

Sentiment Analysis of Public Responses to the Establishment of the Merah Putih Village Cooperative Program Using Naive Bayes and Logistic Regression [Analisis Sentimen Review Awal Pembentukan Program Koperasi Desa Merah Putih Dengan Metode Naive Bayes Dan Logistic Regression]

Muhammad Athallah Hazel Orando¹⁾, Nuril Lutvi Azizah ^{*,2)}

¹⁾Program Studi Ilmu Hukum, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: nurillutviazizah@umsida.ac.id

Abstract. Motivated by the government policy on the Merah Putih Village Cooperative Program, which has generated various public opinions, this study aims to analyze public sentiment toward the initial review of the program's establishment based on comments on the YouTube platform. The research data were obtained using the YouTube Data API, consisting of 11,821 Indonesian-language comments collected from June 10, 2026, to July 4, 2026. The methods used in this study are Naive Bayes and Logistic Regression. The labeling results show that negative sentiment dominated at 65.7%, followed by neutral sentiment at 19.0% and positive sentiment at 15.3%. Model evaluation was conducted using accuracy, precision, recall, and F1-score across several data-splitting scenarios. The testing results indicate that Logistic Regression achieved the best performance, with an accuracy of 0.81 in the 90:10 data split scenario. Therefore, Logistic Regression is considered more effective in classifying public sentiment toward the Merah Putih Village Cooperative Program.

Keywords - sentiment analysis; Koperasi Desa Merah Putih; YouTube comments; Naive Bayes; Logistic Regression

Abstrak. Dilatarbelakangi mengenai kebijakan pemerintah Program Koperasi Desa Merah Putih yang menimbulkan berbagai macam opini dari masyarakat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap review awal pembentukan Program Koperasi Desa Merah Putih berdasarkan komentar pada platform YouTube. Data penelitian diperoleh menggunakan YouTube Data API dengan jumlah 11.821 komentar berbahasa Indonesia yang dikumpulkan pada periode 10 Juni 2026 hingga 4 Juli 2026. Metode yang digunakan dalam penelitian ini adalah Naive Bayes dan Logistic Regression. Hasil pelabelan menunjukkan bahwa sentimen negatif mendominasi sebesar 65,7%, diikuti sentimen netral sebesar 19,0% dan sentimen positif sebesar 15,3%. Evaluasi model dilakukan menggunakan accuracy, precision, recall, dan F1-score pada beberapa skenario pembagian data. Hasil pengujian menunjukkan bahwa Logistic Regression memperoleh performa terbaik dengan akurasi sebesar 0,81 pada skenario pembagian data 90:10. Dengan demikian, Logistic Regression dinilai lebih efektif dalam mengklasifikasikan sentimen publik terhadap Program Koperasi Desa Merah Putih.

Kata Kunci - analisis sentimen; Koperasi Desa Merah Putih; komentar YouTube; Naive Bayes; Logistic Regression

I. PENDAHULUAN

Program strategis pemerintah dalam bentuk program Koperasi Desa Merah Putih, memiliki tujuan untuk meningkatkan ekonomi kerakyatan berbasis desa, didasarkan pada Instruksi Presiden Nomor 9 Tahun 2025 tentang Percepatan Pembentukan Koperasi Desa/Kelurahan Merah Putih., dengan target pendirian, pengembangan, dan revitalisasi 80.000 koperasi di desa dan kelurahan [1]. Kopdes Merah Putih diarahkan sebagai lembaga ekonomi yang dapat menjalankan layanan kebutuhan pokok, simpan pinjam, logistik, pergudangan, apotek, klinik, serta penguatan rantai distribusi lokal. Dengan demikian, koperasi diposisikan sebagai instrumen pemerataan pembangunan, kemandirian ekonomi, dan peningkatan daya tawar masyarakat desa [2]

