



Efektivitas IndoBERT pada Klasifikasi Sentimen Evaluasi Dosen: Studi Komparatif *Support Vector Machine* dan Naive Bayes

Tiffany Nabarian ¹, Maryam Hasnaa' Syamila ^{1*}, Salman El Farisi ¹, Ananto Dwi Saputro ²

¹ Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri, Indonesia

² Direktorat Kebijakan Sumber Daya Manusia Keamanan Siber dan Sandi, Deputi Bidang Strategi dan Kebijakan Keamanan Siber dan Sandi, Badan Siber dan Sandi Negara, Indonesia

* Korespondensi: mary22067ti@student.nurulfikri.ac.id

Sitasi: T. Nabarian, M. H. Syamila, S. El Farisi, and A. D. Saputro, "Efektivitas IndoBERT pada Klasifikasi Sentimen Evaluasi Dosen: Studi Komparatif Support Vector Machine dan Naive Bayes", *Jurnal Teknologi Informasi Dan Multimedia*, vol. 8, no. 2, pp. 281-292, 2026. <https://doi.org/10.35746/jtim.v8i2.984>

Diterima: 06-03-2026

Direvisi: 02-04-2026

Disetujui: 09-02-2026



Copyright: © 2026 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Abstract: Sentiment analysis of student feedback plays an important role in evaluating the quality of teaching and learning processes in higher education. Qualitative comments in Student Evaluation of Teaching (SET) provide deeper insights than numerical ratings. However, they are expressed in unstructured textual form, making large-scale analysis difficult to conduct consistently and systematically. Therefore, Natural Language Processing (NLP) approaches are required to automatically identify sentiment tendencies within student comments. This study aims to compare the performance of Gaussian Naive Bayes and Support Vector Machine (SVM) algorithms for classifying sentiment in SET comments using IndoBERT-based text embeddings. The research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, including stages of data understanding, data preparation, modeling, evaluation, and deployment. Text comments were pre-processed and transformed into numerical vectors using IndoBERT Sentence-BERT embeddings to capture contextual semantic relationships between words. These embeddings were then used as input features for both classification algorithms. Evaluation results show that the IndoBERT + SVM model achieved an accuracy of 93.88%, outperforming IndoBERT + Naive Bayes which obtained 92.40%. The SVM model also demonstrated more balanced precision, recall, and F1-score values across sentiment classes. These findings indicate that SVM is more effective in utilizing high-dimensional contextual embeddings for sentiment classification of student feedback.

Keywords: indobert; naive bayes; sentiment analysis; support vector machine (SVM); student evaluation of teaching (SET)

Abstrak: Analisis sentimen umpan balik mahasiswa merupakan aspek penting dalam mengevaluasi kualitas proses pembelajaran di perguruan tinggi. Komentar kualitatif pada Evaluasi Dosen oleh Mahasiswa (EDOM) memberikan informasi lebih mendalam dibandingkan penilaian kuantitatif, namun berbentuk teks tidak terstruktur sehingga sulit dianalisis secara sistematis dan konsisten ketika jumlah data sangat besar. Oleh karena itu, diperlukan pendekatan komputasional berbasis *Natural Language Processing* (NLP) untuk mengidentifikasi kecenderungan sentimen dalam komentar mahasiswa secara otomatis dan terstruktur. Penelitian ini bertujuan membandingkan performa algoritma Gaussian Naive Bayes dan *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen komentar EDOM menggunakan representasi teks berbasis *embedding* IndoBERT. Proses penelitian mengikuti kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang mencakup tahap pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi model. Teks komentar melalui tahap praproses kemudian direpresentasikan sebagai vektor numerik menggunakan IndoBERT Sentence-BERT untuk menangkap hubungan semantik antar kata secara kontekstual. Vektor *embedding* digunakan sebagai fitur dalam proses klasifikasi menggunakan kedua algoritma. Hasil evaluasi menunjukkan model IndoBERT + SVM memperoleh nilai *accuracy* sebesar 93,88%, lebih tinggi dibandingkan IndoBERT + Naive Bayes sebesar 92,40%,

serta menunjukkan keseimbangan *precision*, *recall*, dan *F1-score* yang lebih baik pada kedua kelas sentimen. Temuan ini menunjukkan bahwa SVM lebih efektif dalam memanfaatkan representasi *embedding* kontekstual berdimensi tinggi untuk klasifikasi sentimen mahasiswa.

Kata kunci: analisis sentimen; evaluasi dosen oleh mahasiswa (EDOM); indobert; naive bayes; *support vector machine* (SVM)

1. Pendahuluan

Evaluasi Dosen oleh Mahasiswa (EDOM) merupakan instrumen penting dalam sistem penjaminan mutu pendidikan tinggi karena menyediakan umpan balik langsung mengenai kualitas proses pembelajaran dari perspektif mahasiswa. Selain skor penilaian kuantitatif, EDOM juga menghasilkan komentar tekstual yang berisi opini, kritik, dan saran terhadap proses pembelajaran di kelas sehingga memberikan gambaran yang lebih mendalam mengenai pengalaman belajar mahasiswa serta aspek pembelajaran yang memerlukan perbaikan [1]. Namun, analisis komentar dalam jumlah besar masih sering dilakukan secara manual sehingga memerlukan waktu yang panjang dan berpotensi menimbulkan inkonsistensi interpretasi antar penilai [2]. Kondisi ini menunjukkan perlunya pendekatan otomatis untuk membantu proses analisis komentar mahasiswa secara lebih efisien dan konsisten.

