

Analisis Sentimen Terhadap Hasil Evaluasi Kinerja Dosen Menggunakan Metode Naïve Bayes

DOI: <http://dx.doi.org/10.35889/jutisi.v15i3.3615>

Creative Commons License 4.0 (CC BY – NC)



Yudho Mulyono^{1*}, Andita Suci Pratiwi²

Teknik Informatika, STMIK Banjarbaru, Banjarbaru, Indonesia

*e-mail *Corresponding Author*: yudhomulyono8@gmail.com

Abstract

Quality assurance in higher education institutions relies heavily on systematic lecturer performance evaluation. Conducting this process manually, however, results in extended processing timelines and opens the door to subjective bias. To address this, the present study proposes a Gaussian Naïve Bayes-based approach for automated sentiment classification of student assessments derived from numerical Likert scale responses. Data were gathered from 575 students at STMIK Banjarbaru throughout the even semester of the 2024/2025 academic year, spanning 17 defined performance indicators. Preprocessing encompassed feature normalization, missing data removal, and rule-based sentiment assignment: scores averaging ≥ 4 were labeled positive, $= 3$ neutral, and ≤ 2 negative. Model training and testing followed a stratified 80:20 partition. Evaluation utilized a confusion matrix alongside a 5-fold cross-validation scheme. The trained model attained 98.26% test accuracy, with precision, recall, and F1-score averaging 98%. Cross-validation confirmed stability at a mean accuracy of 96.87%, establishing the viability of this approach for data-driven academic quality monitoring.

Keywords: *Sentiment Analysis; Lecturer Evaluation; Machine learning; Gaussian Naïve Bayes; Likert Scale*

Abstrak

Penjaminan mutu perguruan tinggi sangat bergantung pada mekanisme evaluasi kinerja dosen yang terstruktur dan berkelanjutan. Namun, pendekatan manual dalam mengolah data hasil survei mahasiswa acapkali memakan waktu cukup panjang dan berpotensi melahirkan penilaian yang tidak konsisten. Penelitian ini mengajukan solusi berbasis *Gaussian Naïve Bayes* untuk mengklasifikasikan sentimen mahasiswa secara otomatis dari data penilaian skala Likert numerik. Sebanyak 575 mahasiswa STMIK Banjarbaru pada semester genap tahun akademik 2024/2025 menjadi sumber data, mencakup 17 indikator performa dosen. Pra-pemrosesan meliputi normalisasi fitur, penghapusan data tidak lengkap, serta pemetaan sentimen berdasarkan aturan ambang batas: rata-rata ≥ 4 dikategorikan positif, $= 3$ netral, dan ≤ 2 negatif. Pembagian data dilakukan dengan proporsi 80:20 secara stratifikasi. Pengujian model mencakup *confusion matrix* dan validasi silang 5-fold. Model mencapai akurasi uji 98,26% dengan presisi, recall, dan F1-score rata-rata 98%. Validasi silang menghasilkan akurasi rata-rata 96,87%, membuktikan kestabilan dan kelayakan pendekatan ini untuk mendukung pengawasan mutu akademik berbasis data.

Kata Kunci: *Analisis Sentimen; Evaluasi Kinerja Dosen; Machine learning; Gaussian Naïve Bayes; Skala Likert*

1. Pendahuluan

Penjaminan mutu di perguruan tinggi menempatkan evaluasi kinerja dosen sebagai salah satu mekanisme paling strategis untuk menjaga dan meningkatkan standar mutu pembelajaran [1]. Do et al. [2] menegaskan bahwa penilaian terhadap dosen merupakan aktivitas inti institusi sekaligus komponen krusial dalam manajemen sumber daya manusia perguruan tinggi, mengingat kualitas pengajaran secara langsung mempengaruhi kapasitas intelektual lulusan dan daya saing institusi dalam jangka panjang. Dalam praktiknya, proses evaluasi dijalankan

melalui survei persepsi mahasiswa yang mengukur beragam dimensi kinerja pengajaran menggunakan instrumen skala Likert [3]. Walaupun pendekatan ini telah diadopsi secara luas, hambatan nyata masih dijumpai pada tahap analisis data, terutama saat volume responden yang besar harus ditangani secara manual [4], [5].

Pemrosesan data survei secara konvensional memiliki sejumlah keterbatasan. Pertama, agregasi dan interpretasi data secara manual memerlukan waktu pemrosesan yang panjang, umumnya berkisar antara dua hingga tiga minggu sejak pengumpulan data hingga publikasi hasil, sehingga menghambat responsivitas institusi dalam menangani isu pedagogis [6], [7]. Kedua, analisis manual rentan terhadap bias subjektivitas evaluator ketika mengubah pola data numerik menjadi kesimpulan kualitatif, yang berpotensi mengorbankan objektivitas evaluasi [8]. Ketiga, format penyajian hasil evaluasi konvensional yang hanya berupa statistik deskriptif dasar tidak mampu memberikan wawasan menyeluruh mengenai kategorisasi kualitas kinerja dosen secara holistik [8].

