Jejakdigital
JURNAL ILMIAH MULTIDISIPLIN
Available online at https://indojurnal.com/index.php/jejakdigital

eISSN 3089-7734; pISSN 3089-7742
Vol. 2, No. 4, Tahun 2026
doi.org/10.63822/hb46z150
Hal. 8892-8902

Evaluasi Komparatif Lintas Platform Analisis Sentimen Berbasis Aspek dengan LDA dan Algoritma Klasifikasi

Ade Sarah Huzaifah[1], R.A. Fattah Adriansyah[2]
Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia[1]
Teknik Informatika, Universitas Mikroskil, Medan, Indonesia[2]

*Email Korespodensi: adesarah@usu.ac.id

Sejarah Artikel:

Diterima 15-07-2026
Disetujui 21-07-2026
Diterbitkan 23-07-2026

ABSTRACT
Aspect-based sentiment analysis is becoming increasingly important for understanding user opinions across various digital platforms; however, differences in text characteristics such as length, formality, and writing style often affect algorithm performance. This study aims to evaluate the effectiveness of a hybrid model that integrates Latent Dirichlet Allocation (LDA) with three classification algorithms Multinomial Naïve Bayes, XGBoost, and AdaBoost across three different platforms: Twitter, Google Play, and Female Daily. The evaluation was conducted using the Precision, Recall, F1-Score, and Accuracy metrics, with F1-Score serving as the primary indicator due to the imbalanced data distribution across aspects. The results show that Multinomial Naïve Bayes is more suitable for short and emotional texts on Twitter, XGBoost excels at technical reviews on Google Play, while AdaBoost is quite adaptive to narrative reviews on Female Daily. Key findings confirm that XGBoost is the most stable algorithm with the highest average F1-Score, while Multinomial Naïve Bayes remains competitive for informal text. This study contributes to the development of aspect-based sentiment analysis methodologies by providing practical guidelines for algorithm selection based on platform characteristics, and opens avenues for further research through dataset expansion, microtext normalization, and the application of deep learning to improve generalization
Keywords*: AdaBoost; Multinomial Naïve Bayes; XGBoost*

ABSTRAK
Analisis sentimen berbasis aspek semakin penting untuk memahami opini pengguna pada berbagai platform digital, namun perbedaan karakteristik teks seperti panjang, formalitas, dan gaya bahasa sering memengaruhi performa algoritma. Penelitian ini bertujuan mengevaluasi efektivitas model hybrid yang mengintegrasikan Latent Dirichlet Allocation (LDA) dengan tiga algoritma klasifikasi, yaitu Multinomial Naïve Bayes, XGBoost, dan AdaBoost, pada tiga platform berbeda: Twitter, Google Play, dan Female Daily. Evaluasi dilakukan menggunakan metrik Precision, Recall, F1-Score, dan Akurasi, dengan F1-Score sebagai indikator utama karena distribusi data antar aspek tidak seimbang. Hasil menunjukkan bahwa Multinomial Naïve Bayes sesuai untuk teks pendek dan emosional di Twitter, XGBoost unggul pada ulasan teknis di Google Play, sedangkan AdaBoost cukup adaptif pada ulasan naratif di Female Daily. Temuan utama menegaskan bahwa XGBoost merupakan algoritma paling stabil dengan rata-rata F1-Score tertinggi, sementara Multinomial Naïve Bayes cukup kompetitif pada teks informal. Penelitian ini berkontribusi pada pengembangan metodologi analisis sentimen berbasis aspek dengan menyediakan panduan praktis pemilihan algoritma sesuai karakteristik platform, serta membuka arah penelitian lanjutan melalui perluasan dataset, normalisasi microtext, dan penerapan deep learning untuk meningkatkan generalisasi.
Katakunci: AdaBoost; Multinomial Naïve Bayes; XGBoost

Bagaimana Cara Sitasi Artikel ini:

