal 3 (Juli–September), yang mengindikasikan adanya fase keraguan atau kritik masyarakat pada awal masa implementasi program sebelum akhirnya kembali stabil di akhir tahun.

4.2.3 Implikasi Teoritis dan Implementasi

1. Implikasi Teoritis: Penelitian ini menegaskan bahwa penggunaan algoritma Naïve Bayes yang dipadukan dengan teknik penanganan data adaptif seperti ADASYN merupakan kebutuhan mutlak (*mandatory*), bukan sekadar opsi dalam penelitian *text mining* berbahasa Indonesia yang melibatkan kelas minoritas ekstrem. Ini mengonstruksi pemahaman bahwa evaluasi klasifikasi teks tidak boleh bertumpu pada *Accuracy* semata, melainkan harus menggunakan rata-rata tertimbang (*weighted average*) dan evaluasi metrik *Recall* pada kelas minoritas.
2. Implikasi Implementasi (Praktis): Secara keilmuan Sistem Informasi, temuan ini menjadi landasan kuat untuk mengimplementasikan Sistem Informasi Pemantauan Sentimen Publik (SIPSP) yang diintegrasikan langsung dengan *dashboard* SIMKOPDES milik Kementerian Koperasi. Sistem ini dapat dibangun menggunakan model Naïve Bayes-ADASYN yang telah dilatih sebagai REST API, guna memberikan peringatan otomatis (*early warning system*) kepada pemangku kebijakan apabila lonjakan opini negatif di media sosial melewati ambang batas toleransi (misalnya >15%). Langkah preventif berbasis data ini akan mentransformasi cara pemerintah merespons dinamika sosial secara *real-time*.

5. KESIMPULAN

Berdasarkan penelitian perbandingan persepsi publik dan pengurus terhadap Program Koperasi Desa Merah Putih yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Hasil Analisis Sentimen: Persepsi terhadap Program Koperasi Desa Merah Putih didominasi oleh sentimen positif, baik dari sisi publik (90,97%) maupun dari sisi pengurus (91,57%). Hal ini menunjukkan adanya penerimaan dan optimisme yang tinggi dari kedua belah pihak terhadap program digitalisasi koperasi ini.
2. Hasil Pemodelan dan Penanganan Data Tidak Seimbang: Model klasifikasi *Multinomial Naïve Bayes* terbukti rentan terhadap fenomena *accuracy paradox* jika dilatih pada data yang tidak seimbang (menghasilkan *Recall* kelas minoritas

sebesar 0,01 meski akurasi mencapai 91,12%). Penerapan teknik *oversampling* ADASYN pada data latih berhasil mengoreksi hal tersebut dengan menaikkan *Recall* sentimen negatif secara drastis menjadi 0,71, menghasilkan model yang jauh lebih sensitif dan rasional dengan akurasi akhir 81,27%.

3. Hasil Uji Komparatif: Uji statistik *Chi-Square* ($p\text{-value} = 0,7289$) dan *Cramer's V* (0,0051) menunjukkan bahwa secara statistik inferensial tidak terdapat perbedaan distribusi persepsi yang signifikan antara masyarakat umum di media sosial X dengan pengurus di lapangan.
4. Kelebihan Penelitian: Penggunaan algoritma Naïve Bayes dipadukan dengan pembobotan TF-IDF sangat efisien dalam komputasi matriks *sparse* berdimensi tinggi. Selain itu, penerapan ADASYN yang diisolasi secara eksklusif hanya pada data latih memastikan evaluasi model tetap sah, tidak bias, dan terhindar dari *data leakage*.
5. Kekurangan Penelitian: Penelitian ini masih memiliki beberapa keterbatasan, di antaranya potensi *social desirability bias* pada pengisian kuesioner pengurus yang dilakukan dalam konteks institusional formal. Selain itu, proses anotasi pelabelan data manual hanya divalidasi oleh satu orang ahli bahasa, dan data opini publik hanya bersumber dari satu platform tunggal (media sosial X).
6. Kemungkinan Pengembangan Selanjutnya: Untuk penelitian di masa mendatang, disarankan untuk membandingkan performa model menggunakan algoritma seperti *Support Vector Machine* (SVM) atau metode *Deep Learning*. Selain itu, sumber data perlu diperluas ke platform lain seperti TikTok, Instagram, atau portal berita daring agar lebih representatif. Secara praktis, luaran algoritma ini sangat potensial untuk dikembangkan menjadi purwarupa Sistem Informasi Pemantauan Sentimen Publik (SIPSP) yang terintegrasi secara *real-time* dengan ekosistem SIMKOPDES milik Kementerian Koperasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer Universitas

