

PEMANFAATAN NATURAL LANGUAGE PROCESSING UNTUK ANALISIS SENTIMEN TERHADAP METODE PEMBELAJARAN DARING: TINJAUAN PUSTAKA

Amelia Ratu Ferisa¹⁾, Rahayu Putri Aulia Simatupang²⁾, Divalya Zahra³⁾, Naura Rahmi Nadira Malau⁴⁾, Ardia Shabrina Ramadani Lubis⁵⁾, Nopri Herayanti Br. Singarimbun⁶⁾, Elsa Sabrina⁷⁾

^{1,2,3,4,5,6,7)} Program Studi Pendidikan Teknologi Informatika dan Komputer,

Fakultas Teknik, Universitas Negeri Medan

Jl. Willem Iskandar Psr V, Medan Estate, Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara

E-mail: ¹ameliaratu135@gmail.com, ²rahayuputriauliasimatupang12@gmail.com,

³divalyazahra20@gmail.com, ⁴naurarahmi98@gmail.com, ⁵ardiashabrina970@gmail.com,

⁶novriphu860@gmail.com, ⁷elsasabrina@unimed.ac.id

ABSTRAK

Pembelajaran daring menghasilkan teks opini dalam jumlah besar, mulai dari cuitan, ulasan aplikasi, hingga umpan balik pada *learning management system* (LMS), sehingga analisis manual tidak lagi memadai. Artikel ini menyajikan tinjauan pustaka atas 36 studi terbitan 2021–2026 yang memanfaatkan *Natural Language Processing* (NLP) untuk analisis sentimen terhadap pembelajaran daring. Studi diseleksi dari 51 studi unik pada tabel ekstraksi data berdasarkan relevansi, kelengkapan metode, dan kualitas pelaporan, lalu disintesis untuk menjawab tiga pertanyaan penelitian mengenai sumber data, metode NLP, serta kecenderungan sentimen. Hasil menunjukkan bahwa 19 dari 32 studi empiris menggunakan algoritma pembelajaran mesin klasik, terutama *Naive Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN), sedangkan model *deep learning* dan *transformer* makin sering dipakai pada studi terbaru dengan akurasi tertinggi 97,12%. Tidak ada satu algoritma yang unggul secara konsisten karena kinerja sangat bergantung pada data dan skema pelabelan. Polaritas sentimen bervariasi menurut konteks: ulasan platform belajar cenderung positif, sedangkan opini tentang perkuliahan daring dan *hybrid* di media sosial lebih banyak negatif atau netral. Tantangan utama meliputi ketidakseimbangan kelas, sulitnya mengenali kelas netral, bahasa informal, dan ketidakseragaman pelabelan. Hasil tinjauan ini menunjukkan bahwa pemilihan metode NLP perlu disesuaikan dengan karakteristik data dan skema pelabelan yang digunakan.

Kata Kunci: analisis sentimen, *natural language processing*, pembelajaran daring, pembelajaran mesin, tinjauan pustaka

ABSTRACT

Online learning generates large volumes of opinion text, ranging from tweets and app reviews to feedback on learning management systems (LMS), which makes manual analysis impractical. This article reviews 36 studies published between 2021 and 2026 that apply Natural Language Processing (NLP) to sentiment analysis of online learning. The studies were screened from 51 unique candidates using relevance, methodological completeness, and reporting quality, and were synthesized around three research questions on data sources, NLP methods, and sentiment tendencies. Of the 32 empirical studies, 19 used classical machine learning, mainly Naive Bayes, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), while deep learning and transformer models appear more often in recent work, with a highest reported accuracy of 97.12%. No single algorithm consistently outperforms the others because performance depends heavily on the data and labeling scheme. Sentiment polarity varies by context: reviews of learning platforms tend to be positive, whereas social media opinions on online and hybrid lectures are more often negative or neutral. Key challenges include class imbalance, difficulty recognizing the neutral class, informal language, and inconsistent labeling. The findings indicate that the choice of NLP methods should be adapted to the characteristics of the data and the labeling scheme used.

Keywords: sentiment analysis, natural language processing, online learning, machine learning, literature review

1. PENDAHULUAN

Transformasi digital pendidikan menjadikan pembelajaran daring bagian yang melekat pada penyelenggaraan pendidikan tinggi, terutama sejak pandemi COVID-19 memaksa institusi beralih ke perkuliahan jarak jauh [1], [2]. Peralihan tersebut melahirkan banjir data teks berupa cuitan di media sosial, ulasan aplikasi, komentar pada forum diskusi, dan umpan balik mahasiswa yang memuat persepsi serta emosi pengguna terhadap metode pembelajaran daring [3], [4]. Menganalisis data sebanyak itu secara manual tidak efisien dan rentan terhadap subjektivitas, sehingga diperlukan pendekatan otomatis yang mampu memahami bahasa alami [5], [6]. *Natural Language Processing* (NLP) menyediakan rangkaian teknik untuk tugas tersebut, mulai dari klasifikasi berbasis pembelajaran mesin dan pendekatan berbasis leksikon hingga model *deep learning* dan *transformer* [4], [6]. Berbagai penelitian di Indonesia maupun mancanegara telah menerapkan NLP pada data Twitter, ulasan Google Play Store, dan sistem manajemen pembelajaran, tetapi hasilnya tersebar dan memakai algoritma, dataset, serta metrik yang berbeda-beda [7] - [9].

Untuk memahami konteks data tersebut, pembelajaran daring dipahami sebagai proses belajar mengajar yang dilaksanakan tanpa tatap muka langsung antara pengajar dan peserta didik, dengan memanfaatkan teknologi informasi dan komunikasi sebagai perantara interaksi [10]. Sistem ini memungkinkan fleksibilitas waktu dan tempat bagi penggunaanya, sehingga tetap bertahan dan berkembang jauh setelah masa pandemi berlalu, terutama melalui platform bimbingan belajar, kursus daring, maupun aplikasi konferensi video yang digunakan institusi pendidikan [10].

Kualitas pembelajaran daring di perguruan tinggi dipengaruhi oleh berbagai faktor, mulai dari kestabilan koneksi internet, kemudahan penggunaan aplikasi, hingga interaksi antara pengajar dan peserta didik [10]. Ketidakpuasan pengguna terhadap salah satu faktor tersebut kerap disampaikan secara terbuka melalui ulasan aplikasi atau media sosial, sehingga diperlukan

pendekatan komputasional yang mampu mengolah opini dalam jumlah besar tersebut secara otomatis dan efisien [10].

Kebutuhan untuk mengolah opini pengguna secara otomatis inilah yang mendorong pemanfaatan NLP, sebuah cabang kecerdasan buatan yang memungkinkan komputer memahami, menafsirkan, dan menghasilkan bahasa manusia secara komputasional [11]. NLP menggabungkan berbagai disiplin, termasuk linguistik komputasional, pembelajaran mesin, dan pembelajaran mendalam, untuk mengubah teks yang tidak terstruktur menjadi representasi yang dapat diproses oleh algoritma [11].

Penerapan NLP mencakup beragam tugas, seperti penerjemahan mesin, peringkasan teks, deteksi spam, hingga penjawaban pertanyaan otomatis [11]. Salah satu cabang penerapan NLP yang paling relevan untuk mengukur persepsi pengguna terhadap suatu layanan atau produk adalah analisis sentimen, yang secara khusus berfokus pada identifikasi opini positif, negatif, atau netral dalam sebuah teks [12].

