

Analisis Sentimen terhadap Koperasi Desa Merah Putih Menggunakan *Random Forest* dan Pelabelan Leksikon

Steven Dermawan

Sistem Informasi, Fakultas Sains dan Teknologi
Universitas Pradita, Tangerang Selatan

*Email steven.dermawan@student.pradita.ac.id

Sejarah Artikel:

Diterima 04-04-2026
Disetujui 11-04-2026
Diterbitkan 13-04-2026

ABSTRACT

This study aims to analyze public sentiment toward the Koperasi Desa Merah Putih program during the period July–October 2025, using 5396 tweets as the primary data source. The research follows a systematic workflow consisting of data crawling, text preprocessing, lexicon-based sentiment labeling, dataset splitting, TF-IDF vectorization, Random Forest model training, evaluation, and word cloud visualization. Sentiment labeling was conducted using a lexicon-based approach and subsequently classified using the Random Forest algorithm. The findings indicate that public opinion is dominated by negative sentiment at 87.19%, followed by positive sentiment at 12.8%, reflecting both expectations for strengthening village economic empowerment and concerns regarding the program's implementation effectiveness. Model evaluation shows an accuracy of 90%, macro precision of 0.91, macro recall of 0.60, and a macro F1-score of 0.63, indicating that although the model performs well overall, class balance still requires improvement. The word cloud visualization highlights dominant terms such as merah, putih, koperasi, desa, and kopdes, underscoring public attention to the program's identity and its relevance to rural economic development. The presence of terms such as economy, community, people, and food reflects expectations for improved welfare, while president, minister, and government indicate the program's connection to national policy agendas. Overall, this study provides a comprehensive understanding of public perceptions and emphasizes the importance of enhancing transparency, public communication, and program implementation effectiveness. Future research is encouraged to employ hybrid labeling methods or deep learning models and to expand the data scope to achieve more accurate and balanced sentiment classification.

Keywords: *Sentiment Analysis; Merah Putih Village Cooperative; Random Forest; TF-IDF; Word Cloud*

ABSTRAK

Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap program Koperasi Desa Merah Putih pada periode Juli–Oktober 2025 dengan menggunakan 5396 tweet sebagai sumber data utama. Proses penelitian mengikuti alur kerja sistematis meliputi crawling data, preprocessing teks, pelabelan sentimen berbasis leksikon, pembagian dataset, vektorisasi TF-IDF, pelatihan model Random Forest, evaluasi, serta visualisasi word cloud. Sentimen dilabeli menggunakan metode lexicon-based dan kemudian diklasifikasikan melalui algoritma Random Forest. Hasil analisis menunjukkan bahwa opini publik didominasi oleh sentimen negatif sebesar 87.19%, diikuti sentimen positif sebesar 12.8%, menggambarkan adanya harapan terhadap penguatan ekonomi desa sekaligus kekhawatiran mengenai efektivitas pelaksanaannya. Evaluasi model menghasilkan akurasi 90%, presisi makro 0,91, recall makro

0,60, dan F1-score makro 0,63, menunjukkan performa klasifikasi yang baik meski keseimbangan antar kelas masih perlu ditingkatkan. Word cloud menampilkan dominasi kata seperti merah, putih, koperasi, desa, dan kopdes, yang menegaskan fokus publik pada identitas program serta relevansinya bagi pembangunan ekonomi desa. Kemunculan istilah ekonomi, masyarakat, rakyat, dan pangan mencerminkan ekspektasi peningkatan kesejahteraan, sementara kata presiden, menteri, dan pemerintah menunjukkan keterkaitan program dengan agenda kebijakan nasional. Secara keseluruhan, penelitian ini memberikan gambaran komprehensif mengenai persepsi publik dan menegaskan pentingnya peningkatan transparansi, komunikasi publik, dan efektivitas implementasi program. Penelitian selanjutnya disarankan menggunakan metode pelabelan hybrid atau model deep learning serta memperluas cakupan data untuk menghasilkan analisis sentimen yang lebih akurat dan seimbang.

