



SENTIMENT ANALYSIS OF TWITTER (X) USERS TOWARD THE KOPERASI DESA MERAH PUTIH PROGRAM USING SUPPORT VECTOR MACHINE AND NAIVE BAYES METHODS

Fauzi¹, Bertha Andini², Muhamad Muslihudin³

^{1,2,3}Prodi Sistem Informasi, Institut Bakti Nusantara, Lampung

^{1,2,3}Jalan Wisma Rini, No.09 Pringsewu, Lampung, Indonesia

E-Mail: drfauziibn@gmail.com, berthaandinii@gmail.com, mmuslihudin@ibnus.ac.id

Article history:

Received: May 20, 2025

Revised: June 8, 2025

Accepted: July 23, 2025

Corresponding authors

drfauziibn@gmail.com

Keywords:

Public Sentiment;

Social Media Analysis;

Text Classification;

Sentiment Distribution;

Short Text Analytics.

Abstract

The emergence of social media as a digital public sphere has opened significant opportunities to capture public opinion directly, including on social programs such as the *Koperasi Desa Merah Putih*. This program has attracted considerable attention on Twitter; however, public perception remains mixed, ranging from support and skepticism to criticism. This study aims to understand public sentiment through text analysis using two classification algorithms: Naive Bayes and Support Vector Machine (SVM). The results show that Naive Bayes performs better in balancing predictions, achieving an AUC of 0.71 and an F1-score of 0.62. In contrast, SVM, despite a slightly higher accuracy (0.69), was only effective in identifying the neutral class. This tendency indicates that Naive Bayes is more capable of capturing opinion variations in short and unstructured texts such as tweets. Furthermore, the sentiment distribution reveals a dominance of neutral opinions, suggesting the public's inclination to share factual information without explicit emotional expression.



This is an open access article under the CC-BY-SA license.

I. INTRODUCTION

Media sosial telah berkembang pesat dalam satu dekade terakhir sebagai platform utama dalam penyebaran informasi, komunikasi publik, hingga pembentukan opini massa [1]. Twitter, sebagai salah satu media sosial yang paling dinamis, memungkinkan penggunanya menyampaikan pendapat secara langsung dalam bentuk teks singkat atau tweet. Dengan lebih dari 400 juta pengguna aktif global, termasuk jutaan dari Indonesia, Twitter menjadi sumber data yang sangat berharga untuk menggambarkan persepsi masyarakat terhadap isu-isu aktual. Melalui analisis terhadap tweet, peneliti dapat memahami kecenderungan opini publik terhadap topik tertentu secara real-time dan berbasis data [2]. Salah satu isu yang menjadi perbincangan publik belakangan ini adalah Koperasi Desa Merah Putih, yang digagas sebagai program pemberdayaan ekonomi berbasis desa, terutama di kawasan timur Indonesia. Koperasi ini hadir sebagai respons atas ketimpangan ekonomi dan perlunya sistem alternatif yang dapat memperkuat struktur ekonomi

masyarakat lokal, khususnya di wilayah Papua. Namun, seperti banyak inisiatif baru, program ini menuai berbagai tanggapan baik dukungan maupun kritik yang ramai diperbincangkan di Twitter. Berdasarkan data tweet yang dikumpulkan pada Januari-Mei 2025, terdapat lebih dari 1000 unggahan yang membahas topik ini secara langsung, menunjukkan urgensi untuk mengetahui bagaimana masyarakat menanggapi inisiatif ini. Meski digagas sebagai program strategis untuk memperkuat ekonomi kerakyatan, Koperasi Desa Merah Putih tidak lepas dari berbagai tantangan dan persoalan di lapangan. Beberapa kritik publik yang terekam di media sosial menyoroti kurangnya transparansi pengelolaan, potensi monopoli bantuan oleh elite lokal, serta keraguan masyarakat terhadap keberlanjutan program [3]. Selain itu, penggunaan nama "Merah Putih" sebagai simbol nasionalisme juga menimbulkan perdebatan mengenai apakah koperasi ini benar-benar dibangun atas dasar kearifan lokal atau hanya alat pencitraan politik belaka. Reaksi masyarakat terhadap isu ini terekam

kelas dalam berbagai unggahan di Twitter, di mana publik tidak hanya menyampaikan dukungan, tetapi juga menyuarakan kekhawatiran dan protes yang layak untuk dianalisis secara sistematis.

