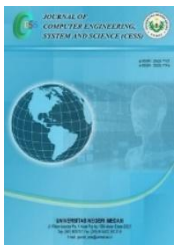
CESS (Journal of Computer Engineering, System and Science)
Vol. 10, No. 2, Juli 2025, pp.347-357

CESS
(Journal of Computer Engineering, System and Science)
Available online: https://jurnal.unimed.ac.id/2012/index.php/cess
ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Analisis Sentimen Pada Teknologi 5G Menggunakan Algoritma Random Forest dan Naïve Bayes dengan Dataset Multibahasa

Sentiment Analysis on 5G Technology Using Random Forest and Naïve Bayes Algorithms with Multilingual Datasets

Muhammad Alwi Nur Fathihah[1*], Amali[2], Annisa Maulana Majid[3]

[1,2,3]Teknik Informatika, Universitas Pelita Bangsa
Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Kab Bekasi, Jawa Barat 17530
Email: [1]alwi.16@mhs.pelitabangsa.ac.id, [2]amali@pelitabangsa.ac.id,
[3]annisa.maulanamajid@pelitabangsa.ac.id

**Corresponding Author*

A B S T R A K

Perkembangan teknologi *5G* sebagai generasi terbaru jaringan nirkabel telah menimbulkan beragam tanggapan publik, baik yang mendukung maupun yang menolak. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap teknologi *5G* berdasarkan komentar pengguna *YouTube* dalam bahasa Indonesia dan Inggris. Data diperoleh menggunakan teknik *web crawling*, kemudian diproses melalui tahapan SEMMA, yang mencakup *preprocessing*, pelabelan sentimen, dan pelatihan model. Dua algoritma yang digunakan adalah *Random Forest* dan *Naïve Bayes*. Evaluasi dilakukan menggunakan *confusion matrix* dan metrik seperti akurasi, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa *Random Forest* memiliki performa yang lebih baik dengan akurasi 94,8% dan mampu mengklasifikasikan sentimen positif dan negatif secara seimbang. Sementara itu, *Naïve Bayes* cenderung bias terhadap sentimen positif dan memiliki kelemahan dalam mendeteksi komentar negatif. Penelitian ini menunjukkan bahwa *Random Forest* lebih andal untuk analisis sentimen multibahasa, khususnya dalam konteks opini publik terhadap teknologi *5G*.

Kata Kunci: *Analisis Sentimen*; *5G*; *Random Forest*; *Naïve Bayes*; *Komentar YouTube*

A B S T R A C T

The development of 5G technology as the latest generation of wireless networks has generated a variety of public responses, both in favor and against. This study aims to analyze public sentiment towards 5G technology based on YouTube user comments in Indonesian and English. Data is obtained using web crawling techniques, then processed through SEMMA stages, which include preprocessing, sentiment labeling, and model training. The two

algorithms used are Random Forest and Naïve Bayes. Evaluation is done using confusion matrix and metrics such as accuracy, precision, recall, and F1-score. The results show that Random Forest performs better with 94.8% accuracy and is able to classify positive and negative sentiments equally. Meanwhile, Naïve Bayes tends to be biased towards positive sentiment and has a weakness in detecting negative comments. This research shows that Random Forest is more reliable for multilingual sentiment analysis, especially in the context of public opinion on 5G technology.

Keywords: *Sentiment Analysis*; *5G*; *Random Forest*; *Naïve Bayes*; *YouTube Comments*