Analisis sentimen adalah proses komputasional untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini, sikap, serta emosi yang terkandung dalam teks terhadap suatu entitas, baik berupa produk, layanan, maupun kebijakan tertentu [12]. Bidang ini terus berkembang seiring meningkatnya volume opini yang dihasilkan pengguna internet melalui ulasan aplikasi, media sosial, dan forum daring [12].

Terdapat tiga pendekatan utama dalam analisis sentimen, yaitu pendekatan berbasis leksikon yang mengandalkan kamus kata bersentimen, pendekatan pembelajaran mesin seperti *Naive Bayes* dan *Support Vector Machine*, serta pendekatan pembelajaran mendalam seperti BERT dan turunannya yang mampu menangkap konteks kalimat secara lebih kaya [12]. Ketiga pendekatan tersebut memiliki kelebihan dan keterbatasan masing-masing, terutama dalam menangani data yang tidak seimbang antar kelas sentimen serta ekspresi bahasa informal yang umum ditemukan pada ulasan pengguna [12],

sehingga pemetaan yang sistematis terhadap studi-studi terdahulu menjadi penting dilakukan.

Pemetaan yang sistematis terhadap studi-studi terdahulu tersebut memerlukan metode yang terstruktur agar hasilnya dapat dipertanggungjawabkan secara ilmiah [13]. *Systematic Literature Review* (SLR) adalah metode penelitian yang bertujuan mengidentifikasi, menilai, dan mensintesis seluruh bukti empiris yang relevan dengan pertanyaan penelitian tertentu melalui prosedur yang eksplisit dan dapat direplikasi, alih-alih mengandalkan penilaian subjektif peneliti semata [13].

Salah satu panduan pelaporan SLR yang paling banyak diadopsi adalah *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA), yang menyediakan alur kerja transparan mulai dari identifikasi jumlah studi yang ditemukan, penyaringan berdasarkan judul dan abstrak, penilaian kelayakan teks penuh, hingga penentuan studi yang benar-benar disertakan dalam tinjauan akhir [14].

Adapun penelitian terdahulu mencakup beragam platform pembelajaran daring, mulai dari Ruangguru dan Duolingo, hingga Coursera dan Zoom Cloud Meetings, dengan metode klasifikasi yang bervariasi dari algoritma klasik seperti *Naive Bayes* dan *Support Vector Machine* hingga model berbasis *transformer* seperti IndoBERT dan ALBERT [15], [16]. Sebagian studi melaporkan bahwa model berbasis pembelajaran mendalam cenderung menghasilkan akurasi yang lebih tinggi dibandingkan metode klasik, terutama dalam menangani kelas sentimen netral yang sering kali menjadi kelas paling sulit diklasifikasi [15], [17], meskipun belum ada algoritma tunggal yang unggul di semua kondisi karena kinerjanya sangat bergantung pada data dan skema pelabelan.

Di sisi lain, sebagian studi juga mengombinasikan beberapa algoritma sekaligus melalui pendekatan *ensemble* untuk meningkatkan performa klasifikasi secara keseluruhan [18]. Meskipun demikian, sebagian besar penelitian yang ditemukan masih berfokus pada evaluasi satu aplikasi secara terpisah, sehingga perbandingan sistematis antar-platform

dan antar-metode dalam satu kerangka tinjauan masih terbatas.

Tinjauan literatur yang ada pun biasanya hanya berfokus pada umpan balik mahasiswa di MOOC atau pada teknik NLP untuk umpan balik instruktur [3], [5], sehingga belum ada pemetaan yang menyatukan beragam algoritma, dataset, dan metrik tersebut. Oleh karena itu, artikel ini meninjau 36 studi terseleksi tahun 2021 sampai 2026 yang memanfaatkan NLP untuk analisis sentimen terhadap pembelajaran daring. Tinjauan ini memetakan sumber data dan platform yang diteliti, membandingkan metode NLP beserta kinerjanya, serta merangkum kecenderungan sentimen dan tantangan yang muncul. Kontribusi artikel ini adalah peta bukti yang terstruktur dan rekomendasi arah penelitian mendatang bagi peneliti maupun pengelola pembelajaran daring.

2. METODE PENELITIAN

2.1 Desain dan Pertanyaan Penelitian

Penelitian ini merupakan tinjauan pustaka dengan pendekatan sintesis naratif dan deskriptif kuantitatif. Pendekatan ini dipilih karena penelitian terdahulu tentang analisis sentimen pada pembelajaran daring tersebar pada banyak platform, dataset, dan metode, sehingga diperlukan pemetaan yang terstruktur dan dapat ditelusuri ulang. Sintesis naratif digunakan karena studi yang ditinjau memakai dataset, pembagian data, dan metrik evaluasi yang berbeda, sehingga hasilnya tidak dapat digabungkan secara statistik. Prosedur tinjauan mengikuti prinsip tinjauan pustaka sistematis [13]. Alur penyaringan studi disajikan dengan gaya diagram alir PRISMA sebagaimana lazim dipakai pada tinjauan sistematis di bidang ini [3]. Tiga pertanyaan penelitian (*research question*, RQ) dirumuskan sebagai berikut.

- RQ1: Platform, sumber data, dan skala data apa yang digunakan dalam analisis sentimen terhadap pembelajaran daring?
- RQ2: Metode dan algoritma NLP apa yang dipakai, dan bagaimana kinerjanya?
- RQ3: Kecenderungan sentimen apa yang ditemukan, dan tantangan apa yang dilaporkan?

Ketiga pertanyaan penelitian tersebut dirumuskan untuk menjawab kesenjangan yang dibahas pada bagian pendahuluan. RQ1 menelaah

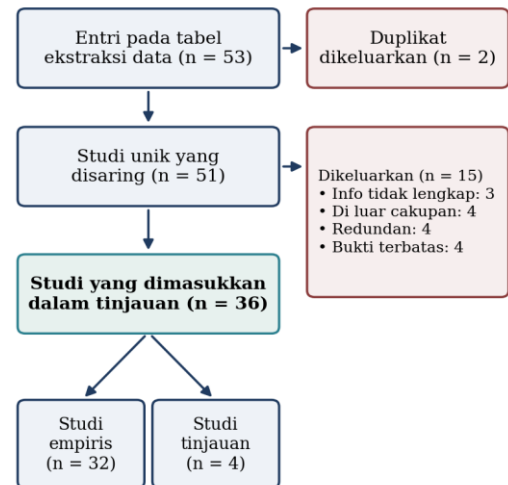
asal dan skala data opini, RQ2 menelaah teknik NLP yang dipakai beserta kinerjanya, dan RQ3 menelaah arah sentimen serta kendala yang dilaporkan. Jawaban RQ1, RQ2, dan RQ3 masing-masing disajikan pada subbagian 3.2, 3.3, dan 3.4 di bagian Hasil dan Diskusi

2.2 Seleksi Studi

Kandidat studi dihimpun dalam tabel ekstraksi data yang memuat 53 entri, dua di antaranya duplikat sehingga tersisa 51 studi unik. Penyaringan dilakukan dalam dua tahap, yaitu pemeriksaan judul dan abstrak, kemudian pemeriksaan kelengkapan metode dan hasil. Kriteria inklusi adalah: (i) menerapkan NLP atau *text mining* untuk analisis sentimen atau opini; (ii) objek kajian berkaitan dengan pembelajaran daring, *hybrid*, *e-learning*, MOOC, atau aplikasi pembelajaran; (iii) terbit pada 2021–2026 sebagai artikel jurnal, prosiding, atau pracetak; dan (iv) melaporkan sumber data, metode, serta temuan yang dapat diekstraksi.