Persoalan tersebut dapat di jawab, dengan menggunakan pendekatan analisis sentimen, yaitu proses mengkaji opini pengguna media sosial berdasarkan isi teks yang mereka unggah. Penelitian ini menggunakan dua algoritma utama yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes*, yang keduanya telah terbukti efektif dalam klasifikasi teks berbahasa alami. Tahapan penelitian mencakup proses scraping data dari Twitter menggunakan keyword "Koperasi Merah Putih", kemudian dilanjutkan dengan pre-processing teks seperti case folding, tokenisasi, stopword removal, stemming, serta pemberian label sentimen (positif, negatif, atau netral) terhadap data. Data kemudian diolah melalui algoritma klasifikasi dan hasilnya dievaluasi menggunakan *Confusion Matrix* untuk mengetahui F1-score dari model yang digunakan [4]. Penelitian sebelumnya mengenai analisis sentimen di media sosial telah banyak dilakukan dengan pendekatan algoritma yang beragam. Penelitian yang dilakukan [2] melakukan analisis sentimen terhadap pengguna Twitter mengenai grup musik BTS menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel linear. Penelitian ini menunjukkan bahwa skenario pembagian data latih dan data uji 90:10 memberikan akurasi terbaik sebesar 92,91%, mengungguli algoritma *Naive Bayes* (NB) yang hanya mencapai 85,93%. [4] mengkaji sentimen masyarakat terhadap kasus KSP Indosurya menggunakan algoritma *Naive Bayes*. Hasilnya menunjukkan dominasi sentimen negatif, dengan alasan metode *Naive Bayes* dipilih karena unggul dalam klasifikasi data kategori dan kemudahan penerapannya dalam data berukuran besar.

Dalam konteks penggunaan metode pembandingan, membandingkan kinerja algoritma *Naive Bayes* dan SVM dalam menganalisis sentimen ulasan pengguna aplikasi *Pluang*. Penelitian ini menyimpulkan bahwa algoritma SVM lebih unggul dibandingkan *Naive Bayes*, dengan akurasi mencapai 99,50% sedangkan *Naive Bayes* hanya 99,25%. Di sisi lain, penelitian yang mengkaji sentimen masyarakat terhadap fitur *paylater* di Twitter menggunakan *Naive Bayes Classifier* dan *TextBlob*, juga menunjukkan dominasi sentimen negatif. Pengujian model menyatakan bahwa *Naive Bayes* lebih akurat dibanding *TextBlob* dengan tingkat akurasi 91% [5]

Melalui beberapa penelitian terdahulu tersebut diharapkan dapat diperoleh gambaran yang objektif mengenai persepsi dan sentimen masyarakat terhadap trend dalam media sosial, mengenai "Koperasi Desa Merah Putih" yang sedang marak di bicarakan oleh pengguna aplikasi x. Penelitian ini bertujuan untuk mengukur dan membandingkan kinerja dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine*, dalam

mengelompokkan opini publik ke dalam kategori sentimen positif, negatif, dan netral. Hasil yang diperoleh diharapkan tidak hanya memberikan kontribusi akademik dalam bidang text mining dan analisis sentimen, tetapi juga menjadi acuan praktis bagi pemerintah, pengelola koperasi, dan pemangku kebijakan dalam mengevaluasi dan mengembangkan program berbasis masyarakat. Selain itu, penelitian ini diharapkan dapat mendorong pemanfaatan media sosial sebagai sumber data alternatif yang kaya untuk memahami respons publik secara real-time dan berbasis bukti.

II. LITERATURE REVIEW

2.1. Konsep Analisis Sentiment Data Mining

Analisis sentimen adalah proses mengidentifikasi dan mengklasifikasikan opini atau emosi dalam sebuah teks, seperti positif, negatif, atau netral. Teknik ini biasanya digunakan untuk memahami pandangan publik terhadap produk, layanan, atau isu tertentu. Data yang digunakan umumnya berasal dari media sosial, ulasan pengguna, atau komentar online. Analisis ini sangat membantu dalam pengambilan keputusan strategis berbasis opini Masyarakat[6], [7].