Perkembangan terkini di bidang *machine learning* dan analisis sentimen menawarkan solusi untuk mengotomasi analisis data evaluasi survei dengan objektivitas yang lebih tinggi dan waktu pemrosesan yang lebih singkat [9], [10]. Analisis sentimen berbasis *machine learning* memungkinkan klasifikasi otomatis persepsi mahasiswa ke dalam kategori sentimen yang telah ditentukan berdasarkan pola numerik dari hasil survei [11], [12]. Pendekatan ini telah terbukti efektif di berbagai domain aplikasi, termasuk analisis media sosial, evaluasi produk, dan analisis sentimen di sektor pendidikan [13].

Penelitian-penelitian terdahulu yang relevan dapat dikelompokkan menjadi dua aliran utama. Aliran pertama berfokus pada analisis sentimen berbasis data teks kualitatif: Kastrati et al. [14] melaksanakan kajian pemetaan sistematis yang membuktikan keunggulan pendekatan *deep learning* pada teks umpan balik mahasiswa, termasuk dalam ekosistem MOOC [15]; Birjali et al. [16] memetakan tantangan dan tren analisis sentimen dan mencatat bahwa tahapan pra-pemrosesan teks yang kompleks kerap menjadi hambatan efisiensi. Fatmawati et al. [17] memperlihatkan keandalan *Naïve Bayes* dalam mengklasifikasikan sentimen ulasan teks evaluasi dosen, sedangkan Sasmita et al. [18] mengembangkan sistem serupa pada institusi yang berbeda. Aliran kedua, yang masih jarang dieksplorasi, berkaitan dengan pemanfaatan data numerik skala Likert sebagai sumber informasi sentimen [16]. Berbeda dengan pendekatan berbasis teks, data numerik terstruktur menawarkan keunggulan dalam konsistensi pengukuran, kemudahan pra-pemrosesan, dan kesesuaian alami dengan asumsi distribusi probabilistik algoritma klasifikasi [17]. Di samping itu, penerapan klasifikasi sentimen tiga kategori pada data numerik terstruktur masih sangat jarang dilakukan, dan belum ada penelitian spesifik yang mengeksplorasi pendekatan ini di lingkungan STMIK Banjarbaru [18]. Kesenjangan inilah yang menjadi landasan kebaruan penelitian ini.

Atas dasar itu, penelitian ini memanfaatkan algoritma *Gaussian Naïve Bayes* untuk menjalankan klasifikasi sentimen dari data numerik survei evaluasi dosen. Pemilihan algoritma ini dilandasi beberapa pertimbangan teknis. Pertama, algoritma ini dirancang khusus untuk menangani data kontinu dengan asumsi distribusi normal yang sesuai dengan karakteristik data Likert yang teragregasi [19]. Kedua, kompleksitas komputasinya rendah dengan kompleksitas waktu $O(nd)$, memungkinkan eksekusi cepat pada perangkat dengan sumber daya terbatas [20]. Ketiga, algoritma ini memerlukan sedikit penyetulan hiperparameter namun tetap menghasilkan kinerja kompetitif pada *dataset* berukuran kecil hingga menengah [21]. Keempat, sifat probabilistik keluaran model memberikan transparansi dalam interpretasi hasil klasifikasi.

Kontribusi utama penelitian ini meliputi: (1) penerapan *Gaussian Naïve Bayes* untuk analisis sentimen berbasis data numerik skala Likert dalam konteks evaluasi kinerja dosen; (2) pengembangan *pipeline* pra-pemrosesan otomatis dengan penanganan nilai yang hilang dan normalisasi yang sistematis; (3) klasifikasi sentimen tiga kategori yang memberikan granularitas lebih tinggi dibandingkan klasifikasi biner; dan (4) evaluasi kinerja model komprehensif menggunakan berbagai metrik dan validasi silang *k-fold* [21].

2. Metodologi

Penelitian ini mengikuti alur metodologis yang sistematis, meliputi pengadaan data, pra-pemrosesan, penerapan algoritma, dan pengukuran kinerja model. Setiap tahap dirancang agar dapat direproduksi dan diverifikasi secara eksperimental sesuai dengan standar praktik terbaik dalam riset *machine learning*.

2.1. Dataset Penelitian

Sumber data utama penelitian ini adalah rekam hasil survei evaluasi kinerja dosen yang diselenggarakan pada semester genap tahun akademik 2024/2025 di STMIK Banjarbaru. Instrumen pengumpulan data berupa kuesioner daring terstruktur dengan skala Likert 5 tingkat, di mana skor 1 menandakan kinerja sangat kurang, skor 2 kurang, skor 3 cukup, skor 4 baik, dan skor 5 sangat baik. Setiap responden mengisi penilaian terhadap 17 butir indikator kinerja dosen, yang secara garis besar mencakup aspek penguasaan materi, strategi pengajaran, kemampuan komunikasi, ketertiban mengajar, serta cara mengevaluasi hasil belajar mahasiswa [22].

Secara keseluruhan, *dataset* memuat 575 entri responden, menghasilkan 9.775 titik data penilaian. Dari 28 butir yang semula terdapat dalam instrumen asli, hanya 17 butir yang lolos seleksi kualitas pascapra-pemrosesan. Ke-17 butir inilah yang selanjutnya difungsikan sebagai variabel prediktor dalam proses pemodelan. Struktur *dataset* terdiri atas kolom identitas mahasiswa, kolom identitas dosen yang dinilai, dan kolom nilai numerik dari masing-masing butir penilaian.