Perkembangan teknologi *Natural Language Processing* (NLP) memungkinkan analisis sentimen dilakukan secara otomatis untuk mengidentifikasi polaritas opini dalam teks. Berbagai penelitian menunjukkan bahwa algoritma klasifikasi seperti Naive Bayes dan *Support Vector Machine* (SVM) memiliki performa yang stabil dan efisien dalam pengolahan data teks serta sering digunakan sebagai metode dasar dalam penelitian komparatif analisis sentimen [3], [4], [5], [6], [7]. Naive Bayes dikenal sebagai algoritma probabilistik yang sederhana dan efektif dalam menangani data teks berdimensi tinggi, sedangkan SVM memiliki kemampuan dalam membentuk batas pemisah optimal pada ruang fitur yang kompleks sehingga mampu menghasilkan performa yang lebih stabil. Perbedaan karakteristik tersebut menjadikan kedua algoritma ini relevan untuk dibandingkan dalam mengevaluasi efektivitas representasi fitur pada tugas klasifikasi sentimen.

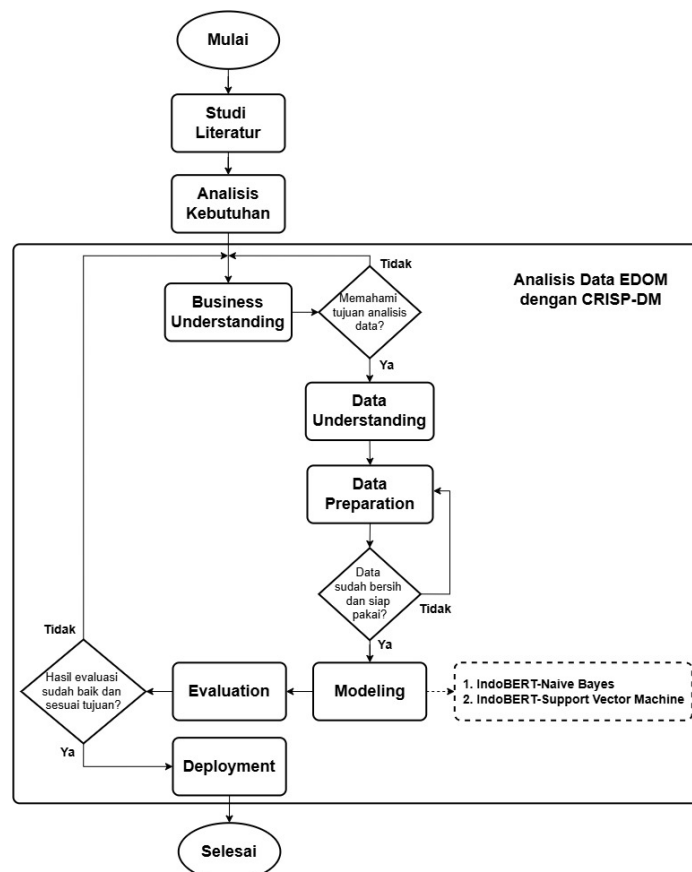
Meskipun model *transformer* seperti IndoBERT dapat meningkatkan performa melalui proses *fine-tuning*, pendekatan tersebut umumnya memerlukan sumber daya komputasi yang lebih besar serta proses pelatihan yang lebih kompleks. Penelitian [9] dan [12] menunjukkan bahwa representasi *embedding* dari model prelatih dapat dimanfaatkan secara efektif sebagai fitur bagi algoritma klasifikasi yang lebih ringan seperti *Support Vector Machine* atau Naive Bayes tanpa perlu melakukan *fine-tuning* secara langsung. Pendekatan ini memungkinkan proses pelatihan yang lebih efisien sekaligus tetap mempertahankan kualitas representasi semantik teks yang dihasilkan oleh model IndoBERT. Selain itu, IndoBERT dirancang khusus untuk bahasa Indonesia sehingga mampu menangkap konteks linguistik, variasi ekspresi informal, serta nuansa makna dalam komentar mahasiswa yang tidak dapat direpresentasikan secara optimal oleh metode berbasis frekuensi.

Meskipun penelitian analisis sentimen telah banyak dilakukan, terdapat *research gap* yang signifikan pada domain Evaluasi Dosen oleh Mahasiswa (EDOM). Pertama, mayoritas penelitian sentimen pada EDOM masih mengandalkan ekstraksi fitur tradisional berbasis frekuensi (seperti TF-IDF) yang gagal menangkap konteks linguistik dan nuansa makna dari komentar mahasiswa [8]. Kedua, meskipun model *transformer* seperti IndoBERT terbukti unggul pada ulasan produk [6], [9], implementasi *fine-tuning* secara

penyempurnaan memerlukan sumber daya komputasi yang besar, sehingga kurang efisien untuk diintegrasikan secara langsung pada sistem akademik perguruan tinggi berskala menengah. Kajian komparatif yang mengevaluasi pemanfaatan representasi *embedding* IndoBERT sebagai input bagi algoritma klasifikasi klasik yang lebih ringan (seperti Naive Bayes dan SVM) khusus pada teks evaluasi pembelajaran masih sangat terbatas. Untuk menjembatani kesenjangan tersebut, penelitian ini bertujuan melakukan analisis komparatif performa algoritma Naive Bayes dan SVM yang diintegrasikan dengan representasi *embedding* IndoBERT. Penelitian ini memberikan dua kontribusi utama. Secara metodologis, penelitian ini menyajikan bukti empiris yang komprehensif (berbasis *accuracy*, *recall*, *precision*, dan *F1-score*) mengenai efektivitas algoritma klasik ketika dikombinasikan dengan *embedding* kontekstual pada domain teks pendidikan. Secara praktis, penelitian ini menghasilkan kontribusi berupa implementasi prototipe sistem analisis sentimen berbasis web yang efisien, ringan secara komputasi, dan dapat dimanfaatkan secara langsung untuk mendukung pengambilan keputusan berbasis data dalam peningkatan kualitas pembelajaran di perguruan tinggi.

2. Bahan dan Metode

Penelitian ini menggunakan pendekatan kuantitatif komparatif untuk mengevaluasi performa algoritma klasifikasi Naive Bayes dan *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen mahasiswa pada sistem Evaluasi Dosen oleh Mahasiswa (EDOM). Perbandingan dilakukan untuk menganalisis perbedaan kemampuan kedua algoritma menggunakan metrik evaluasi kuantitatif guna mengidentifikasi pendekatan paling efektif dalam mengenali polaritas sentimen berbahasa Indonesia dengan memanfaatkan representasi teks berbasis *embedding* IndoBERT. Proses penelitian mengikuti *flowchart Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang menyediakan tahapan analisis data sistematis mulai dari pemahaman masalah hingga implementasi model [10].