1. PENDAHULUAN

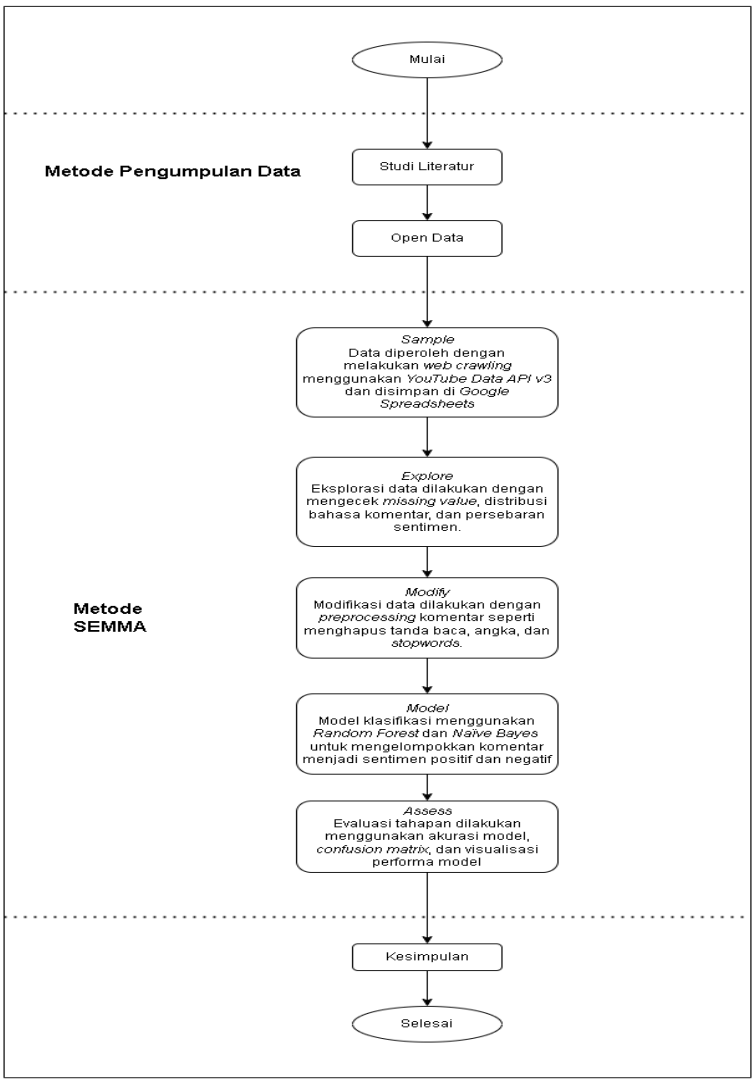
Perkembangan teknologi informasi dan komunikasi mendorong inovasi besar dalam jaringan nirkabel, salah satunya adalah teknologi *5th Generation* (*5G*). Teknologi ini menawarkan kecepatan data tinggi, latensi rendah, dan kapasitas koneksi masif yang memungkinkan penerapan *Internet of Things* (*IoT*), kendaraan otonom, serta layanan *real-time* berbasis *cloud* [1], [2]. Di Indonesia, implementasi *5G* mulai diperkenalkan secara bertahap seiring upaya pemerintah dan penyedia layanan untuk mendukung transformasi digital. Namun demikian, penerapan teknologi *5G* di berbagai negara, termasuk Indonesia, memunculkan respons yang beragam dari masyarakat. Sebagian kalangan menyambut baik kehadiran *5G* karena manfaatnya yang signifikan, sementara sebagian lainnya menyuarakan kekhawatiran, baik terkait biaya implementasi, ketersediaan infrastruktur, hingga isu kesehatan dan privasi. Berbagai tanggapan tersebut banyak disampaikan melalui media sosial dan platform digital, salah satunya *YouTube*, yang menjadi ruang bagi masyarakat untuk mengungkapkan opini mereka dalam bentuk komentar.

Fenomena perbedaan persepsi masyarakat terhadap *5G* menjadi tantangan tersendiri bagi pemerintah dan penyedia layanan. Di satu sisi, *5G* dinilai mampu mendorong transformasi digital secara menyeluruh. Namun, di sisi lain, masih terdapat kekhawatiran masyarakat terhadap dampaknya, baik dari sisi kesehatan, biaya, maupun keamanan data [3], [4]. Sayangnya, hingga kini belum tersedia analisis sistematis dan menyeluruh yang dapat memetakan opini publik terhadap *5G* secara objektif, khususnya dalam konteks multibahasa.

Beberapa penelitian sebelumnya telah menerapkan *sentiment analysis* untuk menilai opini publik terhadap isu tertentu menggunakan algoritma *machine learning.* Algoritma *Random Forest* dan *Naïve Bayes* merupakan dua metode populer yang sering digunakan untuk tugas klasifikasi teks. *Random Forest* dikenal karena stabilitas dan akurasinya dalam menangani data dengan fitur kompleks, sedangkan *Naïve Bayes* unggul dalam hal efisiensi komputasi dan kecepatan [5], [6]. Penelitian oleh Mario dan Suryono [7] menunjukkan bahwa *Random Forest* memiliki akurasi tinggi dalam mengklasifikasikan sentimen komentar *YouTube,* sedangkan penelitian oleh Marlistiana dan Widayanti [8] menunjukkan keunggulan *Naïve Bayes* pada data Twitter bertopik kebijakan publik. Namun, sebagian besar penelitian analisis sentimen di Indonesia cenderung fokus pada satu bahasa atau isu yang berbeda, meninggalkan kesenjangan dalam pemahaman sentimen publik multibahasa terhadap topik teknologi yang spesifik seperti *5G*. Data multibahasa seringkali memiliki karakteristik yang kompleks dan distribusi sentimen yang tidak seimbang, sehingga memerlukan pendekatan yang mampu mengelola variasi tersebut secara efektif.