Huzaifah, A. S., & Adriansyah, R. F. . (2026). Evaluasi Komparatif Lintas Platform Analisis Sentimen Berbasis Aspek dengan LDA dan Algoritma Klasifikasi. Jejak Digital: Jurnal Ilmiah Multidisiplin, 2(4), 8892-8902. https://doi.org/10.63822/hb46z150

PENDAHULUAN

Pertumbuhan eksponensial interaksi digital melalui media sosial, *marketplace*, dan aplikasi seluler telah menciptakan aliran data opini yang masif sekaligus tidak terstruktur. Bagi pelaku bisnis maupun pengembang sistem, memahami sentimen pengguna secara mendalam tidak lagi cukup hanya dengan klasifikasi polaritas sederhana (positif/negatif). Sebaliknya, diperlukan kemampuan untuk mengidentifikasi aspek spesifik yang disukai atau dikeluhkan oleh pengguna. Fenomena ini menempatkan *Aspect-Based Sentiment Analysis* (ABSA) sebagai instrumen krusial dalam ekstraksi opini yang lebih terperinci dibandingkan analisis sentimen konvensional (Jayakody et al., 2024; Shukla, Kumar, Dwivedi, & Singh, 2026).

Namun, penerapan ABSA menghadapi tantangan signifikan terkait variabilitas karakteristik bahasa antar platform. Opini yang diunggah di media sosial seperti Twitter cenderung bersifat informal, pendek, dan penuh singkatan, sementara ulasan pada aplikasi seluler seperti Google Play lebih teknis dan fungsional, dan ulasan pada situs gaya hidup seperti Female Daily cenderung deskriptif serta terstruktur. Variasi ini secara langsung memengaruhi performa algoritma klasifikasi yang digunakan. Seringkali, sebuah algoritma yang unggul pada satu platform belum tentu menunjukkan performa serupa ketika dihadapkan pada karakteristik teks yang berbeda (Negi et al., 2024).

Selain itu, pemilihan *Latent Dirichlet Allocation* (LDA) sebagai ekstraktor fitur utama dalam penelitian ini bukan tanpa alasan. LDA memiliki keunggulan dalam mengidentifikasi distribusi topik tersembunyi dari kumpulan teks yang besar dan tidak terstruktur, sehingga mampu memberikan representasi yang lebih interpretable dibandingkan metode lain seperti *clustering* berbasis K-means atau pendekatan neural murni. Dengan menghasilkan distribusi probabilistik kata terhadap topik dan topik terhadap dokumen, LDA memungkinkan peneliti untuk memahami konteks semantik yang mendasari opini pengguna secara lebih sistematis. Keunggulan ini menjadikan LDA relevan sebagai fondasi dalam analisis aspek, karena setiap aspek dapat dipetakan ke dalam topik yang konsisten dan dapat ditindaklanjuti (Blei, Ng, & Jordan, 2003).

