

ANALISIS SENTIMEN KINERJA TENAGA MEDIS INDONESIA MENGGUNAKAN MODELING ROBERTA DAN METODE MACHINE LEARNING

Rahayun Amrullah Husaini¹, Dadang Priyanto², Galih Hendro Martono³

Ilmu Komputer, Fakultas Teknik, Universitas Bumigora Mataram, Nusa Tenggara Barat 83127, Indonesia

younk@universitasbumigora.ac.id, galih.hendro@universitasbumigora.ac.id², dadang.priyanto@universitasbumigora.ac.id³

Abstrak—Perkembangan teknologi informasi telah menyebabkan munculnya berbagai aplikasi media sosial, seperti X, yang memungkinkan pengguna untuk berbagi informasi dan opini. Namun, media sosial juga dapat berfungsi sebagai platform untuk penyebaran hoaks dan ujaran kebencian. Salah satu tantangan yang dihadapi adalah menentukan apakah komentar pengguna bersifat negatif, netral, atau positif melalui analisis sentimen. Penelitian ini bertujuan untuk membandingkan kinerja berbagai algoritma klasifikasi dalam analisis sentimen terhadap layanan tenaga medis di X, menggunakan algoritma Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), dan Naïve Bayes (NB) dengan pelabelan berbasis model RoBERTa. Data dikumpulkan melalui pustaka tweet-harvest dan diproses menggunakan bahasa pemrograman Python. Hasil penelitian menunjukkan peningkatan akurasi yang signifikan, dengan model mampu mengklasifikasikan opini publik ke dalam kategori positif, negatif, atau netral. Model SVM mencapai akurasi tertinggi sebesar 91,8%, mengungguli model lainnya seperti Random Forest dan Naïve Bayes, serta memberikan wawasan mengenai sentimen terhadap layanan kesehatan pemerintah. Temuan ini memberikan wawasan berharga bagi pembuat kebijakan dalam meningkatkan penyediaan layanan kesehatan dan mengelola persepsi publik terhadap tenaga medis.

Kata kunci—Analisis Sentimen, Tenaga Medis, Machine Learning, RoBERTa.

Abstract—The development of information technology has led to the emergence of various social media applications, such as X, which allow users to share information and opinions. However, social media can also function as a platform for the spread of hoaxes and hate speech. One of the challenges faced is determining whether user comments are negative, neutral, or positive through sentiment analysis. This study aims to compare the performance of various classification algorithms in sentiment analysis of medical personnel services in X, using the Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Naïve Bayes (NB) algorithms with RoBERTa model-based labeling. Data was collected through the tweet-harvest library and processed using the Python programming language. The results showed a significant increase in accuracy, with the model able to classify public opinion into positive, negative, or neutral categories. The SVM model achieved the highest accuracy of 91.8%, outperforming other models such as Random Forest and Naïve Bayes, and provided insight into sentiment towards government health services. These findings provide valuable insights for policymakers in improving the provision of health services and managing public perceptions of medical personnel.

Keywords—Sentiment Analysis, Medical Personnel, Machine Learning, RoBERTa.

I. PENDAHULUAN

Perkembangan teknologi informasi telah melahirkan berbagai aplikasi media sosial, seperti X. Menurut Andreas Kaplan dan Michael Heinlein dalam jurnal mereka (Azmi Fauziah Suanda & Tajul Arifin, 2024), media sosial adalah sekelompok platform perangkat lunak berbasis internet yang dibangun berdasarkan ideologi dan teknologi Web 2.0, yang memungkinkan pengguna untuk membuat dan bertukar "konten buatan pengguna" (user-generated content). Media sosial memungkinkan individu untuk berbagi informasi tanpa perlu interaksi tatap muka dan memberikan kebebasan dalam

mengekspresikan opini. Namun, pengguna media sosial juga dapat menyebarkan informasi yang merugikan, seperti tuduhan, fitnah, hoaks, serta konten yang melibatkan isu sensitif seperti suku, agama, ras, dan antar-golongan (SARA). Dalam ranah media sosial, istilah "ujaran kebencian" semakin menonjol, dipicu oleh perbedaan perspektif di antara berbagai kelompok, termasuk suku, agama, ras, dan kelas sosial (Surapati & Zulkarnain, 2023).

Untuk mengatasi permasalahan ini, pemerintah Indonesia telah menetapkan regulasi terkait kejahatan di media sosial. Dengan volume informasi dan komentar pengguna yang sangat besar, sering kali sulit untuk menentukan apakah suatu

sentimen bersifat positif atau negatif (Dinarti et al., 2024). Oleh karena itu, analisis sentimen diperlukan. Selain itu, opini publik mengenai tenaga medis di media sosial dapat berdampak signifikan terhadap layanan kesehatan. Sentimen negatif dan informasi yang salah tentang tenaga medis dapat menyebabkan ketidakpercayaan terhadap layanan kesehatan, yang berpengaruh pada kepuasan pasien dan hasil kesehatan masyarakat secara keseluruhan. Sebaliknya, sentimen positif dapat meningkatkan kepercayaan dan memperbaiki persepsi terhadap tenaga medis, yang pada akhirnya mendukung penyediaan layanan kesehatan yang lebih baik (Aminda et al., 2024).