Berdasarkan kebutuhan tersebut, penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen terhadap teknologi *5G* berdasarkan komentar pengguna di *YouTube* yang ditulis dalam bahasa Indonesia dan Inggris. Data dikumpulkan menggunakan teknik *web crawling,* kemudian dilakukan proses *text preprocessing,* pelabelan sentimen, dan evaluasi model menggunakan *confusion matrix.* Algoritma yang digunakan adalah *Random Forest* dan *Naïve Bayes*, dengan pendekatan metodologi SEMMA (*Sample, Explore, Modify, Model, Assess*) untuk proses pengolahan data [9]. Nilai kebaruan penelitian ini terletak pada penggunaan dataset komentar multibahasa dalam konteks opini publik terhadap teknologi *5G* di Indonesia, yang masih jarang dijumpai dalam literatur. Penelitian ini secara khusus mengeksplorasi bagaimana algoritma *Random Forest* dan *Naïve Bayes* menangani kompleksitas data multibahasa dan ketidakseimbangan distribusi sentimen, yang merupakan tantangan signifikan dalam analisis sentimen. Hasil penelitian diharapkan dapat memberikan kontribusi nyata bagi pengembangan sistem pendukung keputusan dan penyusunan strategi komunikasi publik berbasis opini masyarakat.

2. METODE PENELITIAN



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan tahapan penelitian mulai dari pengumpulan data hingga evaluasi model. Penelitian ini diawali dengan pengumpulan data berupa komentar *YouTube* yang membahas teknologi *5G*, baik dalam bahasa Indonesia maupun Inggris, untuk melihat persepsi masyarakat terhadap penerapan teknologi tersebut. Proses penelitian secara keseluruhan disajikan pada Gambar 1.

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk membangun model klasifikasi sentimen terhadap komentar *YouTube* yang membahas topik teknologi *5G*. Dua algoritma yang digunakan adalah *Random Forest* dan *Naïve Bayes*, dengan tahapan penelitian mengikuti metode SEMMA, mulai dari pengumpulan data, *preprocessing*, pelabelan sentimen, pelatihan model, hingga evaluasi performa klasifikasi.

2.2. Dataset

Dataset pada penelitian ini diperoleh melalui proses *web crawling* dari komentar publik pada platform *YouTube*, menggunakan *YouTube Data API v3*. Pengambilan data dilakukan terhadap video-video yang secara khusus membahas topik teknologi *5G*, baik dalam konteks global maupun lokal, selama periode tahun 2022 hingga 2024. Data terdiri dari dua bahasa, yaitu Bahasa Indonesia dan Bahasa Inggris, dengan total 37.404 komentar. Setiap komentar hanya berisi satu kolom teks (*textDisplay*) dan pelabelan sentimen dilakukan pada tahap *preprocessing*.

Tabel 1. Komentar dalam Dataset

No	Komentar (Bahasa Inggris)	Komentar (Bahasa Indonesia)
1	4g/4g+ is more than enough. But damn 5g better	tujuannya biar sinyal bagus kenapa gk sinyalnya aja yg di botulin
2	5G wont kill you,it will make karens angry.	Sampek sekarang lampung blm ada 5g
3	5g is dangerous! Don't listen to this guy, He obviously hasn't studied!	jd kesimpulan saya menigan 4G deh harga juga lebih murah, kalau 5G lebih mahal sinyal juga belum merata