Katakunci: Analisis Sentimen; Koperasi Desa Merah Putih; *Random Forest*; *TF-IDF*; *Word Cloud*

Bagaimana Cara Sitasi Artikel ini:

Dermawan, S. (2026). Analisis Sentimen terhadap Koperasi Desa Merah Putih Menggunakan Random Forest dan Pelabelan Leksikon. *Jejak Digital: Jurnal Ilmiah Multidisiplin*, 2(3), 3820-3831. <https://doi.org/10.63822/9p4b3715>

PENDAHULUAN

Koperasi Desa Merah Putih (Kopdes Merah Putih) merupakan program strategis nasional yang dibentuk sebagai upaya pemerintah memperkuat kemandirian ekonomi desa melalui pengembangan koperasi modern berbasis digital (Aprilis & Zulkarnain, 2025). Koperasi ini dirancang sebagai pusat layanan ekonomi desa terintegrasi, mencakup distribusi pangan, simpan-pinjam, logistik, hingga layanan kesehatan, dengan target pembentukan puluhan ribu unit di seluruh Indonesia. Namun, di tengah perluasan implementasinya, persepsi publik terhadap Kopdes Merah Putih masih beragam. Sebagian memandangnya sebagai langkah positif menuju digitalisasi desa, sementara yang lain mengkhawatirkan efektivitas, kesiapan SDM, dan tata kelola. Keragaman respons ini penting dipahami karena keberhasilan koperasi bergantung tidak hanya pada kebijakan, tetapi juga pada penerimaan masyarakat.

Dalam konteks inilah analisis sentimen menjadi alat yang relevan untuk mengidentifikasi kecenderungan opini publik terhadap Kopdes Merah Putih. Analisis sentimen adalah pendekatan komputasional yang memanfaatkan teknik pemrosesan bahasa alami (NLP) untuk mengklasifikasikan emosi, penilaian, atau kecenderungan opini dalam teks, sehingga mampu menggambarkan pola persepsi masyarakat secara lebih objektif (Nurwanda et al., 2024). Melalui metode ini, data teks yang bersifat tidak terstruktur, seperti komentar media sosial, dapat diubah menjadi informasi yang terukur dan dapat dibandingkan (Sholekhah & Muntahanah, 2025). Meskipun isu digitalisasi koperasi dan penguatan ekonomi desa telah banyak dibahas, penelitian yang secara spesifik mengkaji persepsi masyarakat terhadap Kopdes Merah Putih masih sangat terbatas. Informasi yang tersedia umumnya berasal dari laporan pemerintah, publikasi kebijakan, dan artikel populer tanpa analisis ilmiah yang sistematis. Oleh karena itu, penelitian ini penting untuk memberikan pemahaman empiris tentang pandangan masyarakat sekaligus menjadi dasar bagi perumusan kebijakan yang lebih responsif, partisipatif, dan berbasis bukti.

Dalam konteks tersebut, penelitian ini akan menggunakan algoritma Random Forest untuk menganalisis sentimen masyarakat terkait Kopdes Merah Putih. Algoritma Random Forest adalah metode *ensemble learning* yang bekerja dengan membangun sejumlah pohon keputusan (*decision trees*) dan menggabungkan hasil prediksinya untuk memperoleh klasifikasi yang lebih stabil (Ernawati et al., 2025). Pendekatan ini mampu mengurangi risiko *overfitting* karena setiap pohon dilatih menggunakan subset data dan fitur yang berbeda, sehingga menghasilkan model yang lebih robust dalam menangkap variasi pola linguistik (Ningsih et al., 2024). Algoritma ini dikenal memiliki akurasi tinggi dalam klasifikasi teks, khususnya pada dataset dengan pola bahasa yang kompleks seperti opini masyarakat di media sosial (Setiawan et al., 2025).

Penelitian ini merujuk kepada beberapa penelitian terdahulu yang terkait. Pertama, penelitian Putra et al. (2025) terkait analisis sentimen Trump tariff menggunakan Random Forest di X. Hasil penelitian menunjukkan bahwa model Random Forest yang dilatih dengan pembagian data 70% pelatihan dan 30% pengujian serta dievaluasi menggunakan 10-fold cross-validation mampu mengklasifikasikan sentimen dengan akurasi sebesar 87,38%. Kedua, penelitian Fauzi et al. (2025) terkait analisis sentimen koperasi desa merah putih menggunakan Naïve Bayes dan SVM di X. Hasil penelitian menunjukkan bahwa Naive Bayes lebih unggul dalam menyeimbangkan prediksi dan menangkap variasi opini pada teks pendek seperti tweet (AUC 0,71; F1-score 0,62), sementara SVM hanya efektif pada kelas netral, yang juga mendominasi distribusi sentimen publik. Hasil penelitian juga menunjukkan persepsi publik terhadap Koperasi Desa Merah Putih didominasi sentimen netral, disusul positif dan sangat sedikit negatif, yang mencerminkan sikap masyarakat yang rasional dan berhati-hati.

