

Analisis Sentimen Komentar Masyarakat terhadap Posyandu di Media Sosial Menggunakan Metode *Support Vector Machine*

Peni Sriwahyu Natasari¹, Ahmad Mustofa², Audhita Palupi Ekasari³

^{1,3} Sekolah Tinggi Ilmu Administrasi Malang

² Politeknik Negeri Madura

Email: penitasari93@gmail.com¹, mustofaahmad327@gmail.com², dhitarijanto@gmail.com³

Abstrak

Posyandu merupakan lini strategis utama di masyarakat untuk memberikan layanan kesehatan ibu dan anak. Namun, pemanfaatan layanan posyandu oleh ibu masih belum maksimal yang tercermin dari angka prevalensi stunting yang cukup tinggi serta belum meratanya cakupan imunisasi. Kurang optimalnya pemanfaatan layanan posyandu dapat dipengaruhi oleh informasi negatif yang didapat dari sosial media sebagai tempat untuk menyampaikan opini dan pengalaman terhadap posyandu. Oleh karena itu, diperlukan penelitian untuk menganalisis sentimen masyarakat terhadap posyandu di media sosial menggunakan pendekatan pembelajaran mesin, sehingga didapatkan gambaran sentimen secara umum mengenai layanan posyandu.

Penelitian ini memanfaatkan data media sosial dari *Youtube* dan *X* tentang layanan posyandu. Dataset yang didapatkan diproses melalui tahapan pra-pemrosesan dasar, anotasi manual oleh dua annotator independen, serta pengukuran kesepakatan menggunakan *Cohen's Kappa* sebagai upaya mendapatkan dataset yang reliabel. TF-IDF kemudian diaplikasikan untuk mendapatkan vektor numerik. Selanjutnya dilakukan klasifikasi oleh model SVM, dengan *Naïve Bayes*, *Logistic Regression*, dan *Random Forest* sebagai model pembanding.

Hasil penelitian menunjukkan pendekatan SVM menghasilkan performa tertinggi dengan akurasi 81%. Kelas sentimen positif dapat diklasifikasikan dengan baik oleh model, sementara kinerja pada sentimen negatif masih belum optimal. Hal ini diduga disebabkan oleh keterbatasan jumlah data sentimen negatif. Dari hasil pengujian, dapat disimpulkan bahwa analisis sentimen berbasis SVM mampu memberikan gambaran persepsi publik terhadap layanan posyandu secara efektif. Hal ini dapat dimanfaatkan sebagai acuan evaluasi layanan kesehatan masyarakat oleh pemangku kepentingan terkait.

Kata Kunci: analisis sentimen; *machine learning*; posyandu; SVM; media sosial

Abstract

Posyandu serves as a strategic frontline in the community for providing maternal and child health services. However, the utilization of posyandu services by mothers remains suboptimal, as reflected in the relatively high prevalence of stunting and the uneven coverage of immunization. The limited utilization of posyandu services may be influenced by negative information obtained from social media, which functions as a platform for expressing opinions and sharing experiences related to posyandu. Therefore, this study aims to analyze public sentiment toward posyandu on social media using a machine learning approach in order to obtain an overall representation of public sentiment regarding posyandu services.

This study utilizes social media data collected from YouTube and the X platform related to posyandu services. The dataset was processed through basic preprocessing stages, manual annotation by two independent annotators, and inter-annotator agreement measurement using Cohen's Kappa to ensure data reliability. TF-IDF was applied to generate numerical feature vectors, followed by sentiment classification using the SVM model, with Naïve Bayes, Logistic Regression, and Random Forest used as comparative models.

The results indicate that the SVM-based approach achieved the highest performance with an accuracy of 81%. The positive sentiment class was classified effectively by the model, while performance on the negative sentiment class remained less optimal, likely due to the limited number of negative sentiment samples. Overall, the findings demonstrate that SVM-based sentiment analysis can effectively capture public perceptions of posyandu services and can be used as a reference for evaluating community health services by relevant stakeholders.

Keywords: sentiment analysis; machine learning; posyandu; SVM; social media

PENDAHULUAN

Tantangan nasional dalam hal kesehatan ibu dan anak masih menjadi fokus teratas pada capaian pembangunan kesehatan nasional di Indonesia. Hal ini tertuang pada dokumen Rencana Strategis Kemenkes 2025-2029 dimana prioritas pertama terletak pada proses akselerasi penurunan angka stunting serta kesehatan ibu dan anak (Kementerian Kesehatan RI, 2025). Demi mencapai target nasional tersebut, pemerintah akan menguatkan lini posyandu dan puskesmas. Pos Pelayanan Terpadu atau yang populer dengan singkatan posyandu merupakan lapis pertama dalam menyelenggarakan kegiatan pelayanan kesehatan dasar untuk masyarakat. Layanan posyandu diharapkan dapat menaikkan tingkat kesehatan dengan memberikan kemudahan akses bagi ibu hamil, bayi, ibu menyusui, balita, dan lansia.