Dalam rencana pembangunan berkelanjutan Indonesia tahun 2030, terdapat 17 tujuan yang ingin dicapai, salah satunya adalah peningkatan layanan kesehatan dan keadilan oleh pemerintah, yang menunjukkan komitmen kuat pemerintah dalam menangani masalah rendahnya kualitas layanan kesehatan (Hamzah, 2020). Menurut Salim (2020) dalam jurnal ulasan buku (Nugroho, 2021), pembangunan berkelanjutan di Indonesia harus mempertimbangkan keseimbangan antara faktor ekonomi, sosial, dan lingkungan. Tenaga medis adalah individu yang berdedikasi di bidang kesehatan dengan sikap profesional, pengetahuan, dan keterampilan yang diperoleh melalui pendidikan profesi medis, yang memberi mereka wewenang untuk melakukan upaya kesehatan. Menurut Anireon (1984) dalam jurnalnya (Siahaan, 2021), tenaga medis adalah ahli medis yang memiliki fungsi utama memberikan layanan medis berkualitas tinggi kepada pasien menggunakan metode dan teknik yang berdasarkan ilmu dan etika kedokteran yang berlaku serta dapat dipertanggungjawabkan.

Salah satu peran utama tenaga medis adalah memberikan layanan kepada pasien. Namun, meskipun telah memberikan upaya terbaik, layanan ini tetap mendapatkan berbagai persepsi dari pasien (Dinda et al., 2022). Secara umum, fasilitas kesehatan di rumah sakit memiliki visi dan misi untuk meningkatkan kualitas layanan kesehatan, yang tercantum dalam rencana tahunan mereka. Salah satu visi dan misi yang dinyatakan oleh institusi rumah sakit adalah membangun komunikasi kesehatan yang efektif guna mencegah persepsi negatif (stereotip) terhadap layanan kesehatan (Putri Citradika & Satrio, 2021). Untuk mencapai layanan kesehatan yang berkualitas tinggi, tenaga medis telah mengikuti berbagai pelatihan (workshop) guna mengembangkan keterampilan yang dibutuhkan dalam menyampaikan komunikasi kesehatan yang efektif, sehingga dapat meminimalkan persepsi negatif dari pasien terhadap layanan yang diberikan oleh tenaga medis (Singgalen, 2021).

Menurut undang-undang, tenaga medis dikategorikan ke dalam sembilan kelompok. Indonesia memiliki jumlah tenaga medis yang besar, yang menyebabkan adanya beragam persepsi di masyarakat. Secara umum, pasien di Indonesia berasal dari latar belakang sumber daya manusia (SDM) yang beragam (Adila, 2021). Pasien dengan latar belakang SDM yang lebih rendah memiliki pemahaman komunikasi yang berbeda dibandingkan dengan mereka yang memiliki latar belakang SDM lebih tinggi. Perbedaan persepsi ini juga dipengaruhi oleh

tingkat SDM yang bervariasi dalam menilai kualitas layanan kesehatan. Namun, secara umum, pasien memiliki persepsi bahwa terdapat diskriminasi di antara tenaga medis dalam memberikan layanan (Suryanto, 2020). Salah satu permasalahan utama adalah adanya ulasan mengenai tenaga medis. Meskipun mereka telah berupaya memberikan layanan kesehatan berkualitas, persepsi negatif tetap diungkapkan oleh pasien terhadap layanan yang diberikan. Selain itu, beberapa persepsi negatif muncul di media sosial, yang mengindikasikan bahwa tenaga medis cenderung melakukan diskriminasi terhadap pasien, yang secara tidak langsung mencerminkan buruknya kualitas layanan (Asril et al., 2024). Tujuan dari penelitian ini adalah melakukan analisis sentimen menggunakan data dari platform media sosial seperti X. Studi kasusnya berfokus pada layanan kesehatan yang dibahas di media sosial, dengan menggunakan algoritma klasifikasi Bert RoBERTa dan Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), dan Naïve Bayes (NB). Layanan kesehatan dipilih sebagai topik karena merupakan aspek fundamental dan krusial dalam kehidupan manusia. Kepuasan masyarakat terhadap layanan kesehatan sangat berkaitan dengan persepsi dan sikap publik, yang berkontribusi pada kualitas hidup secara keseluruhan. Dalam penelitian ini, ulasan tentang tenaga medis di media sosial diklasifikasikan ke dalam sentimen positif, negatif, dan netral (Ananda et al., 2023).

II. TINJAUAN PUSTAKA

Analisis sentimen merupakan suatu proses yang bertujuan untuk menentukan apakah isi dari suatu kumpulan data memiliki nilai positif, negatif, atau netral. Pendekatan ini, yang juga dikenal sebagai opinion mining, merupakan metode analisis yang berbasis komputasi (Krisdiyanto, 2021). Analisis sentimen penting untuk mengidentifikasi kecenderungan dari suatu pendapat, apakah itu dianggap positif, negatif, atau netral. Salah satu contoh penerapan analisis sentimen dalam kehidupan sehari-hari adalah dalam mengukur tingkat kepuasan konsumen terhadap suatu produk. Analisis sentimen terbagi menjadi tiga kategori, yaitu positif, negatif, dan netral. Polaritas positif (1) mengindikasikan sentimen yang mengarah ke persetujuan terhadap suatu hal. Polaritas negatif (-1) menunjukkan sentimen yang mengarah ke penolakan terhadap suatu hal. Sementara itu, polaritas netral (0) menunjukkan sentimen yang tidak menunjukkan sifat positif atau negatif.

Penelitian mengenai analisis sentimen telah dilakukan dengan berbagai metode, mulai dari pendekatan sederhana hingga teknik berbasis deep learning. Afni and Silmina (2022), membahas analisis sentimen tanpa menggunakan metode pembelajaran mesin tertentu, hanya memisahkan teks menjadi kalimat dan menentukan sentimen positif, negatif, atau netral. Keterbatasan ini menyebabkan akurasinya masih perlu ditingkatkan menggunakan metode yang lebih canggih.

Beberapa penelitian telah mengadopsi model berbasis deep learning, seperti RoBERTa dan LSTM.