Penelitian Darwis, Siskawati, dan Abidin menunjukkan bahwa algoritma Naive Bayes dapat diterapkan dalam analisis sentimen review data Twitter BMKG Nasional dengan mengklasifikasikan opini masyarakat ke dalam tiga kelas, yaitu positif, negatif, dan netral, serta menghasilkan akurasi sebesar 69,97% [3]. Ridwan menjelaskan bahwa Naive Bayes merupakan algoritma pembelajaran mesin untuk masalah klasifikasi yang banyak digunakan pada klasifikasi teks berdimensi tinggi, termasuk analisis sentimen dan penyaringan spam, karena sederhana, efektif, serta mampu membangun model secara cepat [4]. Haq, Octiva, Ayuliana, Nuryanto, dan Suryadi menerapkan Naive Bayes untuk mengidentifikasi hoaks di media sosial dengan tahapan prapemrosesan berupa tokenisasi, ekstraksi fitur, pembersihan teks, dan TF-IDF, sehingga menunjukkan bahwa metode ini sesuai untuk pengolahan data teks dari media sosial [5]. Watratan, Puspita, dan Moeis menjelaskan bahwa Naive Bayes memiliki kelebihan berupa

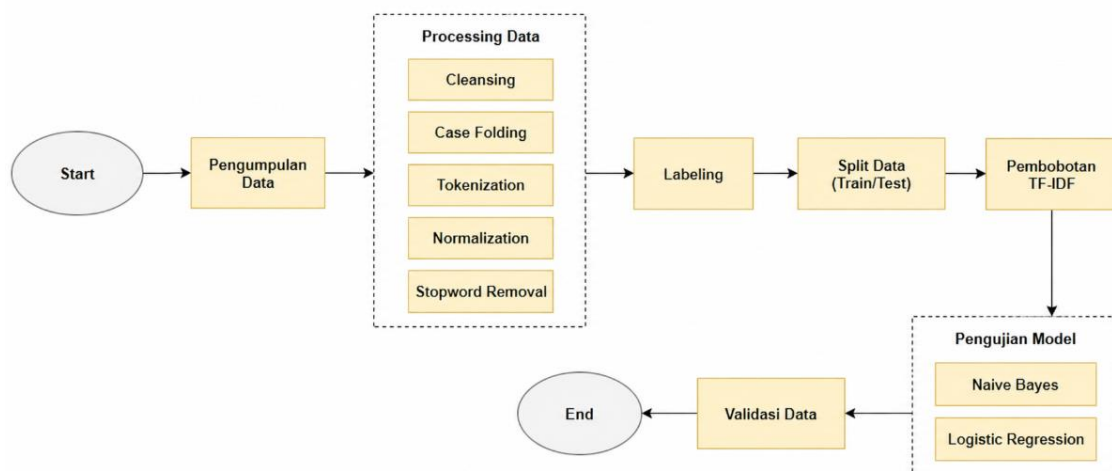
perhitungan yang cepat, efisien, mudah dipahami, dapat digunakan untuk klasifikasi biner maupun multikelas, tetapi memiliki keterbatasan karena asumsi independensi antarvariabel dapat memengaruhi akurasi ketika terdapat hubungan antarfitur [6]. Berdasarkan penelitian terdahulu tersebut, Naive Bayes layak digunakan dalam penelitian analisis sentimen komentar YouTube terhadap pembentukan Program Koperasi Desa Merah Putih karena komentar publik merupakan data teks tidak terstruktur yang perlu diklasifikasikan ke dalam sentimen positif, negatif, dan netral.

Analisis sentimen menjadi metode yang menarik untuk membaca opini masyarakat secara menyeluruh. Dalam penelitian ini, metode Naive Bayes dan Logistic Regression dipilih karena keduanya termasuk algoritma pembelajaran mesin yang banyak digunakan untuk klasifikasi teks [3]. Naive Bayes memiliki keunggulan pada kesederhanaan model, efisiensi komputasi, dan kemampuan bekerja dengan baik pada data teks berbasis probabilitas kata. Logistic Regression unggul dalam interpretabilitas, stabilitas performa, dan kemampuan untuk memodelkan hubungan antara fitur teks dengan kelas sentimen.

Meskipun memiliki tujuan yang ideal, pembentukan Kopdes Merah Putih tidak terlepas dari berbagai permasalahan implementatif [7]. Pembentukan koperasi dalam jumlah besar berpotensi menghadapi kendala kesiapan sumber daya manusia, pemahaman tata kelola koperasi, administrasi kelembagaan, transparansi pengelolaan, serta kesesuaian model usaha dengan potensi lokal desa. Penelitian tentang pembentukan koperasi ini juga menemukan kendala berupa rendahnya literasi kelembagaan, keterbatasan administrasi, dan ketidaksesuaian petunjuk teknis dengan kondisi lokal [8]. Kondisi tersebut dapat memunculkan opini publik yang beragam, mulai dari dukungan, keraguan, hingga kritik. Salah satu ruang digital yang merekam respons awal masyarakat adalah kolom komentar YouTube pada konten review program [2].