Pada tahun 2023 tercatat sebanyak 359.782 posyandu di Indonesia (Kementerian Kesehatan RI, 2024), namun pada praktiknya pemanfaatan layanan posyandu oleh masyarakat masih belum optimal. Hal ini tercermin dari prevalensi stunting yang masih berada pada angka 19,8%, dimana target penurunan prevalensi stunting tersebut pada tahun 2029 ditetapkan sebesar 14,4% (Kementerian Kesehatan RI, 2025). Tidak hanya itu, kasus kematian ibu serta kasus bayi berat lahir rendah (BBLR) masih terjadi di Indonesia, sehingga peran dari posyandu masih harus ditingkatkan.

Bukti lain bahwa pelayanan posyandu tidak digunakan secara maksimal oleh masyarakat tampak pada layanan imunisasi anak. Data terbaru menunjukkan sebanyak 64 dari 100 anak (64%) usia 12-23 bulan telah menerima imunisasi dasar lengkap (IDL), yang artinya 36% lainnya belum menerima IDL (Badan Pusat Statistik, 2025). Meskipun besaran cakupan meningkat, masih ada anak-anak yang belum pernah diimunisasi. Dari data BPS yang sama, diungkap tiga alasan teratas tidak dilakukannya imunisasi sebagai berikut:

- kekhawatiran terhadap efek samping vaksin (74,43%),
- kekhawatiran terhadap kandungan dalam vaksin (41,37%),
- dan keraguan terhadap efektivitas imunisasi (39,21%).

Persepsi orang tua maupun pendamping anak seperti ini dapat dipengaruhi oleh beberapa faktor dan mampu diekstrak dari informasi opini publik. Hal ini diperkuat dari studi terdahulu yang menunjukkan bahwa opini negatif di media sosial terhadap tenaga kesehatan dan layanan kesehatan dapat menurunkan tingkat kepercayaan dan kepatuhan masyarakat terhadap layanan (Paul & Headley-Johnson, 2025; Aminda et al., 2024). Tidak terkecuali sentimen terhadap posyandu, opini negatif terhadap layanan posyandu bisa menjadi salah satu pemicu ketidakhadiran ibu untuk melakukan pemantauan kesehatan anak mereka.

Dalam konteks layanan kesehatan masyarakat, opini yang beredar di media sosial dapat membentuk persepsi kolektif yang berdampak pada tingkat kepercayaan dan partisipasi masyarakat. Hal ini menjadikan analisis sentimen media sosial relevan untuk mengkaji potensi pengaruh opini publik terhadap pemanfaatan layanan posyandu.

Pemilihan media sosial sebagai sumber data untuk dianalisa didasarkan pada karakteristiknya yang terbuka dan dinamis, sehingga memungkinkan masyarakat menyampaikan opini dan pengalaman secara leluasa. Data di dunia digital dapat dihimpun sebanyak mungkin agar dapat menggambarkan sentiment publik secara akurat. Oleh karenanya, sebagai ganti opini yang diambil dengan kuesioner, penelitian ini akan mengambil data dari sosial media yang menggambarkan opini publik secara natural. Data yang diambil bersumber dari data media sosial *Youtube* dan *X* dengan topik bahasan posyandu.

Untuk mengkaji persepsi masyarakat terhadap posyandu berdasarkan data media sosial, digunakan pendekatan analisis sentimen. Analisis sentimen merupakan proses menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi seseorang terhadap sebuah produk, layanan, organisasi, individu, masalah, atau topik (Ananda & Pristyanto, 2021). Analisis terhadap respons individu menjadi penting untuk memperoleh langkah yang efektif

dan strategis dalam menangani suatu permasalahan.

Ekstraksi informasi dari berbagai sumber data teks, yang dikenal sebagai *text mining*, memiliki beberapa metode khusus untuk analisis sentimen. Pendekatan yang sering digunakan yaitu penggunaan *Support Vector Machine* (SVM), *Naïve Bayes*, *Random Forest* dan *Logistic Regression*. Pada umumnya pendekatan SVM memiliki performa yang lebih unggul dari model lainnya (Yu et al., 2010). Oleh sebab itu, penelitian ini akan memfokuskan proses analisis sentimen pada penggunaan SVM. Sedangkan model lainnya akan digunakan sebagai pembandingan.