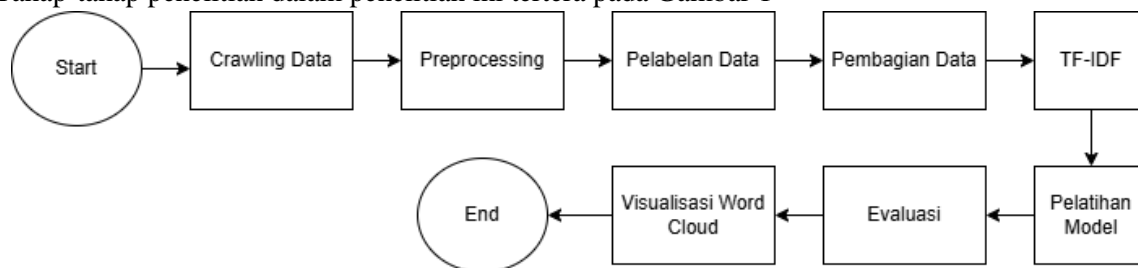
Ketiga, penelitian Manullang et al. (2023) terkait memprediksi hasil calon pemilu presiden menggunakan Lexicon dan Random Forest. Hasil penelitian menunjukkan bahwa metode Lexicon-Based dan Random Forest mampu memprediksi sentimen pengguna Twitter terhadap kandidat Pemilihan Presiden dengan akurasi keseluruhan 88%. Keempat, penelitian Azmi et al. (2023) terkait analisis sentimen user review jamsostek di Google Play Store menggunakan Random Forest dan pelabelan Lexicon. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi klasifikasi review dengan performa model yang tinggi (akurasi 93%, precision 88%, recall 93%, dan F1-score 90%). Kelima, penelitian Nugraha & Papatungan (2025) terkait analisis sentimen video review starlink Indonesia di Youtube menggunakan pelabelan Lexicon dan Random Forest. Hasil penelitian menunjukkan bahwa rasio data 80:20 menghasilkan performa terbaik dengan akurasi 77,4%, precision 0,78, recall 0,77, dan F1-score 0,77.

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa pendekatan analisis sentimen menggunakan algoritma Random Forest yang dikombinasikan dengan pelabelan berbasis Lexicon masih sangat relevan, efektif, dan menunjukkan performa yang konsisten tinggi pada berbagai konteks data teks digital. Namun, kajian yang secara spesifik dan mendalam menyoroti persepsi masyarakat terhadap program Koperasi Desa Merah Putih masih relatif terbatas dan belum banyak dikaji dengan pendekatan *machine learning* berbasis data percakapan publik terkini.

Tujuan dari penelitian ini adalah (1) menganalisis sentimen masyarakat terhadap Kopdes Merah Putih menggunakan algoritma Random Forest; (2) mengevaluasi kinerja model dalam mengklasifikasikan sentimen berbasis teks; dan (3) mengidentifikasi kata kunci dominan yang membentuk persepsi publik terhadap Kopdes Merah Putih melalui visualisasi Word Cloud, guna menggambarkan isu-isu utama yang paling sering muncul dalam percakapan masyarakat. Dengan demikian, hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan literatur mengenai digitalisasi koperasi sekaligus mendukung perumusan kebijakan berbasis data yang lebih responsif terhadap kebutuhan masyarakat desa.

METODE PENELITIAN

Tahap-tahap penelitian dalam penelitian ini tertera pada Gambar 1



Gambar 1. Metodologi Penelitian

Crawling Data

Tahap pertama penelitian ini dimulai dengan proses *crawling* data menggunakan tweet harvest, untuk mengumpulkan teks berbahasa Indonesia dari media sosial X pada bulan Juli–Oktober 2025. Data hasil *crawling* kemudian disimpan dalam format Excel dan diunggah ke Google Collab untuk diproses lebih lanjut. Dataset kemudian dibaca menggunakan pandas dan diperiksa kembali untuk memastikan kesesuaian nama kolom sebelum memasuki tahap analisis.