Rentang tahun 2021–2026 dipilih agar tinjauan mencakup studi selama dan setelah pandemi COVID-19, ketika model *deep learning* dan *transformer* semakin sering dipakai. Artikel jurnal, prosiding, dan pracetak dimasukkan agar studi terbaru tidak terlewat, dengan catatan bahwa pracetak belum tentu melalui telaah sejawat. Keputusan inklusi dan eksklusi dibuat oleh 6 penulis, dan perbedaan penilaian diselesaikan melalui diskusi.

Studi dikeluarkan jika informasi metode atau hasilnya tidak lengkap atau tidak konsisten, datanya bukan teks atau konteksnya di luar pembelajaran daring, isinya redundan dengan studi ulasan aplikasi lain yang serupa, atau bukti kuantitatifnya terbatas. Rinciannya, 3 studi dikeluarkan karena informasi tidak lengkap, 4 karena berada di luar cakupan, 4 karena redundan, dan 4 karena bukti kuantitatifnya terbatas. Dengan begitu, sebanyak 15 studi dikeluarkan sehingga 36 studi dimasukkan dalam tinjauan (Gambar 1).



Gambar 1. Alur seleksi studi.

2.3 Ekstraksi dan Sintesis Data

Dari setiap studi diekstraksi penulis dan tahun, platform yang diteliti, algoritma NLP, dataset, hasil utama, kesimpulan, dan relevansinya terhadap RQ, terutama berdasarkan abstrak serta bagian metode dan hasil. Data disintesis secara naratif dan dihitung frekuensinya per kategori. Karena dataset, pembagian data, dan metrik evaluasi berbeda antarstudi, nilai akurasi tidak digabungkan dalam meta-analisis dan hanya dibandingkan di dalam studi yang sama.

Algoritma NLP yang dilaporkan dikelompokkan menjadi pembelajaran mesin klasik, pendekatan berbasis leksikon, *deep learning*, model *transformer*, *ensemble*, dan pemodelan topik. Polaritas sentimen dikategorikan menjadi positif, negatif, dan netral. Tantangan yang dilaporkan dikelompokkan menjadi ketidakseimbangan kelas dan kelas netral, bahasa informal dan normalisasi, ukuran data, pelabelan, dan validasi, serta biaya komputasi dan generalisasi.

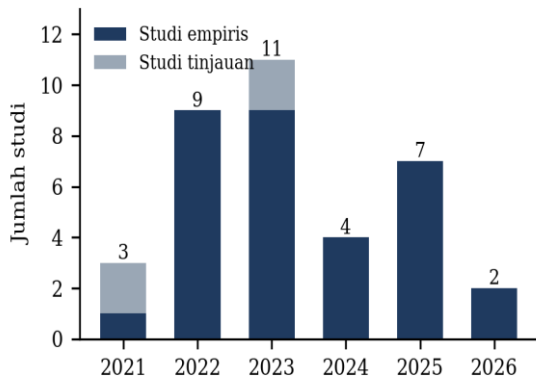
Karena studi melaporkan metrik yang berbeda, seperti akurasi dan F1, nilai kinerja disajikan sesuai metrik aslinya. Hasil sintesis dirangkum dalam tabel ekstraksi (Tabel 1), distribusi studi menurut tahun terbit (Gambar 2), distribusi kategori metode NLP (Gambar 3), dan perbandingan kinerja antarmetode dalam studi yang sama (Tabel 2).

3. HASIL DAN DISKUSI

3.1 Karakteristik Studi

Tinjauan ini mencakup 36 studi, terdiri atas 32 studi empiris dan 4 studi tinjauan [3]–[6].

Ringkasan setiap studi disajikan pada Tabel 1. Distribusi tahun terbit menunjukkan puncak pada 2023 (11 studi) dan 2022 (9 studi), sedangkan 2025 memuat 7 studi (Gambar 2). Studi empiris berasal dari beragam konteks: setidaknya 16 di antaranya secara eksplisit berkonteks Indonesia, sementara yang lain mencakup Amerika Serikat [1], Asia Selatan [19], Yordania [20], Maroko [21], dan dataset universitas di Vietnam [17].



Gambar 2. Distribusi studi menurut tahun terbit

Tabel 1. Ringkasan 36 Studi yang Ditinjau

Ref	Platform / Sumber Data	Metode NLP	Hasil Utama
[1]	Survei dosen fisika AS (364 respons Spring 2020; 134 respons Fall 2020)	Skor sentimen NLP (−1 s.d. 1) dan analisis tematik	Median sentimen 0,2347 (sedikit positif); relatif stabil antarsemester
[2]	Komentar aplikasi pengajaran daring di app market (93.754 komentar)	SnowNLP, ekstraksi kata topik, metode statistik	Evolusi sentimen sebelum dan sesudah COVID-19 teridentifikasi
[3]	MOOC (Coursera, edX, dll.)	Tinjauan sistematis (ML, NLP, DL, leksikon)	440 artikel menjadi 40 artikel final; enam tema utama

[4]	Berbagai platform pembelajaran	Systematic mapping NLP dan deep learning	Tren pergeseran ke deep learning
[5]	Umpan balik mahasiswa kepada instruktur	Tinjauan sistematis NLP	NLP membantu institusi memahami umpan balik secara otomatis
[6]	MOOC, forum diskusi, umpan balik mahasiswa	Tinjauan ML (SVM, KNN, NB, RF) dan DL (LSTM, CNN)	Puluhan studi (2000–2022); analisis sentimen efektif di berbagai kanal
[7]	Twitter (1.260 tweet, 2021–2022)	Naive Bayes, SVM	Akurasi NB 72%, SVM 66%
[8]	Ulasan Coursera di Google Play (2.047 ulasan; Inggris, AS, Indonesia)	Naive Bayes, SVM, KNN, ensemble voting	Akurasi NB 89%, SVM 90%, KNN 88%; ensemble memperbaiki kesalahan model tunggal
[9]	LMS Universitas Mitra Indonesia (forum, komentar, pesan)	Word frequency, LDA, SVM, Naive Bayes	Akurasi 80%, F1 83,33%; netral 60%, positif 25%, negatif 15%
[15]	Ulasan Zoom Cloud Meetings di Play Store (5.000 ulasan)	IndoBERT (fine-tuning)	Akurasi 92%, F1 tertimbang 0,92; F1 netral 0,74