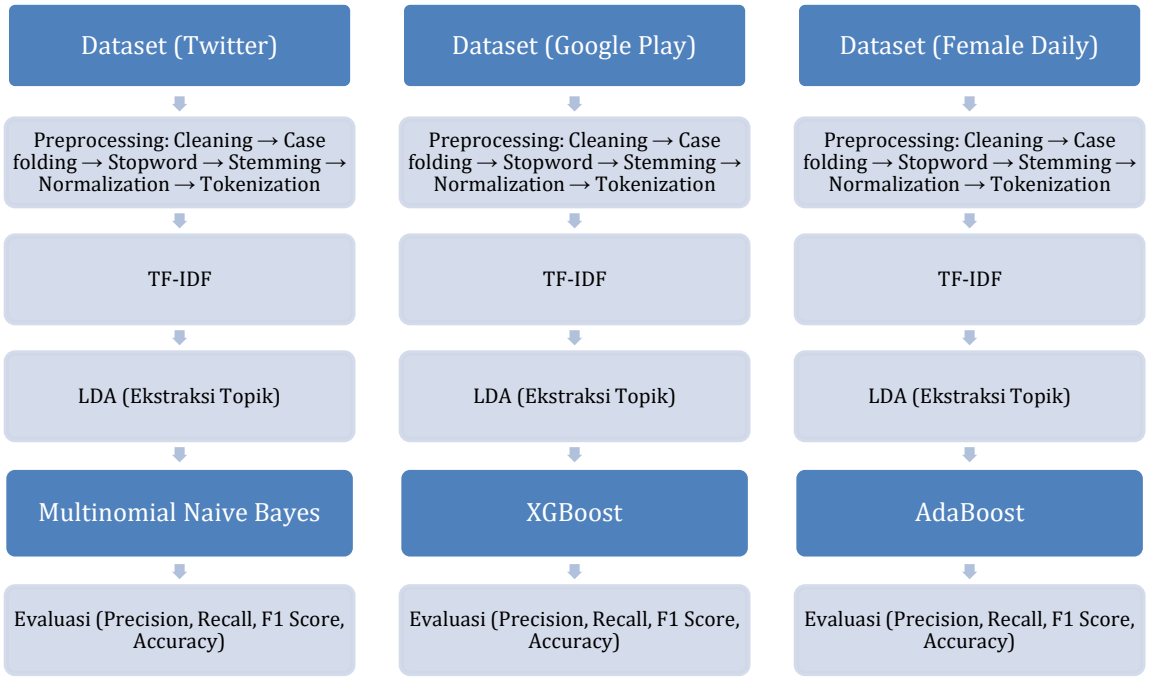
Dalam literatur riset terkini, penggunaan metode *hybrid* yang menggabungkan *Latent Dirichlet Allocation* (LDA) untuk ekstraksi topik dengan algoritma *Machine Learning* untuk klasifikasi sentimen telah menjadi praktik yang mapan. Pendekatan ini terbukti meningkatkan akurasi dan ketahanan model terhadap variasi bahasa, baik melalui integrasi dengan SVM, BiLSTM, maupun Transformer modern (Ziaulla & Biradar, 2026; Said, Smairi, & Abadlia, 2025). Namun, sebagian besar studi masih berfokus pada evaluasi model tunggal di satu domain spesifik, sehingga belum ada kajian komparatif yang secara sistematis menilai efektivitas arsitektur hibrid ini lintas platform dengan pendekatan *aspect-level*. Celah riset ini menimbulkan kebutuhan untuk eksplorasi lebih lanjut, terutama dalam upaya standarisasi performa model agar dapat diandalkan di berbagai konteks digital (Nazir, Saiful, & Al Sohan, 2025).

Berdasarkan celah tersebut, penelitian ini bertujuan untuk membandingkan kinerja model *hybrid* LDA-*Machine Learning* pada berbagai platform digital. Fokus utama studi ini adalah mengevaluasi stabilitas dan efektivitas algoritma melalui metrik *precision, recall, F1-score*, dan akurasi yang diukur secara mendalam di tingkat aspek. Melalui pendekatan komparatif ini, penelitian diharapkan dapat memberikan kontribusi dalam pengembangan arsitektur ABSA yang lebih adaptif dan tangguh dalam menghadapi dinamika data opini di era digital.

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimental komparatif untuk mengevaluasi kinerja arsitektur hybrid LDA–*Machine Learning* dalam melakukan *Aspect-Based Sentiment Analysis* (ABSA). Alur kerja sistem dirancang dalam tiga tahapan utama: (1) *Data Preprocessing*, (2) *Topic Modeling* menggunakan LDA, dan (3) Klasifikasi sentimen per aspek menggunakan algoritma *Machine Learning*. Secara visual, alur sistem ini disajikan pada Gambar 1.



Gambar 1. Flowchart Sistem

Pengumpulan Data dan Karakteristik Platform

Data yang digunakan berasal dari tiga platform digital dengan karakteristik unik. Ringkasan dataset yang digunakan dalam penelitian tersaji dalam tabel 1.

