

Analisis Sentimen Kepuasan Mahasiswa terhadap Fasilitas Kampus ITSB Menggunakan Algoritma SVM, Naïve Bayes, dan Random Forest

Amril Mutoi Siregar^{1*}, Gunawan Witjaksono², Dara Ratu Avyana³, Ade Kurniawan⁴

^{1,2,3}Informatika, Fakultas Digital Desain dan Bisnis, Institut Teknologi Sains Bandung, Bekasi, Indonesia

⁴Sains Data, Fakultas Digital Desain dan Bisnis, Institut Teknologi Sains Bandung, Bekasi, Indonesia

^{1*}amril.s@itsb.ac.id, ²gunawan.witjaksono@itsb.ac.id, ³ratuavyana@gmail.com, ⁴ade.k@itsb.ac.id

Abstrak

Infrastruktur merupakan fasilitas pendidikan yang memainkan peran penting dalam meningkatkan kemandirian dan proses pembelajaran di lembaga pendidikan tinggi. Penilaian tradisional kualitas fasilitas sering membutuhkan durasi yang panjang dan seringkali tidak efisien dalam menghasilkan wawasan objektif. Oleh karena itu, tujuan penyelidikan ini adalah untuk mengevaluasi sentimen kepuasan mahasiswa mengenai kualitas fasilitas kampus di Institut Teknologi Sains Bandung (ITSB) dengan melakukan analisis komparatif terhadap kinerja tiga algoritma pembelajaran yang diawasi: *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Random Forest*. Data penelitian diperoleh melalui distribusi kuesioner kepada mahasiswa, yang kemudian dikategorikan menurut dimensi spesifik fasilitas. Metodologi penelitian mencakup tahapan seperti prapemrosesan data teks, ekstraksi fitur menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF), bersama dengan analisis statistik deskriptif dan inferensial untuk menjelaskan distribusi data. Untuk mengurangi ketidakseimbangan data, *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan di setiap model klasifikasi. Temuan menunjukkan bahwa pemanfaatan SMOTE secara signifikan meningkatkan kinerja klasifikasi semua algoritma yang diperiksa. Algoritma SVM mencapai tingkat akurasi tertinggi 78,57%, melampaui *Naïve Bayes* dan *Random Forest*. Hasil ini menyiratkan bahwa analisis sentimen yang didasarkan pada metodologi pembelajaran mesin dapat berfungsi sebagai mekanisme yang objektif dan efisien untuk menilai kualitas fasilitas kampus, sehingga membantu lembaga pendidikan dalam menetapkan prioritas untuk peningkatan layanan berdasarkan umpan balik mahasiswa dan sebagai dasar pengambilan keputusan pimpinan perguruan tinggi.

Kata kunci: analisis sentimen, fasilitas pembelajaran, *machine learning*, SMOTE, TF-IDF.

1. Pendahuluan

Fasilitas pembelajaran merupakan sebuah komponen yang mawadahi kelancaran proses belajar demi mencapai tujuan pendidikan yang berkualitas (Jumiati et al., 2024). Keberhasilan pengajar dalam menyampaikan ilmu serta efektivitas mahasiswa dalam menerima pendidikan sangat bergantung pada keberadaan fasilitas pembelajaran yang memadai. Fasilitas pembelajaran memengaruhi tingkat kepuasan mahasiswa selama menempuh studi di perguruan tinggi yang mencakup tiga aspek, yaitu: kecukupan, kemudahan, dan kualitas (Julianto et al., 2024).

Kesadaran akan pentingnya peningkatan kualitas fasilitas Institut Teknologi Sains Bandung (ITSB) telah mendorong pihak kampus untuk menyediakan kotak saran dan kritik sebagai metode penilaian mahasiswa terhadap fasilitas. Namun, metode seperti itu dinilai tidak efisien dalam melakukan pengumpulan data dan pembacaan manual sehingga sulit mendapatkan timbal balik dengan cepat. Oleh karena itu, penelitian ini menerapkan sistem kuesioner berbasis pendekatan analisis sentimen yang dirancang otoma-

tis mengelompokkan teks berdasarkan opini (Gunawan et al., 2018).