[16]	Ulasan Coursera dari Kaggle (10.000 ulasan)	ALBERT, augmentasi kontekstual	F1 0,9710 (dengan augmentasi) vs 0,9170	[26]	Twitter kuliah daring (2.275 tweet)	Naive Bayes, TF-IDF	Akurasi 99,87%; negatif 75,68%
[17]	Dataset UIT-VSFC (16.175 komentar mahasiswa)	Embedding FastText/Roberta/ELMo, attention, Bi-LSTM/GRU	Akurasi 95%, F1 96%	[27]	Twitter, Facebook, Instagram (200 opini; kuliah hybrid)	TF-IDF, KNN	Negatif 50%, positif 30%, netral 20%
[18]	Ulasan Ruangguru di Google Play (99.000 ulasan)	Stacking ensemble (RF, SVM, XGBoost; meta LR), TF-IDF	Akurasi 87,5%, AUC 0,945; F1 netral 0,72	[28]	Twitter (76.737 tweet)	Topic modeling BTM, SVM	Lima tema masalah pembelajaran jarak jauh
[19]	Twitter (Asia Selatan)	VADER, TextBlob, LDA	Positif 63,2%, netral 30,5%, negatif 6,3%	[29]	Twitter (± 50.000 tweet)	VADER, Afinn, TextBlob; subjektivitas dan toksisitas	Mayoritas positif; 56,6% least opinionated
[20]	Twitter, pembelajaran hybrid di Yordania (4.000 tweet Arab)	LSTM (GloVe/Word2Vec), AraBERT	F1 0,85 (LSTM+GloVe) vs 0,78 (AraBERT)	[30]	Grup Facebook mahasiswa (200 data)	KNN, 10-fold cross validation	Akurasi 87%, AUC 0,916; negatif 87%
[21]	Forum Coursera dan Moodle (mahasiswa Maroko)	Bayesian CNN-BiLSTM, FastText	Akurasi 97,12%; R = -0,92 dengan dropout	[31]	Forum dan YouTube Universitas Terbuka (801 menjadi 249 komentar)	TF-IDF; NB, SVM, KNN, DT, LR, DL	Akurasi Decision Tree 92,0%
[22]	Twitter (1.825 tweet, Feb–Sep 2020)	KNN, TF-IDF	Akurasi 84,65% (K = 10); positif 56,24%	[32]	Ulasan Ruangguru di Google Play (1.913 ulasan)	Leksikon multilingual (Orange), analisis emosi	Positif 57,02%; emosi joy 46,31%
[23]	Twitter kuliah daring (300 data)	Multinomial Naive Bayes, stopword TBRS	Akurasi 75,8% (TBRS) vs 75,3% tanpa stopword	[33]	Ulasan Duolingo di Google Play (4.793 ulasan setelah pembersihan)	Multinomial Naive Bayes, TF-IDF	Akurasi 92,81%; recall negatif 0,38
[24]	Twitter, kebijakan kuliah daring (350 tweet)	SVM, fitur leksikon, TF-IDF	Akurasi 0,60 vs 0,48 tanpa fitur leksikon	[34]	Ulasan Coursera dari Kaggle (706.828 ulasan)	Random Forest, Word2vec	Akurasi 91,42%; pelatihan ± 20 kali lebih cepat
[25]	Twitter (953.378 tweet, 2020–2022)	Naive Bayes, SVM, TF-IDF	Akurasi SVM 84% vs NB 79,33%; netral 43,28%				

[35]	Ulasan kursus bootcamp di courserereport.com (10.610 ulasan)	CNN, LSTM, BERT, RoBERTa, XLNet; NB, SVM	RoBERTa akurasi 95,5%; SVM terbaik untuk topik (79,8%)
[36]	EDOM STMIK Primakara (172 komentar setelah penyaringan)	KNN, SVM, Naive Bayes	Akurasi KNN 82,48%; sentimen negatif 80,23%
[37]	Survei CME pada LMS, AY 2018–2025 (22.206 umpan balik)	VADER, LSTM	Akurasi LSTM 88,6%; gagal mendeteksi kelas negatif
[38]	E-learning Universitas XYZ (1.995 opini uji)	Complement Naive Bayes, SVM	Akurasi CNB 88,56% vs SVM 74,46%
[39]	Evaluasi kursus (>3.000 mahasiswa, 2014–2017)	CNN-BiLSTM multi-attention, GloVe	Akurasi 78%, F1 77,9%; terbaik dibanding varian LSTM
[40]	Ulasan Google Classroom (950 ulasan)	SVM, seleksi fitur PSO, TF-IDF	Akurasi 79% menjadi 83% dengan PSO
[41]	Survei Google Form mahasiswa (35 opini; 100 kalimat uji)	IndoBERT	Akurasi 87%, F1 89%

3.2 RQ1: Platform, Sumber Data, dan Skala Data

Data yang dianalisis dapat dikelompokkan menjadi tiga sumber utama, yaitu media sosial dan forum (13 studi), ulasan aplikasi atau

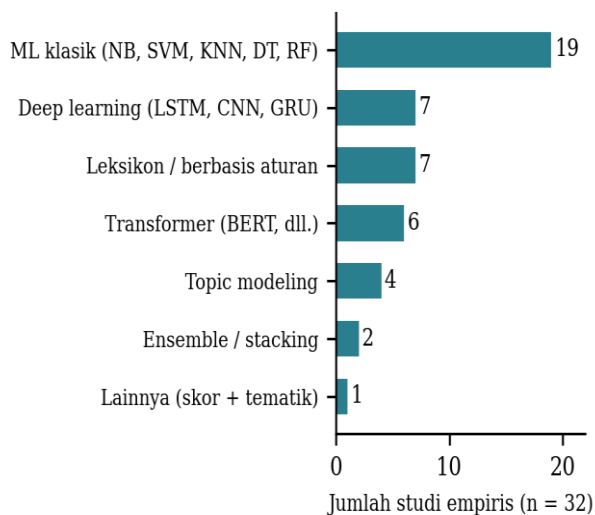
platform belajar (10 studi), serta survei, evaluasi, dan log LMS (9 studi). Twitter merupakan sumber yang paling sering dipakai, muncul pada 11 studi, baik untuk opini terhadap perkuliahan daring di Indonesia [7], [22], [23], [24], [25], [26] dan perkuliahan *hybrid* [27] maupun untuk wacana daring internasional [19], [20], [28], [29]. Facebook dianalisis pada grup mahasiswa [30], sedangkan YouTube dan forum diskusi menjadi sumber pada studi Universitas Terbuka [31].

Ulasan pada toko aplikasi menjadi sumber kedua terbesar. Ulasan Google Play Store dianalisis untuk Zoom [15], Ruangguru [18], [32], Duolingo [33], dan Google Classroom [34], sedangkan ulasan Coursera dipakai dari Google Play [8] dan dari dataset Kaggle [16], [34]. Sumber lain berupa komentar aplikasi pengajaran daring di pasar aplikasi [2] dan situs ulasan kursus [35]. Pada sumber institusional, studi memanfaatkan interaksi mahasiswa di LMS [9], forum Coursera dan Moodle [21], evaluasi dosen oleh mahasiswa [36], survei evaluasi kursus dan mentor [37], kuesioner opini mahasiswa terhadap *e-learning* [38], survei dosen [1], serta data evaluasi kursus [39] dan umpan balik mahasiswa [17].

Skala data sangat beragam. Studi berskala kecil hanya memakai 200 opini [27], [30] hingga 300–350 data [23], [24], sedangkan studi berskala besar menganalisis 93.754 komentar aplikasi pengajaran daring [2], 706.828 ulasan Coursera [34], dan 953.378 tweet [25]. Pelabelan sentimen dilakukan secara manual [25], [26], berbasis leksikon [18], atau berdasarkan skor rating [35], dan perbedaan ini memengaruhi keterbandingan hasil antarstudi.