Gambar 1. Alur Metodologi Penelitian Berbasis CRISP-DM

Sebagaimana ditunjukkan pada Gambar 1, penelitian ini mengikuti tahapan dalam kerangka kerja CRISP-DM yang diawali dengan tahap *business understanding* untuk mendefinisikan kebutuhan analisis sentimen pada komentar EDOM, kemudian dilanjutkan dengan *data understanding* yang mencakup eksplorasi dataset dan pemeriksaan kualitas data. Tahap berikutnya adalah *data preparation* untuk membersihkan dan menyiapkan data sebelum proses pemodelan, diikuti pembangunan model klasifikasi pada tahap *modeling* serta evaluasi performa model pada tahap *evaluation* menggunakan metrik kinerja. Model dengan performa terbaik selanjutnya diimplementasikan pada tahap *deployment* dalam bentuk prototipe sistem analisis sentimen, menggunakan dataset komentar mahasiswa pada sistem EDOM Program Studi Teknik Informatika Sekolah Tinggi Teknologi Terpadu Nurul Fikri yang diperoleh melalui kerja sama internal program studi, dilindungi oleh perjanjian *Non-Disclosure Agreement* (NDA), serta dianonimkan sebelum diproses melalui tahapan *preprocessing* untuk keperluan pemodelan dan evaluasi algoritma klasifikasi.

Tahap *data preparation* bertujuan meningkatkan kualitas dataset sebelum proses pelatihan model klasifikasi. Proses ini mencakup pembersihan teks, normalisasi huruf, penghapusan URL dan karakter non-Latin, serta penyaringan komentar yang terlalu pendek. Selain itu dilakukan pelabelan sentimen otomatis menggunakan model IndoBERT Sentiment (agufsamudra/indo-sentiment-analysis) serta penyesuaian *rule-based* untuk menangani komentar yang tidak memiliki polaritas sentimen yang jelas, sementara penyeimbangan kelas dilakukan melalui teknik *undersampling*. Rangkaian tahapan *preprocessing* yang diterapkan dalam penelitian ini dirangkum pada Tabel 1.

Tabel 1. Tahapan *Preprocessing* Data Komentar EDOM

Tahapan	Deskripsi
Pembersihan data	Menghapus komentar kosong dan duplikat
Normalisasi teks	<i>Lowercase</i> serta penghapusan URL, simbol, emoji, dan karakter non-Latin
Penyaringan komentar	Menghapus komentar dengan jumlah kata kurang dari tiga
Pelabelan sentimen	Pelabelan otomatis menggunakan model IndoBERT Sentiment
Penyesuaian <i>rule-based</i>	Penyesuaian label untuk komentar netral
<i>Balancing</i> dataset	Penyeimbangan kelas menggunakan <i>undersampling</i>

Sebagaimana ditunjukkan pada Tabel 1, proses *preprocessing* menghasilkan dataset yang lebih bersih dan memiliki distribusi kelas yang lebih seimbang. Selanjutnya teks diubah menjadi representasi numerik menggunakan teknik *embedding* berbasis IndoBERT Sentence-BERT (firqaaa/indo-sentence-bert-base) yang menghasilkan vektor berdimensi 768. Representasi kontekstual ini dipilih karena mampu menangkap hubungan semantik antar kata secara lebih baik dibandingkan pendekatan berbasis frekuensi seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) [11], [12], serta telah menunjukkan performa yang baik pada berbagai penelitian analisis sentimen berbahasa Indonesia [6], [7], [9].

Tabel 2. Hasil Validasi Pelabelan Sentimen oleh *Annotator*

Komponen	<i>Annotator 1</i>	<i>Annotator 2</i>
Jumlah Data Sampel	300	300
Jumlah Kesesuaian Label	231	234
<i>Agreement</i> (%)	77%	78%

Untuk memastikan kualitas pelabelan sentimen yang dihasilkan oleh model IndoBERT, dilakukan proses validasi menggunakan penilaian manual oleh dua *annotator* independen. Sebanyak 300 sampel komentar dipilih secara acak dan dibandingkan dengan label yang dihasilkan model, dengan tingkat kesesuaian sebesar 77% pada *annotator* pertama dan 78% pada *annotator* kedua sebagaimana ditunjukkan pada Tabel 2. Hasil tersebut menunjukkan bahwa pelabelan otomatis memiliki tingkat kesesuaian yang baik dengan penilaian manusia. Meskipun demikian, potensi bias dari pelabelan algoritmik tetap menjadi batasan yang perlu dicermati. Dalam konteks EDOM, bias ini umumnya

teridentifikasi pada dua skenario utama: (1) komentar dengan sentimen campuran (*mixed-sentiment*), seperti "Dosen mengajar dengan sangat interaktif, tetapi tugas yang diberikan terlalu membebani", dan (2) penggunaan sarkasme spesifik mahasiswa yang mungkin gagal ditangkap secara presisi oleh model pralatih. Validasi manual oleh *annotator* dalam penelitian ini berfungsi sebagai langkah mitigasi (pengendalian mutu) untuk memastikan bahwa bias algoritmik tersebut bersifat minor dan tidak mendistorsi distribusi kelas secara keseluruhan, sehingga *dataset* tetap dinilai reliabel untuk melatih model klasifikasi.