Analisis sentimen adalah teknik yang digunakan untuk mengekstrak pendapat agar mudah dipahami dan mengelola data teks untuk mengidentifikasi sentimen yang terkandung dalam suatu pendapat. Tujuan dari analisis sentimen adalah untuk menentukan suatu teks memiliki sentimen positif, netral, atau negatif (Putri Ajie & Wardhana, 2025). Penelitian menggunakan tiga algoritma klasifikasi, yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Random Forest* lalu dilakukan perbandingan tingkat akurasi dalam menangani data yang tidak seimbang menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE).

Sejumlah penelitian sebelumnya telah menggunakan analisis sentimen sebagai kerangka metodologis untuk menilai layanan pendidikan dan fasilitas kampus. Penelitian yang dilakukan oleh Honainah menggunakan algoritma *Naïve Bayes* dalam hubungannya dengan penimbangan TF-IDF dan pendekatan berbasis leksikon untuk meneliti persepsi

siswa mengenai fasilitas kampus, mencapai tingkat akurasi 85% melalui metodologi berbasis leksikon. (Honainah, 2026) menggunakan *Support Vector Machine* (SVM) bersama TF-IDF untuk analisis sentimen layanan akademik MyUMSIDA, menghasilkan akurasi 86,92% dan dengan demikian memperkuat kemanjuran SVM dalam klasifikasi teks opini siswa. (Angga et al., 2025) melibatkan analisis komparatif berbagai algoritma klasifikasi sentimen, menggabungkan teknik SMOTE dan TF-IDF dalam evaluasi tinjauan layanan publik, menunjukkan bahwa pendekatan SMOTE secara signifikan meningkatkan kinerja model pada kumpulan data yang tidak seimbang. (Budaya & Suniantara, 2024) melakukan analisis komparatif SVM dan *Naïve Bayes* menggunakan TF-IDF dan SMOTE dalam konteks tinjauan aplikasi digital, di mana SVM menunjukkan kinerja yang unggul setelah penerapan teknik SMOTE (Wahid et al., 2025).

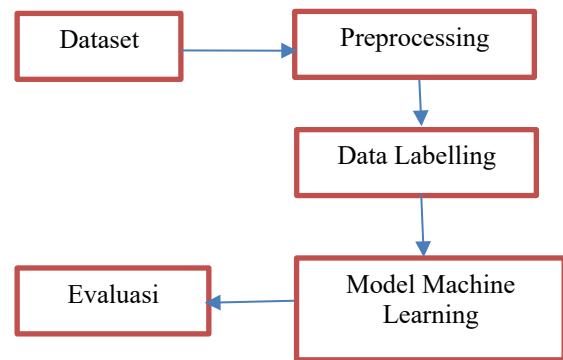
Berdasarkan tinjauan pustaka terhadap penelitian analisis sentimen di lingkungan akademik sebelumnya cenderung berfokus pada ulasan publik secara umum, kinerja dosen, atau hanya menggunakan satu metode klasifikasi tertentu saja (Farizi et al., 2025)(Usman et al., 2024). Meskipun kombinasi algoritma klasifikasi seperti SVM, *Naïve Bayes*, dan *Random Forest* telah banyak diimplementasikan pada berbagai objek komersial, namun belum ditemukan penelitian yang secara spesifik melakukan evaluasi mendalam terhadap fasilitas pembelajaran (Pratama & Jatnika, 2026).

Selain itu, teknik SMOTE yang telah terbukti efektif dalam menangani ketidakseimbangan data pada objek komersial (Dewi & Mailoa, 2023), penggunaannya masih terbatas pada *dataset* kepuasan fasilitas mahasiswa di perguruan tinggi. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan mengintegrasikan analisis aspek fasilitas, komparasi tiga algoritma, dan optimasi SMOTE.

Penelitian ini untuk mengevaluasi dan membandingkan algoritma klasifikasi yang paling efektif sebagai model analisis sentiment kepuasan mahasiswa terhadap fasilitas kampus. Menganalisis pengaruh SMOTE terhadap peningkatan kinerja model dalam mengklasifikasi data sentimen. Menghasilkan informasi berbasis data yang akurat dari sentimen mahasiswa guna optimalisasi fasilitas pembelajaran dan menciptakan lingkungan akademik yang lebih suportif.