SVM merupakan salah satu model *supervised learning* pada dunia pembelajaran mesin yang sering digunakan untuk klasifikasi. Pada konteks analisis sentimen data teks, diperlukan proses mengubah representasi teks ke bentuk numerik agar model SVM dapat melakukan *training*. Proses tersebut dapat dicapai dengan melakukan proses *Term Frequency-Inverse Document Frequency* (TF-IDF) yaitu metode pembobotan kata yang umum digunakan untuk merepresentasikan teks ke dalam bentuk vektor numerik.

Meskipun pendekatan SVM juga telah banyak diterapkan dalam analisis sentimen di sektor kesehatan, kajian yang menitikberatkan pada posyandu masih minim, khususnya yang menggunakan data media sosial yang sifatnya spontan dan masif.

Pada penelitian sebelumnya metode *Naïve Bayes* dan *Support Vector Machine* digunakan untuk menganalisis persepsi masyarakat pada layanan posyandu. Namun, penelitian tersebut masih terbatas dengan menggunakan data yang diperoleh melalui kuesioner untuk mengklasifikasikan pendapat masyarakat ke dalam sentimen positif dan negatif. Temuan riset tersebut mengindikasikan bahwa pendekatan *Naïve Bayes* menghasilkan tingkat akurasi sebesar 68,24%, sedangkan metode SVM menghasilkan tingkat akurasi yang lebih baik, sebesar 87,06% (Siahaan, 2023). Studi tersebut dilakukan hanya pada satu puskesmas. Pada penelitian ini, sumber data kuesioner yang diterapkan pada penelitian sebelumnya digantikan dengan data komentar dari media sosial yang bersifat natural, tanpa adanya intervensi atau pengarahan dari peneliti, sehingga data yang diperoleh mencerminkan opini dan persepsi masyarakat secara spontan. Analisis sentimen melalui data media sosial juga telah banyak digunakan untuk mendapatkan persepsi publik terhadap isu-isu kesehatan, seperti kesehatan mental (Kurniati, 2024), layanan kesehatan nasional (Asyharudin et al., 2022), dan ulasan aplikasi Mobile JKN (Putra, 2025). Namun riset yang khusus menganalisa sentimen terhadap posyandu menggunakan data media sosial masih sangat minim. Berdasarkan latar belakang dan pemaparan kesenjangan riset yang telah

dipaparkan maka rumusan masalah dari penelitian ini yaitu:

1. Bagaimana representasi sentimen publik terhadap layanan posyandu berdasarkan data komentar di media sosial (*YouTube* dan *X*)?
2. Bagaimana kinerja model SVM dalam mengklasifikasikan sentimen komentar terkait posyandu dibandingkan dengan metode machine learning lainnya?
3. Sejauh mana hasil analisis sentimen dapat digunakan sebagai dasar evaluasi layanan posyandu oleh pemangku kepentingan?

Sehingga dengan menjawab rumusan masalah yang ada, penelitian ini bertujuan untuk menganalisis sentimen (positif, negatif, netral) opini masyarakat terhadap posyandu di media sosial dengan pendekatan *machine learning*. Sedangkan tujuan secara khusus yaitu mengembangkan dan mengevaluasi model SVM yang dibandingkan dengan beberapa model *machine learning* lainnya, serta mengkaji kegunaan hasil analisis sentimen untuk dasar evaluasi layanan posyandu.

Penelitian diharapkan dapat memperkaya kajian analisis sentimen berbasis *machine learning* di sektor kesehatan yang dapat membantu proses evaluasi dan peningkatan layanan posyandu berdasarkan persepsi publik di media sosial.

METODE PENELITIAN

Analisis dilakukan menggunakan metode kuantitatif berbasis komputasional (*data-driven research*)

dengan memanfaatkan model *machine learning* untuk mengategorikan opini terhadap layanan posyandu ke sentimen positif, negatif, dan netral. Analisis tidak hanya akan berhenti pada distribusi sentimen namun juga dengan membandingkan beberapa model *machine learning* (*Naïve Bayes*, *Random Forest*, dan *Logistic Regression*). Proses melihat distribusi sentimen dan menganalisa beberapa model sekaligus merupakan desain deskriptif-analitik yang akan dieksekusi menjadi lima tahap.

1. Tahap 1 – Pengumpulan Data

Data yang akan digunakan untuk melatih model SVM merupakan data dari media sosial berupa komentar *Youtube* dan unggahan opini dari platform X. Data komentar *Youtube* akan diperoleh dengan memanfaatkan *Application Programming Interface* (API) resmi dari *Google* dengan kata kunci 'posyandu' untuk memperoleh video terkait pelayanan posyandu. Selain kata kunci 'posyandu', beberapa kata terkait posyandu seperti 'kader', 'imunisasi', 'gizi', 'pmt', 'stunting', 'buku pink' dan 'timbang badan' juga diambil guna mendapatkan komentar berunsur 'posyandu' dari pengguna *Youtube*. Sedangkan pada platform X, data akan diambil menggunakan API resmi dengan kata kunci yang sama yaitu 'posyandu'. Data dari dua platform yang dihimpun akan disatukan sebagai dataset untuk penelitian.