Tabel 1. Ringkasan Karakteristik Dataset Tiap Platform

Platform	Jumlah Dataset (N)	Data latih (80%)	Data Uji (20%)	Fokus Karakteristik Teks
Twitter	1.750	1.400	350	Pendek, informal, banyak slang (Wu et al., 2025)
Google Play	3.500	2.800	700	Teknis, fungsional, narasi proses (Lovenia et al., 2024)
Female Daily	4.000	3.200	800	Naratif, sensorik, deskriptif (Liskowski et al., 2026)

Dataset ini dipilih karena merepresentasikan spektrum opini digital yang luas, mulai dari interaksi media sosial yang volatil hingga ulasan produk yang bersifat naratif dan terstruktur.

Data Preprocessing

Untuk memastikan validitas perbandingan, seluruh dataset diproses menggunakan prosedur seragam untuk meminimalkan bias (Pedregosa et al., 2018; Liu et al., 2025). Tahapan meliputi:

a. **Cleaning Data:** Menghapus karakter yang tidak diperlukan dari teks, seperti tanda baca, simbol, angka, tautan (URL), dan spasi berlebih (*whitespace*) agar teks lebih bersih dari noise.
b. **Case Folding:** Mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menyeragamkan bentuk kata (misalnya, "IndoBERT" dan "indobert" dianggap sama).
c. **Stopword Removal:** Menghapus kata-kata umum yang sering muncul tetapi tidak memiliki informasi atau makna signifikan terhadap analisis sentimen atau konteks (seperti "yang", "dan", "di", "dari").
d. **Stemming:** Mengembalikan kata berimbuhan ke bentuk dasar aslinya dengan menghapus awalan, akhiran, dan sisipan (misalnya, "mengimplementasikan" menjadi "implementasi").
e. **Normalization:** Konversi kata tidak baku (bahasa gaul, singkatan, atau *typo*) ke bentuk baku sesuai Kamus Besar Bahasa Indonesia (KBBI) atau kamus slang lokal.
f. **Tokenization:** Memecah kalimat atau teks panjang menjadi satuan kata atau token-token individual (misalnya, memecah kalimat "Arsitektur RAG sangat efektif" menjadi array token: ['arsitektur', 'rag', 'sangat', 'efektif']) sebagai persiapan penting sebelum tahap ekstraksi fitur atau pemodelan lanjutan.

Pemodelan Topik (*Latent Dirichlet Allocation*)

LDA digunakan sebagai metode ekstraksi fitur untuk mengidentifikasi tiga aspek utama dalam setiap dataset. Dengan menghasilkan distribusi probabilistik kata terhadap topik dan topik terhadap dokumen, LDA mampu memetakan kumpulan opini ke dalam aspek-aspek spesifik yang relevan (Blei et al., 2003). Pendekatan *unsupervised* ini memungkinkan pemisahan opini secara otomatis ke dalam tiga klaster aspek utama, yang kemudian dilabeli secara manual untuk kebutuhan klasifikasi sentimen selanjutnya (Akhmedov et al., 2021; Zhu et al., 2024).

Strategi Klasifikasi

Untuk menangani kompleksitas opini, maka strategi klasifikasi dilakukan dengan mengembangkan model khusus per aspek. Artinya, setiap dataset memiliki tiga model klasifikasi yang terpisah, di mana masing-masing model dilatih khusus untuk mendeteksi sentimen pada satu aspek yang telah diidentifikasi oleh LDA. Tabel 2 menjelaskan ringkasan strategi algoritmanya sebagai berikut:

Tabel 2. Ringkasan Strategi Klasifikasi

Algoritma	Penggunaan	Justifikasi
Multinomial Naïve Bayes	Twitter	Efisien untuk teks pendek & informal (Singh, 2025)
XGBoost	Google Play	Tangguh menangani fitur teknis (Chen & Guestrin, 2016; Hossain & Logofătu, 2025)
AdaBoost	Female Daily	Optimal untuk dataset naratif-kompleks (Freund & Schapire, 2017; Sojanv, 2025)

Metrik Evaluasi

Evaluasi performa menggunakan metrik *Precision*, *Recall*, *F1-score*, dan *Accuracy*. Mengingat dataset opini sering memiliki imbalanced data, maka *F1-score* ditetapkan sebagai metrik prioritas karena

memberikan representasi yang adil antara *Precision* dan *Recall* (Wang et al., 2023; Lee, 2025). Sebagai ancaman validitas, penelitian ini menyadari bahwa variasi volume data (N) antar platform dapat memengaruhi stabilitas model. Oleh karena itu validasi dilakukan dengan membandingkan performa model dalam konteks karakteristik platform yang spesifik.