Tahap *modeling* dilakukan dengan membangun dua model klasifikasi yaitu Gaussian Naive Bayes dan Support Vector Machine (LinearSVC). Naive Bayes merupakan algoritma probabilistik yang efisien untuk klasifikasi teks, sedangkan SVM mampu membentuk batas pemisah optimal pada ruang fitur berdimensi tinggi seperti *embedding* teks. Sejumlah penelitian sebelumnya menunjukkan bahwa SVM menghasilkan performa yang lebih stabil dibandingkan Naive Bayes pada berbagai tugas analisis sentimen [13], [14], [15], meskipun sebagian besar penelitian tersebut masih menggunakan representasi fitur tradisional [3], dan penelitian pada domain EDOM umumnya minim analisis komparatif sistematis [8]. Oleh karena itu, penelitian ini menggunakan pembagian dataset dengan skema *stratified train-test split* (80% data latih dan 20% data uji) dengan konfigurasi model yang ditampilkan pada Tabel 3. Pendekatan *stratified* diterapkan secara khusus untuk memastikan bahwa proporsi kelas sentimen pada data uji benar-benar merepresentasikan distribusi populasi asli. Mengingat ukuran dataset evaluasi yang memadai serta penggunaan fitur *dense embedding* berdimensi tinggi dari IndoBERT, skema *stratified hold-out* ini dinilai secara metodologis telah cukup *robust* untuk menghasilkan estimasi performa model yang stabil tanpa memerlukan beban komputasi eksponensial dari *k-fold cross-validation*. Penerapan *cross-validation* dapat direkomendasikan untuk penelitian lanjutan, khususnya saat menguji generalisasi model pada dataset lintas institusi.

Tabel 3. Konfigurasi Model Klasifikasi

Komponen	Konfigurasi
Model <i>embedding</i>	IndoBERT SBERT (firqaaa/indo-sentence-bert-base)
Dimensi <i>embedding</i>	768
Algoritma 1	Gaussian Naive Bayes
Parameter Naive Bayes	<i>Default</i>
Algoritma 2	LinearSVC
Parameter <i>Support Vector Machine</i>	C = 1.0, max_iter = 5000
Pembagian data	80% data latih, 20% data uji
Target klasifikasi	Sentimen positif dan negatif

Parameter regularisasi C pada model *Support Vector Machine* ditetapkan sebesar 1.0 sebagai nilai *default* yang umum digunakan dalam implementasi LinearSVC. Nilai ini memberikan keseimbangan antara upaya memaksimalkan margin pemisah dan meminimalkan kesalahan klasifikasi pada data latih, khususnya pada data dengan representasi fitur berdimensi tinggi seperti *embedding* IndoBERT. Pemilihan nilai *default* juga bertujuan untuk menjaga konsistensi dan keadilan dalam perbandingan performa antar algoritma tanpa melakukan optimasi hiperparameter yang dapat memengaruhi hasil komparasi.

Performa model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kemampuan klasifikasi pada kedua kelas sentimen. Seluruh proses pengolahan dan evaluasi dilakukan menggunakan Python dengan pustaka Scikit-learn, Transformers, Pandas, dan NumPy pada lingkungan Google Colab dengan dukungan GPU NVIDIA T4 untuk mempercepat ekstraksi *embedding*. Model dengan performa terbaik kemudian diimplementasikan dalam prototipe sistem berbasis Streamlit sebagai bentuk penerapan analisis sentimen dalam evaluasi pembelajaran berbasis data.

3. Hasil

Bagian ini menyajikan hasil evaluasi model klasifikasi sentimen pada komentar Evaluasi Dosen oleh Mahasiswa (EDOM). Analisis mencakup persiapan dataset, representasi fitur menggunakan *embedding* IndoBERT, serta evaluasi performa algoritma Gaussian Naive Bayes dan *Support Vector Machine*. Hasil disajikan dalam bentuk tabel dan visualisasi sebagai dasar pembahasan lebih lanjut.

3.1. Hasil Persiapan Dataset

Tahap awal penelitian dilakukan dengan menggabungkan data komentar Evaluasi Dosen oleh Mahasiswa (EDOM) dari beberapa periode akademik tahun 2024. Proses tersebut menghasilkan dataset awal sebanyak 14.744 baris data yang memuat atribut periode, nama dosen, mata kuliah, program studi, kategori komentar, dan komentar mahasiswa. Kolom komentar menjadi fokus analisis karena berisi umpan balik kualitatif mahasiswa mengenai proses pembelajaran, sedangkan ringkasan proses penyaringan dataset sebelum pemodelan ditunjukkan pada Tabel 4.

Tabel 4. Ringkasan Proses Penyaringan Dataset

Tahapan <i>preprocessing</i>	Jumlah data yang tersisa
Dataset awal	14.744
Penghapusan komentar kosong dan duplikat	14.740
Penghapusan komentar pendek (<3 kata), Pelabelan menggunakan IndoBERT + <i>rule-based</i>	6.247
Penghapusan label tidak valid (<i>unknown</i>)	5.682
<i>Balancing dataset (undersampling)</i>	3.022

Sebagaimana ditunjukkan pada Tabel 4, dataset awal berjumlah 14.744 komentar kemudian dibersihkan dengan menghapus 4 komentar kosong sehingga tersisa 14.740 data, dan penyaringan komentar dengan panjang kurang dari tiga kata mengurangi dataset menjadi 6.247 komentar. Penyaringan ini dilakukan karena komentar dengan jumlah kata yang sangat terbatas umumnya tidak mengandung informasi sentimen yang cukup jelas dan cenderung bersifat umum, sehingga sulit dianalisis secara kontekstual oleh model. Pelabelan menggunakan IndoBERT dengan penyesuaian *rule-based* menghasilkan tiga kategori sentimen, yaitu positif, negatif, dan *unknown*, di mana 565 komentar berlabel *unknown* dihapus sehingga tersisa 5.682 komentar. Selanjutnya dilakukan proses *undersampling* hingga diperoleh dataset akhir sebanyak 3.022 komentar, dengan tujuan menyeimbangkan distribusi kelas, di mana meskipun berpotensi mengurangi sebagian data, pendekatan ini dinilai tetap mempertahankan pola utama karena jumlah data yang tersisa masih cukup besar untuk mendukung proses pembelajaran model.