3.3 RQ2: Metode NLP dan Kinerjanya

Metode yang dipakai dapat dikelompokkan menjadi pembelajaran mesin klasik, pendekatan berbasis leksikon, *deep learning*, model *transformer*, *ensemble*, dan pemodelan topik (Gambar 3). Pembelajaran mesin klasik masih paling dominan, yaitu 19 dari 32 studi empiris, disusul *deep learning* dan leksikon masing-masing 7 studi, serta *transformer* pada 6 studi.



Gambar 3. Distribusi kategori metode NLP pada studi empiris (satu studi dapat masuk lebih dari satu kategori).

Representasi fitur dan praproses. Pembobotan TF-IDF adalah representasi yang paling umum pada pembelajaran mesin klasik [18], [22], [24], [27], [31], sedangkan model saraf memanfaatkan *embedding* seperti GloVe, FastText, atau Word2vec [17], [18], [20], [21]. Pada *Random Forest*, Word2vec memberi akurasi 91,42% dan sedikit melampaui TF-IDF (91,25%) dan TF (90,43%) dengan waktu pelatihan sekitar 20 kali lebih cepat [34]. Perbaikan praproses cenderung berdampak kecil: daftar *stopword* berbasis *term based random sampling* hanya menaikkan akurasi Naive Bayes dari 75,3% menjadi 75,8% [23], dan fitur leksikon menaikkan akurasi SVM dari 0,48 menjadi 0,60 yang tetap tergolong rendah [24]. Seleksi fitur *Particle Swarm Optimization* (PSO) menaikkan akurasi SVM dari 79% menjadi 83% pada ulasan Google Classroom [40].

Pembelajaran mesin klasik. Studi yang membandingkan beberapa algoritma tidak menemukan satu pemenang tetap (Tabel 2). Naive Bayes unggul atas SVM pada tweet perkuliahan daring (72% berbanding 66%) [7] dan pada opini *e-learning* (88,56% berbanding 74,46%) [38]. Sebaliknya, SVM lebih baik pada 953.378 tweet (84% berbanding 79,33%) [25], sedangkan KNN terbaik pada komentar evaluasi dosen (82,48%, dibandingkan SVM 80,69% dan Naive Bayes 68,95%) [36]. Pada ulasan Coursera, SVM (90%), Naive Bayes (89%), dan KNN (88%) berselisih tipis [8], dan *Decision Tree* mencapai akurasi

tertinggi 92,0% pada data forum dan YouTube Universitas Terbuka [31]. Algoritma tunggal juga dapat menghasilkan angka tinggi, misalnya KNN 84,65% [22] dan 87% dengan AUC 0,916 [30] serta Naive Bayes 99,87% [26], meskipun angka yang terakhir perlu diverifikasi dengan validasi independen karena jauh di atas capaian studi lain.

Tabel 2. Perbandingan Kinerja pada Studi yang Membandingkan Beberapa Metode

Ref.	Perbandingan (akurasi atau F1)	Terbaik
[7]	NB 72%; SVM 66%	NB
[8]	SVM 90%; NB 89%; KNN 88%	SVM
[16]	ALBERT F1 0,9710 (augmentasi); 0,9170 (tanpa)	Augmentasi
[18]	Stacking 87,5%; Logistic Regression 86,8%	Stacking
[20]	LSTM+GloVe F1 0,85; AraBERT 0,78; LSTM+Word2Vec 0,68	LSTM+GloVe
[23]	Stopword TBRS 75,8%; tanpa stopwords 75,3%; Tala 73,8%	TBRS
[24]	SVM + leksikon 0,60; tanpa leksikon 0,48	+ Leksikon
[25]	SVM 84%; NB 79,33%	SVM
[31]	DT 92,0% (tertinggi dari NB, SVM, KNN, DT, LR, DL)	DT
[34]	RF: Word2vec 91,42%; TF-IDF 91,25%; TF 90,43%	Word2vec
[35]	Sentimen: RoBERTa 95,5%; topik: SVM 79,8%	RoBERTa; SVM
[36]	KNN 82,48%; SVM 80,69%; NB 68,95%	KNN
[38]	CNB 88,56%; SVM 74,46%	CNB
[39]	CNN-BiLSTM 78% (F1 77,9%) vs LSTM, TD-LSTM, Bi-LSTM	CNN-BiLSTM

Ensemble dan stacking. Penggabungan model memberi perbaikan yang moderat. *Voting* atas Naive Bayes, SVM, dan KNN memperbaiki banyak kesalahan prediksi model tunggal pada ulasan Coursera [8], sedangkan *stacking* dengan

Random Forest, SVM, dan XGBoost mencapai akurasi 87,5% dan AUC 0,945, sedikit di atas model tunggal terbaik (Logistic Regression, 86,8%) [18].

Pendekatan leksikon. Pendekatan ini tidak memerlukan data latih berlabel dan dinilai cukup untuk eksplorasi awal [32], tetapi pada studi yang ditinjau umumnya tidak melaporkan metrik akurasi [19], [29], [32]. Leksikon multilingual pada 1.913 ulasan Ruangguru menemukan 57,02% ulasan positif dengan emosi dominan *joy* [32], sementara VADER dan TextBlob dipakai untuk memetakan opini publik Asia Selatan [19] dan wacana berbasis gender [29]. SnowNLP dipadukan dengan ekstraksi kata topik dan uji statistik untuk membandingkan komentar sebelum dan sesudah pandemi [2], dan VADER dipadukan dengan LSTM pada data evaluasi mentor [37]. Studi Ruangguru tersebut menyarankan pembelajaran mesin atau *deep learning* untuk konteks yang lebih dalam [32].

Deep learning dan transformer. Model saraf dan *transformer* mencatat sebagian angka tertinggi pada studi yang ditinjau. CNN-BiLSTM dengan *multi-attention* dan GloVe mengungguli LSTM, TD-LSTM, dan Bi-LSTM biasa pada evaluasi kursus, meskipun akurasi absolutnya 78% [39]. Model CNN-BiLSTM Bayesian dengan FastText mencapai akurasi 97,12% pada teks panjang dan 96,53% pada teks pendek [21], sedangkan kombinasi *embedding* FastText, RoBERTa, dan ELMo dengan *attention* serta Bi-LSTM mencapai 95% pada 16.175 komentar mahasiswa [17]. Pada model *transformer*, RoBERTa memberi akurasi sentimen terbaik (95,5%) pada ulasan kursus tetapi membutuhkan waktu komputasi lebih lama daripada model tradisional [35], IndoBERT mencapai 92% pada ulasan Zoom berbahasa Indonesia [15] dan 87% pada 100 kalimat uji [41], dan ALBERT dengan augmentasi kontekstual menaikkan F1 dari 0,9170 menjadi 0,9710 [16]. Hasil tidak selalu berpihak pada *transformer*: pada tweet berbahasa Arab, LSTM dengan GloVe (F1 0,85) mengungguli AraBERT (0,78) [20], dan LSTM dengan akurasi 88,6% pada data evaluasi mentor gagal mendeteksi kelas negatif [37].