HASIL DAN PEMBAHASAN

Efektivitas Ekstraksi Aspek LDA

Tahap ekstraksi aspek menggunakan *Latent Dirichlet Allocation* (LDA) bertujuan untuk memetakan opini pengguna ke dalam kategori yang relevan sebelum masuk ke proses klasifikasi sentimen. Hasil ini menunjukkan bagaimana model melihat struktur semantik dari data yang berasal dari tiga platform berbeda. Dengan memanfaatkan distribusi kata yang dominan, LDA mampu mengidentifikasi topik yang konsisten dan merepresentasikan aspek yang menjadi fokus utama pengguna.

Tabel 3. Hasil Identifikasi Aspek melalui LDA pada Berbagai Platform

Platform	Aspek 1 (Keywords Utama)	Aspek 2 (Keywords Utama)	Aspek 3 (Keywords Utama)	Aspek 4 (Keywords Utama)
Twitter (Garuda Indonesia)	Fasilitas, Pesawat, Nyaman, Harga, Bagus.	Refund, Call, Center, Kecewa, Ubah.	Jadwal, Tepat, Delay, Senang, Sedih.	-
Google Play (AdaKami)	Data, Akun, Tolak, Limit, Hapus.	Proses, Cepat, Cair, Mudah, Bantu.	Bayar, Tagih, Tempo, Lunas, Kecewa.	-
Female Daily (Facial Wash)	Jerawat, Breakout, Timbul, Acne, Kulit.	Bersih, Kering, Lembut, Segar, Minyak.	Cocok, Bagus, Harga, Nyaman, Beli.	Tekstur, Gel, Busa, Gentle, Bilas.

Hasil ekstraksi LDA pada Twitter yang mereview Garuda Indonesia memperlihatkan tiga aspek utama yang konsisten dengan karakteristik cuitan singkat dan emosional. Aspek pertama berhubungan dengan fasilitas penerbangan seperti kursi, kebersihan, dan kenyamanan; aspek kedua menyoroti pelayanan, khususnya interaksi dengan call center dan proses refund; sedangkan aspek ketiga berfokus pada jadwal penerbangan, baik ketepatan maupun keterlambatan. Pada Google Play yang mereview aplikasi AdaKami, LDA juga menghasilkan tiga aspek yang lebih teknis. Aspek pertama berkaitan dengan data dan akun pengguna, termasuk limit dan verifikasi; aspek kedua menyoroti proses pencairan dana yang cepat dan mudah; sementara aspek ketiga berhubungan dengan tagihan pembayaran, tempo jatuh tempo, dan keluhan terkait status pelunasan. Berbeda dengan dua platform sebelumnya, ulasan di Female Daily tentang produk *facial wash* menghasilkan empat aspek yang lebih deskriptif dan naratif. Aspek pertama berhubungan dengan efek di kulit seperti jerawat dan *breakout*; aspek kedua menyoroti sensasi penggunaan seperti bersih, segar, dan lembut; aspek ketiga mencakup evaluasi produk secara umum seperti kecocokan, harga, dan kenyamanan; sedangkan aspek keempat lebih detail pada fisik produk seperti tekstur, busa, dan aroma.