Preprocessing Teks

Tahap *preprocessing* bertujuan untuk menstandarkan data teks dan menghilangkan komponen yang tidak relevan agar lebih mudah dianalisis oleh sistem (Aditiya et al., 2022). Proses ini mencakup beberapa langkah utama, yaitu (Hickman et al., 2022):

- a. *Cleaning*: Menghapus URL, angka, tanda baca, mention, hashtag, dan karakter lain yang tidak diperlukan dalam teks.
- b. *Case folding*: Mengonversi seluruh huruf menjadi huruf kecil untuk menghindari perbedaan makna akibat kapitalisasi.
- c. *Normalization*: Menyeragamkan penulisan kata, termasuk memperbaiki karakter berulang seperti “bagusss” menjadi “bagus”.
- d. *Tokenization*: Memecah teks menjadi kata-kata (token) menggunakan `nltk.word_tokenize`.
- e. *Stopword removal*: Menghilangkan kata-kata umum yang kurang bermakna dengan memanfaatkan daftar *stopwords* Bahasa Indonesia dari `stopwords-iso`.
- f. *Stemming*: Mengubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan agar lebih konsisten dalam pemrosesan.

Pelabelan Data dengan Lexicon-Based Sentiment Analysis

Setelah teks melalui tahapan pembersihan, proses pelabelan otomatis dilakukan menggunakan pendekatan *lexicon-based sentiment analysis*. Penelitian ini menggunakan kamus sentimen InSet (Indonesian Sentiment Lexicon) yang diperoleh dari repositori GitHub. Setiap token hasil *preprocessing* dibandingkan dengan kata-kata dalam lexicon; apabila ditemukan kecocokan maka skor sentimen ditambahkan atau dikurangi sesuai bobot yang tertera pada kamus. Akumulasi skor menentukan polaritas sentimen: skor ≥ 0 diklasifikasikan sebagai positif, sedangkan skor < 0 diklasifikasikan sebagai negatif

Pembagian Data Train–Test Split

Dataset yang sudah bersih kemudian dibagi menjadi dua bagian menggunakan `train_test_split`: 80% data untuk proses pelatihan dan 20% untuk pengujian. Teknik *stratified sampling* digunakan untuk memastikan distribusi label sentimen tetap proporsional pada kedua subset data. Pembagian ini penting untuk menilai performa model secara objektif dan menghindari bias yang mungkin muncul jika model hanya dilatih pada pola tertentu.

Tahap Ekstraksi Fitur dengan TF-IDF

Untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning, penelitian ini menggunakan TF-IDF. TF-IDF merupakan salah satu metode dalam analisis teks yang berfungsi untuk menentukan tingkat relevansi sebuah kata di dalam suatu dokumen dibandingkan dengan kumpulan dokumen lainnya (Dermawan & Ayunda, 2025). Komponen TF (*Term Frequency*) menggambarkan seberapa sering suatu kata muncul dalam satu dokumen, sedangkan IDF (*Inverse Document Frequency*) menunjukkan seberapa jarang kata tersebut digunakan di keseluruhan korpus. Kedua nilai ini kemudian dikalikan untuk menghasilkan bobot yang mencerminkan pentingnya kata tersebut dalam konteks analisis (Septiani & Isabela, 2023). Proses ini dilakukan dengan `TfidfVectorizer` yang dilengkapi daftar *stopwords* Bahasa Indonesia, batas maksimal 10.000 fitur, dan `ngram_range` (1,2) agar dapat

menangkap pola unigram dan bigram. Representasi TF-IDF inilah yang menjadi masukan utama dalam pelatihan model Random Forest pada tahap berikutnya.

Pelatihan Model Menggunakan Random Forest

Model klasifikasi yang digunakan adalah Random Forest Classifier dengan criterion berbasis *Gini Index*. Pelatihan model dilakukan melalui Pipeline yang menggabungkan TF-IDF dan Random Forest sekaligus, sehingga memungkinkan pemrosesan data yang konsisten. Selanjutnya, dilakukan *Grid Search Cross Validation* ($cv=10$) untuk mencari kombinasi parameter terbaik, termasuk *n_estimators*, *max_depth*, dan *min_samples_split*. Proses ini menghasilkan model dengan performa optimal berdasarkan metrik akurasi tertinggi pada data latih.