Penelitian Putri and Ardiansyah (2023) membahas analisis sentimen terhadap kemajuan kecerdasan buatan di Indonesia menggunakan BERT dan RoBERTa. Proses penelitian

mencakup pengumpulan data, labeling dataset, analisis data eksploratif, preprocessing dataset, pemodelan, dan evaluasi menggunakan confusion matrix, precision, recall, f1-score, dan akurasi. Penelitian mendatang akan berbeda dari segi objek kasus (tenaga medis) serta proses seleksi fitur teks dan pembobotan, dengan penambahan metode SVM, Random Forest, KNN dan Naïve Bayes untuk meningkatkan akurasi klasifikasi dari model RoBERTa.

Penelitian yang dilakukan oleh (Marta Putri et al., 2023) menggunakan metode LSTM dan RoBERTa untuk menganalisis sentimen masyarakat terhadap kenaikan harga minyak goreng di X. Hasil analisis menunjukkan dominasi sentimen negatif. Dalam penelitian lanjutan, diusulkan untuk menerapkan metode SVM, Random Forest, KNN, Naïve Bayes, serta model RoBERTa untuk mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral.

Studi yang dilakukan oleh (Renaldy Indra Oetama, 2023) membahas akurasi SVM dalam analisis sentimen terhadap dataset X terkait kandidat presiden Indonesia tahun 2019. Studi ini memberikan wawasan tentang faktor-faktor yang mempengaruhi kinerja SVM. Dalam penelitian mendatang, model RoBERTa akan digunakan untuk tokenisasi dan embedding kata pada teks komentar, sementara metode SVM, Random Forest, KNN, dan Naïve Bayes akan digunakan untuk klasifikasi sentimen guna meningkatkan akurasi analisis sentimen.

Hasil penelitian dari (Tan et al., 2022) menunjukkan bahwa penggunaan word embeddings dan teknik SMOTE membantu mengatasi tantangan kelas yang tidak seimbang dalam data ulasan sosial media X tentang maskapai udara AS dan ulasan film IMDb. Penelitian mendatang akan menggunakan SVM sebagai banding untuk mengevaluasi kinerja model RoBERTa dalam klasifikasi sentimen komentar tenaga medis.

Dalam penelitian yang dilakukan oleh (Fikri et al., 2020), performa Naïve Bayes dan SVM dibandingkan dalam analisis sentimen tweet terkait Universitas Muhammadiyah Malang. Hasilnya menunjukkan bahwa Naïve Bayes lebih unggul, tetapi keterbatasannya dalam memahami konteks tetap menjadi tantangan. Oleh karena itu, model berbasis transformer seperti RoBERTa diusulkan untuk meningkatkan akurasi analisis sentimen.

Sebuah penelitian yang dilakukan oleh (Surya Sayogo et al., 2024) melibatkan proses scraping data, evaluasi model, dan visualisasi data terkait ulasan di Play Store mengenai game PUBG Mobile menggunakan algoritma Naïve Bayes. Hasil prediksi cenderung positif dengan 578 sentimen positif dan 232 sentimen negatif. Berdasarkan evaluasi menggunakan confusion matrix, penelitian ini memperoleh hasil 83.95% untuk akurasi, 88.10% untuk precision, dan 89.62% untuk recall. Penelitian mendatang akan mempertimbangkan metode SVM, Random Forest, KNN, dan Naïve Bayes dengan pelabelan RoBERTa untuk membandingkan kinerja serta meningkatkan akurasi analisis sentimen.

Temuan dari penelitian yang dilakukan oleh (Kusuma & Cahyono, 2023) menunjukkan bahwa algoritma KNN dapat mencapai akurasi 82% dalam analisis sentimen e-commerce di Indonesia. Namun, penelitian ini belum mengkombinasikan

dan membandingkan model yang digunakan dengan metode deep learning lain seperti RoBERTa, yang dapat lebih baik dalam memahami konteks teks.

Menurut penelitian yang dilakukan oleh (Pratama et al., 2023) performa Random Forest dan Naïve Bayes dibandingkan dalam analisis sentimen komentar pengguna aplikasi BBM. Hasilnya menunjukkan bahwa Random Forest memiliki akurasi lebih tinggi. Dalam penelitian lanjutan, diusulkan penerapan teknik NLP berbasis transformer untuk meningkatkan akurasi.

Penelitian yang dipublikasikan oleh (Dita Nurmadewi., 2024) membandingkan performa dua algoritma analisis sentimen, yaitu VADER (Valence Aware Dictionary and sEntiment Reasoner) dan RoBERTa (Robustly Optimized BERT Pretraining Approach), menggunakan dataset opini masyarakat terkait perubahan iklim di Twitter. Analisis dilakukan untuk menentukan distribusi sentimen dari tweet yang dianalisis, apakah bersifat positif, negatif, atau netral. Selain itu, penelitian ini mengidentifikasi kata kunci yang paling sering muncul serta melakukan analisis time series untuk melihat distribusi sentimen selama 12 bulan. Hasil penelitian menunjukkan bahwa VADER lebih efektif dalam situasi yang memerlukan tanggapan cepat terhadap perubahan sentimen publik, sementara RoBERTa lebih unggul dalam analisis mendalam pada konten yang lebih kompleks dan ambigu. Dalam penelitian lanjutan, diusulkan analisis sentimen dalam berbagai bahasa, khususnya bahasa Indonesia, untuk menguji efektivitas VADER dan RoBERTa dalam menangani perbedaan struktur bahasa dan ekspresi sentimen.

Kajian yang dilakukan oleh (Permana et al., 2023) membandingkan KNN dan Naïve Bayes dalam analisis sentimen ulasan marketplace, di mana Naïve Bayes menunjukkan performa lebih baik. Namun, penelitian ini belum menerapkan model berbasis transformer seperti RoBERTa atau BERT. Oleh karena itu, pengembangan model yang lebih kompleks diusulkan agar dapat meningkatkan akurasi analisis sentimen di berbagai domain.