Penelitian Ramadhani dan Suryono menunjukkan bahwa Logistic Regression relevan digunakan dalam analisis sentimen karena metode ini termasuk algoritma klasifikasi teks yang dapat digunakan untuk mengelompokkan opini masyarakat ke dalam kategori sentimen, serta pada studi sentimen metaverse menghasilkan akurasi 95% setelah optimasi SMOTE dengan nilai precision 94%, recall 93%, dan F1-score 95% [9]. Moelyono, Setiawan, dan Wijaya menjelaskan bahwa Logistic Regression termasuk metode klasifikasi dalam supervised learning yang dapat dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, sehingga sesuai digunakan untuk mengukur performa model klasifikasi secara kuantitatif [10]. Berdasarkan penelitian terdahulu tersebut, Logistic Regression layak digunakan dalam penelitian analisis sentimen komentar YouTube terhadap pembentukan Program Koperasi Desa Merah Putih karena komentar publik merupakan data teks tidak terstruktur yang perlu diklasifikasikan secara terukur ke dalam sentimen positif, negatif, dan netral.

II. METODE



Gambar 1. Diagram Tahapan Penelitian

2.1 Sumber dan Pengumpulan Data

Data yang dimanfaatkan untuk penelitian ini adalah data sekunder berupa komentar pengguna pada platform YouTube terkait kebijakan Pemerintah terkait Pembentukan Koperasi Desa Merah Putih. Data dikumpulkan dengan memanfaatkan YouTube Data API untuk mengambil komentar secara otomatis dari video yang telah ditentukan, dengan contoh data mentah yang disajikan pada Tabel 1. Komentar berbahasa Indonesia digunakan sebagai data penelitian yang mencerminkan opini masyarakat terhadap kebijakan tersebut. Pengambilan data dilakukan pada

periode 10 Juni 2026 hingga 4 Juli 2026 dengan total 11821 komentar, yang selanjutnya disimpan dalam format XLXS untuk dilanjutkan dalam tahapan pengolahan serta analisis data.

Tabel 1. Contoh Data Komentar Youtube

Video_id	Author	Published_at	Text	Likecount
M	H5A5TvKztw 5	@ariantiutomo280 2026-02-10T21:05:19Z	Ujung2nya lahan korupsi lg....	182
M	H5A5TvKztw	@sisonajeg517 2026-02-10T17:15:35Z	* Lokasi tak Setrategis * SDM tak siap * Harga Tak bisa bersaing	32
M	H5A5TvKztw	@just-for-entertain798 2026-02-10T16:49:20Z	Koperasi nggak bisa dipaksa dari atas..!!!	27

2.2 Preprocessing Data

Tahap preprocessing dilaksanakan dalam menyiapkan data komentar sebelum permodelan dengan tujuan membersihkan dan menyeragamkan teks agar siap diproses oleh algoritma klasifikasi. Tahapan yang diterapkan meliputi case folding untuk mengubah teks menjadi huruf kecil, cleaning untuk menghapus karakter yang tidak relevan, tokenization untuk memecah teks menjadi kata, normalization untuk menyesuaikan kata tidak baku, serta stopword removal untuk menghilangkan kata yang tidak memiliki makna signifikan. Ilustrasi perubahan data teks pada setiap tahapan tersebut disajikan secara detail pada Tabel 3. Secara keseluruhan, rangkaian proses ini bertujuan untuk mengurangi noise dan meningkatkan kualitas fitur teks sehingga dapat diolah secara optimal pada tahap klasifikasi.

Tabel 2. Data Hasil Preprocessing

Tahap Komentar

Data Awal	Ujung2nya lahan korupsi lg....
Case Folding	ujung2nya lahan korupsi lg....
Cleaning	ujungnya lahan korupsi lg
Tokenization	['ujungnya', 'lahan', 'korupsi', 'lg']
Normalization	['akhirnya', 'lahan', 'korupsi', 'lagi']
Stopword Removal	['akhirnya', 'lahan', 'korupsi', 'lagi']

2.3 Labeling

Setelah tahap *preprocessing*, data komentar mengenai Koperasi Desa (Kopdes) dilabeli sentimen secara otomatis menggunakan *Lexicon* ke dalam tiga kategori: positif (1), negatif (-1), dan netral (0). Pelabelan dilakukan dengan mengakumulasi bobot kata berdasarkan kamus sentimen. Komentar berlabel positif jika skor > 0, negatif jika skor < 0, dan netral (0) jika total skor bernilai 0 atau tidak mengandung kata dari kamus positif maupun negatif. Implementasi teknis dan contoh hasil pelabelan yang menunjukkan keterhubungan antara teks, stopword removal, skor, hingga label akhir tersebut disajikan secara detail pada Tabel 2.