2.1.1. Sampel Data Penelitian

Untuk memberikan transparansi dan kemampuan reproduksi penelitian, Tabel 1 menyajikan sembilan contoh representatif dari keseluruhan *dataset* yang mencakup ketiga kelas sentimen. Setiap baris merepresentasikan satu responden dengan nilai penilaian pada 17 indikator kinerja dosen (I1–I17), rata-rata agregat, dan label sentimen yang ditetapkan.

Tabel 1 Sampel Data Penelitian (9 Instans Representatif)

No	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	Rata-rata	Sentimen
1	4	4	5	5	4	5	4	4	4	4	4	5	5	5	5	5	5	4,53	Positif
2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5,00	Positif
3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4,00	Positif
4	3	4	3	4	3	4	4	3	4	3	4	4	3	4	3	4	3	3,53	Netral
5	4	4	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	3,88	Netral
6	3	3	3	3	3	3	4	4	3	5	4	5	5	4	3	3	3	3,59	Netral
7	1	1	2	2	2	2	3	2	1	2	2	2	2	2	2	2	2	1,88	Negatif
8	3	3	2	2	2	2	1	3	1	3	2	3	3	2	3	2	2	2,29	Negatif
9	2	2	2	2	2	2	2	3	1	3	3	3	3	3	3	3	3	2,47	Negatif

2.2. Pra-Pemrosesan Data

Tahap ini bertujuan memastikan kualitas data sebelum digunakan untuk melatih model. Pertama, normalisasi nama kolom dilakukan untuk menstandarkan format penamaan variabel. Kedua, sistem secara otomatis mengidentifikasi kolom penilaian berbasis Likert dengan mendeteksi nilai numerik dalam rentang 1 hingga 5. Ketiga, seluruh kolom penilaian dikonversi ke tipe data numerik untuk memastikan kompatibilitas dengan algoritma *machine learning* yang digunakan.

Penanganan nilai yang hilang (*missing values*) diterapkan menggunakan strategi penghapusan baris data yang tidak lengkap, sesuai dengan praktik terbaik dalam ekstraksi fitur untuk analisis sentimen [23]. Pendekatan ini dipilih untuk menghindari potensi bias dari imputasi nilai pada data evaluasi kinerja. Selanjutnya, nilai agregat dihitung dengan mengambil rata-rata dari 17 aspek penilaian untuk setiap responden, kemudian dipetakan ke dalam tiga kelas sentimen: positif (rata-rata ≥ 4), netral (rata-rata = 3), dan negatif (rata-rata ≤ 2), konsisten dengan penelitian analisis sentimen pada domain lainnya [24].

2.3. Implementasi Algoritma Gaussian Naïve Bayes

Gaussian Naïve Bayes (GNB) merupakan varian dari keluarga *Naïve Bayes* yang dikembangkan untuk mengelola fitur kontinu dengan anggapan bahwa setiap fitur berdistribusi *Gaussian* di dalam masing-masing kelasnya. Fondasi algoritma ini adalah Teorema Bayes yang

dipasangkan dengan asumsi independensi kondisional antar fitur, diformulasikan dalam Persamaan 1:

$$P(C_k | X) = \frac{P(X | C_k) \cdot P(C_k)}{P(X)} \quad (1)$$

di mana $P(x_i | C_k)$ merepresentasikan probabilitas *posterior* kelas C_k yang diberikan vektor fitur X , $P(C_k | X)$ menyatakan *likelihood* fitur yang diberikan kelas, $P(C_k)$ merupakan probabilitas *prior* kelas, dan $P(X)$ berfungsi sebagai konstanta normalisasi. Untuk fitur numerik kontinu, *likelihood* dihitung menggunakan fungsi densitas probabilitas distribusi *Gaussian* (Persamaan 2):

$$P(x_i | C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (2)$$

di mana μ_k menyatakan rata-rata dan σ_k^2 merepresentasikan varians fitur x_i untuk kelas C_k , yang diestimasi dari data pelatihan menggunakan estimasi kemungkinan maksimum [20]. Implementasi teknis menggunakan *Python* bersama pustaka *Scikit-learn*, memanfaatkan kelas *GaussianNB* yang dilatih dengan metode *fit* terhadap matriks fitur X (17 indikator kinerja dosen) dan vektor label y (kelas sentimen hasil pelabelan ambang batas).

2.4. Pembagian Data dan Strategi Evaluasi

Pemisahan *dataset* menjadi porsi latih dan uji mengikuti skema *stratified split* dengan perbandingan 80:20. Stratifikasi diterapkan agar distribusi kelas sentimen terjaga proporsional pada kedua subset, meminimalkan risiko bias dalam pelatihan maupun pengujian model [24]. Porsi latih sebesar 460 instans dimanfaatkan untuk mengestimasi parameter μ_k dan σ_k^2 melalui *maximum likelihood estimation*. Porsi uji sebesar 115 instans digunakan untuk mengukur kemampuan generalisasi model pada data yang belum pernah diproses sebelumnya.

Evaluasi dilaksanakan melalui dua mekanisme yang saling melengkapi: (1) pengukuran metrik pada data uji menggunakan *confusion matrix* sebagai landasan; dan (2) validasi silang *5-fold* yang membagi keseluruhan *dataset* menjadi lima bagian, lalu secara bergantian menggunakan satu bagian sebagai data uji dan empat bagian lainnya sebagai data latih [4]. Nilai rata-rata akurasi beserta standar deviasinya dari kelima *fold* digunakan sebagai indikator kestabilan model.