2.3. Metode SEMMA

Metode yang digunakan dalam penelitian ini adalah metode SEMMA (*Sample, Explore, Modify, Model, Assess*), sebuah metodologi yang dikembangkan oleh *SAS Institute.* Metode ini dirancang untuk mengeksplorasi, memilih, dan mentransformasi variabel prediktif yang relevan, membangun model prediksi, serta mengevaluasi tingkat akurasi model yang dihasilkan. Proses ini terdiri dari lima tahap, yaitu: *Sample* (mengumpulkan data dan informasi), *Explore* (mengeksplorasi data untuk menemukan pola dan hubungan), *Modify* (memodifikasi dan mengelompokkan variabel), *Model* (membangun model prediktif), dan *Assess* (mengevaluasi performa model) [10]. Berikut merupakan tahapan metode SEMMA dalam pada penelitian ini:

1) *Sample*
   Dataset diperoleh dengan melakukan *web crawling* terhadap komentar-komentar pada video *Youtube* yang membahas topik teknologi *5G.* Proses pengambilan data dilakukan menggunakan *YouTube Data API v3* serta Pustaka pendukung seperti *google-api-python-*

client dan *youtube-comment-downloader*. Komentar yang berhasil dikumpulkan berjumlah 37.404, yang terdiri dari 25.748 komentar berbahasa Inggris dan 11.656 komentar berbahasa Indonesia. Data kemudian disimpan dalam format *CSV* dan diimpor ke *Google Colab* untuk proses selanjutnya.

2) *Explore*
Pada tahap ini eksplorasi dilakukan dengan menampilkan struktur dataset untuk mengetahui jumlah entri, jenis kolom, serta informasi awal mengenai tipe data dan kelengkapan isinya. Proses ini bertujuan untuk memastikan data dapat dibaca dengan benar oleh sistem dan tidak mengandung nilai kosong pada kolom utama.

3) *Modify*
Pada tahap ini dilakukan *preprocessing* data teks untuk mempersiapkan data sebelum pelatihan model. Proses ini meliputi:
- *Cleaning*: menghapus *URL*, karakter non-alfabet, angka, simbol, dan elemen tidak relevan dengan ekspresi reguler.
- *Case Folding*: mengubah seluruh huruf menjadi huruf kecil.
- *Stopword Removal*: menghapus kata-kata umum yang tidak penting. Bahasa Inggris menggunakan pustaka *nltk,* dan bahasa Indonesia menggunakan pustaka *Sastrawi*.
- *Lemmatization*: mengembalikan kata ke bentuk dasarnya. Bahasa Inggris menggunakan *WordNetLemmatizer*, sedangkan komentar berbahasa Indonesia dinormalisasi menggunakan pendekatan leksikal dari *Sastrawi*.

Setelah teks dibersihkan, proses pelabelan sentimen dilakukan. Untuk komentar berbahasa Inggris digunakan metode *lexicon-based* dari pustaka *TextBlob* dengan nilai polaritas. Untuk komentar berbahasa Indonesia, pelabelan dilakukan secara semi-otomatis dengan bantuan leksikon sentimen yang dikembangkan dari korpus umum Bahasa Indonesia, kemudian diverifikasi manual oleh dua annotator. Jika terjadi perbedaan pelabelan antara annotator, diskusi dilakukan untuk mencapai konsensus. Seluruh komentar diklasifikasikan menjadi dua sentimen, yaitu positif dan negatif. Kategori sentimen ditentukan sebagai berikut:
- Positif jika skor polaritas ≥ 0.
- Negatif jika skor polaritas < 0.

4) *Model*
Pada tahap ini membangun model klasifikasi menggunakan dua algoritma *supervised learning,* yaitu *Random Forest* dan *Naïve Bayes.* Kedua algoritma ini dipilih karena keunggulannya dalam melakukan klasifikasi teks dan telah terbukti efektif dalam berbagai penelitian analisis sentimen. Data komentar yang telah dibersihkan dan diberi label sentimen digunakan sebagai masukan ke dalam proses pelatihan model. Kedua model dibangun menggunakan pustaka *scikit-learn* di lingkungan *Google Colab.* Setelah proses pelatihan selesai, model digunakan untuk memprediksi sentimen dari komentar yang dianalisis. Hasil klasifikasi dari masing-masing algoritma kemudian dievaluasi untuk mengetahui tingkat akurasi dan kualitas prediksi terhadap sentimen positif dan negatif [11].

5) *Assess*
Evaluasi model dilakukan menggunakan *Confusion Matrix,* serta metrik evaluasi seperti akurasi, *precision, recall, dan F1-score.* Proses evaluasi bertujuan untuk membandingkan performa kedua algoritma dalam mengklasifikasikan komentar ke dalam kategori

sentimen positif dan negatif. Selain itu, ditampilkan pula visualisasi *confusion matrix* dan metrik evaluasi untuk memberikan gambaran komparatif performa model.