Tabel 3.Data Hasil Labeling

Komentar	Stopword_removal	Sentiment_Score	Label
Koprasi ladang korupsi masal pejabat....!!!	['koperasi', 'ladang', 'korupsi', 'masal', 'pejabat']	-1	-1
Openingnya kepanjangan mba,	['openingnya', 'kepanjangan', 'mba']	0	0
Lebih baik koperasi seperti ini dari pada koperasi seperti yg ada saatini ..lintah darat berkedok'seperti', 'koperasi', 'seperti', 'koperasi'	['baik', 'koperasi', 'seperti']	1	1

'saatini', 'lintah', 'darat',
'berkedok', 'koperasiiii']

2.4 Split Data

Tahap selanjutnya melibatkan pembagian dataset ke dalam dua komponen, diantaranya data latih serta data uji, dalam tiga scenario rasio 70%-30%, 80%-20%, 90%-20%. Data latih dimanfaatkan dalam membangun serta melatih model untuk mengenali pola sentiment, akan tetapi data uji digunakan dalam mengevaluasi kinerja model terhadap data yang belum pernah diproses sebelumnya. Penggunaan beberapa variasi rasio ini bertujuan untuk menilai kestabilan dan keandalan model dalam menganalisis teks.

Pembobotan TF-IDF

Secara matematis, pembobotan TF-IDF dirumuskan sebagai berikut:

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

dengan:

TF(t,d) adalah frekuensi kemunculan term t dalam dokumen d

IDF(t) mempresentasikan inverse document frequency dari term t

Nilai IDF dihitung menggunakan persamaan:

$$IDF(t) = \log \left(\frac{N}{df(t)} \right) \quad (2)$$

Keterangan :

N adalah total jumlah dokumen

df(t) adalah jumlah dokumen yang memuat term t

Pada tahap pembobotan fitur, teks pada data latih dan data uji diubah menjadi representasi vektor numerik menggunakan metode TF-IDF. Metode ini memberikan nilai bobot pada setiap kata sesuai tingkat kepentingannya dalam dokumen, sehingga dapat digunakan sebagai fitur input pada algoritma klasifikasi.

2.5 Pembobotan TF-IDF

a. Naïve Bayes

Algoritma Naïve Bayes menggunakan pendekatan probabilistik dalam proses klasifikasi dengan mengacu pada Teorema Bayes serta mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya. Secara matematis, Teorema Bayes dinyatakan sebagai berikut:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (3)$$

Keterangan:

P(C|X) merupakan probabilitas kelas C dengan kondisi fitur X.

P(X|C) merupakan probabilitas fitur X pada kondisi kelas C.

P(C) merupakan probabilitas prior dari kelas C.

P(X) merupakan probabilitas dari fitur X.

Dalam proses klasifikasi, nilai P(X) bersifat konstan sehingga tidak memengaruhi perbandingan antar kelas dan dapat diabaikan. Penentuan sentimen dilakukan dengan memilih kelas yang memiliki nilai probabilitas posterior tertinggi.

b. Logistic Regression

Logistic Regression menggunakan pendekatan fungsi logistik untuk memodelkan probabilitas suatu data termasuk ke dalam kelas tertentu. Proses klasifikasi dilakukan dengan menghitung nilai probabilitas menggunakan fungsi sigmoid. Secara matematis, fungsi tersebut dinyatakan sebagai berikut:

$$P(Y = 1|X) = \frac{1}{(1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)})} \quad (4)$$

Keterangan:

$P(Y=1|X)$ merupakan probabilitas data termasuk ke dalam kelas tertentu.

β_0 merupakan konstanta (intercept).

$\beta_1, \dots, \beta_n$ merupakan koefisien masing-masing fitur.

$X_1, \dots, X_n$ merupakan nilai fitur.

e merupakan bilangan eksponensial.