2.5. Metrik Evaluasi

Kinerja model diukur menggunakan empat metrik standar klasifikasi multi-kelas yang diturunkan dari *confusion matrix*. Akurasi mencerminkan proporsi prediksi yang tepat dari keseluruhan prediksi, dinyatakan dalam Persamaan 3:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Presisi mengukur ketepatan model dalam memprediksi kelas tertentu sebagai positif, dinyatakan pada Persamaan 4:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall atau sensitivitas mengukur kemampuan model menemukan seluruh instans positif yang sesungguhnya, ditunjukkan pada Persamaan 5:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1-score menggabungkan presisi dan *recall* menjadi satu angka melalui rata-rata harmonik, sebagaimana tertulis pada Persamaan 6:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Dalam persamaan-persamaan tersebut, TP (*True Positive*) merujuk pada instans positif yang benar-benar teridentifikasi, TN (*True Negative*) pada instans negatif yang benar teridentifikasi, FP (*False Positive*) pada instans negatif yang salah diprediksi positif, dan FN (*False Negative*) pada instans positif yang luput dari deteksi.

3. Hasil dan Pembahasan

Bagian ini memaparkan luaran eksperimen penerapan *Gaussian Naïve Bayes* pada data evaluasi kinerja dosen, dilengkapi dengan diskusi mendalam atas tiap-tiap temuan. Cakupan

analisis meliputi profil statistik *dataset* pascapra-pemrosesan, kinerja model pada data pengujian, validasi kestabilan model lintas partisi data, interpretasi pola kesalahan prediksi, serta perbandingan hasil dengan publikasi terdahulu yang relevan.

3.1. Karakteristik Dataset

Setelah tahapan pra-pemrosesan yang komprehensif, *dataset* penelitian terdiri dari 575 instans dengan 17 fitur numerik. Distribusi sentimen menunjukkan bahwa 423 instans (73,57%) diklasifikasikan sebagai sentimen positif, 130 instans (22,61%) sebagai sentimen negatif, dan 22 instans (3,83%) sebagai sentimen netral. Distribusi ini mengungkapkan adanya ketidakseimbangan kelas yang signifikan dengan dominasi kelas positif, yang merupakan karakteristik umum dalam survei evaluasi kinerja dosen [24].

Statistik deskriptif menunjukkan rata-rata nilai agregat berada di angka 3,72 dengan standar deviasi 0,58, serta rentang data dari 1,82 hingga 4,94. Persebaran nilai yang cenderung mengelompok pada sisi tengah hingga atas mencerminkan iklim penilaian yang umumnya positif di kalangan mahasiswa terhadap kinerja dosen. *Dataset* yang demikian memiliki variasi yang memadai untuk melatih dan menguji model *Gaussian Naïve Bayes* secara representatif.

3.2. Kinerja Klasifikasi pada Data Uji

Pengujian model *Gaussian Naïve Bayes* terhadap 115 instans data uji menghasilkan akurasi keseluruhan sebesar 98,26%, yang berarti 113 dari total 115 data berhasil diklasifikasikan ke dalam kelas yang tepat. Capaian ini memperlihatkan bahwa model mampu menggeneralisasi pengetahuan yang diperoleh dari data latih ke data baru secara memuaskan, setara dengan temuan pada studi-studi prediksi kinerja mahasiswa berbasis *data mining*.

Analisis yang lebih rinci menggunakan presisi, *recall*, dan *F1-score* per kelas menunjukkan kinerja yang kompetitif dibandingkan pendekatan *deep learning* untuk analisis sentimen [25]. Rincian nilai metrik per kategori sentimen disajikan pada Tabel 2.

Tabel 2 Hasil Klasifikasi Model *Gaussian Naïve Bayes*

Kategori Sentimen	Presisi (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
Negatif	96,00	96,00	96,00
Netral	100,00	100,00	100,00
Positif	99,00	99,00	99,00
Rata-rata Makro	98,00	98,00	98,00
Rata-rata Tertimbang	98,00	98,00	98,00
Akurasi Keseluruhan	98,26%		

Secara keseluruhan, model menunjukkan kinerja yang sangat baik dan seimbang di semua kelas. Kategori positif dan negatif masing-masing mencapai nilai presisi, *recall*, dan *F1-score* yang tinggi (96–99%), mengindikasikan kemampuan model dalam mengidentifikasi sentimen ekstrem secara akurat. Kategori netral mencapai presisi dan *recall* sempurna sebesar 100%, membuktikan bahwa meskipun jumlah instans kelas ini relatif kecil, model tetap mampu membedakannya dengan tepat. Nilai rata-rata makro dan tertimbang untuk ketiga metrik masing-masing mencapai 98%, menandakan kinerja model tidak didominasi kelas mayoritas dan relatif bebas dari bias klasifikasi.

Meskipun tingkat kesalahan sangat rendah, analisis terhadap dua instans yang salah diklasifikasikan mengungkap pola yang informatif. Instans pertama memiliki rata-rata agregat tepat 4,00 dengan nilai yang sepenuhnya seragam pada angka 4, kecuali satu indikator yang bernilai 3 (I16) dan satu yang bernilai 5 (I11). Kondisi ini menciptakan ambiguitas batas keputusan antara kelas Positif dan Netral, karena nilai rata-rata berada persis di ambang klasifikasi. Model memprediksinya sebagai Netral, sementara label aktualnya adalah Positif. Instans kedua memiliki rata-rata 2,59 dengan sebagian besar nilai pada angka 2–3, namun terdapat beberapa indikator bernilai 4 (I10, I11, I12, I13, I14) yang menciptakan pola skor heterogen. Model memprediksinya sebagai Negatif, padahal label aktualnya adalah Netral.