2.4. Random Forest

Random Forest adalah algoritma pembelajaran mesin berbasis *ensemble* yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini menggabungkan sejumlah pohon keputusan (*decision tree*) yang dibangun secara acak untuk menghasilkan prediksi akhir yang lebih akurat dan stabil [12], [13]. Pada penelitian ini, *Random Forest* digunakan untuk mengklasifikasikan komentar *YouTube* menjadi sentimen positif dan negatif, dengan memanfaatkan pustaka *scikit-learn* pada platform *Google Colaboratory*.

2.5. Naïve Bayes

Naïve Bayes merupakan algoritma pembelajaran mesin yang mengandalkan perhitungan probabilitas untuk melakukan klasifikasi, terutama dalam konteks bahasa alami (*Natural Language Processing*). Algoritma ini menerapkan *Teorema Bayes* dengan asumsi independensi antar fitur, sehingga menghitung kemungkinan suatu data termasuk dalam kelas tertentu berdasarkan distribusi kata dalam teks [14], [15]. Pada penelitian ini, *Naïve Bayes* digunakan untuk mengklasifikasikan komentar *YouTube* ke dalam dua kategori sentimen, yaitu positif dan negatif. Implementasi dilakukan menggunakan pustaka *scikit-learn* pada platform *Google Colaboratory*. Secara umum, metode ini dapat dirumuskan melalui persamaan berikut:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)}$$

Keterangan:

X = Sampel data yang memiliki *class* (label) yang tidak diketahui.
C = Hipotesis bahwa X adalah data *class* (label).
P(C) = Probabilitas hipotesis C.
P(X) = Peluang dari data sampel yang diamati (probabilitas C).
P(X|C) = Probabilitas berdasarkan kondisi pada hipotesis.

3. HASIL DAN PEMBAHASAN

3.1. Data Awal dan Seleksi

Data komentar *YouTube* mengenai teknologi *5G* yang digunakan dalam penelitian ini dikumpulkan menggunakan *YouTube Data API v3* dengan bantuan pustaka pendukung, seperti *google-api-python-client* dan *youtube-comment-downloader*. Total data yang diperoleh adalah 37.404 komentar, yang terdiri dari 25.748 komentar berbahasa Inggris dan 11.656 komentar berbahasa Indonesia, sesuai dengan periode pengambilan data pada tahun 2022 hingga 2024.

Data mentah masih mengandung *noise* berupa *URL,* angka, symbol, serta kata-kata umum yang tidak relevan untuk analisis sentimen. Oleh karena itu, dilakukan proses *preprocessing* yang meliputi tahapan *cleaning, case folding, stopword removal,* dan *lemmatization.*

Contoh hasil setiap tahap *preprocessing* ditampilkan pada Tabel 2.

Tabel 2. Hasil *Preprocessing*

No	Komentar Asli	Setelah *Cleaning*	Setelah *Case Folding*	Setelah *Stopword Removal*	Setelah *Lemmatization*
1	4g/4g+ is more than enough. But damn 5g better	4g4g is more than enough But damn g better	4g4g is more than enough but damn g better	4g4g enough damn g better	4g4g enough damn g better
2	tujuannya biar sinyal bagus kenapa gk sinyalnya aja yg di botulin	tujuannya biar sinyal bagus kenapa gk sinyalnya aja yg di botulin	tujuannya biar sinyal bagus kenapa gk sinyalnya aja yg di botulin	tuju sinyal bagus kenapa gk sinyal aja botul	tuju sinyal bagus kenapa gk sinyal aja botul

3.2. Pelabelan Sentimen

Setelah tahap *preprocessing,* Komentar kemudian dilabeli sentimen berdasarkan skor polaritas:

- Bahasa Inggris menggunakan pustaka *TextBlob.*
- Bahasa Indonesia dilakukan secara semi-otomatis, dengan verifikasi manual.