Nilai probabilitas hasil perhitungan kemudian dibandingkan dengan ambang batas (*threshold*) yang telah ditentukan. Data akan diklasifikasikan sebagai kelas positif jika nilainya melebihi ambang batas atas, dan kelas negatif jika berada di bawah ambang batas bawah. Sementara itu, data yang nilainya berada di antara kedua ambang batas tersebut diklasifikasikan sebagai kelas netral.

2.1. Validasi Data

Evaluasi kinerja model dilakukan untuk menilai kemampuan dua algoritma dalam mengklasifikasikan sentimen ulasan aplikasi. Proses evaluasi ini menggunakan *confusion matrix* dan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, serta *F1-score*.

Accuracy digunakan untuk mengetahui tingkat ketepatan model dalam melakukan klasifikasi, dengan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision menunjukkan ketepatan model saat memprediksi kelas positif, yang dihitung dengan rumus:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall menggambarkan kemampuan model dalam mengidentifikasi data yang benar-benar termasuk kelas positif, dengan rumus:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

F1-score merupakan gabungan dari *precision* dan *recall* dalam bentuk rata-rata harmonis, yang dirumuskan sebagai:

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Keterangan :

- TP = Jumlah data dari kelas target (Positif / Netral / Negatif) yang berhasil diprediksi dengan benar oleh model.
- TN = Jumlah data dari kelas lain yang berhasil diprediksi secara benar bukan sebagai kelas target.
- FP = Jumlah data dari kelas lain yang salah diprediksi oleh model sebagai kelas target.

FN = Jumlah data dari kelas target yang salah diprediksi oleh model sebagai kelas lain.

III. HASIL DAN PEMBAHASAN

3.1. Scrapping

Dataset sebanyak 11821 ulasan 10 Juni 2026 hingga 4 Juli 2026 dikumpulkan melalui YouTube Data API dan dilabeli secara otomatis menggunakan Lexicon. Hasil pelabelan menunjukkan sentimen positif sebesar 15,3% (1.804 ulasan), sentimen negatif sebesar 65,7% (7.767 ulasan), dan sentiment netral sebesar 19,0% (2.250 ulasan).

3.2 Preprocessing

Tahap preprocessing data dilakukan untuk mengubah komentar mentah menjadi data teks yang bersih dan terstruktur melalui lima tingkatan utama. Teks diawali dengan *case folding* untuk menyeragamkan huruf menjadi *lowercase*, *cleansing* untuk menghapus tanda baca, serta *tokenizing* untuk memotong kalimat menjadi kata tunggal. Teks tersebut kemudian melewati tahap *normalization* guna mengoreksi kata tidak baku (seperti "klo" menjadi "kalo"),

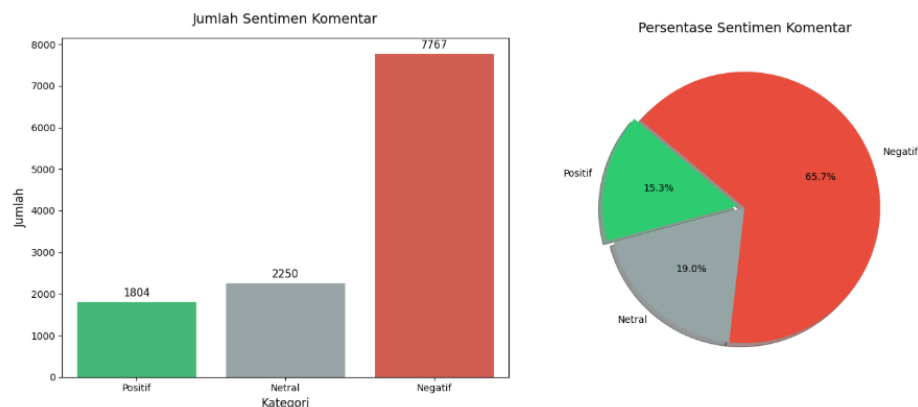
dan diakhiri dengan *stopword removal* untuk membuang kata hubung yang tidak informatif. Contoh bisa dilihat di tabel 2. Dapat dilihat juga wordcloud sebelum dan sesudah preprocessing pada gambar dibawah ini.



Gambar 2. Word cloud sebelum dan setelah preprocessing

3.3 Labeling

Setelah tahap preprocessing, data komentar mengenai Koperasi Desa (Kopdes) dilabeli sentimen secara otomatis menggunakan *Lexicon* ke dalam tiga kategori: positif (1), negatif (-1), dan netral (0). Pelabelan dilakukan dengan mengakumulasi bobot kata berdasarkan kamus sentimen. Contohnya dapat dilihat di tabel 3. Untuk hasil labeling jumlah dan presentase bisa dilihat di gambar bawah ini.