Tabel 5. Contoh Perubahan Teks Komentar Sebelum dan Sesudah Pembersihan

komentar	komentar_cleaned
terimakasih atas satu semesternya 🙏❤️	terimakasih atas satu semesternya
Mohon maaf pak saat menjelaskan lebih pelan lagi karena kami suka tertinggal codingannya @刺晶	mohon maaf pak saat menjelaskan lebih pelan lagi karena kami suka tertinggal codingannya
Simulasi Kerja Nyata 竊💎 Ikut hackathon, buat proyek kolaboratif, atau rewiu kode teman.	simulasi kerja nyata ikut hackathon buat proyek kolaboratif atau rewiu kode teman

Contoh perubahan teks sebelum dan sesudah proses pembersihan ditunjukkan pada Tabel 5. Proses ini meliputi normalisasi huruf, penghapusan simbol, emoji, dan karakter non-teks untuk mengurangi *noise* tanpa mengubah makna utama komentar. Setelah tahap pembersihan selesai dilakukan, data siap digunakan pada proses representasi fitur menggunakan *embedding* IndoBERT.

3.2. Hasil Representasi Fitur Menggunakan IndoBERT

Tahap berikutnya adalah merepresentasikan teks komentar menjadi fitur numerik menggunakan model IndoBERT Sentence-BERT. Proses ekstraksi fitur menghasilkan vektor *embedding* berdimensi 768 untuk setiap komentar yang merepresentasikan hubungan semantik antar kata dalam kalimat. Berdasarkan hasil ekstraksi fitur, data latih direpresentasikan dalam matriks berukuran 2417×768 , sedangkan data uji berukuran 605×768 , yang kemudian digunakan sebagai input bagi algoritma klasifikasi sentimen pada tahap evaluasi model.

3.3. Hasil Evaluasi Model Naive Bayes

Model pertama yang diuji dalam penelitian ini adalah Gaussian Naive Bayes. Model dilatih menggunakan *embedding* IndoBERT dari data latih dan selanjutnya diuji pada data uji. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *recall*, *precision*, dan *F1-score* untuk menilai kemampuan klasifikasi model. Ringkasan hasil evaluasi tersebut disajikan pada Tabel 6.

Tabel 6. Hasil Evaluasi Model Gaussian Naive Bayes

Metrik	Nilai
<i>Accuracy</i>	0,9240
<i>Precision</i> (Negatif)	0,8954
<i>Precision</i> (Positif)	0,9571
<i>Recall</i> (Negatif)	0,9604
<i>Recall</i> (Positif)	0,8874
<i>F1-score</i> (Negatif)	0,9268
<i>F1-score</i> (Positif)	0,9210

Berdasarkan Tabel 6, model Naive Bayes memperoleh nilai *accuracy* sebesar 92,40% pada data uji dengan nilai *F1-score* pada kedua kelas sentimen di atas 0,92 yang menunjukkan performa klasifikasi yang cukup stabil. Hasil tersebut menunjukkan bahwa sebagian besar komentar mahasiswa dapat diklasifikasikan dengan benar oleh model, meskipun masih terdapat beberapa kesalahan prediksi. Kesalahan tersebut mengindikasikan keterbatasan model probabilistik dalam menangkap hubungan kompleks antar fitur *embedding*, sementara distribusi prediksi terhadap label aktual disajikan pada Tabel 7.

Tabel 7. Confusion Matrix Model Gaussian Naive Bayes

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	291	12
Aktual Positif	34	268

Sebagaimana terlihat pada Tabel 7, mayoritas komentar berhasil diklasifikasikan dengan benar oleh model Naive Bayes. Kesalahan klasifikasi masih terjadi pada sejumlah kecil data uji, terutama pada komentar positif yang diprediksi sebagai negatif. Hal ini menunjukkan bahwa model Naive Bayes masih memiliki keterbatasan dalam memanfaatkan hubungan antar fitur *embedding* secara lebih kompleks.

3.4. Hasil Evaluasi Model Support Vector Machine

Model kedua yang digunakan dalam penelitian ini adalah *Support Vector Machine* (LinearSVC). Model dilatih menggunakan *embedding* IndoBERT yang sama sehingga perbandingan performa dapat dilakukan secara objektif. Evaluasi model dilakukan menggunakan metrik klasifikasi yang sama dengan model sebelumnya. Ringkasan hasil evaluasi model SVM ditunjukkan pada Tabel 8.

Tabel 8. Hasil Evaluasi Model Support Vector Machine

Metrik	Nilai
<i>Accuracy</i>	0,9388

Metrik	Nilai
<i>Precision</i> (Negatif)	0,9524
<i>Precision</i> (Positif)	0,9260
<i>Recall</i> (Negatif)	0,9241
<i>Recall</i> (Positif)	0,9536
<i>F1-score</i> (Negatif)	0,9380
<i>F1-score</i> (Positif)	0,9396

Berdasarkan Tabel 8, model SVM memperoleh nilai *accuracy* 93,88%, lebih tinggi dibandingkan model Naive Bayes. Nilai *recall*, *precision*, dan *F1-score* pada kedua kelas sentimen juga menunjukkan keseimbangan performa yang baik. Hal ini menunjukkan bahwa SVM mampu memanfaatkan representasi *embedding* IndoBERT secara lebih efektif dalam memisahkan kelas sentimen. Distribusi prediksi model terhadap label aktual ditampilkan dalam bentuk *confusion matrix* pada Tabel 9.

Tabel 9. *Confusion Matrix* Model Support Vector Machine

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	280	23
Aktual Positif	14	288

Sebagaimana ditunjukkan pada Tabel 9, jumlah prediksi yang benar jauh lebih dominan dibandingkan kesalahan klasifikasi. Hal ini menunjukkan bahwa model SVM memiliki kemampuan generalisasi yang baik terhadap data uji. Hasil tersebut menunjukkan bahwa kombinasi *embedding* IndoBERT dan algoritma SVM mampu menangkap pola sentimen komentar mahasiswa secara lebih konsisten.