Singaperbangsa Karawang atas fasilitas yang diberikan, Bapak Dr. Ade Andri Hendriadi, S.Si., M.Kom. dan Bapak Apriade Voutama, M.Kom. selaku dosen pembimbing atas arahnya, serta Kementerian Koperasi dan seluruh pengurus Koperasi Desa Merah Putih atas dukungan data dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Kementerian Sekretariat Negara R.I., “Instruksi Presiden Republik Indonesia Nomor 9 Tahun 2025 Tentang Percepatan Pembentukan Koperasi Desa/Kelurahan Merah Putih,” 2025.
- [2] L. Nursinggah, Ruuhwan, and T. Mufizar, “ANALISIS SENTIMEN PENGGUNA APLIKASI X TERHADAP PROGRAM MAKAN SIANG GRATIS DENGAN METODE NAÏVE BAYES CLASSIFIER,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4336.
- [3] D. D. Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [4] M. Afriansyah, J. Saputra, V. Y. P. Ardhana, and Y. Sa’adati, “ALGORITMA NAIVE BAYES YANG EFISIEN UNTUK KLASIFIKASI BUAH PISANG RAJA BERDASARKAN FITUR WARNA,” *Journal of Information Systems Management and Digital Business (JISMDB)*, vol. 1, no. 2, Jan. 2024.
- [5] D. P. Nuria, U. Enri, and Y. Umaidah, “Analisis Sentimen Maxim dengan Perbandingan Chi Square dan MI pada Naive Bayes,” *Jurnal Pendidikan Tambusai*, vol. 8, no. 1, pp. 16070–16081, 2024.
- [6] N. P. A. Noviyanti and I. G. N. A. C. Putra, “Penerapan Multinomial Naïve Bayes dan Chi-Square pada Analisis Sentimen Makan Bergizi Gratis,” *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 3, no. 4, Aug. 2025.
- [7] Fauzi, B. Andini, and M. Muslihudin, “SENTIMENT ANALYSIS OF TWITTER (X) USERS TOWARD THE KOPERASI DESA MERAH PUTIH PROGRAM USING SUPPORT VECTOR MACHINE AND NAIVE BAYES METHODS,” *Technology Acceptance Model) Jurnal TAM*, vol. 16, no. 1, pp. 50–64, Jul. 2025, [Online]. Available: <https://jurnal.ftikomibn.ac.id/index.php/JurnalTam/index>
- [8] M. K. Zuhanda, L. Permata, Hartono, E. Ongko, and Desniarti, “Impact of Adaptive Synthetic on Naïve Bayes Accuracy in Imbalanced Anemia Detection Datasets,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 1, pp. 85–93, Feb. 2025, doi: 10.29207/resti.v9i1.6031.
- [9] A. Nurhopipah and C. Magnolia, “PERBANDINGAN METODE RESAMPLING PADA IMBALANCED DATASET UNTUK KLASIFIKASI KOMENTAR PROGRAM MBKM,” *Jurnal Publikasi Ilmu Komputer dan Multimedia*, vol. 1, no. 2, pp. 9–22, Jan. 2022.
- [10] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, “ANALISIS SENTIMEN MENGGUNAKAN METODE LEXICON-BASED DAN SUPPORT VECTOR MACHINE PADA PRESIDEN DAN WAKIL PRESIDEN INDONESIA PERIODE 2024–2029,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5604.
- [11] N. Amalia, T. Suprpti, and G. Dwilestari, “ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAKSANAAN KURIKULUM MBKM,” *Jurnal Teknik Elektro dan Informatika*, vol. 18, no. 1, May 2023.
- [12] A. Safira and F. N. Hasan, “ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, Jan. 2023.
- [13] M. Mujahid *et al.*, “Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering,” *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00943-4.
- [14] M. W. Hadinata, T. A. Y. Siswa, and W. J. Pranoto, “Analisis Kecelakaan Lalu Lintas Berbasis Decision Tree Terintegrasi Forward Selection dan Metode ADASYN,” *JURNAL BUFFER INFORMATIKA*, vol. 11, no. 2, Oct. 2025, [Online]. Available: <https://journal.fkom.uniku.ac.id/index.php/buffer>
- [15] A. Heryana, “UJI CHI SQUARE,” May 2020. doi: <https://doi.org/10.13140/RG.2.2.23266.15047>.