

ANALISIS SENTIMEN REVIEW AWAL PEMBENTUKAN PROGRAM KOPERASI DESA MERAH PUTIH DENGAN METODE NAÏVE BAYES DAN LOGISTIC REGRESSION

Oleh:

Muhammad Athallah Hazel Ornando 221080200135

Dosen pembimbing: Nuril Lutvi Azizah, S.Si., M.Si.

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Agustus, 2026



www.umsida.ac.id



[umsida1912](#)



[umsida1912](#)



[universitas
muhammadiyah
sidoarjo](#)



[umsida1912](#)

Latar Belakang

- Program Koperasi Desa Merah Putih merupakan program pemerintah berdasarkan Instruksi Presiden Nomor 9 Tahun 2025 yang menargetkan pembentukan, pengembangan, dan revitalisasi 80.000 koperasi desa atau kelurahan untuk memperkuat ekonomi kerakyatan dan meningkatkan kesejahteraan masyarakat desa.
- YouTube menjadi salah satu media yang merekam respons awal masyarakat terhadap pembentukan Program Koperasi Desa Merah Putih melalui kolom komentar pada konten-konten yang membahas atau mengulas program tersebut.
- Analisis sentimen terhadap komentar YouTube dipilih untuk mengetahui respons atau persepsi masyarakat terhadap pembentukan Koperasi Desa Merah Putih.

Batasan Masalah

1. Penelitian ini hanya menggunakan data komentar dari platform YouTube yang membahas program Koperasi Merah Putih.
2. Data komentar yang digunakan merupakan komentar berbahasa Indonesia.
3. Sentimen komentar diklasifikasikan pada penelitian ini terdiri atas tiga kategori, yaitu sentimen positif, sentimen negatif, dan netral.
4. Proses pengambilan data dilakukan menggunakan teknik web scraping pada komentar YouTube.