Evaluasi Model pada Data Uji

Model terbaik hasil *Grid Search* kemudian diuji pada subset data uji untuk mengukur performa generalisasi. Evaluasi dilakukan dengan menghitung akurasi, *classification report* (precision, recall, F1-score), serta *confusion matrix*. Nilai-nilai tersebut didapat dari persamaan berikut.

$$\text{Akurasi} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\% \quad (5)$$

$$\text{Presisi} = \text{TP} / (\text{TP} + \text{FP}) \times 100\% \quad (6)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \times 100\% \quad (7)$$

$$\text{F1 Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall}) \times 100\% \quad (8)$$

Keterangan dari rumus-rumus di atas (Putri et al., 2024):

- TP artinya True Positif,
- TN artinya True Negatif,
- FP artinya False Positif, dan
- FN artinya False Negatif

Hasil evaluasi ini memberikan gambaran sejauh mana model mampu mengklasifikasikan sentimen secara benar pada data yang belum pernah dilihat sebelumnya.

Visualisasi Word Cloud

Sebagai bagian dari eksplorasi data, penelitian juga menghasilkan Word Cloud untuk memvisualisasikan kata-kata yang paling sering muncul dalam dataset. Visualisasi ini berguna untuk memahami pola bahasa yang dominan serta memberikan perspektif kualitatif mengenai kecenderungan narasi dalam teks.

HASIL DAN PEMBAHASAN

Dataset Awal

Dataset yang digunakan terdiri atas sekumpulan teks yang diperoleh melalui proses *crawling* dan telah disimpan dalam format Excel. Setelah proses pembacaan data dilakukan, terdapat 5412 tweet dataset yang menunjukkan variasi struktur kalimat dan keragaman ekspresi linguistik khas media sosial, termasuk keberadaan URL, mention, angka, dan karakter non-alfabet. Merujuk tabel 1, terlihat beberapa dataset awal yang ada.

Tabel 1. Sampel Data Mentah

No	Tweet
1	@ElvinaNara Efek baik hadir lagi di desa setelah kopdes Merah Putih Tuban operasi normal layanan pun makin beragam & manfaatnya langsung terasa
.....
5412	Mohon atensinya.@DPR_RI dan Bpk.@prabowo Kacau..!!Baru Sehari Diresmikan Presiden Prabowo Kopdes Merah Putih ditutup. KDMP Program dipaksakan dgn waktu singkat Pemdes dan Peng.Kopdes pada kebingungan krn juklas juknis KDMP blm jg diberikan.. https://t.co/eZSRjexGy5

Preprocessing Text

Prapemrosesan data merupakan tahap penting dalam analisis sentimen terhadap Koperasi Desa Merah Putih. Data tweet dibersihkan melalui beberapa tahapan, meliputi pembersihan teks, *case folding*, tokenisasi, penghapusan *stopword*, normalisasi, dan *stemming*. Proses ini bertujuan menghilangkan elemen yang tidak relevan serta mempertahankan kata-kata bermakna agar model dapat mengenali pola sentimen secara lebih akurat. Secara teknis, seluruh teks diubah menjadi huruf kecil, URL, simbol, angka, tanda baca, dan karakter berulang dihapus. Setelah itu dilakukan tokenisasi menggunakan *word_tokenize* dan penghapusan *stopword* Bahasa Indonesia berdasarkan kamus *stopwords-iso*. Hasil *preprocessing* tertera pada Tabel 2.

Tabel 2. Sampel Data Mentah

Tahapan	Tweet
Case folding	@elvinaranara efek baik hadir lagi di desa setelah kopdes merah putih tuban operasi normal layanan pun makin beragam & manfaatnya langsung terasa
Cleaning	efek baik hadir lagi di desa setelah kopdes merah putih tuban operasi normal layanan pun makin beragam manfaatnya langsung terasa
Tokenisasi	['efek','baik','hadir','lagi','di','desa','setelah','kopdes','merah','putih','tuban','operasi','normal','layanan','pun','makin','beragam','manfaatnya','langsung','terasa']
Stopword Removal	['efek','baik','hadir','desa','kopdes','merah','putih','tuban','operasi','normal','layanan','makin','beragam','manfaatnya','langsung','terasa']
Normalisasi + Stemming	['efek','baik','hadir','desa','kopdes','merah','putih','tuban','operasi','normal','layan','makin','ragam','manfaat','langsung','rasa']
Hasil Akhir	efek baik hadir desa kopdes merah putih tuban operasi normal layanan makin ragam manfaat langsung rasa