2. Tahap 2 – Pra Pemrosesan Data

Preprocessing data merupakan tahap pembersihan untuk mengatasi data yang tidak lengkap, tidak konsisten, ataupun *noise* yang cenderung mengakibatkan akurasi menurun. Pembersihan data teks dapat dilakukan dengan beberapa proses. Pada penelitian ini dataset dikenakan penghapusan *Uniform Resource Locator* (URL), penghapusan mention @, penghapusan spasi berlebih, serta proses *case folding*.

Selain itu untuk menjamin setiap data unik, dilakukan penghapusan komentar yang duplikat.

3. Tahap 3 – Anotasi Data

Dataset yang sudah terkumpul baik dari platform X dan *Youtube* akan diambil sebagian untuk dianotasi. Data yang akan diberi label diambil secara acak. Anotasi akan dilakukan manual oleh minimal dua annotator independen. Setiap opini/komentar akan dianotasi satu label: positif, negatif, atau netral. Untuk mengukur kesepakatan antar annotator akan diukur menggunakan *Cohen's Kappa*.

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Keterangan:

- P_o = proporsi kesepakatan aktual antar annotator
- P_e = proporsi kesepakatan yang terjadi secara kebetulan

Nilai kesepakatan minimal bernilai 0.6 dimana data dianggap reliabel. Jika data tidak mencapai angka tersebut,

akan ditambahkan data yang akan dilabel hingga nilai kesepakatan ≥ 0.6 . Pada data yang telah dianotasi, peneliti hanya menggunakan data yang memiliki kesesuaian label antara dua annotator. Apabila hasil evaluasi model menunjukkan tingkat akurasi yang belum optimal, maka data dengan perbedaan label akan didiskusikan kembali oleh kedua annotator hingga diperoleh kesepakatan akhir terhadap label yang diberikan.

4. Tahap 4 – Transformasi Teks ke Numerik

Sebuah model pembelajaran mesin belajar pada data numerik, tidak terkecuali model seperti SVM yang akan digunakan untuk mengklasifikasikan sentimen data teks. Data teks akan ditransformasikan ke dalam bentuk numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). *Term Frequency* akan menghitung seberapa sering kata muncul pada satu opini. Sedangkan *Inverse Document Frequency* akan memberi bobot lebih kecil pada kata yang umum digunakan, dan lebih besar pada istilah spesifik. Proses ekstraksi fitur dilakukan menggunakan *TfidfVectorizer* dengan n-gram satu hingga dua kata untuk menangkap konteks kata dan frasa. Selanjutnya, hasil vektorisasi digunakan sebagai masukan bagi model SVM dengan kernel linear untuk melakukan klasifikasi sentimen.

5. Tahap 5 – Pemodelan

Model utama pada riset ini menggunakan SVM dengan model pembandingan *Naïve Bayes*, *Random Forest*, dan *Logistic Regression*. SVM digunakan sebagai model utama karena dari studi literatur yang telah dilakukan, SVM memiliki performa yang lebih unggul dalam menangani data teks berdimensi tinggi dan data dengan jumlah fitur yang besar, seperti pada kasus analisis sentimen.

Sementara itu, model *Naïve Bayes*, *Random Forest*, dan *Logistic Regression* digunakan sebagai model pembandingan untuk mengevaluasi performa SVM secara objektif dan memastikan bahwa model yang digunakan memberikan hasil klasifikasi yang optimal pada data opini masyarakat.

6. Tahap 6 – Prediksi Sentimen pada Data Tidak Berlabel

Model terlatih dengan performa terbaik akan digunakan untuk memprediksi sentimen pada data tidak berlabel. Data tidak berlabel diproses menggunakan tahapan pra-pemrosesan yang sama dengan data latih, kemudian diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Hasil prediksi ini digunakan untuk memperoleh gambaran sentimen masyarakat secara menyeluruh serta mendukung analisis persepsi publik terhadap layanan posyandu.

7. Tahap 7 – Evaluasi Model

Data yang telah dianotasi akan dibagi menjadi data latih dan data uji dengan perbandingan 80% data *training* dan 20% data *testing*. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi performa model.

Setelah model dilatih dengan data latih, model akan melakukan prediksi menggunakan data *testing*. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kapabilitas model dalam mengategorikan sentimen. Selanjutnya *confusion matrix* juga akan diberikan untuk melihat jumlah prediksi benar maupun salah pada setiap kelas sentimen.