2. Metode

Penelitian menggunakan metode kuantitatif berbasis data teks untuk menganalisis sentimen mahasiswa terhadap kepuasan fasilitas pembelajaran Institut Teknologi Sains Bandung (ITSB). Dalam penelitian ini, analisis statistik dari data teks menjadi representasi numerik melalui ekstraksi fitur TF-IDF, kemudian pelatihan model dengan algoritma SVM,



Gambar 1. Alur Penelitian

Naïve Bayes, dan *Random Forest*, serta perbandingan SMOTE terhadap akurasi ketiga model. Gambar 1 merupakan alur penelitian yang digunakan untuk menyelesaikan penelitian ini.

2.1 Dataset

Pengumpulan data melalui penyebaran kuesioner digital dengan teknik *simple random sampling* dalam skala yang terbatas hanya kepada mahasiswa aktif ITSB dari berbagai program studi. Proses pengumpulan data dilakukan selama satu bulan delapan hari, dimulai sejak 2 November 2025 hingga 10 Desember 2025.

Data kuesioner mengambil jawaban dengan tipe skala numerik dan teks berupa isian komentar. Penelitian ini berfokus pada data teks karena dinilai mampu memberikan informasi lebih spesifik yang didasarkan dari opini setiap responden sehingga pola kepuasan mahasiswa dapat dianalisis secara mendalam menggunakan ekstraksi fitur dan *supervised learning*. Dari seleksi tersebut, terkumpul sebanyak 220 data mentah yang akan masuk ke tahap *preprocessing* untuk kemudian dilakukan pembagian antara data latih dan data uji.

2.2 Preprocessing

Preprocessing melibatkan seluruh data mentah yang terpilih untuk menjadi data bersih. Tujuan *preprocessing* adalah membersihkan data dari elemen yang tidak relevan dan mengumpulkan informasi yang lebih akurat sehingga menambah efektivitas algoritma yang diterapkan (Gunawan et al., 2018). Langkah-langkah *preprocessing* sebagai berikut: **Case Folding**: Tahap mengubah seluruh alfabet pada teks komentar menjadi huruf kecil untuk menghindari duplikasi analisis terhadap kata yang sama dengan kapitalisasi berbeda (Sari et al., 2025). **Penghapusan tanda baca angka, dan simbol**: Pembersihan dari karakter non alfabet untuk mengurangi karakter yang tidak dibutuhkan (Sari et al., 2025). **TOKENISASI**: Pemisahan kata dari kalimat menjadi bentuk tunggal. Tahap ini mempermudah pembobotan dan ekstraksi makna. **Stopwords Removal**: Mengha-

pus konjungsi dan kata kata lain yang tidak memengaruhi kepentingan teks. pada proses ini peneliti melakukan persyaratan pada daftar kata yang dihapus, yaitu mempertahankan kata negasi seperti 'tidak', 'kurang', 'belum', dan sejenisnya karena dinilai penting dalam deteksi sentimen dan menghindari perubahan makna (misalnya: kata "kurang bersih" hanya terdeteksi "bersih" karena *stopword* menghapus negasi). **Stemming**: Pemisahan setiap kata dari imbuhan (normalisasi), menjadikan setiap kata berbentuk dasar dan mudah untuk dilakukan analisis (Fitriyana et al., 2023).

2.3 Data Labelling

Pelabelan data menggunakan pendekatan *Aspect-Based Sentiment Analysis* (ABSA). Pendekatan ini digunakan untuk mencegah hilangnya informasi krusial dengan membedah sentimen mahasiswa pada setiap aspek fasilitas secara spesifik (Alfaridzi et al., 2025). Dengan menghubungkan label sentimen langsung ke aspek tertentu, sistem dapat memberikan nilai pertimbangan yang lebih akurat bagi pihak institusi dalam melakukan tinjauan mendalam terhadap fasilitas tersebut.