Data mining adalah sebuah disiplin ilmu yang berfokus pada penemuan pola, pengetahuan, atau informasi tersembunyi dari kumpulan data yang sangat besar, dengan memanfaatkan teknik statistik, matematika, dan kecerdasan buatan. Teknik-teknik utama dalam data mining antara lain klasifikasi, regresi, asosiasi, klustering, deteksi anomali, dan reduksi dimensi. Data Mining atau sering juga disebut *Knowledge Discovery in Database* (KDD) adalah sebuah bidang ilmu yang banyak membahas tentang pola sebuah data. Serangkaian proses guna mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining[8][9], [10]. Dalam penelitian ini, teknik yang di pakai yaitu klasifikasi untuk mengidentifikasi analisis sentiment pengguna aplikasi x dalam program "Koperasi Desa Merah Putih", mulai dari pengumpulan data (scrapper), pre-processing data (cleaning, case folding, normalisasi kata, tokenizing, stopword removal/filtering, dan stemming), pebelan data, splitting dataset memisahkan data training dan data testing, selanjutnya melakukan algoritma SVM dan *Naive Bayes*.

2.2. Tools Orange

Orange Data Mining adalah sebuah aplikasi open-source yang dirancang untuk analisis data dan pembelajaran mesin berbasis visualisasi interaktif. Aplikasi ini memungkinkan pengguna melakukan proses data mining tanpa perlu menulis kode program yang kompleks, karena menggunakan antarmuka drag-and-drop. Dengan pendekatan ini, Orange sangat membantu peneliti, pelajar, dan praktisi dalam mengeksplorasi data, membangun

model, dan memvisualisasikan hasil analisis secara intuitif [8].



Gambar 1. Tools Orange

Orange menyediakan berbagai widgets atau tools yang mencakup seluruh tahapan proses data mining, mulai dari input data, pra-proses, visualisasi, pemodelan, evaluasi, hingga penyimpanan hasil. Beberapa widgets penting untuk input data seperti File yang mendukung berbagai format data (CSV, Excel). Sementara untuk preprocessing tersedia tools seperti *Select Columns* untuk memilih fitur, *Impute* untuk mengisi data hilang, dan *Normalize* untuk standarisasi data. Visualisasi data dapat dilakukan dengan Scatter Plot, Box Plot, dan Heat Map yang memudahkan pemahaman pola dan hubungan antar variabel. Teknik data mining membantu dalam menemukan pengetahuan tersembunyi dalam tim data yang dapat digunakan untuk menganalisis dan memprediksi perilaku di masa depan. Klasifikasi adalah salah satu metode penambangan catatan yang menetapkan label kelas ke sekumpulan kasus yang tidak diklasifikasikan.

III. RESEARCH METHODS

3.1. Metode Pengumpulan Data

Metode pengumpulan data merupakan hal penting dalam penelitian dan merupakan strategi atau metode yang di gunakan oleh peneliti dalam mengumpulkandata yang di butuhkan dalam penelitiannya. Pada tahap ini dilakukan proses pengumpulan data menggunakan API dari twitter dimana prosesnya akan dibantu dengan library pada bahasa pemrograman python yaitu library tweepyc [2]. Total dataset yang digunakan pada aplikasi x terhadap analisis sentimen opini “Koperasi Merah Putih” pada x sebanyak 437 tweet menggunakan 1 keyword yaitu “Koperasi Merah Putih” dengan periode pengumpulan data pada 04 Juli 2025. Selanjutnya proses penyimpanan data crawling ke dalam format CSV, seluruh data crawling dari Twitter disimpan dalam format file CSV berdasarkan keyword pada tahap sebelumnya[11].