8. Tahap 8 –Analisa Hasil

Tahap ini merupakan bagian krusial dimana peneliti akan menganalisa hasil kinerja model klasifikasi sentimen dan melihat pola kesalahan klasifikasi di setiap kelas sentimen. Selanjutnya, performa model utama dibandingkan dengan model pembanding untuk menentukan metode yang paling optimal. Model yang memiliki performa paling unggul akan digunakan untuk mengevaluasi data yang tidak berlabel sehingga didapat persepsi publik terhadap layanan posyandu. Hal ini digunakan untuk memberikan gambaran agregat mengenai citra posyandu di mata masyarakat tanpa memerlukan survei

konvensional yang memakan biaya dan waktu.

HASIL DAN PEMBAHASAN

Hasil

Pada proses pengambilan data dari platform X dan Youtube, telah terkumpul sebanyak 6674 komentar Youtube dan 489 unggahan opini X. Ketimpangan data antara platform terjadi karena pada API X dibatasi pengambilan sebanyak 100 data tiap bulannya. Pengambilan data ini dibatasi menurut paket berlangganan pada API X, dimana peneliti hanya mampu berlangganan pada paket dasar yang tidak berbayar.

Setelah dilakukan pengecekan terhadap sejumlah komentar Youtube secara acak, terdapat komentar yang tidak relevan dengan layanan posyandu seperti opini yang berunsur dukungan terhadap tokoh politik. Data tersebut kemudian dihapus sehingga didapatkan total sebanyak 4255 data.

Selanjutnya, penyaringan tahap dua dilakukan menggunakan pendekatan *keyword-based filtering* untuk memastikan relevansi komentar dengan topik posyandu dan layanan kesehatan ibu dan anak. Kata kunci yang digunakan mencakup istilah yang berkaitan dengan posyandu, kader kesehatan serta aktivitas dan sasaran layanan posyandu, seperti imunisasi, penimbangan balita, gizi, dan stunting. Pendekatan ini digunakan untuk meminimalkan data tidak relevan (*noise*) dan meningkatkan fokus analisis terhadap konteks penelitian. Proses

penyaringan tahap dua ini menghasilkan 2256 data dengan jumlah komentar *Youtube* sebanyak 1767 data dan 489 data unggahan opini dari X. Distribusi dataset berdasarkan platform terlihat pada Tabel 1.

Tabel 1. Distribusi Data Berdasarkan Platform

Platform	Jumlah
<i>Youtube</i>	1767
X	489
Total	2256

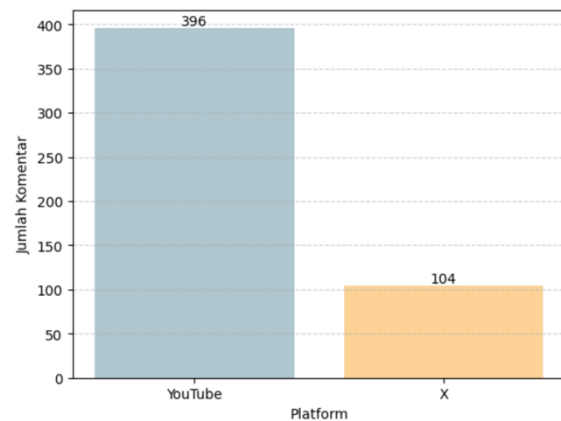
Data yang telah disaring kemudian dikenakan *preprocessing* sederhana dengan menghapus URL, mention @, spasi berlebih, serta proses *case folding*. Lihat beberapa contoh hasil pembersihan dari teks asli pada Tabel 2.

Tabel 2. Contoh Hasil dari Proses Pembersihan Teks

Teks	Hasil Pembersihan
@a*****d	posyandu
Posyandu ternyata masi eksis, tinggal dikasih upgrade biar sesuai kebutuhan masyarakat modern	ternyata masi eksis, tinggal dikasih upgrade biar sesuai kebutuhan masyarakat modern
Lah, MBG dibagiin di posyandu juga? https://t.co/2p26WQ****	lah, mbg dibagiin di posyandu juga?

Masuk pada tahap proses *labelling*. Pada proses anotasi, dataset yang telah bersih diambil sampel secara acak sebanyak 500 opini untuk dianotasi oleh dua annotator independen. Distribusi dari

500 sampel data yang dianotasi terlihat pada Gambar 1.



Gambar 1. Distribusi Data Komentar untuk Proses Pelabelan

Setelah 500 data berhasil diberi label, diperlukan pengukuran kesepakatan hasil anotasi oleh dua penilai dengan menggunakan *Cohen's kappa*. Hal ini diperlukan untuk menentukan tingkat reliabilitas dataset.