Dari berbagai penelitian tersebut, terlihat bahwa RoBERTa dan model berbasis transformer lainnya semakin banyak digunakan karena kemampuannya dalam memahami konteks dan nuansa teks dengan lebih baik dibandingkan metode klasik seperti SVM, KNN, atau Naïve Bayes. Oleh karena itu, penelitian mendatang diusulkan untuk mengeksplorasi kombinasi model pembelajaran mesin dan deep learning guna meningkatkan akurasi analisis sentimen dalam berbagai domain.

III. METODE PENELITIAN

Penelitian ini melibatkan beberapa tahapan utama. Tahap pertama adalah pengumpulan data, di mana data dikumpulkan dari media sosial X menggunakan teknik *web scraping*. Setelah data terkumpul, tahap berikutnya adalah preprocessing data yang mencakup beberapa langkah penting seperti cleaning data, case folding, tokenizing, stopwords removal, dan stemming untuk memastikan data dalam format yang siap digunakan. Setelah preprocessing selesai, data kemudian dilabeli menggunakan kombinasi metode RoBERTa untuk ekstraksi

fitur dan VADER untuk analisis sentimen. Proses pelabelan ini bertujuan untuk mengklasifikasikan data ke dalam kategori sentimen tertentu. Selanjutnya, dilakukan pembuatan model klasifikasi dengan menerapkan algoritma Support Vector Machine (SVM), Random Forest (RF), *K-Nearest Neighbors* (KNN), dan *Naïve Bayes* (NB) menggunakan pelabelan RoBERTa.

Tahap akhir dari penelitian ini adalah evaluasi model. Model yang telah dibangun dievaluasi berdasarkan beberapa metrik performa utama, yaitu akurasi, presisi, recall, dan F1-score. Evaluasi ini bertujuan untuk menentukan efektivitas model dalam mengklasifikasikan sentimen dengan tingkat keakuratan yang optimal.

IV. HASIL DAN PEMBAHASAN

Proses scraping data dilakukan dengan menggunakan library *tweet-harvest*. Library ini digunakan untuk mengambil data ulasan pengguna sosial media X. Data yang digunakan dalam penelitian ini berjumlah total 6.162 data. Data dikumpulkan dalam periode Januari 2022 hingga November 2024 dengan menggunakan hashtag Kinerja Tenaga Medis, Pelayanan Kesehatan, Perawat Indonesia, Medis, Dokter Indonesia, Rumah Sakit Indonesia, Pengalaman Pasien, Kritik Pelayanan Kesehatan, dan Evaluasi Tenaga Medis. scraping data ulasan ditunjukkan pada Gambar 1.

```
# Install Node.js (Because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | gpg --dearmor | sudo tee /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/v$NODE_MAJOR/stable stable main" | sudo tee /etc/apt/sources.list.d/nodesource.list
!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
Show hidden output
# Crawl Data
filename = 'PengalamanPasien.csv'
search_keyword = 'pengalaman pasien'
limit = 2000
!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab
```

Gambar 1. Source Code Scraping Data Ulasan

Setelah data ulasan berhasil diperoleh dan disaring, langkah selanjutnya adalah menyimpan data tersebut dalam format CSV. Data ini akan digunakan sebagai dataset dalam penelitian.

Pemrosesan awal data ulasan diawali dengan menjalankan fungsi *caseandclean()* untuk melakukan normalisasi huruf dan pembersihan teks, fungsi *split()* untuk memisahkan kata-kata dalam teks, fungsi *normalisasikata()* untuk memperbaiki kesalahan ejaan, fungsi *removestopword()* untuk menghilangkan kata-kata umum yang tidak memiliki makna signifikan, serta fungsi *stemmer.stem()* untuk mengubah kata menjadi bentuk dasarnya. Kode sumber untuk pengambilan data ulasan ditampilkan pada Gambar 2.

Dataset yang telah melewati tahap preprocessing kemudian diberi label menggunakan tiga pendekatan utama, yaitu RoBERTa, *Feature Extraction*, dan VADER. Dalam pembagian data, sebanyak 20% dari total dataset akan digunakan sebagai data uji (*test set*), sementara 80% sisanya dialokasikan sebagai data latih (*training set*) untuk melatih model.

```
def casefolding(text):
    text = str(text).lower()
    text = ''.join(re.sub('([#][A-Za-z0-9_]+|([~`@=^`-?]{1})|(\w+:\w+/\w+)+|<.*?>|https?:\/\/.*[\r\n]*,*', text).split(' '))
    text = re.sub("(\d+)", "", text)
    return text

def tokenizing(text):
    text = nltk.tokenize.word_tokenize(text)
    return text

def filtering(tokens):
    filter_ = [t for t in tokens if t not in new_stopwords]
    return " ".join(filter_)

def stemming(filtered_tokens):
    stemmed = stemmer.stem(filtered_tokens)
    return stemmed

def check_positive_negative(stemming_words):
    words = stemming_words.split() # Split kata - kata dari hasil stemming
    positive = 0 # Declarasi positive
    negative = 0 # Declarasi Negative
    for word in words:
        if positive < 1 if len([n for n in positive_data.to_numpy() if n == word]) > 0 else 0: # Check apakah kata positive atau negative
            positive += 1
        if negative < 1 if len([n for n in negative_data.to_numpy() if n == word]) > 0 else 0: # Check apakah kata positive atau negative
            negative += 1
    # Jika data tidak positive ataupun negative di skip
```

Gambar 2. Source Code Preprocessing Data Ulasan

Setelah data berhasil diberi label, langkah berikutnya adalah mengklasifikasikannya menggunakan algoritma *Support Vector Machine (SVM)*, *Random Forest (RF)*, *K-Nearest Neighbors (KNN)*, dan *Naïve Bayes (NB)*. Evaluasi performa model dilakukan dengan menghitung tingkat akurasi menggunakan *confusion matrix*. Berikut merupakan pembahasan mengenai *confusion matrix* yang digunakan dalam proses evaluasi, termasuk perhitungan *true positive (TP)*, *true negative (TN)*, *false positive (FP)*, dan *false negative (FN)* untuk menilai sejauh mana model dapat mengklasifikasikan data dengan benar.