Pelabelan Sentimen Menggunakan Lexion-Based

Pelabelan sentimen dilakukan menggunakan lexicon InSet yang memuat daftar kata dengan skor positif dan negatif. Setiap token dalam teks diberi nilai sesuai entri kamus, kemudian skor akumulatif

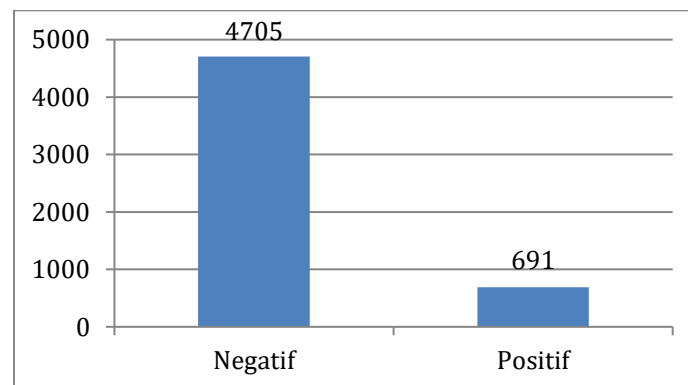
digunakan untuk menentukan label sentimen akhir. Merujuk Gambar 4, distribusi label hasil lexicon-based menunjukkan pola tertentu sesuai kecenderungan isi teks. Penerapan kamus lexicon terbukti mampu menghasilkan label yang konsisten dan sistematis tanpa intervensi manual. Label “Positif” diberikan ketika skor ≥ 0 , sedangkan skor < 0 menghasilkan label “Negatif”. Hasil pelabelan ini kemudian dipakai sebagai label target pada tahap pemodelan *machine learning*.

Tabel 3. Pelabelan Sentimen

No	Tweet	Skor	Label
1	efek baik hadir desa kopdes merah putih tuban operasi normal layanan makin ragam manfaat langsung rasa	1	Positif
2	mohon atensi bpk prabowo kacau baru hari resmi presiden prabowo kopdes merah putih tutup kdmp program paksa waktu singkat pemdes peng kopdes bingung juklas juknis kdmp belum beri	-12	Negatif

Klasifikasi Data

Sebelum tahap penerapan algoritma, perlu dilakukan analisis statistik awal terhadap data untuk memperoleh gambaran umum mengenai distribusi sentimen. Statistik data yang mencakup jumlah kategori positif dan negatif dapat dilihat pada Gambar 2.



Gambar 2. Klasifikasi Data

Dari awal sejumlah 5.412 tweet yang dikumpulkan, dilakukan proses penyaringan untuk menghapus data yang redundan sehingga tersisa total 5.396 tweet yang dianalisis. Hasil analisis menunjukkan bahwa sentimen negatif mendominasi dengan 4.705 tweet atau 87,19%, sedangkan sentimen positif hanya berjumlah 691 tweet atau 12,81%. Komposisi ini memperlihatkan ketimpangan yang cukup signifikan antara kedua kategori sentimen. Secara keseluruhan, proporsi tersebut menggambarkan bahwa persepsi publik terhadap program Koperasi Desa Merah Putih pada periode penelitian cenderung didominasi oleh respons negatif.

Penerapan Algoritma Random Forest

Pelatihan model dilakukan dengan Random Forest Classifier menggunakan *criterion* = gini. Parameter model dioptimalkan menggunakan GridSearchCV yang mengevaluasi kombinasi *n_estimators*, *max_depth*, dan *min_samples_split*. Hasil Grid Search (lihat Tabel 4) menunjukkan kombinasi hiperparameter terbaik dengan skor akurasi tertinggi pada validasi silang. Parameter inilah yang kemudian dipilih sebagai model final.