3.5. Perbandingan Performa Model

Perbandingan performa kedua model klasifikasi dilakukan menggunakan metrik evaluasi *accuracy*, *recall*, *precision*, dan *F1-score*. Evaluasi dilakukan untuk mengukur kemampuan model dalam mengklasifikasikan dua kelas sentimen, yaitu positif dan negatif. Ringkasan hasil evaluasi kedua model ditampilkan pada Tabel 10.

Tabel 10. Perbandingan Performa Model Klasifikasi

Model	<i>Accuracy</i>	<i>Precision</i> Negatif	<i>Recall</i> Negatif	<i>F1-score</i> Negatif	<i>Precision</i> Positif	<i>Recall</i> Positif	<i>F1-score</i> Positif
IndoBERT + SVM	0,9388	0,9524	0,9241	0,9380	0,9260	0,9536	0,9396
IndoBERT + Naive Bayes	0,9240	0,8954	0,9604	0,9268	0,9571	0,8874	0,9210

Berdasarkan Tabel 10, kedua model menunjukkan performa klasifikasi yang baik dalam menganalisis sentimen komentar mahasiswa. Model IndoBERT + SVM memperoleh nilai *accuracy* tertinggi sebesar 93,88%, sedangkan model IndoBERT + Naive Bayes memperoleh *accuracy* sebesar 92,40% dengan nilai *F1-score* yang lebih seimbang pada kedua kelas sentimen. Perbedaan juga terlihat pada nilai *precision* dan *recall*, di mana model Naive Bayes memiliki *recall* lebih tinggi pada kelas negatif namun *precision* lebih rendah, sedangkan model SVM menunjukkan keseimbangan metrik yang lebih baik sehingga dipilih sebagai model dengan performa terbaik dalam penelitian ini.

3.6. Analisis Kesalahan (Error Analysis)

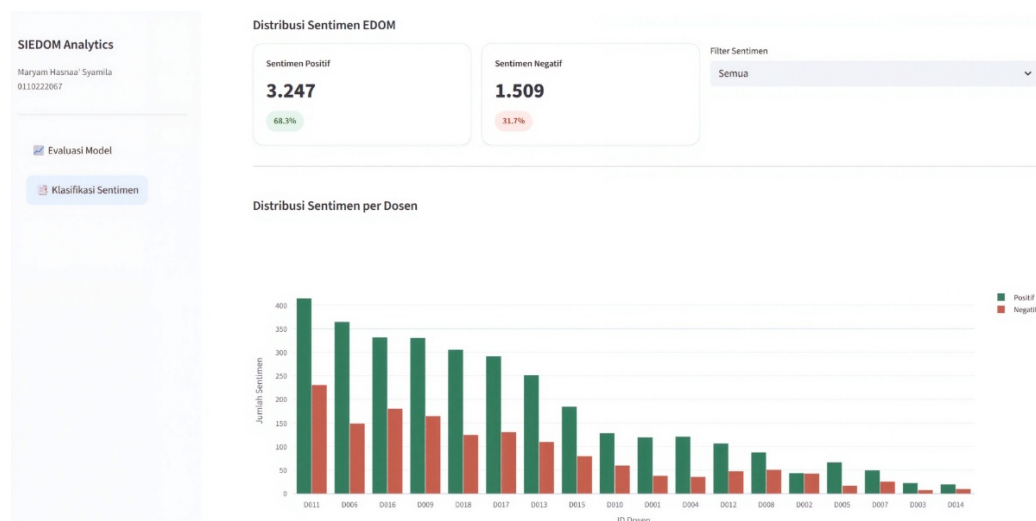
Sebagai bagian dari analisis lebih lanjut, dilakukan analisis kesalahan (*error analysis*) terhadap hasil prediksi model terbaik yaitu *Support Vector Machine* (SVM). Berdasarkan hasil evaluasi, terdapat 37 kesalahan dari total 605 data uji dengan *error rate* sebesar 6,12%, yang menunjukkan bahwa sebagian besar prediksi model telah sesuai dengan label aktual. Analisis terhadap kesalahan prediksi menunjukkan bahwa *error* umumnya terjadi pada komentar yang bersifat ambigu atau mengandung lebih dari satu sentimen dalam satu kalimat sehingga sulit ditangkap secara kontekstual oleh model. Selain itu, kesalahan juga ditemukan pada komentar dengan struktur bahasa yang tidak baku, yang

menunjukkan bahwa meskipun model memiliki performa tinggi, masih terdapat keterbatasan dalam memahami nuansa makna pada teks yang kompleks.

Untuk mendukung analisis perbandingan performa model, dilakukan uji statistik menggunakan McNemar test untuk mengevaluasi perbedaan hasil prediksi antara model SVM dan Naive Bayes. Hasil pengujian menunjukkan nilai p -value sebesar 0,2429, yang mengindikasikan bahwa perbedaan performa antara kedua model belum menunjukkan perbedaan yang cukup kuat secara statistik pada tingkat signifikansi 0,05. Meskipun demikian, model SVM secara konsisten memperoleh nilai metrik evaluasi yang lebih tinggi dibandingkan Naive Bayes pada seluruh metrik yang digunakan. Hal ini menunjukkan bahwa SVM tetap memberikan performa yang lebih baik secara empiris dalam mengklasifikasikan sentimen komentar mahasiswa. Hal ini menunjukkan bahwa kedua model memiliki performa yang relatif sebanding dalam memanfaatkan representasi embedding IndoBERT.

3.7. Implementasi Prototipe Sistem Klasifikasi

Sebagai bentuk penerapan hasil penelitian, model klasifikasi diimplementasikan dalam prototipe sistem berbasis web menggunakan Streamlit. Sistem memanfaatkan model terbaik yaitu *Support Vector Machine* dengan representasi *embedding* IndoBERT untuk mengklasifikasikan sentimen komentar mahasiswa secara otomatis serta menyajikan hasil analisis dalam bentuk visualisasi. Antarmuka sistem yang dikembangkan ditunjukkan pada Gambar 2 yang menampilkan *dashboard* hasil klasifikasi sentimen EDOM.