Tabel 1. Pembagian Dataset

Data review	Total
Data Traing	4928
Data Testing	1233

Tabel 1. di atas menunjukkan pembagian data ke dalam dataset pelatihan dan pengujian, yang kemudian digunakan dalam proses evaluasi.

1. Evaluasi Metode *Support Vector Machine (SVM)*

Evaluasi dilakukan dengan menggunakan *confusion matrix* untuk mengukur performa model berdasarkan metrik akurasi, presisi, *recall*, dan *F1-Measure*. Berikut merupakan *confusion matriks* *Support Vector Machine (SVM)* pada tabel 2.

Tabel 2. Tabel Confusion Matriks SVM

Actual	Predicted Negative	Predicted Neutral	Predicted Positive
Negative	20	37	3
Neutral	3	1057	12
Positive	4	42	55

Berdasarkan tabel 2. *confusion matriks* *Support Vector Machine (SVM)* di atas, sebanyak 55 ulasan positif berhasil diprediksi dengan benar, sementara 4 ulasan positif diklasifikasikan sebagai negatif, 42 ulasan positif

diklasifikasikan sebagai netral, dan 20 ulasan negatif berhasil diprediksi dengan akurat. Secara keseluruhan, terdapat 1.233 ulasan yang diklasifikasikan menggunakan algoritma SVM. Dalam penelitian ini, pengujian akurasi digunakan untuk mengevaluasi kinerja metode klasifikasi dengan menerapkan rumus persamaan untuk Akurasi, *Presisi*, *Recall*, dan *F1-Measure*. Berikut adalah hasil perhitungan berdasarkan *confusion matriks* diatas.

$$\text{Accuracy} = \frac{20 + 1057 + 55}{1233} = \frac{1132}{1233} = 0.918 \times 100\% = 91.8\%$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{55}{55 + 42} = \frac{55}{97} = 0.567 \times 100\% = 56.7\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{55}{55 + 4} = \frac{55}{59} = 0.932 \times 100\% = 93.2\%$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.567 \times 0.932}{0.567 + 0.932} = 0.705 \times 100\% = 70.5\%$$

2. Evaluasi Metode Random Forest (RF)

Random Forest (RF) adalah metode klasifikasi berbasis *ensemble learning* yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan kestabilan prediksi. Pada penelitian ini, RF digunakan untuk mengklasifikasikan data ulasan pengguna terhadap tenaga medis di Indonesia. Performa model dievaluasi berdasarkan *confusion matrix*, yang kemudian dianalisis menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-Score* untuk memahami sejauh mana model dapat mengidentifikasi sentimen dengan benar. Berikut merupakan *confusion matriks Random Forest* (RF) pada tabel 3.

Tabel 3. Tabel *Confusion Matriks Random Forest*

Actual	Predicted: Negatif	Predicted: Netral	Predicted: Positif
Negatif	2	58	0
Netral	0	1072	0
Positif	2	77	22

Berdasarkan table 3. *confusion matriks Random Forest* (RF) di atas, sebanyak 2 ulasan negatif berhasil diprediksi dengan benar, sementara 58 ulasan negatif salah diklasifikasikan sebagai netral, dan tidak ada ulasan negatif yang dikategorikan sebagai positif. Untuk ulasan netral, sebanyak 1.072 ulasan diklasifikasikan dengan benar tanpa kesalahan klasifikasi ke kategori negatif atau positif. Sementara itu, untuk ulasan positif, sebanyak 22 ulasan berhasil diprediksi dengan tepat, 2 ulasan salah diklasifikasikan sebagai negatif, dan 77 ulasan dikategorikan sebagai netral. Dari total 1.233 ulasan yang diklasifikasikan menggunakan algoritma *Random*

Forest, penelitian ini mengevaluasi kinerja klasifikasi menggunakan metrik evaluasi seperti Akurasi, *Presisi*, *Recall*, dan *F1-Score*. Berikut adalah hasil perhitungan berdasarkan *confusion matriks Random Forest*.

$$\text{Accuracy} = \frac{2 + 1072 + 22}{1233} = \frac{1096}{1233} = 0.889 \times 100\% = 88.9\%$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{22}{22 + 0} = \frac{22}{22} = 1.0 \times 100\% = 100\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{22}{22 + 77} = \frac{22}{99} = 0.222 \times 100\% = 22.2\%$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{1.0 \times 0.222}{1.0 + 0.222} = 2 \times \frac{0.222}{1.222} = 0.363 \times 100\% = 36.3\%$$

3. Evaluasi Metode K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) merupakan algoritma klasifikasi berbasis kedekatan atau jarak antar titik data. Metode ini bekerja dengan mengidentifikasi sejumlah tetangga terdekat suatu data untuk menentukan kelasnya. Pada penelitian ini, KNN diterapkan untuk mengklasifikasikan ulasan terkait tenaga medis di Indonesia. Evaluasi dilakukan dengan *confusion matrix* untuk menganalisis performa model berdasarkan akurasi, *presisi*, *recall*, dan *F1-Measure*. Berikut merupakan *confusion matriks K-Nearest Neighbors* (KNN) pada tabel 4.