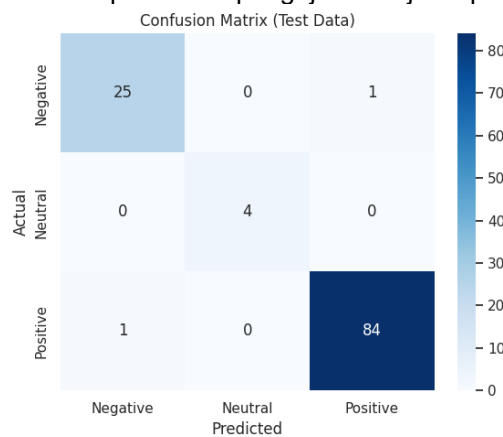
Kedua kasus ini secara konsisten menunjukkan bahwa kesalahan prediksi terjadi pada instans yang nilai rata-ratanya berada di dekat ambang batas antarkelas (sekitar 2,5 dan 4,0), yakni zona abu-abu di mana distribusi probabilitas antarkelas saling berdekatan. Tabel 3 menyajikan rincian kedua instans yang salah diklasifikasikan tersebut.

Tabel 3 Instans yang Salah Diklasifikasikan oleh Model

No	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	Rata-rata	Label Aktual	Prediksi
1	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	3	4	4,00	Positif	Netral
2	2	3	2	2	2	2	2	3	2	4	3	3	3	3	3	2	3	2,59	Netral	Negatif

3.3. Analisis Confusion Matrix

Confusion matrix digunakan untuk menganalisis pola kesalahan klasifikasi yang dihasilkan model *Gaussian Naïve Bayes* pada data pengujian. Analisis ini memberikan wawasan mendetail mengenai hubungan antara kelas aktual dan kelas yang diprediksi. Visualisasi *confusion matrix* dari hasil klasifikasi pada data pengujian disajikan pada Gambar 1.



Gambar 1 Confusion matrix Klasifikasi Gaussian Naïve Bayes pada Data Uji

Berdasarkan Gambar 1, model menunjukkan akurasi klasifikasi yang sangat tinggi di semua kategori sentimen. Pada kelas Negatif, 25 dari 26 instans teridentifikasi benar; 1 instans tersalur ke prediksi Positif. Kelas Netral mencapai klasifikasi sempurna dengan seluruh 4 instans terdeteksi tanpa kesalahan. Pada kelas Positif, 84 dari 85 instans terklasifikasi tepat, sementara 1 instans tercatat sebagai Negatif. Yang menarik, seluruh kesalahan prediksi hanya terjadi di antara kelas Positif dan Negatif, tanpa ada satu pun misklasifikasi yang melibatkan kelas Netral. Hal ini menandakan model memiliki kapasitas yang baik untuk mengenali batas antara kategori ambivalen tersebut meski representasinya dalam *dataset* sangat terbatas.

3.4. Validasi Ketahanan Model Menggunakan Validasi Silang

Ketahanan model diuji menggunakan skema validasi silang *5-fold* pada keseluruhan *dataset*. Prosedur ini mempartisi data menjadi lima subset; dalam setiap iterasi, satu subset digunakan sebagai data uji sementara empat subset yang tersisa berfungsi sebagai data latih. Hasilnya dapat dilihat pada Tabel 4.

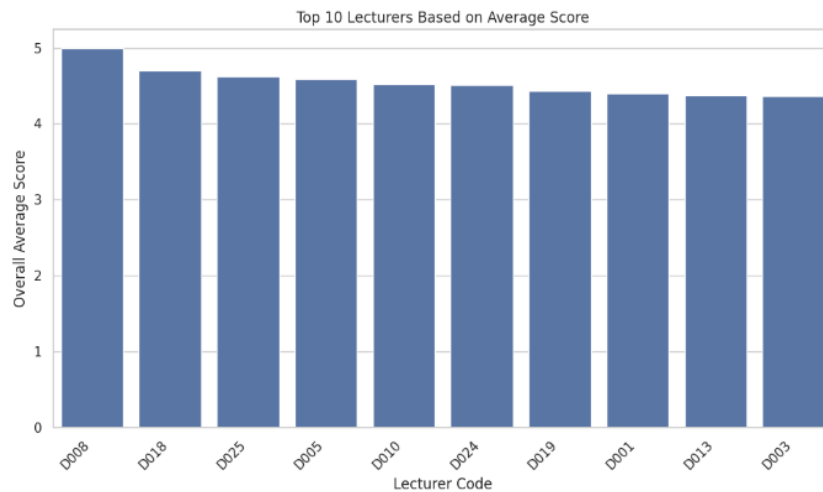
Tabel 4 Hasil Validasi Silang Lima-Fold

Fold	Akurasi (%)
Fold 1	98,26
Fold 2	97,39
Fold 3	95,65
Fold 4	95,65
Fold 5	97,39
Rata-rata ± SD	96,87 ± 1,15