Pemodelan topik. LDA dan BTM dipakai untuk menjelaskan apa yang dibicarakan, bukan hanya polaritasnya. Pemodelan BTM pada 76.737

tweet yang dilanjutkan klasifikasi SVM mengidentifikasi lima tema masalah pembelajaran jarak jauh, yaitu pendidikan, sosial, teknologi, kesehatan, serta sikap atau etika [28], sedangkan LDA dipadukan dengan analisis sentimen untuk melacak topik pada tweet Asia Selatan [19] dan pola keterlibatan mahasiswa di LMS [9].

3.4 RQ3: Kecenderungan Sentimen dan Faktor Terkait

Polaritas dominan bergantung pada objek dan sumber data. Sentimen positif dominan pada studi yang menganalisis wacana publik atau ulasan platform: 63,2% pada tweet Asia Selatan [19], 56,24% pada tweet berbahasa Indonesia [22], 57,02% pada ulasan Ruangguru [32], dan 55% pada ulasan Zoom [15]. Dosen fisika di Amerika Serikat juga cenderung sedikit positif dengan median skor 0,2347 [1], dan tweet tentang pembelajaran daring dalam analisis berbasis gender mayoritas positif [29]. Pada ulasan aplikasi, dominasi positif sering sangat kuat dan disertai ketimpangan kelas, misalnya 85,34% ulasan Duolingo [33] dan 91,5% ulasan kursus *bootcamp* [35].

Sebaliknya, opini tentang pelaksanaan perkuliahan daring cenderung negatif atau netral. Sentimen negatif mencapai 87% pada grup Facebook mahasiswa [30], 80,23% pada evaluasi dosen oleh mahasiswa [36], dan 75,68% pada tweet kuliah daring [26]. Pada perkuliahan *hybrid*, komentar negatif mencapai 50% pada media sosial [27], dan sekitar 40% tweet berbahasa Arab mengekspresikan kemarahan, kebencian, atau kesedihan sementara hanya 13% menunjukkan kebahagiaan [20]. Analisis 953.378 tweet menunjukkan sentimen netral terbesar (43,28%) diikuti negatif (32,91%), dengan kenaikan sentimen negatif pada awal tahun ajaran baru dan saat kebijakan *hybrid* diterapkan [25]. Interaksi mahasiswa di LMS juga didominasi sentimen netral (60%) [9].

Pola ini menunjukkan bahwa ulasan platform belajar dan wacana publik umum lebih positif, sedangkan opini yang langsung berkaitan dengan pengalaman kuliah daring dan *hybrid* lebih kritis. Namun, perbedaan tersebut tidak dapat dianggap sebagai efek metode pembelajaran semata karena juga dipengaruhi oleh sumber data, pemilihan sampel, dan skema

pelabelan. Pelabelan berbasis rating pada ulasan kursus, misalnya, menghasilkan 91,5% ulasan positif [35].

Keluhan yang muncul relatif konsisten dan bersifat teknis atau fungsional. Ulasan Ruangguru menyoroti gangguan teknis [32], opini mahasiswa pada *e-learning* menilai fitur BigBlueButton dan koneksi server paling negatif meskipun delapan dari 12 fitur bersentimen positif [38], ulasan Google Classroom mengeluhkan *error*, sulitnya mengumpulkan tugas, dan notifikasi terlambat [40], dan ulasan Duolingo mengeluhkan sistem hati atau energi serta iklan [33].

Beberapa studi mengaitkan sentimen dengan luaran pembelajaran. Skor sentimen pada forum Coursera dan Moodle berkorelasi negatif kuat ($R = -0,92$) dengan jumlah *dropout* mingguan sehingga berpotensi menjadi indikator dini pembelajar berisiko [12]. NLP juga dipakai untuk mengukur keterlibatan mahasiswa di LMS [9], mengevaluasi mentor dan materi [39], serta memantau perubahan kebutuhan pengguna sebelum dan sesudah pandemi [2]. Tinjauan sistematis mencatat bahwa analisis sentimen membantu pendidik memperbaiki metode mengajar [6], memetakan enam tema utama pada umpan balik MOOC termasuk prediksi *dropout* [3], dan membantu institusi memahami umpan balik secara otomatis [5], serta bahwa penggunaan *deep learning* makin dominan [3], [4].

3.5 Tantangan dan Keterbatasan

Ketidakseimbangan kelas dan kelas netral. Kelas minoritas dan kelas netral adalah kelemahan yang paling sering dilaporkan. Akurasi LSTM 88,6% pada evaluasi mentor menutupi kegagalan mendeteksi kelas negatif [37], dan Naive Bayes pada ulasan Duolingo mencapai akurasi 92,81% tetapi *recall* kelas negatif hanya 0,38 [33]. Kelas netral juga paling sulit dikenali oleh IndoBERT (F1 0,74) [15] dan *stacking ensemble* (F1 0,72) [18]. Tanpa augmentasi, ALBERT hanya mengklasifikasikan empat sampel netral dengan benar, dan augmentasi kontekstual menaikkan F1 dari 0,9170 menjadi 0,9710 [16].

Bahasa informal dan normalisasi. Kata tidak baku, *slang*, dan kata di luar kamus menurunkan akurasi pendekatan berbasis leksikon [24], dan perbaikan *stopword* saja belum

cukup sehingga peneliti menyarankan normalisasi kata tidak baku [23].

Ukuran data, pelabelan, dan validasi. Beberapa studi memakai sampel kecil, misalnya 200 opini [27], [30], dan ada yang kehilangan banyak data saat penyaringan, seperti 801 komentar menjadi 249 [31]. Skema pelabelan (manual, leksikon, rating) dan pembagian data (misalnya 70:30, 80:20, dan 90:10 [25]) berbeda antarstudi sehingga perbandingan langsung sulit dilakukan. Pelaporan metrik per kelas menjadi penting karena akurasi dapat menyesatkan pada data yang timpang, sebagaimana terlihat pada kasus di atas.

Biaya komputasi dan generalisasi. Model *transformer* membutuhkan waktu komputasi lebih lama [35], meskipun ALBERT dilaporkan memiliki waktu pelatihan tercepat [16]. Selain itu, hasil pada satu bahasa tidak selalu berlaku pada bahasa lain, seperti terlihat pada perbandingan LSTM dan AraBERT untuk bahasa Arab [20] dan pemakaian terjemahan bahasa Inggris pada dataset umpan balik mahasiswa Vietnam [17].

3.6 Arah Penelitian Mendatang dan Implikasi

Berdasarkan sintesis di atas, terdapat beberapa arah penelitian yang layak dikembangkan. Pertama, studi berbahasa Indonesia perlu memanfaatkan model *transformer* seperti IndoBERT [15], [41] dengan perbandingan yang adil terhadap pembelajaran mesin klasik pada dataset yang sama dan berukuran besar. Kedua, ketidakseimbangan kelas perlu ditangani secara sistematis, misalnya dengan augmentasi kontekstual [16] serta pelaporan *F1-macro* dan metrik per kelas. Ketiga, analisis sentimen berbasis aspek dapat menerjemahkan keluhan menjadi tindakan perbaikan, misalnya pada fitur konferensi video dan koneksi server [38] atau sistem hati dan iklan [33]. Keempat, dataset terlabel manual yang terbuka dengan kesepakatan antar-anotator akan memperkuat keterbandingan antarstudi.