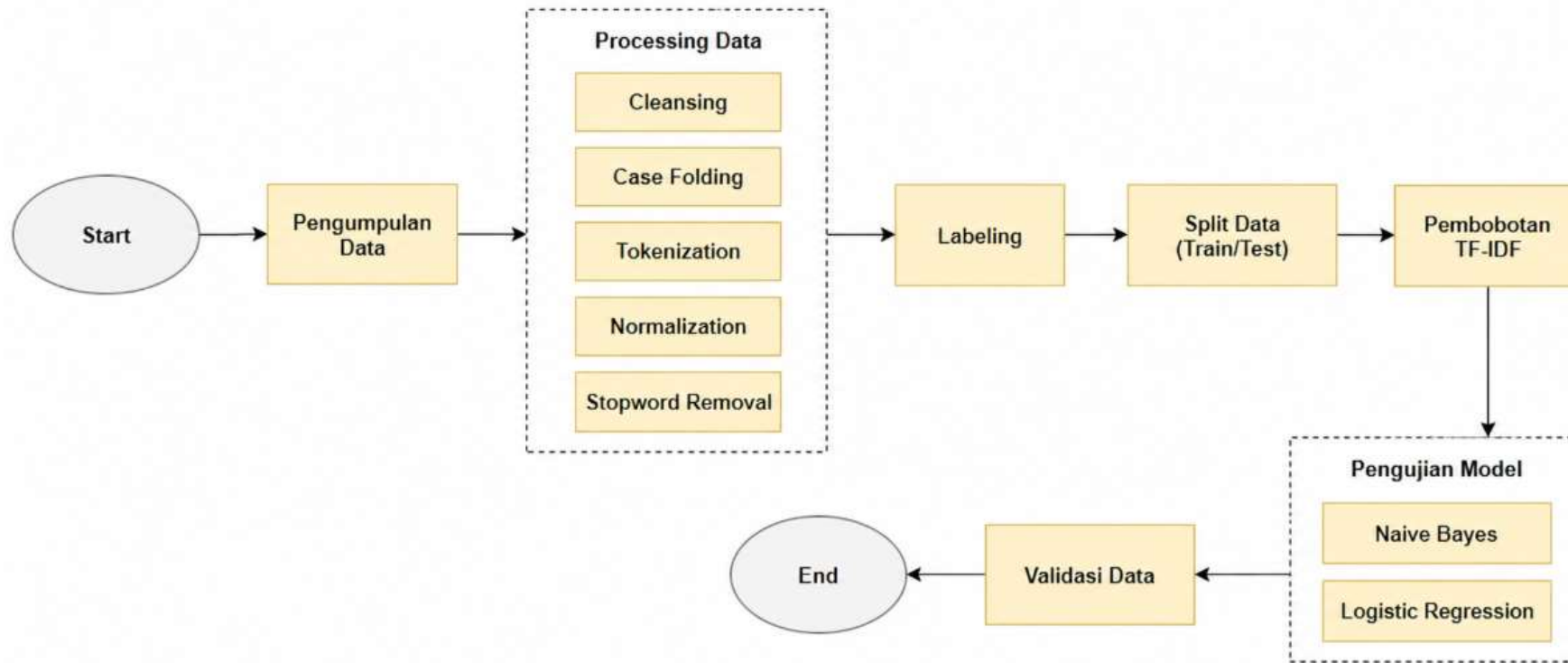
Batasan Masalah

5. Data komentar diambil dari platform YouTube menggunakan 5 video yang di upload oleh channel berita, yang terdiri dari satu video Kompas.com, satu video Tempodotco, satu video dari Hersubeno Point, satu video dari Fristian Greic Media, dan satu video dari Liputan6.
6. Jumlah dataset yang didapat kurang lebih sejumlah 11.821 data comments dengan 1 video diberikan batasan 6000 komentar.
7. Dataset merupakan data komentar yang hanya bertipe text tidak termasuk emoji.

Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui hasil performa metode Naïve Bayes dan Logistic Regression dalam analisis sentimen komentar Youtube terhadap terhadap awal pembentukan koperasi merah putih.

Tahapan Penelitian



1. Pengumpulan Data

Penelitian ini dilakukan secara online dengan data komentar pengguna yang dikumpulkan dari platform YouTube. 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). Setiap langkah penelitian—mulai dari pengumpulan data, pengolahan data, hingga analisis—dilakukan secara mandiri menggunakan perangkat komputer dan koneksi internet.

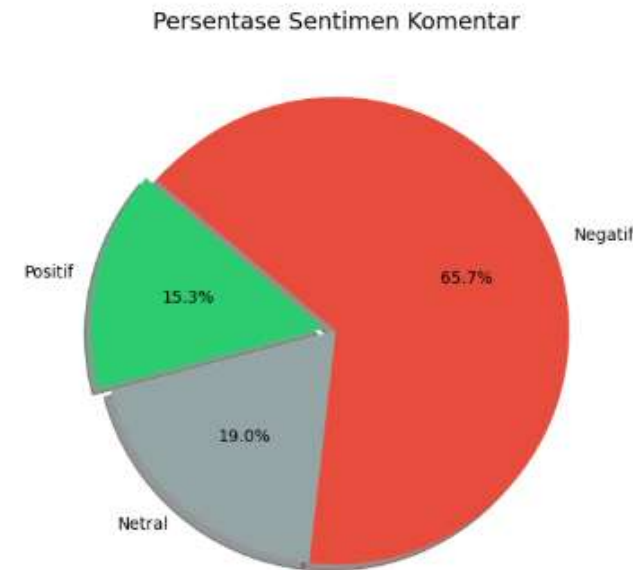
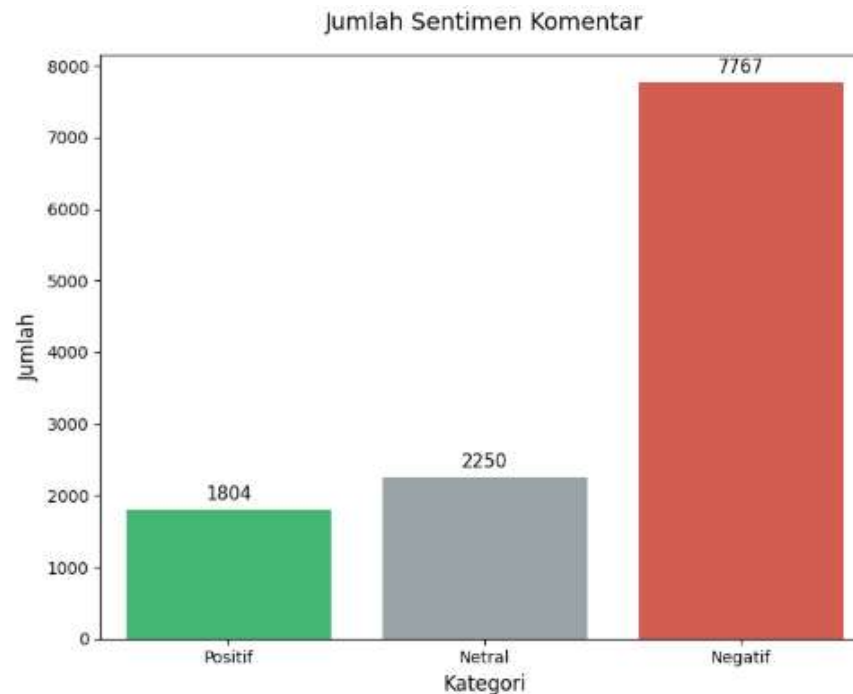
2. Preprocessing Data