Tabel 4. Kombinasi Hiperparameter Terbaik

Parameter	Nilai Terbaik
<i>n_estimators</i>	200
<i>max_depth</i>	30
<i>min_samples_split</i>	5

Selanjutnya, evaluasi terhadap data uji menunjukkan bahwa model mampu melakukan generalisasi dengan cukup baik. *Classification report* menampilkan nilai precision, recall, dan F1-score untuk masing-masing kelas sentimen (lihat Gambar 3), yang memberikan gambaran lebih mendalam mengenai kualitas prediksi model. *Confusion matrix* (lihat Gambar 4) memperlihatkan jumlah prediksi benar dan salah pada masing-masing kelas.

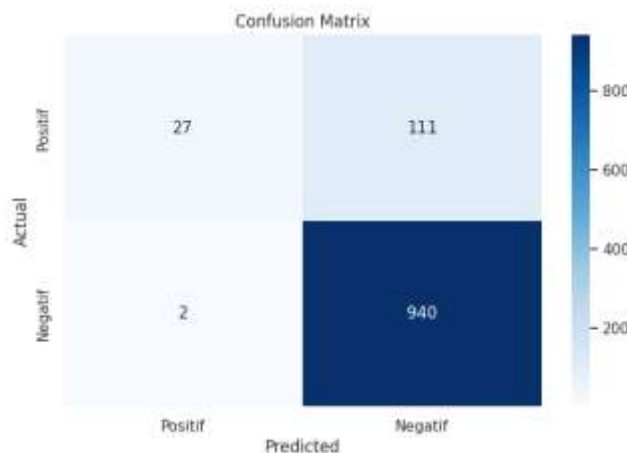
```

Classification Report:
              precision    recall  f1-score   support

   Negatif      0.89      1.00      0.94      942
   Positif      0.93      0.20      0.32      138

 accuracy              0.90      1080
 macro avg              0.91      0.60      0.63      1080
 weighted avg          0.90      0.90      0.86      1080
    
```

Gambar 3. Classification Report



Gambar 4. Confusion Matrix

KESIMPULAN

Penelitian ini menganalisis sentimen masyarakat terhadap Koperasi Desa Merah Putih pada periode Juli–Oktober 2025 menggunakan 5396 tweet sebagai sumber data. Sentimen diberi label menggunakan metode lexicon-based dan kemudian diklasifikasikan dengan algoritma Random Forest. Hasilnya menunjukkan bahwa mayoritas opini publik cenderung negatif sebesar 4705 tweet atau 87.19%, disusul sentimen positif sebesar 691 tweet atau 12.81%, yang mencerminkan dinamika persepsi masyarakat, mulai dari optimisme terhadap potensi pemberdayaan desa hingga kekhawatiran terkait implementasi di lapangan.

Dari sisi performa, model Random Forest menghasilkan akurasi 90%, presisi (macro) 0,91, recall (macro) 0,60, dan F1-score (macro) 0,63. Kinerja ini menandakan bahwa model memiliki kemampuan umum yang baik dalam mengklasifikasikan sentimen, meskipun keseimbangan antar kelas masih perlu ditingkatkan, terutama pada kelas dengan recall rendah. Meski demikian, model tetap mampu memberikan gambaran representatif tentang pola opini masyarakat.

Visualisasi Word Cloud memperlihatkan bahwa kata yang paling sering muncul adalah merah (5.303 kali), putih (5.288 kali), koperasi (4.806 kali), desa (3.715 kali), dan kopdes (1.420 kali). Dominasi istilah tersebut menegaskan bahwa perhatian publik berfokus langsung pada program Kopdes Merah Putih sebagai simbol kebangsaan sekaligus instrumen pembangunan ekonomi desa. Kehadiran kata ekonomi, masyarakat, rakyat, dan pangan menandakan adanya harapan terhadap peningkatan kesejahteraan dan ketahanan pangan, sedangkan kemunculan presiden, menteri, dan pemerintah memperlihatkan bahwa program ini dipersepsikan sebagai bagian dari agenda kebijakan nasional.