Dalam penelitian ini, teknik pengambilan sampel yang digunakan adalah teknik sampling probabilitas. Teknik ini dipilih karena setiap elemen populasi yang digunakan memiliki probabilitas yang sama untuk terpilih menjadi sampel. Sampel adalah sebagian populasi yang diambil untuk dijadikan objek penelitian. [12], [13] jika jumlah subjek

kurang dari 100, maka sebaiknya menggunakan semua subjek agar penelitian tersebut merupakan jenis penelitian populasi. Selanjutnya, jika jumlah subjek lebih dari 100, dapat diambil 10% - 15% atau 20% - 25% Ekaresta.

3.2. Metode Pengembangan Sistem

Selanjutnya penerapan algoritma yang mana pada penelitian ini digunakan 2 algoritma berbeda dalam pemodelan analisis sentimennya, yaitu Metode Support Vector Mechine dan Metode Navive Bayes :

3.2.1 Algoritma SVM

Dalam penelitian ini, pengembangan sistem analisis sentimen dilakukan dengan menerapkan dua algoritma pembelajaran mesin yang berbeda, yaitu Support Vector Machine (SVM) dan Naive Bayes. Pemilihan dua algoritma ini bertujuan untuk melakukan perbandingan performa klasifikasi dalam menentukan sentimen positif atau negatif dari suatu data teks. Masing-masing algoritma memiliki pendekatan matematis dan statistik yang berbeda, yang akan dijelaskan lebih rinci berikut ini.

Algoritma SVM bertujuan untuk membangun sebuah hyperplane optimal yang memisahkan data menjadi dua kelas dengan margin terbesar. SVM bekerja sangat baik pada data dengan dimensi tinggi dan digunakan untuk masalah klasifikasi biner seperti analisis sentimen. Fungsi klasifikasi SVM secara umum dapat dituliskan dalam bentuk:

$$f(x) = w^T x + b$$

Dimana:

1. $f(x)$ adalah fungsi klasifikasi
2. w adalah vektor bobot, misalnya = 0,7
3. X adalah vektor fitur input, misalnya $X=[0,5]$
4. b adalah bias, misalnya $b=0.2$

Contoh Perhitungan:

$$f(x) = w^T x + b = [0.7, -0.5] \cdot [4, 1] + 0.2 = 2.5$$

$$\text{sign} = +1 \text{ (positif)}$$

$$f(x) = [0.7, -0.5] \cdot [1, 5] + 0.2 = -1.6$$

$$\text{sign} = -1 \text{ (negatif)}$$

Fungsi klasifikasi dalam klasifikasi SVM diwakili oleh $f(x)$. Fungsi $f(x)$ memprediksi kelas target y berdasarkan fitur input x , dan w adalah vektor bobot, X adalah vektor masukan fitur, dan b adalah bias [14]–[17].

3.2.2 Algoritma Naive Bayes

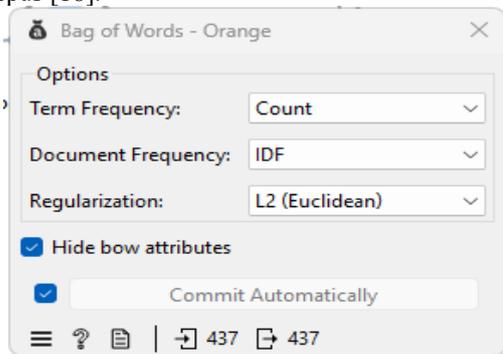
Naive Bayes merupakan metode pengklasifikasian yang paling terkenal dengan menggunakan probabilitas yang sederhana. Defenisi lain dari naive bayes adalah metode machine learning yang menggunakan proses perhitungan probabilitas dan statistika. Metode ini dikemukakan oleh ilmuwan Inggris yaitu Thomas Bayes untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Dalam memahami suatu algoritma Naive Bayes ,berikut adalah rumus

Resampling yaitu mengatasi ketidakseimbangan data dengan mengurangi data mayoritas dalam dataset dan Data Proporsional mengatasi ketidakseimbangan data dengan membagi semua data menjadi 1 proporsi yang sama.

4.1.4 Vektorisasi Teks (Text Vectorization)

Setelah data selesai dipreproses, tahap selanjutnya adalah melakukan vektorisasi teks menggunakan *widget Bag of Words* pada Orange Data Mining. Widget ini mengubah teks menjadi representasi numerik yang dapat digunakan dalam algoritma *machine learning*. Pada tahap ini, setiap kata dalam korpus teks diubah menjadi vektor berdasarkan frekuensi kemunculannya.