Tabel 4. *Confusion Matrix KNN*

	Predicted: Negative	Predicted: Neutral	Predicted: Positive
Actual: Negative	30	2	28
Actual: Neutral	308	303	461
Actual: Positive	32	3	66

Berdasarkan tabel 4. *confusion matriks K-Nearest Neighbors* (KNN) di atas, sebanyak 30 ulasan negatif berhasil diprediksi dengan benar, sementara 2 ulasan negatif salah diklasifikasikan sebagai netral, dan 28 ulasan negatif dikategorikan sebagai positif. Selanjutnya, untuk ulasan netral, sebanyak 303 ulasan berhasil diprediksi dengan benar, namun 308 ulasan netral salah diklasifikasikan sebagai negatif, dan 461 ulasan netral dikategorikan sebagai positif. Sementara itu, untuk ulasan positif, sebanyak 66 ulasan positif berhasil diprediksi dengan benar, 32 ulasan positif salah diklasifikasikan sebagai negatif, dan 3 ulasan positif dikategorikan sebagai netral. Dari total 1.233 ulasan yang

diklasifikasikan menggunakan algoritma KNN, penelitian ini mengevaluasi kinerja klasifikasi menggunakan metrik evaluasi seperti Akurasi, *Presisi*, *Recall*, dan *F1-Score*. Berikut adalah hasil perhitungan berdasarkan *confusion matriks* KNN.

$$\text{Accuracy} = \frac{30 + 303 + 66}{1233} = \frac{1096}{1233} = 0.324 \times 100\% = 32.4\%$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{66}{66 + 32 + 461} = \frac{66}{555} = 0.119 \times 100\% = 11.9\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{66}{66 + 32 + 3} = \frac{66}{101} = 0.653 \times 100\% = 65.3\%$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.119 \times 0.653}{0.119 + 0.653} = 2 \times \frac{0.777}{0.772} = 0.200 \times 100\% = 20.0\%$$

4. Evaluasi Metode *Naïve Bayes* (NB)

Naïve Bayes (NB) adalah metode klasifikasi berbasis probabilistik yang sering digunakan dalam pemrosesan bahasa alami karena kesederhanaan dan efisiensinya dalam menangani data berskala besar. Dalam penelitian ini, *Naïve Bayes* digunakan untuk mengklasifikasikan sentimen ulasan terhadap tenaga medis di Indonesia. Evaluasi dilakukan dengan *confusion matrix* untuk mengukur kinerja model berdasarkan metrik akurasi, *presisi*, *recall*, dan *F1-Measure*. Berikut merupakan *confusion matriks* *Naïve Bayes* (NB) pada tabel 5.

Tabel 5. *Confusion Matrix Naïve Bayes*

Predicted	Negative	Neutral	Positive
Actual: Negative	27	17	16
Actual: Neutral	147	715	210
Actual: Positive	12	16	73

Berdasarkan table 5. *confusion matriks* *Naïve Bayes* (NB) di atas, terdapat 27 ulasan negatif yang berhasil diprediksi dengan benar, sementara 17 ulasan negatif diklasifikasikan sebagai ulasan netral, dan 16 ulasan negatif dikategorikan sebagai ulasan positif. Selain itu, sebanyak 715 ulasan netral berhasil diprediksi dengan benar, sedangkan 147 ulasan netral salah diklasifikasikan sebagai ulasan negatif, dan 210 ulasan netral dikategorikan sebagai ulasan positif. Untuk ulasan positif, sebanyak 73 ulasan berhasil diprediksi dengan benar, sementara 12 ulasan positif salah diklasifikasikan sebagai ulasan negatif, dan 16 ulasan positif dikategorikan sebagai ulasan netral. Penelitian ini menggunakan pengujian akurasi sebagai metode untuk mengevaluasi kinerja pendekatan

klasifikasi dengan menerapkan rumus Akurasi, *Presisi*, *Recall*, dan *F-measure*. Berikut adalah hasil perhitungan klasifikasi berdasarkan *confusion matriks* *Naïve Bayes*.

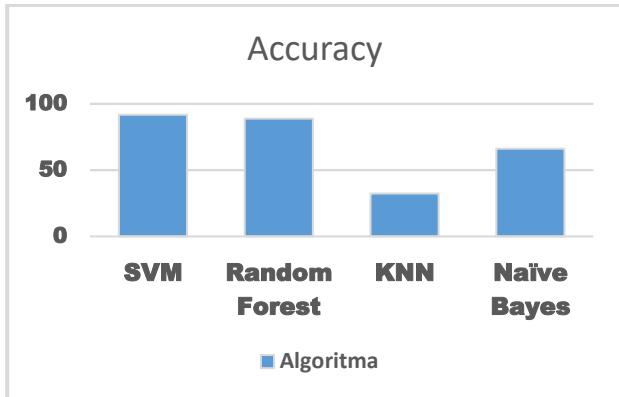
$$\text{Accuracy} = \frac{27 + 715 + 73}{1233} = \frac{815}{1233} = 0.661 \times 100\% = 66.1\%$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{75}{73 + 12} = \frac{75}{85} = 0.859 \times 100\% = 85.9\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{73}{73 + 16} = \frac{73}{89} = 0.820 \times 100\% = 82.0\%$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.859 \times 0.820}{0.859 + 0.820} = 2 \times \frac{0.704}{1.679} = 0.419 \times 100\% = 41.9\%$$