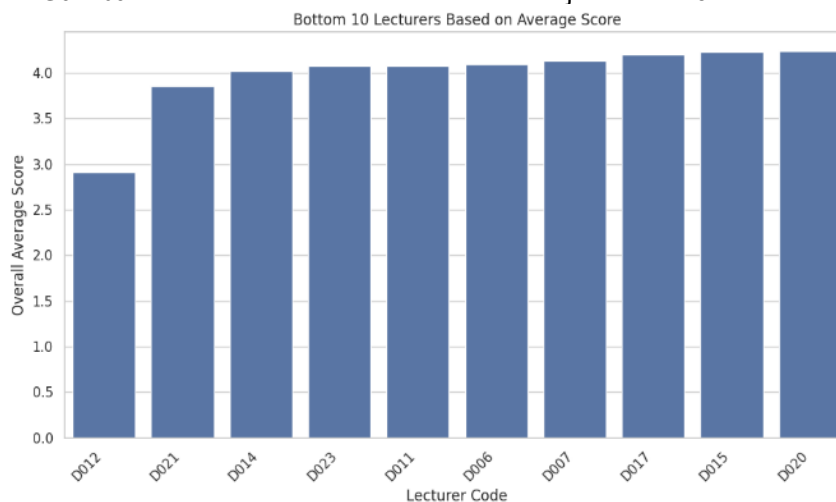
Berdasarkan Tabel 4, model *Gaussian Naïve Bayes* mencapai akurasi rata-rata 96,87% dengan standar deviasi 1,15%. Nilai standar deviasi yang relatif kecil mengindikasikan kinerja model yang stabil dan konsisten pada berbagai pembagian data, membuktikan tidak adanya indikasi *overfitting*. Fluktuasi akurasi antar *fold* yang tidak signifikan mencerminkan kemampuan generalisasi yang memadai di berbagai komposisi data, sejalan dengan implementasi *machine learning* pada konteks prediksi kinerja pendidikan [20]. Hasil ini memperkuat temuan dari evaluasi *holdout*, memberikan keyakinan lebih terhadap keandalan model.

3.5. Visualisasi Kinerja Dosen Berdasarkan Nilai Rata-rata

Visualisasi kinerja dosen disajikan untuk memberikan wawasan komparatif yang lebih intuitif mengenai hasil evaluasi berdasarkan nilai rata-rata penilaian mahasiswa. Gambar 2 menampilkan sepuluh dosen dengan nilai rata-rata keseluruhan tertinggi dan terendah. Identitas dosen ditampilkan dalam format kode dosen untuk menjaga anonimitas dan memenuhi pertimbangan etika penelitian.



Gambar 2 Nilai Rata-rata Keseluruhan Kinerja Dosen 10 Terbaik



Gambar 3 Nilai Rata-rata Keseluruhan Kinerja Dosen 10 Terendah

Berdasarkan Gambar 2 dan 3, tampak adanya variasi dalam nilai rata-rata kinerja dosen, meskipun seluruh dosen dalam kelompok sepuluh teratas memiliki skor di atas rata-rata keseluruhan *dataset*. Hal ini menunjukkan bahwa persepsi mahasiswa terhadap kinerja dosen secara umum berada dalam kategori positif dengan perbedaan antar dosen yang tidak terlalu ekstrem. Pemanfaatan visualisasi berbasis kode dosen memungkinkan hasil penelitian digunakan sebagai bahan evaluasi internal institusi tanpa mengungkap identitas pribadi dosen, sesuai dengan praktik terbaik dalam penyampaian umpan balik mahasiswa yang efektif [26].

Secara keseluruhan, visualisasi ini melengkapi hasil analisis kuantitatif sebelumnya dengan menyediakan representasi grafis yang mudah dipahami oleh pihak manajemen akademik.

3.6. Pembahasan

Hasil penelitian ini secara konsisten memperkuat dan memperluas temuan-temuan pada sejumlah studi terdahulu yang berkaitan. Fatmawati et al. [17] melaporkan akurasi 87,5% menggunakan *Naïve Bayes* pada data teks evaluasi dosen, sedangkan penelitian ini mencapai 98,26% dengan data numerik skala Likert sebuah peningkatan yang signifikan yang mengonfirmasi bahwa data numerik terstruktur lebih mudah dimodelkan oleh algoritma probabilistik dibandingkan data teks yang memerlukan pra-pemrosesan kompleks. Temuan ini sejalan pula dengan Daqiqil et al. [7] yang menyimpulkan bahwa pendekatan berbasis *transformer* unggul pada data teks namun dengan biaya komputasi jauh lebih tinggi; penelitian ini menunjukkan bahwa untuk data numerik, algoritma yang lebih sederhana seperti *Gaussian Naïve Bayes* sudah cukup menghasilkan performa kompetitif bahkan superior. Dalam konteks yang lebih luas, Grimalt-Alvaro dan Usart [10] mengidentifikasi bahwa analisis sentimen untuk penilaian formatif di pendidikan tinggi masih menjadi area yang berkembang; penelitian ini memberikan bukti empiris bahwa pendekatan berbasis data numerik merupakan alternatif yang layak dan efisien.