Proses pelabelan mempertimbangkan aspek fasilitas serta polaritas kata (positif, negatif, atau netral). Peneliti menyiapkan daftar fasilitas sebagai aspek penilaian yang berisi 7 kategori fasilitas kampus meliputi: WiFi, AC, ruang kelas, komputer, lab komputer, e-Campuz, dan lab penelitian.

2.4 Model Machine Learning

Exploratory Data Analysis (EDA) adalah proses menampilkan hasil identifikasi pola dan sebaran data yang bertujuan mengeksplorasi dan memperkuat analisis. EDA mencakup penanganan *missing value*, statistik deskriptif terhadap distribusi label sentimen serta pengklasifikasian aspek, dan statistik inferensial (uji hipotesis *Chi-Square* beserta hasilnya) (Radhi et al., 2022).

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik ekstraksi fitur data yang memberikan bobot pada suatu kata berdasarkan frekuensi kemunculannya. Frekuensi kemunculan data dalam komentar menunjukkan kepentingan kata tersebut. Ekstraksi fitur bertujuan untuk mengoptimalkan kinerja model dengan hanya melatih kata yang bersifat informatif (Atmajaya et al., 2023).

Supervised learning adalah metode *machine learning* yang melatih komputer menggunakan *dataset* dengan input dan output yang diketahui (telah berlabel) (Hokijuliandy et al., 2023). SVM adalah algoritma klasifikasi yang banyak diterapkan pada penelitian berbasis data teks karena dapat mengenali pola data secara efektif (Atmajaya et al., 2023). *Random Forest* melibatkan metode *decision tree* untuk klasifikasi dalam *Supervised Learning*. Algoritma

ini mempunyai kinerja yang baik dalam melakukan klasifikasi dengan tingkat error yang cukup rendah (Sun et al., 2024).

SMOTE digunakan sebagai teknik mengatasi ketidakseimbangan kelas agar distribusi data lebih merata. Pada penelitian ini, SMOTE bekerja dengan cara menyintensis sampel dari kelas minoritas di antara label sentimen (positif, netral, atau negatif) untuk meningkatkan performa model klasifikasi (Triandes Sinaga et al., 2025).

2.5 Evaluasi

Evaluasi setiap model menggunakan *confusion matrix* akan memberikan gambaran kinerja model saat melakukan prediksi kepada data uji yang bertujuan menilai kinerja model dalam melakukan klasifikasi sentimen.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

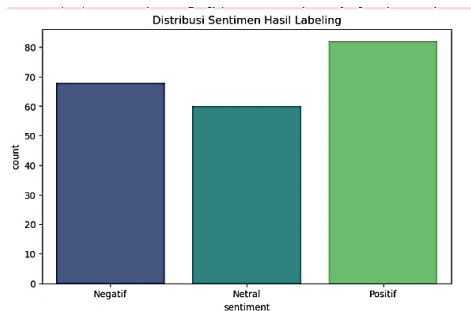
$$F1-Score = 2 \times \frac{Recall \times Precision}{Precision + Recall} \quad (4)$$

Matriks evaluasi kinerja model terdiri dari *accuracy* (mengukur keefektifan model), *precision* (rasio jumlah *true positive* dan jumlah total prediksi positif), *recall* (rasio jumlah *true positive* dan jumlah total aktual positif), dan *f1-Score* (keseimbangan antara *precision* dan *recall*) (Hokijuliandy et al., 2023).

3. Hasil dan Pembahasan

Data yang terkumpul akan melalui tahap *pre-processing* kemudian diberikan label sentimen yang mempertimbangkan aspek fasilitas serta kata bermuatan positif, negatif, dan netral. Pembersihan data dilakukan dengan menghapus kolom kosong, menghasilkan 210 data valid dari total awal 220. Kemudian dilakukan *melting* setiap kolom sehingga menjadi tabel yang menghasilkan 1 kolom 210 baris untuk dilanjutkan ke tahap EDA.