Berdasarkan hasil evaluasi, SVM menunjukkan akurasi tertinggi sebesar 91,8% dengan *recall* yang sangat baik, yaitu 93,2%, yang menunjukkan kemampuannya dalam mengklasifikasikan ulasan positif dengan benar. Namun, presisinya hanya mencapai 56,7%, yang mengindikasikan bahwa model ini masih melakukan beberapa kesalahan klasifikasi, terutama dalam memprediksi ulasan positif. Di sisi lain, *Random Forest* mencapai presisi sempurna sebesar 100%, tetapi *recall*-nya relatif rendah di angka 22,2%, yang berarti banyak ulasan positif tidak teridentifikasi dengan benar. Akurasi keseluruhannya mencapai 88,9%, tetapi nilai *F1-Measure* hanya 36,3%, menunjukkan ketidakseimbangan antara presisi dan *recall*. KNN, yang memiliki performa terburuk, hanya mencapai akurasi sebesar 32,4% dengan presisi yang sangat rendah, yaitu 11,9%, meskipun *recall*-nya cukup tinggi di angka 65,3%. Kinerja yang buruk ini menunjukkan bahwa model KNN tidak efektif untuk dataset ini. Sementara itu, *Naïve Bayes* menawarkan performa yang lebih seimbang dengan akurasi sebesar 66,1%, presisi 85,9%, dan *recall* 82,0%. Meskipun akurasinya tidak setinggi SVM, model ini memberikan keseimbangan yang baik antara presisi dan *recall*, dengan nilai *F1-Measure* sebesar 41,9%, menjadikannya pilihan yang cukup solid. Secara keseluruhan, SVM dan *Random Forest* menunjukkan kinerja terbaik, dengan SVM memiliki akurasi tertinggi, sementara *Naïve Bayes* memberikan hasil yang lebih seimbang. KNN, yang memiliki performa paling buruk, tidak direkomendasikan untuk digunakan pada dataset ini. Grafik di bawah ini menggambarkan evaluasi kinerja tenaga kesehatan di Indonesia menggunakan metode klasifikasi dengan model *Roberta*. Evaluasi ini membandingkan akurasi dari empat algoritma klasifikasi *Support Vector Machine* (SVM), *Random Forest* (RF), *K-Nearest Neighbors* (KNN), dan *Naïve Bayes* (NB) dengan pelabelan *RoBERTa*. Visualisasi dari data yang disajikan dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Hasil Perbandingan Evaluasi Algoritma

V. PENUTUP

Berdasarkan pengujian yang telah dilakukan, penelitian ini berhasil mengevaluasi kinerja tenaga medis di Indonesia menggunakan metode klasifikasi berbasis model *Roberta*. Dengan menerapkan algoritma klasifikasi seperti *Support Vector Machine* (SVM), *Random Forest* (RF), *K-Nearest Neighbors* (KNN), dan *Naïve Bayes* (NB), analisis sentimen terhadap ulasan yang dikumpulkan dari media sosial X telah memberikan wawasan mengenai pola dan kecenderungan sentimen terkait layanan kesehatan.

Hasil evaluasi menunjukkan bahwa SVM memiliki performa terbaik dengan akurasi 91,8% dan *recall* 93,2%, meskipun presisinya masih terbatas di angka 56,7%, yang mengindikasikan adanya kesalahan klasifikasi pada ulasan positif. Sebaliknya, *Random Forest* mencapai presisi tertinggi sebesar 100%, tetapi memiliki *recall* yang rendah sebesar 22,2%, sehingga banyak ulasan positif yang tidak teridentifikasi dengan baik. Sementara itu, KNN menunjukkan performa paling rendah dengan akurasi hanya 32,4%, menjadikannya metode yang kurang efektif dalam menangani dataset ini. *Naïve Bayes* memberikan keseimbangan yang lebih baik dengan akurasi 66,1%, presisi 85,9%, dan *recall* 82,0%.

Secara keseluruhan, penelitian ini membuktikan bahwa model *Roberta* dapat digunakan secara efektif dalam menganalisis sentimen terhadap kinerja tenaga medis. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi bagi pengambil kebijakan dalam meningkatkan kualitas layanan kesehatan di Indonesia.