Konsistensi hasil ini menunjukkan bahwa LDA mampu beradaptasi dengan variasi bahasa dan domain. Pada platform yang cenderung singkat dan emosional seperti Twitter, aspek yang muncul lebih langsung terkait pengalaman pengguna. Sebaliknya, pada platform yang lebih teknis seperti Google Play, aspek yang dihasilkan lebih fungsional dan berhubungan dengan reliabilitas sistem. Sementara itu, pada platform yang naratif seperti Female Daily, aspek yang muncul lebih kaya dan mendalam, mencerminkan

pengalaman sensorik dan evaluasi personal. Hal ini menegaskan bahwa LDA tidak hanya sekadar memisahkan kata, tetapi juga menangkap struktur semantik yang relevan dengan konteks masing-masing.

Selain itu, penting untuk dicatat bahwa kata kunci yang ditampilkan merupakan hasil distribusi probabilistik dengan bobot tertinggi, bukan hasil *labeling* manual. Dengan demikian, pembentukan aspek dilakukan secara objektif dan konsisten berdasarkan data. Temuan ini sejalan dengan penelitian Setiadi et al. (2024) yang menunjukkan efektivitas LDA dalam mengidentifikasi aspek produk pada ulasan Amazon dengan tingkat akurasi tinggi, serta studi Hidayati (2023) yang menekankan peran LDA dalam meningkatkan kualitas analisis aspek pada ulasan hotel berbahasa Indonesia. Kedua penelitian ini memperkuat validitas metodologi yang digunakan, sekaligus menegaskan bahwa hasil ekstraksi aspek dapat dianggap sahih dan siap digunakan sebagai unit analisis dalam tahap klasifikasi sentimen berikutnya.

Analisis Performa Model per Aspek

Tahap berikutnya setelah ekstraksi aspek adalah mengevaluasi performa model klasifikasi sentimen pada setiap aspek yang dihasilkan LDA. Evaluasi dilakukan menggunakan empat metrik utama: *Precision*, *Recall*, *F1-Score*, dan Akurasi. Karena distribusi data antar aspek tidak selalu seimbang, *F1-Score* menjadi indikator utama yang digunakan untuk menilai efektivitas model dalam menangani kasus *True Positive* dan *False Negative* secara bersamaan.

Tabel 4. Evaluasi Performa Model Klasifikasi per Aspek pada Berbagai Platform

Platform	Aspek	Precision	Recall	F1-Score	Akurasi
Twitter (Multinomial Naïve Bayes)	Aspek 1	0.73	0.72	0.72	0.73
	Aspek 2	0.77	0.69	0.72	0.80
	Aspek 3	0.83	0.82	0.83	0.86
	Rata-rata	**0.78**	**0.74**	**0.76**	**0.80**
Google Play (XGBoost)	Aspek 1	0.90	0.87	0.88	0.92
	Aspek 2	0.77	0.75	0.75	0.93
	Aspek 3	0.89	0.78	0.83	0.90
	Rata-rata	**0.85**	**0.80**	**0.82**	**0.92**
Female Daily (AdaBoost)	Aspek 1	0.69	0.68	0.64	0.70
	Aspek 2	0.77	0.74	0.76	0.80
	Aspek 3	0.75	0.66	0.69	0.78
	Aspek 4	0.78	0.67	0.71	0.85
	Rata-rata	**0.75**	**0.69**	**0.70**	**0.79**

Hasil evaluasi menunjukkan adanya variasi performa antar aspek dan antar platform. Pada Twitter, model *Multinomial Naïve Bayes* (MNB) bekerja cukup baik karena teks singkat dengan kosakata emosional yang relatif konsisten. Aspek jadwal penerbangan memiliki *F1-Score* tertinggi (0.83), yang menunjukkan bahwa distribusi kata seperti "delay", "tepat", dan "jadwal" lebih sering muncul dan mudah dipetakan oleh MNB. Sebaliknya, aspek pelayanan lebih rendah karena kosakata yang digunakan lebih variatif ("call center", "refund", "respon"), sehingga probabilitas kata antar kelas menjadi lebih tipis dan *smoothing* α berperan penting untuk mencegah probabilitas nol. Hal ini sesuai dengan teori MNB yang sensitif terhadap kata jarang muncul, tetapi tetap mampu menggeneralisasi dengan bantuan *smoothing*.