Gambar 3. Jumlah dan Presentase labeling

3.4 Splitting dan Pengujian Model

Pemodelan klasifikasi ini memanfaatkan dua algoritma, yaitu Naive Bayes dan Logistic Regression, melalui tiga skenario rasio pembagian data, yakni 70:30, 80:20, dan 90:10 sebagaimana disajikan pada Tabel 4. Berdasarkan hasil eksperimen tersebut, Logistic Regression secara konsisten menghasilkan akurasi tertinggi di seluruh skenario pengujian, dengan nilai puncak sebesar 0,82 pada rasio 90:10. Di sisi lain, Naive Bayes memberikan performa yang stabil dengan tingkat akurasi pada kisaran 0,67–0,68. Temuan ini menegaskan bahwa Logistic Regression merupakan model yang paling efektif dan unggul dalam memetakan sentimen publik terkait Koperasi Desa (Kopdes) Merah Putih secara optimal.

Tabel 4. Hasil perbandingan Naïve dan Logistic dengan tiga pembagian data

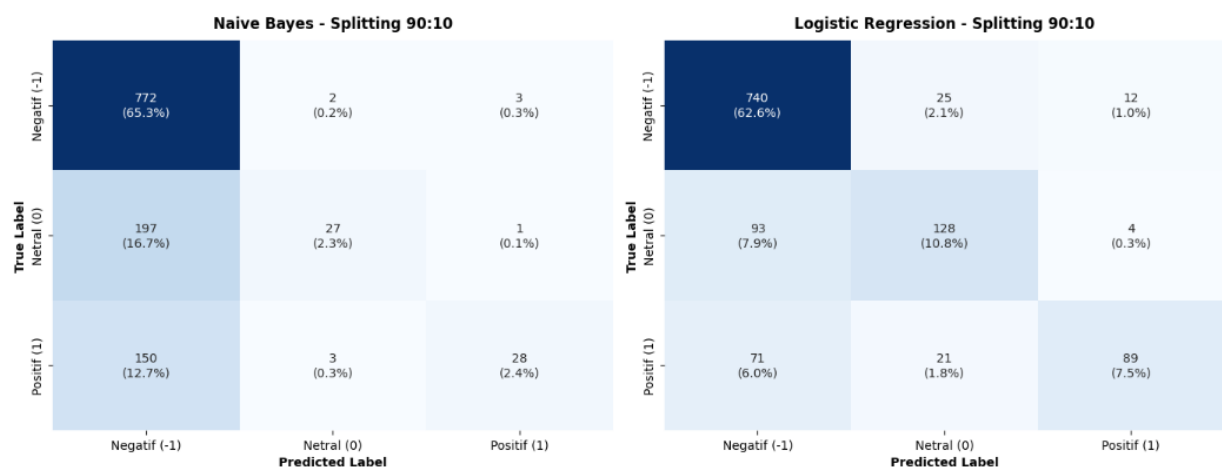
Model	Rasio	Accuracy	Precision	Recall	F1-Score
-------	-------	----------	-----------	--------	----------

Naive Bayes	70:30	0.67	0.72	0.67	0.56
Logistic Regression	70:30	0.81	0.81	0.81	0.80
Naive Bayes	80:20	0.67	0.71	0.67	0.57
Logistic Regression	80:20	0.81	0.80	0.81	0.79
Naive Bayes	90:10	0.67	0.74	0.67	0.56
Logistic Regression	90:10	0.81	0.82	0.81	0.80

Pemodelan Dapat disimpulkan model terbaik adalah Logistic Regression dengan rasio 90:10 yang menunjukkan skor akurasi sebesar 0,81.

3.5 Confussion matrix rasio terbaik

Pengujian kinerja algoritma Naive Bayes dan Logistic Regression dievaluasi menggunakan confusion matrix ukuran 3 x 3 pada tiga skenario pembagian data (data splitting), yaitu 90:10, 80:20, dan 70:30 untuk mengklasifikasikan kelas sentimen negatif (-1), netral (0), dan positif (1). Confusion matrix digunakan untuk mengukur ketepatan prediksi model dengan membandingkan label aktual (True Label) terhadap hasil prediksi (Predicted Label). Berdasarkan hasil pengujian pada seluruh skenario, algoritma Logistic Regression dengan split data 90:10 mencatatkan performa terbaik dengan nilai akurasi mencapai 0,82 (82%), yang menunjukkan keunggulannya dalam memprediksi ketiga kelas sentimen secara optimal dibandingkan kombinasi skenario lainnya.