Cohen's kappa yang dilakukan pada penelitian ini menghasilkan skor sebesar 0.6289 yang dapat dikategorikan sebagai kesepakatan substansial atau baik. Sehingga dapat disimpulkan bahwa dataset yang akan digunakan untuk melatih model telah konsisten, reliabel dan layak untuk proses pembelajaran *machine learning*. Peneliti kemudian mengambil data yang memiliki kesamaan label untuk digunakan pada proses *training* dan *testing* dengan total data 383 opini.

Tahap berikutnya yaitu mengubah data teks menjadi vektor numerik dengan menerapkan TF-IDF. Hasil numerik pada penggunaan *TfidfVectorizer* digunakan sebagai masukan data proses *training* dan *testing* dengan parameter $n\text{-gram} = (1,2)$ untuk

mendapatkan istilah satu kata dan gabungan dua kata terkait sentimen. Data berlabel yang digunakan sebanyak 383 komentar, yang kemudian dibagi menjadi data latih dan data uji dengan rasio 80 banding 20. Lihat Tabel 3 untuk detail persebaran data berlabel tiap kelas sentimen. Dengan pembagian tersebut, sebanyak 306 data digunakan sebagai data latih, sedangkan 77 data digunakan sebagai data uji.

Tabel 3. Distribusi Data Berlabel

Sentimen	Jumlah Data
Positif	205
Netral	117
Negatif	61
Total	383

Proses pelatihan dan pengujian dilakukan menggunakan empat algoritma machine learning, yaitu SVM sebagai model utama serta *Naïve Bayes*, *Logistic Regression*, dan *Random Forest* sebagai model pembanding. Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Berdasarkan hasil pengujian, diperoleh performa model sebagaimana ditunjukkan pada Tabel 4.

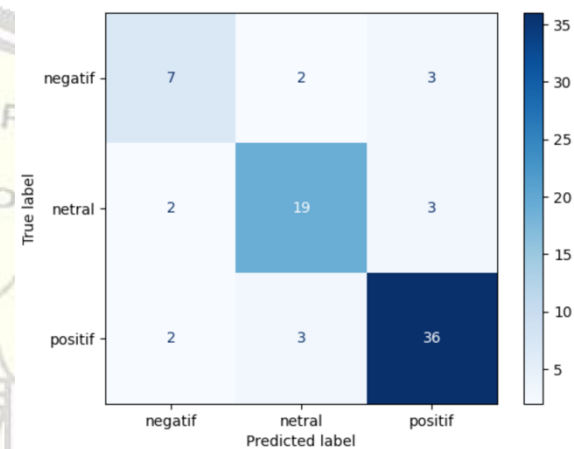
Tabel 4. Perbandingan Akurasi Model Klasifikasi Sentimen

Model	Akurasi
<i>Support Vector Machine</i> (SVM)	0,81
<i>Logistic Regression</i>	0,78
<i>Naïve Bayes</i>	0,77
<i>Random Forest</i>	0,75

Tabel 4 menunjukkan bahwa model SVM menghasilkan akurasi sebesar

0,81; diikuti oleh *Logistic Regression* sebesar 0,78; *Naïve Bayes* sebesar 0,77; dan *Random Forest* sebesar 0,75.

Untuk melihat performa klasifikasi model secara lebih rinci pada setiap kelas sentimen, dilakukan evaluasi menggunakan *confusion matrix*. *Confusion matrix* menggambarkan jumlah prediksi yang benar dan salah pada proses klasifikasi di setiap kelas sentimen yang terlihat pada Gambar 2.



Gambar 2. Confusion Matrics Model SVM
 Evaluasi performa per kelas juga dilihat untuk mengetahui sebaik apa model SVM memprediksi sentimen pada masing-masing kategori kelas yang dapat dilihat pada Tabel 5.

Tabel 5. Evaluasi Performa Model SVM per Kelas Sentimen

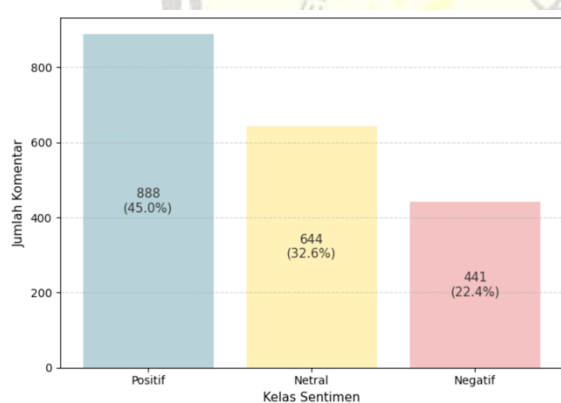
Kelas	Preci sion	Rec all	F1- sco re	Suppo rt
Negatif	0,64	0,58	0,61	12
Netral	0,79	0,79	0,79	24
Positif	0,86	0,88	0,87	41