1. **Case Folding** → mengubah seluruh komentar YouTube ke dalam huruf kecil agar bentuk teks lebih seragam dan memudahkan.
2. **Cleaning** → menghapus elemen yang tidak diperlukan dalam data teks, seperti URL, tanda baca, simbol, dan karakter lain yang dapat mengganggu proses analisis.
3. **Tokenization** → memecah teks atau kalimat menjadi bagian-bagian kecil yang disebut "token". ['ujungya', 'lahan', 'korupsi', 'lg']
4. **Normalization** → mengubah kata tidak baku menjadi bentuk baku, misalnya "lg" → "lagi".
5. **Stopword Removal** → menghapus kata umum yang tidak relevan seperti "yang", "di", "atau", "ke".

Tahapan ini memastikan teks dalam bentuk yang lebih rapi dan konsisten untuk analisis

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.



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.

4. Splitting dan Pengujian Model

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

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.

5. Confussion matrix rasio terbaik

Naive Bayes - Splitting 90:10

True Label	Negatif (-1)	Netral (0)	Positif (1)
	772 (65.3%)	2 (0.2%)	3 (0.3%)
	197 (16.7%)	27 (2.3%)	1 (0.1%)
Negatif (-1)	150 (12.7%)	3 (0.3%)	28 (2.4%)
Predicted Label			

Logistic Regression - Splitting 90:10

True Label	Negatif (-1)	Netral (0)	Positif (1)
	740 (62.6%)	25 (2.1%)	12 (1.0%)
	93 (7.9%)	128 (10.8%)	4 (0.3%)
Negatif (-1)	71 (6.0%)	21 (1.8%)	89 (7.5%)
Predicted Label			

6. Kesimpulan

Data penelitian terdiri atas 11.821 komentar yang telah melalui preprocessing dan labeling berbasis lexicon. Hasil analisis menunjukkan sentimen negatif mendominasi sebesar 65,7% (7.767 komentar), diikuti sentimen netral 19,0% (2.250 komentar) dan positif 15,3% (1.804 komentar). Hal ini menunjukkan bahwa tanggapan masyarakat terhadap Program Koperasi Desa Merah Putih cenderung berupa kritik, keraguan, dan kekhawatiran. Dari pengujian Naive Bayes dan Logistic Regression, performa terbaik diperoleh Logistic Regression pada pembagian data 90:10 dengan akurasi 0,81, sehingga dinilai lebih efektif dalam mengklasifikasikan sentimen masyarakat.

Jadwal Penelitian

JADWAL PENELITIAN

Tabel 3.8. Jadwal Penelitian

No	Kegiatan	Waktu				
		1	2	3	4	5
1	Penulisan proposal					
2	Seminar dan bimbingan proposal					
3	Penelitian dan tindakan					
4	Analisis dan bimbingan hasil penelitian					
5	Ujian skripsi					

Daftar Pustaka

- [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 Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," vol. IV, no. September, pp. 15–21.
- [5] J. M. Polgan et al., "Algoritma Naïve 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.
- [7] S. H. Junaidi, T. Sjah, and K. Budastra, "Mungkinkah Koperasi Merah Putih Berhasil," vol. 4, no. November, pp. 422–434, 2025.

Daftar Pustaka

- [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 Naïve 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 Naïve Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification Evaluasi Komparatif Kinerja Algoritma Naïve 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 Naïve 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 NAÏVE BAYES," vol. 20, no. 1, pp. 9–16, 2023.

Daftar Pustaka

- [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 Naïve 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.

Daftar Pustaka

- [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.



**SKRIPSI TANPA KEKASIH,
UNTUNGNYA ADA CHRISTY MY OSHI,
SEKIAN DAN TERIMAKASIH**