Hasil distribusi data yang telah diberikan label menunjukkan dominasi sentimen positif sebesar 39.05% diikuti dengan sentimen negatif 32.38% dan netral 28.57% dengan keseluruhan data setelah *pre-processing* adalah 210 data. Sentimen positif menunjukkan kebutuhan mahasiswa akan ketersediaan fasilitas sudah terpenuhi, namun perlu dilakukan peninjauan terkait jumlah selisih 6.67 poin persentase terhadap sentimen negatif yang dapat menginterpretasikan bahwa mahasiswa menilai fasilitas yang tersedia belum sepenuhnya dapat dimaksimalkan dalam aktivitas pembelajaran. Gambar 2 merupakan visualisasi distribusi *dataset* dan Gambar 3 visualisasi sentiment dengan WordCloud.



Gambar 2. Distribusi Sentimen Menggunakan BarChart



Gambar 3. Distribusi Sentimen Menggunakan WordCloud

3.1 Analisis Deskriptif

Pada analisis deskriptif, setiap label sentimen direpresentasi menjadi data numerik untuk menghitung nilai statistik. Pembagian data sebagai berikut: label 0 negatif, label 1 positif dan 2 netral.

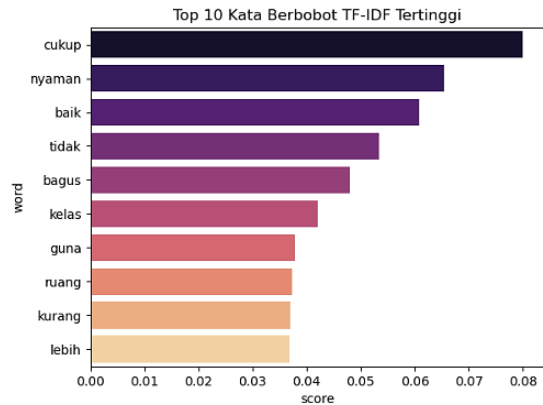
Tabel distribusi rata-rata penilaian terhadap fasilitas menunjukkan AC sebagai aspek yang memiliki skor mean terendah (0.33). Sementara itu, fasilitas kelas adalah aspek dengan komentar terbanyak dan skor mean tertinggi (1.04). Tabel 1 Visualisasi Sentimen.

Tabel 1. Distribusi Sentimen per Fasilitas

Sentimen fasilitas	Negatif	Netral	Positif	Total	Mean skor
Ac	6	0	3	9	0.33
komputer	8	4	4	16	0.75
Wifi	7	4	9	20	0.85
Lab riset	3	3	3	9	1.00
Lab Kom	10	10	22	42	1.00
Kelas	31	36	35	102	1.00
Lainnya	3	3	6	12	1.00

Berdasarkan hasil analisis deskriptif, terdapat beberapa fasilitas dengan skor mean hampir mendekati 1.0 tetapi pada data aktual mendapatkan banyak sentimen negatif, contohnya adalah WIFI. Hal ini diperkirakan terjadi karena proses labelling data difokuskan pada kata kunci positif, netral, dan negatif dengan tidak berfokus pada kalimat saran atau permintaan yang menunjukkan kebutuhan akan perbaikan kualitas fasilitas.

Aspek lab komputer dan komputer yang memiliki gap sebesar 0.25 pada skor mean menunjukkan mahasiswa cukup puas menggunakan lab komputer sebagai kelas yang mendukung kegiatan pembelajaran namun terdapat keluhan teknis pada beberapa komputer, seperti adanya *mouse* dan *keyboard* yang rusak dan komputer lambat memproses.



Gambar 4. Top 10 Bobot TF-IDF Tertinggi

3.3 Analisis Ekstraksi Fitur

Hasil TF-IDF menunjukkan kata “cukup” memiliki bobot tertinggi dalam keseluruhan data. Nilai TF-IDF ini sudah memenuhi aspek yang dibutuhkan namun belum mencapai persepsi emosional mahasiswa. Munculnya kata “nyaman” dan “baik” pada urutan teratas setelahnya memperkuat validitas distribusi sentimen positif. Gambar 4 merupakan peringkat top 10 kata kata dalam *dataset*.

Berikut komentar yang menyoroti aspek keterseediaan fasilitas telah terpenuhi, tetapi aspek kepuasan atau ketepatan fungsionalitas masih butuh peningkatan: “Fasilitas sudah baik, mungkin perlu perawatan rutin pada peralatan yang ada agar tetap berfungsi optimal dan awet digunakan”. “Menurut saya, tata letak kursi di lab-kom kurang nyaman untuk pembelajaran berbasis papan tulis karena pandangan terhalang dan ruang gerak terbatas. Untuk materi yang fokus pada penjelasan, ruang kelas biasa lebih efektif.