Rata-rata (Macro)	0,76	0,75	0,76	77
--------------------------	-------------	-------------	-------------	-----------

Tahap akhir yaitu memprediksi data tidak berlabel dengan jumlah 1973 data untuk mengetahui dominasi sentimen opini masyarakat terhadap layanan posyandu. SVM model yang telah dilatih menghasilkan prediksi seperti yang tertera pada Tabel 6 dan Gambar 3.

Tabel 6. Hasil Prediksi Sentimen pada Data Opini Tidak Berlabel

Sentimen	Jumlah	Persentase
Positif	888	45,0%
Netral	644	32,6%
Negatif	441	22,4%
Total	1.973	100%



Gambar 3. Distribusi Hasil Prediksi Sentimen Komentar Tidak Berlabel

Pembahasan

Berdasarkan hasil uji, model SVM menghasilkan akurasi tertinggi sebesar 81%, diikuti oleh *Logistic Regression*, *Naïve Bayes* dan *Random Forest* dengan hasil akurasi berturut-turut sebesar 78%, 77%, dan 75%.

Hasil ini menunjukkan bahwa SVM memiliki performa yang lebih tinggi

dibandingkan model pembanding dalam mengklasifikasikan sentimen opini masyarakat terhadap posyandu di media sosial. Jika dilihat dari *confusion matrix*, model SVM memiliki performa yang baik dalam mengklasifikasikan sentimen positif dan netral, namun masih memerlukan peningkatan pada klasifikasi sentimen negatif.

Evaluasi lebih lanjut dilakukan dengan menganalisis performa model pada masing-masing kelas sentimen. Berdasarkan hasil *classification report*, model SVM menunjukkan performa terbaik pada kelas sentimen positif, dengan nilai *precision* sebesar 0,86; *recall* 0,88; dan *f1-score* 0,87. Hal ini mengindikasikan bahwa model mampu mengidentifikasi sentimen positif dengan tingkat ketepatan dan kelengkapan yang tinggi.

Pada kelas sentimen netral, model SVM juga menunjukkan performa yang stabil dengan nilai *precision* dan *recall* masing-masing sebesar 0,79. Namun, performa pada kelas sentimen negatif relatif lebih rendah, dengan nilai *f1-score* sebesar 0,61. Penurunan performa ini dipengaruhi oleh jumlah data sentimen negatif yang lebih sedikit dibandingkan kelas lainnya, sehingga model memiliki keterbatasan dalam mempelajari pola sentimen negatif secara optimal.

Hasil pengujian menunjukkan bahwa *Support Vector Machine* merupakan model dengan performa terbaik pada penelitian ini. Keunggulan SVM tidak hanya terlihat dari nilai akurasi tertinggi, tetapi juga dari keseimbangan performa antar kelas sentimen. Hal ini

sejalan dengan karakteristik SVM yang efektif dalam menangani data berdimensi tinggi dan bersifat sparse, seperti data teks hasil transformasi TF-IDF.

Penggunaan representasi fitur TF-IDF dengan n-gram (1,2) memungkinkan model untuk menangkap makna kata tunggal maupun frasa pendek yang

PENUTUP

Kesimpulan

Berdasarkan hasil proses latih dan uji model, dapat disimpulkan bahwa pendekatan *machine learning* menggunakan SVM berbasis TF-IDF mampu memberikan gambaran yang cukup akurat mengenai sentimen masyarakat terhadap posyandu di media sosial. Dominasi performa pada sentimen positif dan netral menunjukkan bahwa sebagian besar opini masyarakat bersifat mendukung atau informatif, meskipun masih terdapat sentimen negatif yang perlu diperhatikan.

Hasil ini dapat dimanfaatkan sebagai dasar evaluasi layanan posyandu, khususnya dalam mengidentifikasi isu-isu yang menimbulkan sentimen negatif di masyarakat. Dengan memahami pola sentimen publik, pemangku kebijakan dan tenaga kesehatan dapat merumuskan strategi komunikasi dan peningkatan layanan yang lebih tepat sasaran.