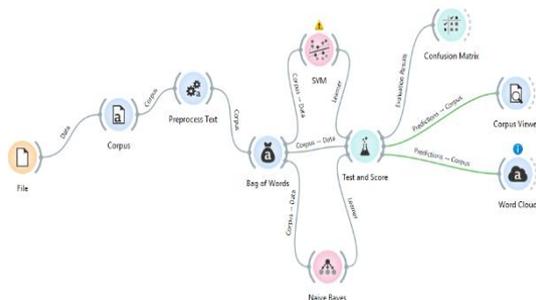
Proses ini penting karena algoritma klasifikasi tidak dapat memproses data dalam bentuk teks mentah. Dengan menggunakan Count Vectorizer atau TF-IDF Vectorizer, sistem dapat memahami bobot dari setiap kata dalam teks berdasarkan frekuensinya dalam dokumen dan dalam seluruh korpus [18].



Gambar 3. Konfigurasi Widget Bag of Words

4.1.5 Pelatihan dan Evaluasi Model Klasifikasi

Langkah selanjutnya adalah pelatihan model proses klasifikasi data dilakukan menggunakan Orange Data Mining dengan membangun alur kerja (workflow) analisis sentimen. Workflow ini terdiri dari beberapa komponen utama seperti pemrosesan teks, pemilihan fitur, pelatihan model, evaluasi model, dan visualisasi hasil klasifikasi. Gambar berikut menunjukkan alur lengkap dari proses tersebut:



Gambar 4. Workflow Analisis Sentimen

Widget yang digunakan meliputi Naive Bayes dan SVM untuk klasifikasi, Test & Score untuk evaluasi performa model, *Confusion Matrix* untuk

mengetahui akurasi klasifikasi sentimen. Model dilatih menggunakan data yang telah diproses dan divalidasi menggunakan metode Cross Validation dengan 10-fold. Evaluasi dilakukan berdasarkan akurasi, precision, recall, dan F1-score.

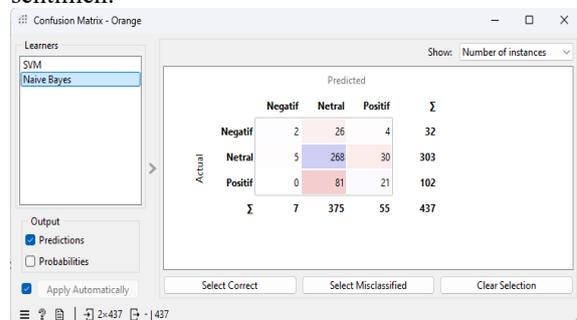
Tabel 1. Hasil Evaluasi Model

Model	AUC	CA (Accuracy)	F1	Precision	Recall
SVM	0.56	0.69	0.56	0.48	0.69
Naive Bayes	0.71	0.66	0.62	0.61	0.67

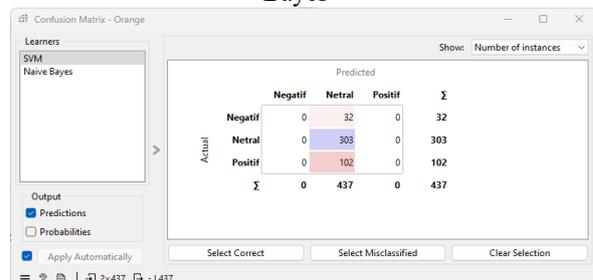
Hasil menunjukkan bahwa Naive Bayes memiliki AUC dan F1 yang lebih tinggi daripada SVM, menunjukkan kinerja keseluruhan yang lebih baik dalam membedakan kelas dan menyeimbangkan presisi/recall. SVM memiliki akurasi (CA) yang sedikit lebih tinggi daripada Naive Bayes, tetapi F1-score dan Precision-nya lebih rendah, yang bisa mengindikasikan bahwa SVM mungkin memiliki masalah dengan prediksi positif palsu (False Positives) atau positif benar (True Positives) yang kurang seimbang.