Kategori sentimen ditentukan sebagai berikut:

- Positif jika skor polaritas ≥ 0
- Negatif jika skor polaritas < 0

Tabel 3. Hasil *Preprocessing* dan Penentuan Kategori Sentimen

No	Hasil *Preprocessing*	*Sentiment score*	Kategori
1	4g4g enough damn g better	0,350000	Positif
2	tuju sinyal bagus kenapa gk sinyal aja botul	0,0	Positif

Tabel 4. Distribusi Kategori Sentimen pada Dataset Multibahasa

Bahasa	Sentimen Positif	Persentase Positif	Sentimen Negatif	Persentase Negatif	Total Komentar
Inggris	21.224	82.4%	4.524	17.6%	25.748
Indonesia	11.363	97.5%	293	2.5%	11.656
Gabungan	32.587	87.1%	4.817	12.9%	37.404

Tabel 4 menunjukkan dominasi komentar positif dalam dataset, terutama pada data berbahasa Indonesia. Distribusi sentimen yang tidak seimbang ini dapat mempengaruhi kinerja model dalam mengklasifikasikan komentar negatif.

3.3. Pelatihan dan Evaluasi Model

Setelah pelabelan sentimen, data dibagi menjadi data latih (80%) dan data uji (20%) menggunakan *train_test_split*. Komentar yang telah melalui *preprocessing* dikonversi ke bentuk numerik menggunakan *CountVectorizer.* Dua model klasifikasi yang digunakan adalah *Random Forest* dan *Naïve Bayes*. Model *RandomForest* dilatih menggunakan *RandomForestClassifier* dari pustaka *scikit-learn*, sementara Naïve *Bayes* menggunakan *MultinomialNB.* Setelah pelatihan, kedua model dievaluasi menggunakan data uji dengan metrik evaluasi berupa *confusion matrix,* akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi ini digunakan untuk mengukur sejauh mana akurasi model dalam mengklasifikasikan komentar ke dalam dua kategori sentimen, yaitu positif dan negatif.

3.4. Hasil Evaluasi

Evaluasi dilakukan untuk mengukur kinerja algoritma dalam mengklasifikasikan komentar ke dalam kategori positif dan negatif. Dua model uji, yaitu *Random Forest* dan *Naïve Bayes*. Evaluasi dilakukan menggunakan fungsi *classification_report* dan *accuracy_score* dari Pustaka *scikit-learn*, dengan data dibagi menggunakan *train_test_split* rasio 80:20. Komentar yang telah melalui proses *preprocessing* dan pelabelan sentimen dikonversi ke bentuk numerik menggunakan *CountVectorizer* sebelum pelatihan model.

Tabel 5. Hasil Evaluasi Algoritma *Random Forest*

Kategori	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	83%	75%	78%	950
Positif	96%	98%	97%	6.531
Accuracy			94,8%	7.481
Macro avg	90%	86%	88%	7.481
Weighted avg	95%	95%	95%	7.481

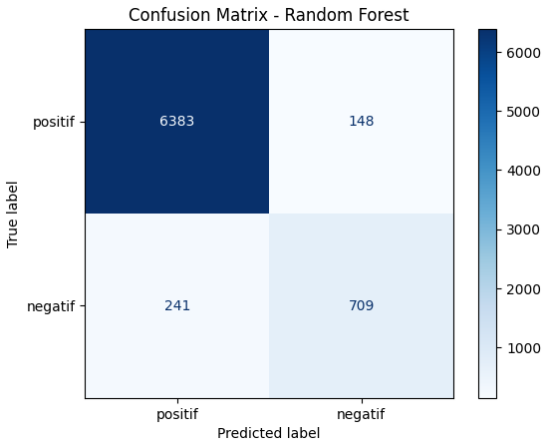
Tabel 6. Hasil Evaluasi Algoritma *Naïve Bayes*

Kategori	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	86%	27%	41%	950
Positif	90%	99%	95%	6.531
Accuracy			90,2%	7.481
Macro avg	88%	63%	68%	7.481
Weighted avg	90%	90%	88%	7.481

Tabel 5 dan Tabel 6 menyajikan metrik evaluasi algoritma *Random Forest* dan *Naïve Bayes* yang mencakup presisi, *recall*, *F1-score*, dan akurasi. Seluruh nilai dihitung secara otomatis menggunakan fungsi *classification_report* dari pustakan *scikit-learn*, berdasarkan hasil klasifikasi terhadap data uji. Dengan demikian, baik nilai maupun proses akurasi telah disampaikan secara eksplisit dalam hasil penelitian ini.