Saran

Penelitian selanjutnya disarankan untuk melakukan *preprocessing* teks

sering muncul dalam opini masyarakat, sehingga meningkatkan kemampuan model dalam membedakan sentimen. Kombinasi TF-IDF dan SVM terbukti mampu menghasilkan performa klasifikasi yang lebih stabil dibandingkan *Naïve Bayes* dan *Random Forest*.

yang lebih komprehensif, seperti normalisasi kata tidak baku, penghapusan emoji, serta penggunaan *stopword* khusus Bahasa Indonesia agar performa model dapat ditingkatkan, khususnya dalam mendeteksi sentimen negatif yang masih memiliki nilai *recall* relatif rendah. Selain itu saat ini model untuk analisis sentimen berkembang dengan pesat. Diperlukan model canggih dan terbaru (IndoBERT) sebagai pembanding untuk meningkatkan performa, sehingga persepsi masyarakat terhadap layanan posyandu dapat diekstrak dengan lebih akurat.

DAFTAR PUSTAKA

- Aminda, R. S., Asri, N., Damanik, M. A., Mawarti, C., Fahriza, D., Nur Hanifah, F., & Humaira, Z. (2024). Pengaruh diskriminasi harga rumah sakit Jakarta terhadap pelayanan tenaga kesehatan untuk meningkatkan kesejahteraan keluarga miskin. *Jurnal Pengabdian Masyarakat UIKA Jaya*, 2(1). <https://doi.org/10.32832/jpmuj.v2i1.2160>
- Ananda, F. D., & Pristyanto, Y. (2021). Analisis sentimen pengguna Twitter terhadap layanan internet

- provider menggunakan algoritma support vector machine. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 407–416. <https://doi.org/10.30812/matrik.v20i2.1130>
- Asyharudin, A., Kusumawati, N., Maspupah, U., Sari, D. R. F., Hamzah, A., Lukito, D., & Saputra, D. D. (2022). Comparing algorithm for sentiment analysis in healthcare and social security agency (BPJS Kesehatan). *Techno Nusa Mandiri*, 19(1), 31–37.
- Badan Pusat Statistik. (2025). *Profil statistik kesehatan 2025 (Nomor Publikasi 04200.25029; Nomor Katalog 4201005)*. Badan Pusat Statistik. Diakses dari <https://www.bps.go.id/id/publication/2025/12/12/7d17daec8d62c852fc354945/profil-statistik-kesehatan-2025.html>
- Kementerian Kesehatan Republik Indonesia. (2024). *Profil Kesehatan Masyarakat Indonesia Tahun 2023*. Direktorat Jenderal Kesehatan Masyarakat. Diakses dari https://kesprimkom.kemkes.go.id/assets/uploads/contents/other/Profil_Kesehatan_Masyarakat_Tahun_2023.pdf
- Kementerian Kesehatan Republik Indonesia. (2025). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 12 Tahun 2025 tentang Rencana Strategis Kementerian Kesehatan Tahun 2025-2029*. Diakses dari <https://www.kemkes.go.id/id/permenkes-ri-no-12-tahun-2025-tentang-rencana-strategis-kementerian-kesehatan-tahun-2025-2029>
- Kementerian Kesehatan Republik Indonesia. (2025, 26 Mei). *SSGI 2024: Prevalensi stunting nasional turun menjadi 19,8%*. Diakses dari <https://kemkes.go.id/id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198>
- Kurniati, K. (2025). Pemanfaatan tools Brand24 untuk menganalisis sentimen publik terkait isu kesehatan mental pada platform media sosial. *Jurnal Teknologi dan Sistem Tertanam*, 6(2). <https://doi.org/10.33365/jtst.v6i2.449>
- Paul, B., & Headley-Johnson, S.A. (2025). The impact of social media on health behaviors: A systematic review. *Healthcare*, 13(21), 2763. <https://doi.org/10.3390/healthcare13212763>
- Putra, D. N. A. (2025). Sentiment analysis of National Health Security mobile application review using machine learning. *Jurnal Jaminan Kesehatan Nasional*. <https://doi.org/10.53756/jjkn.v4i2.269>
- Siahaan, A. W. (2023). *Analisis sentimen informasi layanan posyandu di Puskesmas Rumbai menggunakan metode naïve Bayes dan metode support vector machine (Skripsi S1)*. Universitas Lancang Kuning, Pekanbaru. Retrieved from <http://repository.unilak.ac.id/id/eprint/4382>
- Yu, H., Huang, X., Hu, X., & Cai, H. (2010). A comparative study on data mining algorithms for

ISSN : 2722 - 9831 (Online)
ISSN : 2715 - 9817 (Printed)

individual credit risk evaluation.
*In Proceedings of the 2010
International Conference on
Management of e-Commerce and e-
Government (ICMeCG) (pp. 35-38).*

Chengdu, China: IEEE.
[https://doi.org/10.1109/ICMeC
G.2010.16](https://doi.org/10.1109/ICMeC
G.2010.16)