Dari perspektif kontribusi ilmiah, penelitian ini memberikan setidaknya tiga sumbangan konseptual bagi pengembangan bidang *machine learning* terapan di pendidikan tinggi. Pertama, penelitian ini membuktikan bahwa data penilaian skala Likert yang selama ini diperlakukan sebatas statistik deskriptif dapat dimanfaatkan secara langsung sebagai fitur untuk klasifikasi sentimen tanpa konversi ke teks, menyederhanakan *pipeline* analisis secara substansial. Kedua, skema pelabelan sentimen tiga kelas (Positif, Netral, Negatif) yang dikembangkan berdasarkan ambang batas nilai rata-rata terbukti menghasilkan granularitas diagnostik yang lebih tinggi dibandingkan pendekatan biner, sehingga memberikan informasi yang lebih actionable bagi manajemen institusi dalam menindaklanjuti hasil evaluasi. Ketiga, akurasi validasi silang *5-fold* sebesar 96,87% dengan standar deviasi 1,15% mengonfirmasi stabilitas dan generalisabilitas model, menjadikannya kandidat yang kuat untuk implementasi nyata sebagai modul pendukung keputusan dalam sistem manajemen mutu akademik. Secara keseluruhan, penelitian ini mengisi kesenjangan yang diidentifikasi oleh Jazuli et al. [16] tentang kurangnya eksplorasi data numerik Likert dalam konteks analisis sentimen evaluasi pendidikan, sekaligus memperluas cakupan penelitian Sasmita et al. [18] ke konteks institusi yang berbeda dengan metodologi yang lebih terstandar.

4. Simpulan

Penelitian ini berhasil membuktikan kelayakan *Gaussian Naïve Bayes* sebagai solusi otomatisasi analisis sentimen pada data evaluasi kinerja dosen berbasis skala Likert numerik. Model yang dibangun mampu mengklasifikasikan respons mahasiswa ke dalam tiga kategori sentimen Positif, Netral, dan Negatif dengan akurasi 98,26% pada data uji dan akurasi rata-rata 96,87% melalui validasi silang *5-fold*. Kinerja ini mengkonfirmasi efektivitas algoritma dalam memproses data evaluasi yang bersifat numerik dan terstruktur tanpa memerlukan pra-pemrosesan teks yang kompleks. Telaah *confusion matrix* mengungkap bahwa kesalahan prediksi sangat minim dan hanya terjadi pada instans yang nilai rata-ratanya berada di dekat ambang batas antarkelas, yakni zona ambiguitas di mana distribusi probabilitas antarkelas saling berdekatan. Visualisasi berbasis kode dosen memungkinkan hasil riset dimanfaatkan sebagai bahan audit mutu internal tanpa mengorbankan privasi individu, menjadikan sistem ini siap untuk diintegrasikan sebagai modul pendukung keputusan dalam platform akademik institusi.

Pengembangan ke depan dapat diarahkan pada penggabungan data numerik dengan data teks terbuka untuk meraih pemahaman sentimen yang lebih komprehensif. Perbandingan dengan metode ensemble atau algoritma klasifikasi lanjutan juga berpotensi mendorong capaian akurasi yang lebih tinggi. Dari sisi implementasi, sistem ini dapat diintegrasikan ke dalam platform akademik sebagai modul pendukung keputusan evaluatif, sehingga perguruan tinggi dapat merespons dinamika mutu pengajaran secara lebih cepat dan berbasis bukti.