3.4 Analisis Model

Splitting data dilakukan dengan pembagian data latih 80% dan data uji 20%. Penerapan SMOTE pada ketiga algoritma klasifikasi menunjukkan adanya peningkatan akurasi dalam mengenali data sentimen mahasiswa. Tabel 2 merupakan hasil algoritma sebelum melakukan rekayasa SMOTE.

Tabel 2. Performa Algoritma sebelum SMOTE

Algoritma	Akurasi %	Presisi %	Recal %	F1-Score %
SVM	76.19	82.22	76.19	75.76
Naive Bayes	59.52	42.94	59.52	49.74
Random Forest	61.90	64.62	61.90	58.99

Dapat dilihat dari hasil pelatihan model dengan SMOTE, algoritma SVM, *Naive Bayes*, dan *Random Forest* mengalami kenaikan yang sama sebesar 2.381%. Hal ini dapat dipengaruhi karena jumlah data uji yang jumlahnya cukup sedikit, yaitu 20% dari total 210 data. Implementasi SMOTE pada ketiga algoritma menunjukkan peningkatan akurasi dalam

Tabel 3. performa algoritma Sesudah SMOTE

Algoritma	Akurasi %	Presisi %	Recal %	F1-Score %
SVM	78.57	83.25	78.57	78.49
Naive Bayes	61.90	46.46	61.90	61.90
Random Forest	64.29	63.42	64.29	64.29

mengenali pola sentimen mahasiswa. SVM mencatat akurasi tertinggi sebesar 78.57% dan akurasi terendah. Tabel 3 merupakan hasil algoritma sesudah melakukan rekayasa SMOTE.

Naive Bayes sebesar 52.24% karena kesulitan mengenali keseluruhan label netral. Hal itu bisa terjadi karena distribusi kata pada label netral beririsan dengan label lainnya. Meskipun akurasi meningkat, terdapat penurunan *precision* pada algoritma *Random Forest* dari semula 64.62% menjadi 63.42%.

Berdasarkan hasil analisis kinerja model pada penelitian ini, algoritma SVM menjadi yang terbaik diantara kedua algoritma lainnya dalam mengenali label sentimen diikuti dengan *Random Forest* dan *Naive Bayes*.

Penerapan SMOTE digunakan agar model tidak hanya pintar menebak kelas mayoritas saja, tetapi juga akurat dalam mengenali kelas minoritas yang ditunjukkan pada kinerja SVM dan *Random Forest*. SMOTE pada *Naive Bayes* tidak berhasil menebak kelas netral tetapi bekerja meningkatkan *False Positive*. Tabel 4 merupakan hasil peningkatan kinerja algoritma.

Tabel 4. Peningkatan performa algoritma

Algoritma	Akurasi Tanpa SMOTE	Akurasi Dengan SMOTE	Peningkatan (%)
SVM	0.7619	0.7857	2.381
Naive Bayes	0.5952	0.6190	2.381
Random Forest	0.6190	0.6429	2.381

Hasil ini juga dibuktikan dalam penelitiannya menyatakan bahwa SMOTE cukup baik meningkatkan akurasi dan presisi pada *machine learning* menggunakan algoritma SVM. Meskipun pada beberapa kasus SMOTE bisa menurunkan nilai statistik tersebut, namun teknik ini efektif menangani ketidakseimbangan kelas dalam klasifikasi sentimen.

Hasil penelitian ini konsisten dengan menegaskan bahwa Synthetic Minority Over-sampling Technique (SMOTE) secara signifikan meningkatkan kemanjuran klasifikasi pada kumpulan data yang tidak seimbang dengan menghasilkan contoh sintesis dalam kelas minoritas, sehingga memfasilitasi peningkatan generalisasi model (Honainah, 2026). Selanjutnya, penelitian yang dilakukan oleh Douzas dan Bacao menunjukkan bahwa integrasi SMOTE dengan metodologi klasifikasi, seperti Support Vector Machines (SVM), mampu secara nyata meningkatkan akurasi dan metrik ingatan dalam contoh klasifikasi teks dan sentimen.