REFERENSI

- Adila, N. T. H. (2021). The Hubungan Infeksi Saluran Pernafasan Akut dengan Kejadian Stunting. *Jurnal Ilmiah Kesehatan Sandi Husada*, 10(1), 273–279. <https://doi.org/10.35816/jiskh.v10i1.605>
- Afni, S. V. N., & Silmina, E. P. (2022). Sentiment Analysis of Public on The COVID-19 (Corona Virus Disease 2019) Vaccination Moderna and Sinovac Vaccine Using Naïve Bayes. *Conference SENATIK STT Adisutjipto Yogyakarta*, 7, 103–112. <https://doi.org/10.28989/senatik.v7i0.451>
- Aminda, R. S., Asri, N., Damanik, M. A., Mawarti, C., Fahriza, D., Hanifah, F. N., & Humaira, Z. (2024). PENGARUH DISKRIMINASI HARGA RUMAH SAKIT JAKARTA TERHADAP PELAYANAN TENAGA KESEHATAN UNTUK MENINGKATKAN KESEJAHTERAAN KELUARGA MISKIN. *Jurnal Pengabdian Masyarakat UIKA Jaya: SINKRON*, 2(1), 72. <https://doi.org/10.32832/jpmuj.v2i1.2160>
- Ananda, R., Damayanti, R., & Maharja, R. (2023). Tingkat Kepuasan Masyarakat terhadap Kinerja Pelayanan Kesehatan. *Jurnal Keperawatan Profesional (KEPO)*, 4(1), 9–17. <https://doi.org/10.36590/kepo.v4i1.570>
- Asril, A., Nikmawati, S., Ariadi, A., & Budianti, D. (2024). Pemberian Informasi & Edukasi “Bersama Kita Peduli Menghilangkan Stigma Negatif Epilepsi” Peringatan Hari Peduli Epilepsi Internasional tahun 2024. *Jurnal Pengabdian Masyarakat Kesehatan (JURABDIKES)*, 2(1), 13–17. <https://doi.org/10.56260/jurabdikes.v2i1.153>
- Azmi Fauziah Suanda, & Tajul Arifin. (2024). Signifikansi Hadis Riwayat Dalam Menangani Konten Negatif Di Media Sosial. *Khatulistiwa: Jurnal Pendidikan Dan Sosial Humaniora*, 4(2), 288–298. <https://doi.org/10.55606/khatulistiwa.v4i2.3576>
- Dinarti, N. S., Salsabila, S. R., & Herlambang, Y. T. (2024). Dilema Etika dan Moral dalam Era Digital: Pendekatan Aksiologi Teknologi terhadap Privasi Keamanan, dan Kejahatan Siber. *Daya Nasional: Jurnal Pendidikan Ilmu-Ilmu Sosial Dan Humaniora*, 2(1), 8–16. <https://doi.org/10.26418/jdn.v2i1.74931>
- Dinda, D. A. A. R., Sri Wulandari, & Aries Widjyoko. (2022). Analisis Mutu Rekam Medik Pasien Penyakit Bronkopneumonia Di Rumah Sakit PKU Muhammadiyah Surakarta Tahun 2021. *Journal Health Information Management Indonesian (JHIMI)*, 1(2), 44–52. <https://doi.org/10.46808/jhimi.v2i1.27>
- Dita Nurmadewi, Z. F. J. (2024). Perbandingan Performa Algoritma VADER dan RoBERTa pada Twitter. 13, 1547–1557.
- Fikri, M. I., Sabrina, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Hamzah, A. A. (2020). Pembangunan Sistem Informasi Rawat Jalan Berbasis Web dengan Fitur Mobile pada Puskesmas Tarok Kota Payakumbuh. 1, 7–8. <https://doi.org/https://doi.org/10.25077/TEKNOSI.v3i3.2017.353-359>
- Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32. <https://doi.org/10.24014/coreit.v7i1.12945>
- Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor. 8(3), 302–307.

- Marta Putri, D., Purbolingga, Y., Wajhi Akramunnas, B., & Rahmawati, A. (2023). Prediksi dan Analisis Sentimen Masyarakat Indonesia Terhadap Kenaikan Harga Minyak Goreng di Twitter menggunakan Metode LSTM dan RoBERTa. *Jurnal APTEK*, 15(2), 134–141.
- Nugroho, J. (2021). Ulasan Buku: 90 Tahun Prof. Emil Salim Pembangunan Berkelanjutan: Menuju Indonesia Tinggal Landas 2045. *Syntax Literate ; Jurnal Ilmiah Indonesia*, 6(2), 853. <https://doi.org/10.36418/syntax-literate.v6i2.2243>
- Permana, A. O., Saepudin, S., Studi, P., Informasi, S., Putra, U. N., Riau, U. M., Online, B., & Neighbor, K. (2023). *Jurnal Computer Science and Information Technology (CoSciTech) Comparison of the K-Nearst Neighbor and Naïve Bayes Algorithms on the Sentiment Analysis Shopee Application*. 4(1), 25–32.
- Pratama, M. L., Via, Y. V., & Mandyratha, E. P. (2023). Analisis Performansi Naive Bayes Dan Random Forest Terhadap Sentimen Kenaikan Harga BBM di Indonesia. *Scan : Jurnal Teknologi Informasi Dan Komunikasi*, 18(1). <https://doi.org/10.33005/scan.v18i1.3837>
- Putri Citradika, D., & Satrio, D. (2021). IMPLEMENTASI BALANCED SCORECARD DI RUMAH SAKIT UMUM DAERAH KAB. BATANG. *RISTEK : Jurnal Riset, Inovasi Dan Teknologi Kabupaten Batang*, 6(1), 38–46. <https://doi.org/10.55686/ristek.v6i1.108>
- Putri, N. A. R., & Ardiansyah. (2023). Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa. *Jurnal Sains Dan Informatika*, 9(2), 136–145. <https://doi.org/10.34128/jsi.v9i2.649>
- Renaldy Indra Oetama, 2023, L. (2023). ANALISIS SENTIMEN WARGANET TWITTER TERHADAP PEMILIHAN PRESIDEN INDONESIA TAHUN 2019 MENGGUNAKAN SUPPORT VECTOR MACHINE. 4(1), 88–100.
- Siahaan, R. (2021). The Effect of Work Discipline and Work Motivation on the Performance of Medical Personnel. *International Journal of Multidisciplinary Research and Analysis*, 04(03), 268–274. <https://doi.org/10.47191/ijmra/v4-i3-08>
- Singgalen, Y. A. (2021). Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review. *Journal of Information Systems and Informatics*, 3(2), 278–302. <https://doi.org/10.33557/journalisi.v3i2.125>
- Surapati, U., & Zulkarnain, A. Y. (2023). Implementasi Metode Naïve Bayes Untuk Mendeteksi Hate Speech Pada Twitter. *INTECOMS: Journal of Information Technology and Computer Science*, 6(2), 830–837. <https://doi.org/10.31539/intecom.v6i2.7678>
- Surya Sayogo, D., Irawan, B., & Bahtiar, A. (2024). Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3314–3319. <https://doi.org/10.36040/jati.v7i6.8178>
- Suryanto, H. (2020). Analisis Beban Kerja dan Kebutuhan Sumber Daya Manusia Petugas Rekam Medis Puskesmas Adan-adan Kabupaten Kediri. *Jurnal Rekam Medis Dan Informasi Kesehatan*, 3(1), 29–35. <https://doi.org/10.31983/jrmik.v3i1.5514>
- Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network. *IEEE Access*, 10, 21517–21525. <https://doi.org/10.1109/ACCESS.2022.3152828>