4.1.6 Visualisasi Hasil

Visualisasi dilakukan untuk menggambarkan distribusi sentimen dan efektivitas model klasifikasi. Widget Confusion Matrix dan Data Table digunakan untuk melihat klasifikasi benar dan salah, serta widget Distributions untuk melihat sebaran prediksi sentimen.



Gambar 5. Visualisasi Confusion Matrix Naive Bayes



Gambar 6. Visualisasi Confusion Matrix SVM

Hasil visualisasi menunjukkan opini terhadap “Koperasi Merah Putih” bahwa sentimen Netral merupakan kategori sentimen yang paling dominan dalam dataset, mencakup hampir setengah dari keseluruhan tweet. Diikuti oleh sentimen Positif, dan

yang terakhir adalah sentimen Negatif. Ini mengindikasikan bahwa sebagian besar pengguna cenderung menyatakan opini yang tidak condong ke arah positif atau negatif secara eksplisit, atau tweet mereka mungkin berisi informasi faktual tanpa emosi yang jelas.

4.2 Pembahasan

4.2.1 Peran TF-IDF dan Interpretasi Hasil

TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan dalam proses vektorisasi teks untuk memberikan bobot pada kata-kata yang unik dan bermakna dalam konteks dokumen, serta mengurangi pengaruh kata-kata umum yang sering muncul di seluruh dokumen. Dengan TF-IDF, kata yang jarang muncul tetapi sangat relevan terhadap topik akan memiliki bobot lebih tinggi dibandingkan kata yang sering muncul tetapi kurang bermakna (seperti “dan”, “yang”, “di”). Pendekatan ini memberikan kontribusi penting dalam meningkatkan akurasi klasifikasi, karena model dapat fokus pada kata-kata penting yang mewakili sentimen.

Pada kasus analisis sentimen terhadap Koperasi Merah Putih, TF-IDF membantu dalam membedakan ekspresi opini dari pernyataan faktual. Kata-kata seperti “bagus”, “lambat”, “terbantu” kemungkinan memiliki bobot tinggi dalam klasifikasi sentimen positif atau negatif. Sebaliknya, kata-kata netral seperti “proses”, “informasi”, atau “data” mungkin sering muncul namun dengan bobot TF-IDF rendah karena tersebar luas di banyak dokumen. Hal ini turut menjelaskan mengapa lebih banyak tweet diklasifikasikan sebagai netral banyak pengguna menggunakan kata netral tanpa ada kata bermuatan emosi yang cukup kuat untuk dimasukkan ke dalam kategori positif atau negatif.

4.2.2 Evaluasi Kinerja Model Klasifikasi

Evaluasi terhadap dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Naive Bayes*, dilakukan menggunakan metrik AUC (Area Under Curve), Accuracy (CA), Precision, Recall, dan F1-score. Berdasarkan Tabel 4, model Naive Bayes memperoleh AUC sebesar 0,71 dan F1-score sebesar 0,62, yang berarti mampu menangkap keseimbangan antara precision dan recall lebih baik dibanding SVM yang memiliki AUC 0,56 dan F1-score 0,56. Meskipun SVM sedikit lebih unggul dalam hal akurasi keseluruhan (0,69 dibandingkan 0,66), namun kinerja F1 dan precision-nya lebih rendah, yang menunjukkan bahwa prediksi SVM kurang stabil dalam mengklasifikasikan sentimen positif dan negatif secara akurat.

Kelebihan Naive Bayes dalam konteks ini dapat dikaitkan dengan kesesuaian model terhadap data teks yang bersifat sparsity tinggi, seperti tweet yang cenderung pendek dan langsung. Naive Bayes bekerja dengan asumsi independensi antar fitur, yang ternyata cukup efektif dalam menangani

representasi vektor kata seperti TF-IDF atau *Bag of Words*. Hal ini selaras dengan penelitian oleh (Nurhaliza Agustina et al., 2024), yang menyatakan bahwa model Naive Bayes seringkali lebih unggul dalam tugas klasifikasi teks pendek, terutama jika data tidak seimbang.