Kelima, pengaitan sentimen dengan keterlibatan dan *dropout* [9], [21] serta analisis longitudinal pada pembelajaran *hybrid* [20], [25], [27] masih terbuka untuk dieksplorasi lebih lanjut. Bagi pengelola pembelajaran daring, hasil tinjauan ini menunjukkan bahwa pemantauan sentimen otomatis dapat melengkapi evaluasi

formal, asalkan hasilnya ditafsirkan bersama metrik per kelas dan konteks data.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Tinjauan atas 36 studi menunjukkan bahwa NLP telah dimanfaatkan secara luas untuk memahami opini terhadap pembelajaran daring. Terkait RQ1, sumber data terbanyak adalah media sosial dan forum (13 studi), khususnya Twitter, diikuti ulasan aplikasi atau platform belajar (10 studi) serta survei dan log LMS (9 studi), dengan skala data dari 200 hingga ratusan ribu entri. Terkait RQ2, pembelajaran mesin klasik dominan (19 dari 32 studi empiris), sedangkan *deep learning* dan *transformer* mencatat sebagian akurasi tertinggi (hingga 97,12%), tetapi tidak ada algoritma yang selalu terbaik karena kinerja bergantung pada data, pelabelan, dan proporsi kelas. Terkait RQ3, ulasan platform dan wacana publik cenderung positif, sedangkan opini terhadap pelaksanaan perkuliahan daring dan *hybrid* lebih sering negatif atau netral, dengan keluhan yang berpusat pada masalah teknis dan fungsional. Tinjauan ini memiliki keterbatasan. Seleksi bertumpu pada tabel ekstraksi data yang disusun dari abstrak, metode, dan hasil setiap studi, tidak dilakukan meta-analisis, dan perbandingan akurasi hanya bermakna di dalam studi yang sama.

4.2 Saran

Berdasarkan hasil tinjauan, peneliti selanjutnya disarankan memanfaatkan model *transformer* berbahasa Indonesia seperti IndoBERT dan membandingkannya secara adil dengan pembelajaran mesin klasik pada dataset yang sama dan berukuran besar, mengingat tidak ada algoritma yang selalu terbaik. Ketidakseimbangan kelas perlu ditangani secara sistematis, misalnya augmentasi data, dan dilaporkan dengan F1-macro serta metrik per kelas agar akurasi yang tinggi tidak menutupi kegagalan pada kelas negatif dan netral. Peneliti juga perlu menyusun dataset acuan yang dilabeli manual, terbuka, dan disertai kesepakatan antar-anotator, serta melaporkan skema pelabelan dan pembagian data secara seragam agar hasil antarstudi dapat dibandingkan, sambil

mengembangkan analisis berbasis aspek dan analisis longitudinal pada pembelajaran *hybrid*. Bagi pengelola pembelajaran daring, pemantauan sentimen otomatis dapat melengkapi evaluasi formal asalkan ditafsirkan bersama metrik per kelas dan konteks data, dengan prioritas perbaikan pada fitur dan stabilitas layanan karena keluhan pengguna berpusat pada masalah teknis dan fungsional.

DAFTAR PUSTAKA

- [1] C. Green, E. Brewe, and J. Mellen, "Sentiment and thematic analysis of faculty responses: Transition to online learning," *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 1, p. 10151, 2024, doi: 10.1103/PhysRevPhysEducRes.20.010151.
- [2] H. Kang, J. Zhang, and J. Kang, "Analysis of online education reviews of universities using NLP techniques and statistical methods," *Wirel. Commun. Mob. Comput.*, p. 8730117, 2022, doi: 10.1155/2022/8730117.
- [3] F. Dalipi, K. Zdravkova, and F. Ahlgren, "Sentiment analysis of students' feedback in MOOCs: A systematic literature review," *Front. Artif. Intell.*, vol. 4, p. 728708, 2021, doi: 10.3389/frai.2021.728708.
- [4] 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," *Appl. Sci.*, vol. 11, no. 9, p. 3986, 2021, doi: 10.3390/app11093986.
- [5] A. S. Sunar and M. S. Khalid, "Natural language processing of student's feedback to instructors: A systematic review," *IEEE Trans. Learn. Technol.*, 2023, doi: 10.1109/TLT.2023.3330531.
- [6] Khanam, "Sentiment analysis of user reviews in an online learning environment: Analyzing the methods and future prospects," *Eur. J. Educ. Pedagog.*, vol. 4, no. 2, 2023, doi: 10.24018/ejedu.2023.4.2.531.
- [7] F. D. Samuel, P. D. Atika, and S. Setiawati, "Analisis sentimen masyarakat terhadap

- perkuliahan daring di Twitter menggunakan algoritma Naive Bayes dan Support Vector Machine,” *J. Students’ Res. Comput. Sci.*, 2023, doi: 10.31599/6691v571.
- [8] J. P. Hidayat and I. Nurhaida, “Analisis sentimen pada ulasan LMS pembelajaran menggunakan metode natural language processing,” *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 1, pp. 565–575, 2025, doi: 10.29100/jipi.v10i1.5899.
- [9] B. Hartantom, H. D. Yunita, F. Fahurian, F. Dirayati, T. Winarko, and I. Marliana, “Evaluasi keterlibatan mahasiswa dalam lingkungan pembelajaran daring menggunakan NLP dan analisis sentimen,” *J. Algoritma.*, vol. 22, no. 1, 2025, doi: 10.33364/algoritma/v.22-1.2154.
- [10] I. Yolanda, Suarman, and G. Haryana, “Analisis Faktor Penentu Kualitas Pembelajaran Dalam Jaringan di Perguruan Tinggi,” *J. PAJAR (Pendidikan dan Pengajaran)*, vol. 6, no. 4, pp. 1005–1014, 2022, doi: 10.33578/pjr.v6i4.8644.
- [11] D. Khurana, A. Koli, and K. Khatter, “Natural language processing: State of the art, current trends and challenges,” *Multimed. Tools Appl.*, vol. 82, pp. 3713–3744, 2022, doi: 10.1007/s11042-022-13428-4.
- [12] M. Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends,” *Knowledge-Based Syst.*, vol. 226, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.
- [13] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering,” 2007.
- [14] M. J. Page, J. E. McKenzie, and P. M. Bossuyt, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, 2021, doi: 10.1136/bmj.n71.
- [15] N. Herlinawati, I. Ariyati, Samudi, and A. Y. Asih, “Analisis sentimen Zoom Cloud Meetings di Play Store menggunakan metode IndoBERT,” *J. Sains Inform. Terap.*, vol. 5, no. 1, pp. 243–251, 2026.
- [16] S. D. Pamuja and Noviyanto, “Optimasi analisis sentimen ulasan platform pendidikan daring menggunakan arsitektur ALBERT dan teknik augmentasi kontekstual,” *J. Keilmuan Teknol. Inf.*, vol. 1, no. 2, pp. 34–38, 2025, [Online]. Available: <https://ejournal.umkla.ac.id/index.php/jkti/article/view/2256>
- [17] S. Z. Malik *et al.*, “Attention-aware with stacked embedding for sentiment analysis of student feedback through deep learning techniques,” *PeerJ Comput. Sci.*, vol. 10, p. e2283, 2024, doi: 10.7717/peerj-cs.2283.
- [18] D. Sofia and P. Sekarpuji, “Penerapan metode stacking ensemble untuk analisis sentimen pada ulasan aplikasi Ruangguru,” *IDEALIS Indones. J. Inf. Syst.*, vol. 8, no. 2, pp. 248–257, 2025, [Online]. Available: <https://jom.fti.budiluhur.ac.id/IDEALIS/article/view/3559>
- [19] A. R. F. Shafana and S. M. Safnas, “Does technology assist to continue learning during pandemic? A sentiment analysis and topic modeling on online learning in South Asian region,” *Soc. Netw. Anal. Min.*, 2022, doi: 10.1007/s13278-022-00899-4.
- [20] E. Qaqish, A. Aranki, and W. Etaiwi, “Sentiment analysis and emotion detection of post-COVID educational tweets: Jordan case,” *Soc. Netw. Anal. Min.*, vol. 13, p. 39, 2023, doi: 10.1007/s13278-023-01041-8.
- [21] K. Benabbes, M. Naimi, B. Hmedna, K. Housni, A. Zellou, and A. El Mezouary, “Unveiling the reasons behind learners’ dropout from educational platforms: Analyzing sentiment intensity throughout texts,” *Smart Learn. Environ.*, 2025, doi: 10.1186/s40561-025-00394-1.
- [22] Supriyanto, “Penerapan algoritma K-NN untuk analisis sentimen publik terhadap pembelajaran daring,” *JATIKA*, vol. 4, no. 1, 2023, doi: 10.33365/jatika.v4i1.2468.
- [23] Hidayat and Pramudita, “Analisis sentimen terhadap pembelajaran daring pasca pandemi Covid-19 menggunakan metode IndoBERT,” *Inf. Manag. Educ. Prof.*, vol.