Gambar 4. Confussion matrix rasio terbaik (90:10)

VII. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, analisis sentimen terhadap komentar YouTube mengenai pembentukan Program Koperasi Desa Merah Putih menunjukkan bahwa opini masyarakat terbagi menjadi sentimen positif, negatif, dan netral. Data yang digunakan berjumlah 11.821 komentar dan telah melalui tahap preprocessing serta labeling berbasis lexicon. Hasil pelabelan sentimen menunjukkan bahwa mayoritas komentar masyarakat terhadap pembentukan Program Koperasi Desa Merah Putih berada pada kategori sentimen negatif. Dari total 11.821 komentar yang dianalisis, sebanyak 7.767 komentar atau 65,7% termasuk dalam sentimen negatif. Hal ini menunjukkan bahwa sebagian besar masyarakat memberikan tanggapan berupa kritik, keraguan, ketidaksetujuan, atau kekhawatiran terhadap program tersebut. Selanjutnya, sentimen netral menempati posisi kedua dengan jumlah 2.250 komentar atau 19,0%. Komentar netral umumnya tidak menunjukkan dukungan maupun penolakan secara jelas, melainkan bersifat informatif, bertanya, atau hanya menanggapi secara umum. Sementara itu, sentimen positif memperoleh jumlah paling rendah, yaitu 1.804 komentar atau 15,3%, yang menunjukkan adanya sebagian masyarakat yang mendukung, menyetujui, atau menaruh harapan terhadap keberhasilan Program Koperasi Desa Merah Putih.

Pengujian menggunakan algoritma Naive Bayes dan Logistic Regression menunjukkan bahwa Logistic Regression memberikan performa terbaik pada skenario pembagian data 90:10 dengan skor akurasi sebesar 0,81. Dengan demikian, Logistic Regression dapat dinyatakan sebagai model yang lebih efektif dalam mengklasifikasikan sentimen masyarakat terhadap program Koperasi Desa Merah Putih.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Allah SWT. Dosen pembimbing penulis ibu Nuril Lutvi Azizah, S.Si., M.Si., Kepala program studi Informatika ibu Ade Eviyanti, S.Kom., M.Kom., Dosen penguji, keluarga penulis yang telah mensupport penulis selama ini, rekan-rekan penulis yang telah membantu, serta penyemangat penulis dalam menempuh Pendidikan S1 yaitu Nn. Angelina Christy. Juga Universitas Muhammadiyah Sidoarjo atas dukungan fasilitas laboratorium komputer untuk menunjang pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada bagian administrasi kampus atas bantuan dan dukungan administratif selama proses penelitian berlangsung.