Daftar Referensi

- [1] A. Bhowmik, Noorhuzaimi Mohd Noor, M. S. U. Miah, and D. Karmaker, "Aspect-based Sentiment Analysis Model for Evaluating Teachers' Performance from Students' Feedback," *AIUB Journal of Science and Engineering (AJSE)*, vol. 22, no. 3, pp. 287–294, Dec. 2023, doi: 10.53799/ajse.v22i3.921.
- [2] Q. H. Do, V. T. Tran, and T. T. Tran, "Evaluating lecturer performance in Vietnam: An application of fuzzy AHP and fuzzy TOPSIS methods," *Heliyon*, vol. 10, no. 11, p. e30772, Jun. 2024, doi: 10.1016/j.heliyon.2024.e30772.
- [3] F. Quansah, A. Cobbinah, K. Asamoah-Gyimah, and J. E. Hagan, "Validity of student evaluation of teaching in higher education: a systematic review," *Front. Educ. (Lausanne)*, vol. 9, Feb. 2024, p. no. 1329734, doi: 10.3389/educ.2024.1329734.
- [4] J. L. A. Altuna, "Analysis of the use of digital technologies in education through sentiment analysis," *Educ. Inf. Technol. (Dordr)*, vol. 27, no. 7, pp. 10067–10089, Sep. 2022, doi: 10.1007/s10639-022-11012-9.
- [5] A. F. Meghji, N. A. Mahoto, Y. Asiri, H. Alshahrani, A. Sulaiman, and A. Shaikh, "Early detection of student degree-level academic performance using educational *data mining*," *PeerJ Comput. Sci.*, vol. 9, p. e1294, Mar. 2023, doi: 10.7717/peerj-cs.1294.
- [6] A. Göçen, M. M. Ibrahim, and A. U. I. Khan, "Public attitudes toward higher education using sentiment analysis and topic modeling," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 83, Nov. 2024, doi: 10.1007/s44163-024-00195-4.
- [7] I. Daqiqil ID, H. Saputra, S. Syamsudhuha, R. Kurniawan, and Y. Andriyani, "Sentiment analysis of student evaluation feedback using transformer-based language models," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 2, p. 1127, Nov. 2024, doi: 10.11591/ijeecs.v36.i2.pp1127-1139.
- [8] S. Röhl, H. Bijlsma, and M. Schwichow, "Can feedback from students to teachers improve different dimensions of teaching quality in primary and secondary education? A hierarchical meta-analysis," *Educ. Assess. Eval. Account.*, vol. 37, no. 1, pp. 35–71, Feb. 2025, doi: 10.1007/s11092-024-09450-9.
- [9] M. Fargues, S. Kadry, I. A. Lawal, S. Yassine, and H. T. Rauf, "Automated Analysis of Open-Ended Students' Feedback Using Sentiment, Emotion, and Cognition Classifications," *Applied Sciences*, vol. 13, no. 4, p. 2061, Feb. 2023, doi: 10.3390/app13042061.
- [10] C. Grimalt-Álvaro and M. Usart, "Sentiment analysis for formative assessment in higher education: a systematic literature review," *J. Comput. High. Educ.*, vol. 36, no. 3, pp. 647–682, Dec. 2024, doi: 10.1007/s12528-023-09370-5.
- [11] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational *data mining* to predict students' academic performance: A survey study," *Educ. Inf. Technol. (Dordr)*, vol. 28, no. 1, pp. 905–971, Jan. 2023, doi: 10.1007/s10639-022-11152-y.
- [12] P. S. Ghatora, S. E. Hosseini, S. Pervez, M. J. Iqbal, and N. Shaukat, "Sentiment Analysis of Product Reviews Using *Machine learning* and Pre-Trained LLM," *Big Data and Cognitive Computing*, vol. 8, no. 12, p. 199, Dec. 2024, doi: 10.3390/bdcc8120199.
- [13] P. Ren, L. Yang, and F. Luo, "Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis," *Educ. Inf. Technol. (Dordr)*, vol. 28, no. 1, pp. 797–814, Jan. 2023, doi: 10.1007/s10639-022-11151-z.
- [14] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani, "Sentiment Analysis of Students' Feedback with NLP and *Deep learning*: A Systematic Mapping Study," *Applied Sciences*, vol. 11, no. 9, p. 3986, Apr. 2021, doi: 10.3390/app11093986.
- [15] H. S. S. N. R. N. A. Hasan, "Sentiment analysis of students' feedback on MOOCs: A systematic literature review," *IEEE Access*, vol. 10, pp. 45660–45682, 2022, doi: 10.1109/ACCESS.2022.3168341.
- [16] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl. Based. Syst.*, vol. 226, p. 107134, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.
- [17] S. Fatmawati, M. R. A. Prasetya, S. E. Prastya, and S. P. Cipta, "Penerapan *Naïve Bayes* dalam Analisis Sentimen untuk Evaluasi Kinerja Pengajaran Dosen," *Jurnal SAINTEKOM*, vol. 15, no. 1, pp. 82–93, Mar. 2025, doi: 10.33020/saintekom.v15i1.765.

- [18] S. P. G. A. D. D. G. H. Sasmita, "Sentiment analysis system for lecturer performance evaluation using Naïve Bayes," *Jurnal Sains dan Teknologi*, vol. 11, no. 2, pp. 451–462, Okt. 2022, doi: 10.23887/jstundiksha.v11i2.49071.
- [19] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Applied Sciences*, vol. 15, no. 1, p. 172, Dec. 2024, doi: 10.3390/app15010172.
- [20] S. Nehe, "Sentiment analysis of public opinion using Naïve Bayes," *Jutisi: Jurnal Teknologi Informasi dan Sistem Informasi*, vol. 10, no. 1, pp. 1–11, Jan. 2024, doi: 10.35957/jutis.v10i1.6212.
- [21] M. Yağcı, "Educational *data mining*: prediction of students' academic performance using *machine learning* algorithms," *Smart Learning Environments*, vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
- [22] F. Thabtah, "Machine learning for sentiment classification: Recent advances," *Expert Syst. Appl.*, vol. 182, Art. no. 115333, Nov. 2021, doi: 10.1016/j.eswa.2021.115333.
- [23] Y. K. M. N. S. O. Alginahi, "Sentiment analysis techniques for educational data: A review," *IEEE Access*, vol. 9, pp. 61932–61946, 2021, doi: 10.1109/ACCESS.2021.3073700.
- [24] N. A. Semaary, W. Ahmed, K. Amin, P. Pławiak, and M. Hammad, "Enhancing *machine learning*-based sentiment analysis through feature extraction techniques," *PLoS One*, vol. 19, no. 2, p. e0294968, Feb. 2024, doi: 10.1371/journal.pone.0294968.
- [25] M. H. Bin Roslan and C. J. Chen, "Predicting students' performance in English and Mathematics using *data mining* techniques," *Educ. Inf. Technol. (Dordr.)*, vol. 28, no. 2, pp. 1427–1453, Feb. 2023, doi: 10.1007/s10639-022-11259-2.
- [26] S. Al-Hadhrami, T. Vinko, T. Al-Hadhrami, F. Saeed, and S. N. Qasem, "Deep learning-based method for sentiment analysis for patients' drug reviews," *PeerJ Comput. Sci.*, vol. 10, p. e1976, Apr. 2024, doi: 10.7717/peerj-cs.1976.