Gambar 2. Tampilan Sistem Klasifikasi Sentimen EDOM

Sebagaimana ditunjukkan pada Gambar 2, *dashboard* menyajikan ringkasan distribusi sentimen dalam bentuk kartu statistik dengan jumlah sentimen positif 3.247 (68,3%) dan negatif 1.509 (31,7%). Sistem juga menampilkan grafik distribusi sentimen per dosen untuk membantu pengguna mengidentifikasi variasi persepsi mahasiswa terhadap proses pembelajaran. Selain itu, tersedia fitur penyaringan sentimen dan unduh hasil klasifikasi dalam format Excel untuk mendukung analisis lanjutan.

4. Pembahasan

Hasil penelitian menunjukkan bahwa kedua algoritma mampu mengklasifikasikan sentimen komentar mahasiswa dengan performa yang baik. Model Naive Bayes memperoleh *accuracy* sebesar 92,40%, sedangkan *Support Vector Machine* (SVM) mencapai 93,88% dengan nilai *precision*, *recall*, dan *F1-score* yang lebih seimbang pada kedua kelas sentimen. Untuk memberikan gambaran perbedaan perilaku klasifikasi kedua model, contoh perbandingan hasil prediksi ditampilkan pada Tabel 11.

Tabel 11. Contoh Perbandingan Klasifikasi Model pada Kasus SVM Benar dan Naive Bayes Salah

Komentar Mahasiswa	Label Aktual	Klasifikasi Naive Bayes	Klasifikasi SVM
yang semangat ngajarnya pak jangan terlalu kaku	negatif	positif	negatif
matkul yang menantang dan menarik	positif	negatif	positif
untuk pemweb sangat baik dalam menjelaskan dan selalu seru untuk codingannya	positif	negatif	positif

Sebagai bagian dari analisis kualitatif, Tabel 11 menampilkan beberapa contoh komentar mahasiswa yang berhasil diklasifikasikan dengan benar oleh model *Support Vector Machine* (SVM) namun keliru oleh model Naive Bayes. Kesalahan pada Naive Bayes umumnya terjadi pada komentar dengan konteks kalimat yang lebih kompleks sehingga asumsi independensi fitur pada algoritma tersebut kurang mampu menangkap hubungan semantik antar kata dalam representasi *embedding*. Sebaliknya, SVM yang memanfaatkan *embedding* kontekstual IndoBERT mampu membentuk batas pemisah yang lebih optimal pada ruang fitur berdimensi tinggi sehingga menghasilkan prediksi sentimen yang lebih stabil.

Performa yang lebih baik pada model SVM dapat dijelaskan secara teoritis melalui kemampuannya dalam membentuk *hyperplane* optimal yang memaksimalkan margin antar kelas pada ruang fitur berdimensi tinggi. Representasi *embedding* IndoBERT menghasilkan fitur kontekstual yang kompleks sehingga hubungan antar fitur tidak lagi bersifat independen, yang menjadi keterbatasan utama pada algoritma Naive Bayes. SVM mampu menangkap pola distribusi data yang lebih kompleks melalui pendekatan pemisahan berbasis margin, sehingga lebih efektif dalam memanfaatkan representasi semantik dari *embedding*. Oleh karena itu, kombinasi IndoBERT dan SVM memberikan keunggulan dalam mengklasifikasikan teks dengan struktur bahasa yang lebih kompleks dibandingkan pendekatan probabilistik sederhana.

Temuan penelitian ini sejalan dengan berbagai penelitian sebelumnya yang menunjukkan bahwa algoritma *Support Vector Machine* (SVM) cenderung menghasilkan performa yang lebih stabil dan akurat dibandingkan Naive Bayes dalam analisis sentimen [12], [13], [14], [15]. Namun, sebagian besar penelitian tersebut masih menggunakan representasi fitur tradisional seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) atau pendekatan distribusional seperti Word2Vec [3], [10], serta pada domain Evaluasi Dosen oleh Mahasiswa (EDOM) umumnya hanya menerapkan satu algoritma tanpa analisis komparatif antar model [8]. Oleh karena itu, penelitian ini memberikan kontribusi empiris dengan menunjukkan bahwa kombinasi IndoBERT dan SVM mampu meningkatkan kualitas klasifikasi sentimen pada komentar mahasiswa. Selain itu, pendekatan ini juga berpotensi dikembangkan lebih lanjut melalui *Explainable Artificial Intelligence* (XAI) atau analisis berbasis aspek untuk meningkatkan transparansi evaluasi pembelajaran.

Selain itu, model SVM juga menunjukkan kemampuan generalisasi yang baik, yang ditunjukkan oleh performa tinggi pada data uji yang tidak digunakan selama proses pelatihan. Hal ini mengindikasikan bahwa model mampu mempertahankan kinerja secara konsisten pada data baru di luar data latih. Penggunaan pembagian data secara *stratified* turut membantu menjaga distribusi kelas sehingga evaluasi model menjadi lebih representatif terhadap kondisi data sebenarnya. Meskipun demikian, penerapan teknik evaluasi tambahan seperti *cross-validation* dapat dipertimbangkan pada penelitian selanjutnya untuk memperoleh estimasi performa yang lebih *robust*.