4. Kesimpulan

Penelitian ini menunjukkan bahwa pemanfaatan analisis sentimen berbasis pembelajaran mesin efektif dalam menilai tingkat kepuasan siswa dengan fasilitas kampus di ITS B. Penerapan teknik TF-IDF untuk ekstraksi fitur, bersamaan dengan penerapan SMOTE, telah dibuktikan untuk meningkatkan kemanjuran model klasifikasi. Di antara tiga algoritma yang dievaluasi, *Support Vector Machine* (SVM) menunjukkan kinerja yang unggul, mencapai tingkat akurasi tertinggi 78,57%. Temuan ini mendukung pernyataan bahwa metodologi berbasis data dapat menggantikan teknik evaluasi tradisional yang kurang efektif, sementara secara bersamaan menawarkan landasan yang lebih tidak memihak bagi institusi dalam memprioritaskan peningkatan kualitas fasilitas pendidikan.

Penelitian masa depan, analisis sentimen dapat diperluas dengan menggunakan rangkaian data yang lebih luas dan beragam, sehingga memungkinkan model untuk menghasilkan prediksi yang lebih tepat dan representatif. Investigasi selanjutnya juga dapat menggabungkan teknik pembelajaran mendalam seperti Memori Jangka Panjang (LSTM), Representasi Encoder Bidirectional dari Transformers (BERT), atau model berbasis transformator lainnya untuk menambah kapasitas model untuk memahami konteks bahasa alami.

Daftar Pustaka

- Alfaridzi, D. D., Pudjiantoro, T. H., & Santikarama, I. (2025): *Implementation of aspect-based sentiment analysis for outdoor gear product reviews on the Tokopedia e-commerce platform (case study: Eiger Adventure)*, Holistic Science, 5(2), 173–184. <https://doi.org/10.56495/hs.v5i2.1230>
- Angga, A. W. S., Hamzah, H. S., & Rohman, R. D. (2025): *Analisis sentimen layanan perwalian mahasiswa UMSIDA menggunakan metode Support Vector Machine (SVM)*, JSAI (Journal Scientific and Applied Informatics), 8(1), 93–99. <https://doi.org/10.36085/jsai.v8i1.7387>
- Atmajaya, D., Febrianti, A., & Darwis, H. (2023): *Metode SVM dan Naive Bayes untuk analisis sentimen ChatGPT di Twitter*, The Indonesian Journal of Computer Science, 12(4). <https://doi.org/10.33022/ijcs.v12i4.3341>
- Budaya, I. G. B. A., & Suniantara, I. K. P. (2024): *Comparison of sentiment analysis algorithms with SMOTE oversampling and TF-IDF implementation on Google reviews for public health centers*, MALCOM: Indonesian Journal of Machine Learning and Computer Science, 4(3), 1077–1086. <https://doi.org/10.57152/malcom.v4i3.1459>
- Dewi, T. A., & Mailoa, E. (2023): *Perbandingan implementasi metode SMOTE pada algoritma Sup-*