Secara keseluruhan, dapat disimpulkan bahwa meskipun SVM memiliki akurasi sedikit lebih tinggi, model Naive Bayes menunjukkan performa yang lebih konsisten dan seimbang dalam mengklasifikasikan ketiga kelas sentimen (positif, netral, dan negatif), sebagaimana terlihat dari nilai AUC dan F1-score yang lebih baik. Naive Bayes juga lebih adaptif terhadap karakteristik data teks media sosial yang umumnya pendek, tidak terstruktur, dan memiliki tingkat kemunculan kata yang tinggi namun tidak merata. Sementara itu, SVM tampak kurang efektif dalam menangani distribusi data yang tidak seimbang dan cenderung bias terhadap kelas dominan, seperti yang terlihat dari hasil confusion matrix yang menunjukkan bahwa semua tweet diklasifikasikan sebagai netral. Oleh karena itu, dalam konteks analisis sentimen terhadap tweet tentang koperasi, Naive Bayes lebih direkomendasikan karena mampu memberikan hasil klasifikasi yang lebih representatif dan tidak terjebak pada satu kelas dominan saja.

4.2.3 Analisis Visualisasi dan Distribusi Sentimen

Hasil visualisasi melalui *Confusion Matrix* dan *Distributions* memperlihatkan bahwa sentimen Netral mendominasi klasifikasi, diikuti oleh sentimen Positif, dan terakhir Negatif. Kondisi ini memperkuat kesimpulan bahwa sebagian besar pengguna menyampaikan opini secara netral atau menyampaikan informasi tanpa adanya ekspresi emosi yang jelas. Pola ini cukup umum dalam studi analisis sentimen media sosial, terutama ketika topik yang dibahas adalah organisasi atau lembaga formal seperti koperasi, di mana pengguna lebih sering menyampaikan laporan, pertanyaan, atau fakta, bukan opini eksplisit.

Fenomena dominasi sentimen netral dan positif ini juga tercermin dalam studi oleh [19], [20] yang menemukan bahwa opini masyarakat terhadap program pemerintah di media sosial cenderung netral karena masyarakat lebih tertarik pada informasi dibandingkan ekspresi perasaan. Selain itu, distribusi sentimen dapat dipengaruhi oleh persepsi publik terhadap koperasi yang belum tentu menimbulkan reaksi emosional kuat kecuali pada kasus tertentu. Oleh karena itu, hasil ini menunjukkan pentingnya pendekatan klasifikasi yang mampu menangkap konteks kalimat dan tidak hanya mengandalkan kata kunci tertentu.

V. DISCUSSION

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes menunjukkan performa yang lebih

stabil dibandingkan *Support Vector Machine* (SVM) dalam analisis sentimen terhadap tweet mengenai Koperasi Desa Merah Putih. Naive Bayes menghasilkan nilai AUC sebesar 0,71 dan F1-score sebesar 0,62, yang menunjukkan keseimbangan dalam klasifikasi antara presisi dan recall. Meskipun akurasi keseluruhan SVM sedikit lebih tinggi, model ini mengalami bias yang besar terhadap kelas netral dan gagal mendeteksi sentimen positif maupun negatif. Oleh karena itu, dalam konteks data teks pendek dan distribusi sentimen yang tidak seimbang, *Naive Bayes* lebih direkomendasikan sebagai metode klasifikasi yang efektif dan representatif.

Secara umum, distribusi sentimen pengguna media sosial terhadap program Koperasi Desa Merah Putih didominasi oleh sentimen netral, yang menunjukkan kecenderungan publik menyampaikan opini secara faktual dan tidak emosional. Sentimen positif juga muncul dalam jumlah signifikan, sedangkan sentimen negatif menjadi yang paling sedikit ditemukan dalam dataset. Hasil ini menggambarkan bahwa persepsi masyarakat terhadap koperasi tersebut cenderung berhati-hati dan rasional, serta dipengaruhi oleh informasi yang bersifat netral. Dengan demikian, analisis sentimen ini dapat dijadikan acuan oleh pengambil kebijakan dalam mengevaluasi persepsi publik, serta membuktikan bahwa pendekatan data mining berbasis algoritma klasifikasi efektif digunakan untuk memahami opini masyarakat secara real-time.