5. Kesimpulan

Penelitian ini membandingkan performa algoritma Gaussian Naive Bayes dan *Support Vector Machine* (SVM) dalam klasifikasi sentimen komentar Evaluasi Dosen oleh Mahasiswa menggunakan representasi *embedding* IndoBERT. Hasil evaluasi menunjukkan bahwa model IndoBERT + SVM memperoleh performa lebih tinggi dengan *accuracy* 93,88% dibandingkan IndoBERT + Naive Bayes sebesar 92,40%, serta memiliki keseimbangan nilai *recall*, *precision*, dan *F1-score* yang lebih baik pada kedua kelas sentimen. Temuan ini menunjukkan bahwa SVM lebih efektif dalam memanfaatkan representasi *embedding* berdimensi tinggi dari IndoBERT untuk klasifikasi sentimen komentar mahasiswa dan dapat diimplementasikan dalam sistem sederhana untuk mendukung evaluasi pembelajaran berbasis data. Meskipun demikian, penelitian ini masih memiliki keterbatasan, di antaranya penggunaan pelabelan otomatis yang berpotensi menimbulkan bias serta belum diterapkannya teknik evaluasi seperti *cross-validation*, sehingga penelitian selanjutnya dapat mengkaji pendekatan evaluasi yang lebih *robust* serta pengembangan model menggunakan metode *Explainable Artificial Intelligence* (XAI) atau analisis berbasis aspek untuk meningkatkan interpretabilitas hasil.

Referensi

- [1] A. Sasmita, G. A. Pradnyana, and D. G. H. Divayana, "Pengembangan Sistem Analisis Sentimen untuk Evaluasi Kinerja Dosen Universitas Pendidikan Ganesha dengan Metode Naive Bayes," *JST (Jurnal Sains dan Teknologi)*, vol. 11, no. 2, pp. 451–462, Oct. 2022, <https://doi.org/10.23887/jstundiksha.v11i2.44384>.
- [2] N. Ferdiansyah and A. Solichin, "Analisis Sentimen terhadap Pembelajaran Dosen Berdasarkan Data Kritik Saran Mahasiswa Menggunakan Metode Naive Bayes," *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, vol. 19, no. 2, 2022, Sep. 2022. Accessed: Oct. 07, 2025. <https://doi.org/10.36080/bit.v19i2.2041>
- [3] J. Saputra, L. Maryani, Rahmaddeni, D. Wulandari, and W. Eka, "Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE," *INSTEK (Jurnal Informatika Sains dan Teknologi)*, vol. 10, no. 1, 2025, <https://doi.org/10.24252/instek.v10i1.54889>.
- [4] J. O. Leandro and M. I. Fianty, "Evaluation of Sentiment Analysis Methods for Social Media Applications: A Comparison of Support Vector Machines and Naive Bayes," *JOIV (International Journal on Informatics Visualization)*, vol. 9, no. 2, Mar. 2025, <https://doi.org/10.62527/joiv.9.2.2905>.
- [5] N. Z. B. Jannah and Kusnawi, "Comparison of Naive Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces," *Sinkron (Jurnal & Penelitian Teknik Informatika)*, vol. 8, no. 2, pp. 727–733, Apr. 2024, <https://doi.org/10.33395/sinkron.v8i2.13559>.
- [6] S. Aras, M. Yusuf, R. Ruimassa, E. A. B. Wambrauw, and E. B. Palalangan, "Sentiment Analysis on Shopee Product Reviews Using IndoBERT," *Journal of Information Systems and Informatics*, vol. 6, no. 3, pp. 1616–1627, Sep. 2024, <https://doi.org/10.51519/journalisi.v6i3.814>.
- [7] Tarwoto, R. Nugroho, N. Azka, and W. S. R. Graha, "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT," *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 2, pp. 495–505, Apr. 2025, <https://doi.org/10.35870/jtik.v9i2.3340>.
- [8] A. N. C. Putra, S. F. C. Haviana, and Bedie'ah, "Sentimen Analisis Komentar Mahasiswa EDOM Dengan Metode Support Vector Machine (SVM)," in *SERIMA-CE (Seminar Riset Mahasiswa-Computer & Electrical)*, 2023. Accessed: Oct. 22, 2025. <https://jurnal.unissula.ac.id/index.php/serima/article/view/30626>
- [9] G. T. Fadilah, L. Muflikhah, and R. S. Perdana, "Analisis Sentimen Produk Hijab Pada E-Commerce Tokopedia Menggunakan Algoritma Support Vector Machine dan IndoBERT Embedding," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, pp. 2548–964, Jan. 2025, Accessed: Nov. 14, 2025. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14390>
- [10] A. R. Hanum, E. S. Pramukantoro, and D. P. Kartikasari, "Studi Perbandingan Kinerja TF-IDF dan IndoBERT untuk Rekomendasi Resep Berdasarkan Ketersediaan Bahan Makanan Berbasis Website," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 10, Aug. 2025, Accessed: Oct. 22, 2025. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15384>
- [11] J. F. Tantoro and I. D. M. B. A. Darmawan, "Klasifikasi Berita Berdasarkan Kategori Menggunakan Convolutional Neural Network dengan IndoBERT," *JNATIA (Jurnal Nasional Teknologi Informasi dan Aplikasinya)*, vol. 3, no. 4, Aug. 2025, <https://doi.org/10.24843/JNATIA.2025.v03.i04.p20>.

-
- [12] M. M. Dakwah, A. A. Firdaus, F. Furizal, and R. Faresta, "Sentiment Analysis on Marketplace in Indonesia Using Support Vector Machine and Naïve Bayes Method," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, p. 39, Feb. 2024, <https://doi.org/10.26555/jiteki.v10i1.28070>.
- [13] A. Supian, B. T. Revaldo, N. Marhadi, L. Efrizoni, and Rahmadden, "Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara," *Jurnal Ilmiah Informatika (JIF)*, vol. 12, no. 1, Mar. 2024, <https://doi.org/10.33884/jif.v12i01.8721>.
- [14] H. Hariyadi, D. Firdo, and M. H. Al Rafi, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Ulasan Aplikasi Canva," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 261–269, Mar. 2024, <https://doi.org/10.33395/jmp.v13i1.13568>.
- [15] P. R. Sari, D. R. Indah, E. Rasywir, M. A. Firdaus, and G. Athalina, "Comparison of Naive Bayes and SVM Algorithms for Sentiment Analysis of PUBG Mobile on Google Play Store," *SISTEMASI (Jurnal Sistem Informasi)*, vol. 13, no. 6, Nov. 2024, <https://doi.org/10.32520/stmsi.v13i6.4814>.