REFERENSI

- [1] P. Arthayoga, L. Fanggi, and B. Evangelista, "Kedudukan Hukum Koperasi Merah Putih," vol. 5, no. 1, pp. 1367–1378, 2025.
- [2] P. S. Manajemen, F. Ekonomi, and U. Muhammadiyah, "TRANSFORMASI EKONOMI DESA MELALUI KOPERASI MERAH," vol. 12, 2025.
- [3] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1, pp. 131–145.
- [4] A. Ridwan, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," vol. IV, no. September, pp.
- [5] J. M. Polgan et al., "Algoritma Naive Bayes untuk Mengidentifikasi Hoaks di Media Sosial," vol. 13, pp. 2020–2025, 2024.
- [6] A. F. Watratan, A. P. B, D. Moeis, S. Informasi, and S. P. Makassar, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," vol. 1, no. 1, pp.
- [7] S. H. Junaidi, T. Sjah, and K. Budastra, "Mungkinkah Koperasi Merah Putih Berhasil," vol. 4, no. November, pp. 422–434, 2025.
- [8] V. N. Mei, R. Juanda, T. U. Sabilla, W. Kurniawati, and C. Fransisca, "Implementasi Program Koperasi Merah Putih dalam Mendorong Kemandirian Ekonomi Desa manfaat yang merata bagi masyarakat perdesaan (Nugroho & Setyowati , 2023).," no. April, 2026.
- [9] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naive Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," vol. 8, no. April, pp. 714–725, 2024, doi: 10.30865/mib.v8i2.7458.
- [10] T. Utary et al., "Komparasi Support Vector Machine dengan Logistic Regression terhadap klasifikasi pada data asma," vol. 9, no. 1, pp. 1–8, 2026.
- [11] M. Monica and A. Purwanto, "Comparative Performance Evaluation of Naive Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification Evaluasi Komparatif Kinerja Algoritma Naive Bayes dan Logistic Regression dalam Klasifikasi Sentimen Komentar YouTube Berbahasa Indonesia," vol. 6, no. April, pp. 934–943, 2026.
- [12] A. S. Wijaya, A. Fahmi, Y. Tyas, and C. Pramudi, "Sentiment Classification of Health Education YouTube Comments Using IndoBERT Embeddings with Logistic Regression and Naive Bayes," vol. 10, no. 3, pp. 3005–3020, 2026.
- [13] A. A. Ningtyas, A. Solichin, and R. Pradana, "ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON PREDICTION OF ECONOMIC RECESSION IN 2023 USING THE NAIVE BAYES," vol. 20, no. 1, pp. 9–16, 2023.
- [14] C. A. Misrun, E. Haerani, M. Fikry, and E. Budianita, "Jurnal Computer Science and Information Technology (CoSciTech) naive bayes classifier method," vol. 4, no. 1, pp. 207–215, 2024.
- [15] S. F. Huwaida, R. Kusumawati, and B. Isnaini, "Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naive Bayes," vol. 6, no. 1, pp. 26–39, 2024, doi: 10.37905/jji.v6i1.24718.
- [16] D. Prastyo, I. H. Mursyidin, and D. Irawan, "Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten," vol. 7, no. 2, 2024, doi: 10.32877/bt.v7i2.1833.
- [17] A. Maulana, D. Dyantono, and R. E. Putra, "Perbandingan Sent2vec TF-IDF Logistic Regression dan Word2vec CNN pada hasil Sentiment Analysis Youtube Comment," vol. 05, pp. 63–72, 2023.

- [18] H. S. Mulyono and U. Saprudin, “Efektivitas Logistic Regression dalam Analisis Sentimen Berbahasa Indonesia pada Komentar YouTube tentang Isu,” vol. 6, no. 3, pp. 1547–1555, 2025.
- [19] D. Marganingsih, H. Oktavianto, and G. Abdurrahman, “Analisis Sentimen Komentar Youtube Masterchef Indonesia Menggunakan Algoritma Support Vector Machine Dan Gaussian Naïve Bayes Sentiment Analysis of YouTube Comments on Masterchef Indonesia Using Support Vector Machine and Gaussian Naïve Bayes Algorithms,” vol. 5, no. 1, pp. 16–26, 2025, doi: 10.59395/jitp.v5i1.117.
- [20] R. B. Aprianto, A. Eviyanti, H. Setiawan, I. Ratna, and I. Astutik, “DEDI MULYADI DENGAN METODE MACHINE LEARNING (STUDI KASUS : KENAKALAN REMAJA),” vol. 10, no. 2, pp. 2997–3004, 2026.
- [21] Z. Y. Burnama, M. A. Rosid, and N. L. Azizah, “Analisis Sentimen Pada Komentar Youtube Dalam Turnamen MPL Season 13 Dengan Metode Ensemble Machine Learning Sentiment Analysis on YouTube Comments in MPL Season 13 Tournament Using Ensemble Machine Learning Method”.
- [22] A. Eviyanti and N. Ariyanti, “ANALYSIS OF APRIORI AND K-NEAREST NEIGHBOR (KNN) ALGORITHM IN RECOMMENDING APPROPRIATE LEARNING,” vol. 20, no. 1, pp. 557–572, 2026.
- [23] V. N. Juli, A. Jl, R. Madya, G. Anyar, K. Gn, and J. Timur, “Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM Muhammad Fernanda Naufal Fathoni Eva Yulia Puspaningrum Andreas Nugroho Sihananto menghasilkan F1 sebesar 80 %. Penelitian lain dilakukan (Muhammadi et al ., 2022) dengan VADER sehingga menjadikan peneliti tertarik membuat perbandingan kombinasi lexicon InSet dan VADER dengan SVM pada analisa sentimen Rohingya sosial media X untuk mengetahui manakah performa yang lebih baik .,” vol. 1, no. 3, 2024.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.