REFERENCE

- [1] A. Sholehah and A. S. Adinda, "Peran Media dalam Publisitas Politik Prabowo Subianto di TikTok," *-*, vol. 5, no. 1, pp. 1426–1433, 2025.
- [2] T. Safitri, Y. Umaidah, and I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 28–35, 2023.
- [3] A. Clairine, E. I. D. Lestari, E. N. Wiyono, and M. Wildan R, "Ekspresi Keresahan Pemuda melalui Media Sosial: Studi Kritik terhadap Narasi Indonesia Gelap," *Moderasi J. Stud. Ilmu Pengetah. Sos.*, vol. 6, no. 1, pp. 36–51, 2025.
- [4] O. Sembiring, E. Rulfin Tiara Kiu, and K. Nisa Meiah Ngafidin, "Analisis Sentimen Twitter Menggunakan Algoritma Naive Bayes Pada Kasus Koperasi Simpan Pinjam Indosurya," *Appl. Inf. Technol. Comput. Sci.*, vol. 2, no. 1, pp. 9–18, 2025.
- [5] A. Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023.
- [6] M. I. Maulana, E. Budianita, M. Fikry, and F. Yanto, "Klasifikasi Sentiment Ulasan Aplikasi Sausage Man Menggunakan VADER Lexicon dan Naive Bayes Classifier," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, pp. 485–492, 2023.
- [7] I. S. K. Idris, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023.
- [8] dan S. I. Rulin Swastika, Siti Mukodimah, Ferry Susanto, Muhamad Muslihudin, *Implementasi Data Mining (Clustering, Association, Prediction, Estimation, Classification)*, 1st ed. Indramayu: CV. Adanu Abimata, 2023.
- [9] K. Rismayanti, Fera Damayanti, "Penerapan Data Mining Algoritma C4.55 Dalam Menentukan Rekam Jejak Kinerja Dosen STT Harapan Medan," *J. Penelit. Tek. Inform.*, vol. 3, no. 1, pp. 99–104, 2018.
- [10] M. Muslihudin, "Analisis Prediksi Mahasiswa Tidak Tepat Waktu Menyelesaikan Studi Dengan Menggunakan Metode Algoritma C 4.5 (Studi Kasus: STMIK Pringsewu)," IBI Darmajaya, 2015.
- [11] D. F. Rahman, "Analisis ChatGPT tweet menggunakan EDA dan sentiment analysis: Data pengguna Twitter di Indonesia." 2023.
- [12] A. Suharsimi, *Prosedur Penelitian Suatu Penekatan Praktek*. Jakarta: Rineka Cipta, 1998.
- [13] S. Arikunto, *PROSEDUR PENELITIAN Suatu Pendekatan Praktik Edisi Revisi 2010*. Jakarta: Rineka Cipta, 2010.
- [14] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background," *Int. J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 605–610, 2013.
- [15] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naive Bayes Algorithm Comparison for User Sentiment Analysis on Twitter," *Sinkron*, vol. 7, no. 2, pp. 671–678, 2022.
- [16] W. E. N. Fenghua, X. Jihong, H. E. Zhifang, and G. Xu, "Stock Price Prediction Based on SSA and SVM," *Procedia - Procedia Comput. Sci.*, vol. 31, pp. 625–631, 2014.
- [17] H. Yu, R. Chen, and G. Zhang, "A SVM Stock Selection Model within PCA," *Procedia - Procedia Comput. Sci.*, vol. 31, pp. 406–412, 2014.
- [18] D. Eko, H. Pramono, and F. K. Ikhsan, "Pemodelan Enterprise Architecture Untuk Menentukan Sistem Informasi Di Perguruan Tinggi Menggunakan Metode Eap (Studi Kasus : Perguruan Tinggi Dharma Wacana Metro)," vol. XIX, no. April, pp. 434–438, 2020.
- [19] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, 2024.
- [20] T. Mustaqim, "Sentiment Analysis Opini Pelantikan Kabinet Pemerintah Indonesia Tahun 2019 Menggunakan Vader Dan Random Forest," Universitas Negeri Semarang, 2020.