Confusion Matrix untuk masing-masing model adalah:

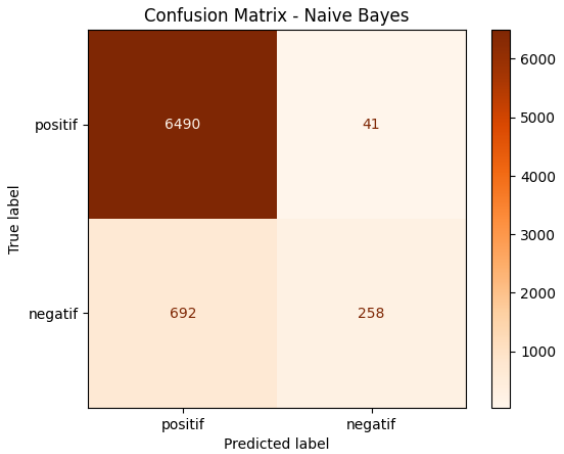
a) *Random Forest*



Gambar 2. *Confusion Matrix Random Forest*

Confusion matrix pada Gambar 2 menunjukkan bahwa algoritma *Random Forest* mampu mengklasifikasikan komentar positif dan negatif secara seimbang. Dari 915 komentar negatif, sebanyak 709 diklasifikasikan dengan benar, sedangkan 206 salah diklasifikasikan sebagai positif. Sementara itu, dari 6.566 komentar positif, sebanyak 6.383 berhasil dikenali dengan tepat, dan hanya 183 yang salah. Hasil ini mencerminkan akurasi tinggi dan kemampuan *Random Forest* dalam menangani data multibahasa dengan distribusi sentimen yang tidak seimbang.

b) *Naïve Bayes*



Gambar 3. *Confusion Matrix Naïve Bayes*

Confusion matrix pada Gambar 3 menunjukkan bahwa algoritma *Naïve Bayes* cenderung bias terhadap sentimen positif. Dari 915 komentar negatif, hanya 267 yang berhasil

diklasifikasikan dengan benar, sementara 648 komentar lainnya salah diklasifikasikan sebagai positif. Sebaliknya, dari 6.566 komentar positif, sebanyak 6.490 diklasifikasikan dengan benar, menunjukkan akurasi tinggi pada kelas positif. Hasil ini mengindikasikan bahwa meskipun *Naïve Bayes* efektif dalam mengenali komentar positif, performa nya kurang optimal dalam mendeteksi komentar negatif pada data multibahasa.

3.5. Analisis Perbandingan Algoritma

Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* memiliki performa yang lebih baik dibandingkan *Naïve Bayes* dalam mengklasifikasikan sentimen komentar *YouTube* multibahasa*. Random Forest* mampu mengenali komentar positif dan negatif secara seimbang dengan akurasi dan *f1-score* yang tinggi. Sebaliknya, *Naïve Bayes* menunjukkan kecenderungan bias terhadap komentar positif dan memiliki *recall* yang rendah pada kelas negatif. Hal ini menunjukkan bahwa *Random Forest* lebih andal dalam menangani distribusi data yang tidak seimbang, sedangkan *Naïve Bayes* lebih cocok untuk data yang seimbang dan homogen.

4. KESIMPULAN

Penelitian ini berhasil membangun model analisis sentimen terhadap komentar *YouTube* yang membahas teknologi *5G* menggunakan algoritma *Random Forest* dan *Naïve Bayes*. Proses dilakukan melalui tahapan SEMMA, mulai dari pengumpulan data multibahasa, *preprocessing*, pelabelan sentimen, hingga pelatihan dan evaluasi model. Hasil evaluasi menunjukkan bahwa *Random Forest* memiliki performa lebih baik dibandingkan *Naïve Bayes*, khususnya dalam mengklasifikasikan komentar negatif secara lebih akurat. *Naïve Bayes* cenderung bias terhadap komentar positif dan kurang optimal pada data yang tidak seimbang. Dengan demikian, *Random Forest* direkomendasikan sebagai algoritma yang lebih efektif untuk analisis sentimen multibahasa pada topik teknologi *5G*.

DAFTAR PUSTAKA

[1] Muhamad Rizky, Selpi Amanda Fadillah, Juniwan Juniwan, Muhamad Yusuf Habibi, dan Didik Aribowo, "Perkembangan Teknologi Jaringan 5G di Indonesia," *Jupit. Publ. Ilmu Keteknikan Ind. Tek. Elektro Dan Inform.*, vol. 2, no. 3, hlm. 58–68, Apr 2024, doi: 10.61132/jupiter.v2i3.279.