- port Vector Machine (SVM) dalam analisis sentimen opini masyarakat tentang Mixue, Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 4(3), 849–855. <https://doi.org/10.35870/jimik.v4i3.289>
- Farizi, G. H., Voutama, A., & Hakim, A. D. (2025): *Analisis sentimen mahasiswa terhadap ketersediaan fasilitas di Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang*, Jurnal Informatika dan Teknik Elektro Terapan, 13(3). <https://doi.org/10.23960/jitet.v13i3.6544>
- Fitriyana, V., Hakim, L., Novitasari, D. C. R., & Asyhar, A. H. (2023): *Analisis sentimen ulasan aplikasi Jamsostek Mobile menggunakan metode Support Vector Machine*, Jurnal Buana Informatika, 14(1), 40–49. <https://doi.org/10.24002/jbi.v14i01.6909>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018): *Sistem analisis sentimen pada ulasan produk menggunakan metode Naive Bayes*, Jurnal Edukasi dan Penelitian Informatika (JEPIN), 4(2), 113. <https://doi.org/10.26418/jp.v4i2.27526>
- Hokijuliandy, E., Napitupulu, H., & Firdaniza, F. (2023): *Analisis sentimen menggunakan metode klasifikasi Support Vector Machine (SVM) dan seleksi fitur Chi-Square*, SisInfo: Jurnal Sistem Informasi dan Informatika, 5(2), 40–49. <https://doi.org/10.37278/sisinfo.v5i2.670>
- Honainah. (2026): *Penerapan algoritma Naive Bayes dengan pembobotan TF-IDF dan pendekatan lexicon untuk analisis sentimen opini mahasiswa terhadap fasilitas kampus*, JSAI (Journal Scientific and Applied Informatics), 9(1), 71–80. <https://doi.org/10.36085/jsai.v9i1.9836>
- Julianto, A., Siregar, N. S., & Hadi, K. A. (2024): *Analisis kepuasan mahasiswa ditinjau dari kinerja dosen dan fasilitas pembelajaran di STIT Al-Quraniyah Manna Bengkulu*, Muallimun: Jurnal Kajian Pendidikan dan Keguruan, 4(2), 119–140. <https://doi.org/10.23971/muallimun.v4i2.8935>
- Jumiati, S., Riyanto, Y., Izzati, U. A., Khamidi, A., Hariyati, N., & Rifqi, A. (2024): *Pengaruh motivasi belajar dan fasilitas pembelajaran terhadap prestasi akademik siswa*, Journal of Education Research, 5(2), 2371–2378. <https://doi.org/10.37985/jer.v5i2.1116>
- Pratama, D. H., & Jatnika, I. (2026): *Pengembangan model analisis sentimen aplikasi fintech dengan algoritma klasifikasi dan TF-IDF*, Jurnal Sains Sistem Informasi, 4(1), 10–17. <https://doi.org/10.31602/jssi.v4i1.21234>
- Putri Ajie, Z., & Wardhana, A. (2025): *Analisis sentimen negatif publik pada tagar #KaburAjaDulu di media sosial “X.”*, Jurnal Audiens, 6(3), 430–442. <https://doi.org/10.18196/jas.v6i3.608>
- Radhi, M., Amalia, A., Sitompul, D. R. H., Sinurat, S. H., & Indra, E. (2022): *Analisis big data dengan metode exploratory data analysis (EDA) dan metode visualisasi menggunakan Jupyter Notebook*, Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA), 4(2), 23–27. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475>
- Sari, R. Y., Nazuwatussyah'diyah, N., & Mardhiyatna, M. (2025): *Analisis sentimen digital publik terhadap perguruan tinggi negeri menggunakan Google Review: Studi kasus ITERA*, Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence), 5(2), 298–305. <https://doi.org/10.55382/jurnalpustakaai.v5i2.1163>
- Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024): *An improved random forest based on the classification accuracy and correlation measurement of decision trees*, Expert Systems with Applications, 237, 121549. <https://doi.org/10.1016/j.eswa.2023.121549>
- Triandes Sinaga, T., Sirait, K. B., Pangaribuan, J. J., Barus, O. P., & Romindo. (2025): *Analisis kualitas wine menggunakan machine learning dengan pendekatan SMOTE dan seleksi fitur*, INSOLOGI: Jurnal Sains dan Teknologi, 4(3), 656–668. <https://doi.org/10.55123/insologi.v4i3.5436>
- Usman, N. S. B., Rumana, N. A., Putra, D. H., & Fannya, P. (2024): *Analisis sentimen pada kolom saran indeks kepuasan mahasiswa terhadap kinerja dosen*, Jurnal Manajemen Pendidikan, 9(2), 241–252. <https://doi.org/10.34125/jmp.v9i2.344>
- Wahid, Y. A., Sanatang, & Andayani, D. D. (2025): *Performance comparison of SVM and Naive Bayes for Indonesian-language sentiment analysis on Free Fire reviews using TF-IDF and SMOTE*, Journal of Embedded Systems, Security and Intelligent Systems, 674–689. <https://doi.org/10.59562/jessi.v6i4.10818>