Berdasarkan keseluruhan temuan, penelitian ini merekomendasikan peningkatan transparansi, sosialisasi, dan strategi komunikasi publik agar masyarakat memperoleh pemahaman yang lebih komprehensif tentang tujuan dan manfaat Kopdes Merah Putih. Implementasi di tingkat desa juga perlu diperkuat secara merata agar program ini memberikan dampak nyata, bukan sekadar wacana. Untuk penelitian selanjutnya, disarankan penggunaan metode pelabelan hybrid atau penerapan model *deep learning* seperti LSTM dan BERT, disertai perluasan jumlah data serta rentang waktu analisis, dan pemanfaatan teknik seperti SMOTE atau *class weighting* guna mengatasi permasalahan ketidakseimbangan kelas.

DAFTAR PUSTAKA

- Aditiya, P., Enri, U., & Maulana, I. (2022). Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 1020–1028. <https://doi.org/10.30865/jurikom.v9i4.4673>
- Azmi, T. A. U., Hakim, L., Novitasari, D. C. R., & Utami, W. D. (2023). Application Random Forest Method for Sentiment Analysis in Jamsostek Mobile Review. *Telematika: Jurnal Informatika Dan Teknologi Informasi*, 20(1), 116–127. <https://doi.org/10.31515/telematika.v20i1.8868>
- Dermawan, S., & Ayunda, A. T. (2025). Sentiment Analysis of Coretax on Social Media X Using Naive Bayes, SVM, and LSTM for Service Improvement. *Journal of Applied Informatics and Computing*, 9(6), 3177–3190. <https://doi.org/10.30871/jaic.v9i6.11063>
- Ernawati, A., Khairul, Z. S., Iqbal, M., & Nasution, D. (2025). Penerapan Data Mining Untuk Klasifikasi Penduduk Miskin Di Kabupaten Labuhanbatu Menggunakan Random Forest Dan K-Nearest

- Neighbors. *Bulletin of Information Technology (BIT)*, 6(2), 23–35. <https://doi.org/10.47065/bit.v5i2.1783>
- Fauzi, Andini, B., & Muslihudin, M. (2025). SENTIMENT ANALYSIS OF TWITTER (X) USERS TOWARD THE KOPERASI DESA MERAH PUTIH PROGRAM USING SUPPORT VECTOR MACHINE AND NAIVE BAYES METHODS. *Jurnal TAM (Technology Acceptance Model)*, 16(1), 50–64.
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25(1), 114–146. <https://doi.org/10.1177/1094428120971683>
- Manullang, O., Prianto, C., & Harani, N. H. (2023). Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest. *Jurnal Ilmiah Informatika*, 11(02), 159–169.
- Ningsih, W., Alfianda, B., Rahmaddeni, R., & Wulandari, D. (2024). Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 556–562. <https://doi.org/10.57152/malcom.v4i2.1253>
- Nugraha, I. A., & Papatungan, I. V. (2025). Analisis Sentimen Video Review Starlink Indonesia Menggunakan Pendekatan Lexicon Dictionary Bahasa Indonesia dan Algoritma Random Forest. *INNOVATIVE: Journal Of Social Science Research*, 5(3), 612–626.
- Nurwanda, Suarna, N., & Prihartono, W. (2024). PENERAPAN NLP (NATURAL LANGUAGE PROCESSING) DALAM ANALISIS SENTIMEN PENGGUNA TELEGRAM DI PLAYSTORE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1841–1846.
- Putra, A. M. H., Novandya, A., & Mulyadi. (2025). Analisis Sentimen ‘Trump Tarif’ pada Platform X Menggunakan Metode Random Forest Classifier. *Jurnal Komputer, Informasi Dan Teknologi*, 5(2), 1–11.
- Putri, C. M., Afdal, M., Novita, R., & Mustakim, M. (2024). Perbandingan Evaluasi Kernel Support Vector Machine dalam Analisis Sentimen Chatbot AI pada Ulasan Google Play Store. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(3), 1236–1245. <https://doi.org/10.32493/jtsi.v7i3.41354>
- Septiani, D., & Isabela, I. (2023). Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks. *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia*, 1(2), 81–88.
- Setiawan, I., Yasin, I. F., & Desianti, Y. T. (2025). Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naïve Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa Menggunakan Student Depression Dataset. *JURNAL ILMU KOMPUTER DAN TEKNOLOGI (IKOMTI)*, 6(1), 47–58.
- Sholekhah, A., & Muntahanah. (2025). Perbandingan Naïve Bayes dan Support Vector Machine Dalam Analisa Sentimen Tentang Penyitaan Aset Koruptor di Twitter. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 981–989.