[2] Mesya Nandawani Manik dan Rayyan Firdaus, "Tantangan Dan Peluang Jaringan 5G Dalam Meningkatkan Operasional Perusahaan," *J. Manuhara Pus. Penelit. Ilmu Manaj. Dan Bisnis*, vol. 2, no. 3, hlm. 203–209, Jun 2024, doi: 10.61132/manuhara.v2i3.1026.

[3] A. Amali, "Perbandingan Algoritma Sentimen Analisis media data Twitter Pilgub Jabar 2018," *Pelita Teknol.*, vol. 15, no. 1, hlm. 26–36, Apr 2020, doi: 10.37366/pelitatekno.v15i1.298.

[4] D. A. Pamungkas dan U. D. Soer, "Analisis Sentimen Publik Terhadap Polusi Udara di Kota Jakarta: Perbandingan Algoritma Support Vector Machine, Naive Bayes, dan Random Forest," vol. 13, no. 3.

[5] N. Wijaya dan E. S. Panjaitan, "Analisis Sentimen Ulasan Aplikasi Instagram di Google Play Store: Pendekatan Multinomial Naive Bayes dan Berbasis Leksikon," *Build. Inform. Technol. Sci. BITS*, vol. 6, no. 2, hlm. 921–929, Sep 2024, doi: 10.47065/bits.v6i2.5615.

[6] O. N. Julianti, N. Suarna, dan W. Prihartono, "Penerapan Natural Language Processing Pada Analisis Sentimen Judi Online di Media Sosial Twitter," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 3, hlm. 2936–2941, Mei 2024, doi: 10.36040/jati.v8i3.9613.

[7] Christ Mario dan Ryan Randy Suryono, "Public Sentiment Analysis on Dirty Vote Movie on YouTube using Random Forest and Naïve Bayes," *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 1, hlm. 111–122, Mar 2025, doi: 10.35314/ev9j2g33.

[8] S. M. Tiana, "Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Ditutupnya Fitur Belanja Pada Tiktok Dengan Menggunakan Naïve Bayes Classifier Dan Random Forest Classifier," *J-Com J. Comput.*, vol. 4, no. 1, hlm. 76–86, Mar 2024, doi: 10.33330/j-com.v4i1.3140.

[9] D. D. Kurniawan dan D. Indrayana, "Metode Naïve Bayes Untuk Klasifikasi Sentimen Tweet Pemain Naturalisasi Tim Nasional Senior Sepak Bola Indonesia," vol. 9, no. 2, 2025.

[10] V. S. Steviana dan A. B. Kusdinar, "Implementasi Naïve Bayes untuk Klasifikasi Rekomendasi Bursa Kerja Khusus Di SMKN 1 Sukalarang," vol. 11, no. 1, 2025.

[11] T. D. Putra dan D. Oktafiani, "Klasifikasi Sentimen Postingan Sosial Media Menggunakan Machine Learning Random Forest dan Naïve Bayes".

[12] D. Mualfah, A. Prihatin, R. Firdaus, dan Sunanto, "Analisis Sentimen Masyarakat Terhadap Kasus Pembobolan Data Nasabah Bank BSI Pada Twitter Menggunakan Metode Random Forest Dan Naïve Bayes," *J. FASILKOM*, vol. 13, no. 3, hlm. 614–620, Jan 2024, doi: 10.37859/jf.v13i3.6478.

[13] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, dan M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Inform. Web Eng.*, vol. 3, no. 1, hlm. 67–75, Feb 2024, doi: 10.33093/jiwe.2024.3.1.5.

[14] A. Karimah, G. Dwilestari, dan M. Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik di Platform Youtube Dengan Metode Naive Bayes," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 1, hlm. 767–737, Mar 2024, doi: 10.36040/jati.v8i1.8373.

[15] A. Amali, D. Maulana, E. Widodo, A. Firmansyah, dan M. Danny, "Sentiment Analysis of Bekasi Floods Using SVM and Naive Bayes with Advanced Feature Selection," *Brill. Res. Artif. Intell.*, vol. 4, no. 1, hlm. 362–371, Jul 2024, doi: 10.47709/brilliance.v4i1.4268.