- 8, no. 2, 2023, doi: 10.51211/imbi.v8i2.2719.
- [24] Utami and Artana, "Text mining dalam analisis sentimen pembelajaran daring di masa pandemi Covid-19 menggunakan K-Nearest Neighbor," *Jinteks*, 2022.
- [25] Sediyo, "Penerapan algoritma K-Nearest Neighbors pada analisis sentimen metode pembelajaran daring di Universitas Kristen Wira Wacana Sumba," *Jointer*, 2021.
- [26] Wang and Yang, "Educational big data analytics using sentiment analysis for student requirement analysis on courses," *Scalable Comput. Pract. Exp.*, vol. 25, no. 5, 2024, doi: 10.12694/scpe.v25i5.3024.
- [27] I. B. K. S. Arnawa, "Analisis sentimen pada media sosial terhadap perkuliahan hybrid menggunakan algoritma TF IDF dan K Nearest Neighbor," *J. Sist. dan Inform.*, vol. 17, no. 1, 2023, doi: 10.30864/jsi.v17i1.495.
- [28] G. Rullyana and R. Triandari, "Analisis sentimen ulasan pengguna Indonesia terhadap platform pembelajaran digital Ruangguru menggunakan algoritma leksikal multilingual," *J. Komtika (Komputasi dan Inform.)*, vol. 9, no. 2, pp. 263–275, 2025, doi: 10.31603/komtika.v9i2.15110.
- [29] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, "Pembentukan daftar stopword menggunakan term based random sampling pada analisis sentimen dengan metode Naive Bayes (Studi kasus: Kuliah daring di masa pandemi)," *Nuansa Inform.*, vol. 17, no. 2, 2022, doi: 10.25134/ilkom.v17i2.11.
- [30] M. J. Aufa and A. Qoiriah, "Analisis sentimen pengguna platform belajar online Coursera menggunakan Random Forest dengan metode ekstraksi fitur Word2vec," *JINACS (Journal Informatics Comput. Sci.)*, vol. 4, no. 2, pp. 244–255, 2022, doi: 10.26740/jinacs.v4n02.p244-255.
- [31] A. Koufakou, "Deep learning for opinion mining and topic classification of course reviews," 2023.
- [32] Soriano and Maddalora, "Towards effective distance education implementation: Utilizing descriptive analytics, opinion mining, and sentiment analysis for online education mentors and learning materials," *Business, Educ. Soc. Sci. Technol.*, vol. 1, no. 2, 2025, doi: 10.69478/BEST2025v1n2a022.
- [33] M. A. A. Tratama, F. S. Murmita, D. A. Maulana, C. Renata, R. Ailsa, and F. N. Hasan, "Analisis sentimen keefektifan belajar bahasa asing di aplikasi Duolingo dengan algoritma Naive Bayes," *SUDO J. Tek. Inform.*, vol. 5, no. 1, pp. 53–68, 2026, doi: 10.56211/sudo.v5i1.1354.
- [34] J. Fernando and Fathoni, "Analisis sentimen pada e-learning Universitas XYZ menggunakan metode Naive Bayes Classifier dan Support Vector Machine," *JUTI J. Ilm. Teknol. Inf.*, vol. 21, no. 2, pp. 81–97, 2023, [Online]. Available: <https://juti.if.its.ac.id/index.php/juti/article/view/1147>
- [35] G. A. Mursianto, D. Widiyanto, and B. T. Wahyono, "Analisis sentimen ulasan pengguna pada aplikasi Google Classroom menggunakan metode SVM dan seleksi fitur PSO," *J. Inform.*, vol. 18, no. 3, 2022.
- [36] H. Alhazmi, "Detection of students' problems in distance education using topic modeling and machine learning," *Futur. Internet*, vol. 14, no. 6, p. 170, 2022, doi: 10.3390/fi14060170.
- [37] N. E. Oktaviana, Y. A. Sari, and Indriati, "Analisis sentimen terhadap kebijakan kuliah daring selama pandemi menggunakan pendekatan lexicon based features dan Support Vector Machine," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 357–362, 2022, doi: 10.25126/jtiik.202295625.
- [38] K. M. A. Candrayani, I. M. A. D. Suarjaya, and A. A. K. A. C. Wiranatha, "Analisis sentimen pembelajaran daring era pandemi COVID-19 menggunakan Naive Bayes dan Support Vector Machine," *Temat. J. Teknol. Inf. Komun.*, vol. 10, no. 1, pp. 47–53, 2023, doi: 10.38204/tematik.v10i1.1274.
- [39] I. B. G. Sarasvananda, D. Selivan, M. L. Radhitya, and I. N. T. A. Putra, "Analisis sentimen pada pembelajaran daring di

- Indonesia melalui Twitter menggunakan Naive Bayes Classifier,” *SINTECH J.*, vol. 5, no. 2, pp. 227–233, 2022.
- [40] N. Thakur, S. Cui, K. Khanna, V. Knieling, Y. N. Duggal, and M. Shao, “Investigation of the gender-specific discourse about online learning during COVID-19 on Twitter using sentiment analysis, subjectivity analysis, and toxicity analysis,” *Computers*, vol. 12, no. 11, p. 221, 2023, doi: 10.3390/computers12110221.
- [41] R. Fatmasari, N. Purnomo, S. A. Putra, W. Gata, and N. K. Wardhani, “Pengujian pelabelan otomatis dataset kualitas pembelajaran daring Universitas Terbuka di forum dan YouTube,” *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, 2024, doi: 10.29100/jipi.v9i3.5231.