Pada Google Play, XGBoost menunjukkan performa yang sangat stabil dengan rata-rata *F1-Score* 0.82. Hal ini dapat dijelaskan oleh sifat XGBoost yang mampu menangani interaksi fitur kompleks dan variasi panjang teks. Aspek data dan akun memperoleh skor tertinggi karena ulasan terkait verifikasi dan limit akun menggunakan kosakata teknis yang konsisten. Namun, aspek proses pencairan sedikit lebih

rendah karena adanya variasi narasi pengguna yang mencampurkan pengalaman positif dan negatif dalam satu ulasan, sehingga pohon keputusan perlu membagi fitur lebih banyak untuk mencapai klasifikasi yang tepat.

Sementara itu, pada Female Daily, performa model AdaBoost relatif lebih rendah dibandingkan dua platform lainnya, dengan rata-rata *F1-Score* 0.70. Hal ini dapat dijelaskan oleh sifat ulasan yang lebih naratif dan deskriptif, sehingga kosakata yang digunakan untuk menggambarkan efek produk lebih beragam dan kadang tumpang tindih antar aspek. Menariknya, aspek tekstur memiliki skor lebih tinggi dibanding aspek efek kulit, karena deskripsi tekstur ("gel", "busa", "lembut") lebih konsisten dibandingkan narasi tentang jerawat atau breakout yang sering kali subjektif. Ini sesuai dengan cara kerja AdaBoost yang memperkuat *weak learners* pada pola kosakata konsisten, tetapi kesulitan ketika variasi kata terlalu tinggi.

Secara keseluruhan, hasil ini menegaskan bahwa algoritma *Boosting* (XGBoost dan AdaBoost) lebih unggul dalam menjaga stabilitas performa antar aspek dibandingkan MNB, yang cenderung sensitif terhadap distribusi kata tidak seimbang. Namun, performa terbaik tetap bergantung pada karakteristik platform: MNB cukup efektif untuk teks pendek dan emosional di Twitter, XGBoost unggul pada data teknis di Google Play, sedangkan AdaBoost cukup adaptif untuk data deskriptif di Female Daily. Temuan ini sejalan dengan penelitian Chen & Guestrin (2016) yang menunjukkan keunggulan XGBoost dalam menangani data kompleks dengan interaksi fitur yang tinggi, serta studi Hidayati (2023) yang menekankan efektivitas algoritma *Boosting* dalam meningkatkan akurasi analisis sentimen pada teks dengan variasi konteks.

Analisis Komparatif Algoritma Lintas Platform

Untuk memberikan gambaran yang lebih ringkas dan komprehensif, performa model klasifikasi dibandingkan secara agregat lintas platform. Ringkasan ini menekankan rata-rata *F1-Score* sebagai indikator utama, karena distribusi sentimen antar aspek tidak seimbang.

Tabel 5. Ringkasan Performa Algoritma Lintas Platform

Platform	Algoritma Klasifikasi	Rata-rata F1-Score (Agregat)	Kesimpulan Kinerja
Twitter	Multinomial Naïve Bayes	0.75	Efektif untuk teks pendek dan emosional, memanfaatkan distribusi frekuensi kata, meski sensitif terhadap kosakata jarang.
Google Play	XGBoost	0.82	Unggul pada data teknis dengan kosakata konsisten, stabil dalam menangani interaksi fitur kompleks.
Female Daily	AdaBoost	0.70	Adaptif pada ulasan naratif, namun performa menurun pada aspek dengan kosakata beragam dan tumpang tindih.

Isi tabel memperlihatkan bahwa XGBoost pada Google Play memiliki rata-rata *F1-Score* tertinggi, diikuti oleh MNB pada Twitter, dan AdaBoost pada Female Daily. Hal ini menunjukkan bahwa algoritma *Boosting* lebih unggul pada platform dengan teks kompleks atau teknis, sementara MNB masih cukup kompetitif pada teks pendek dan emosional.

Analisis lebih lanjut menegaskan bahwa kesesuaian algoritma sangat dipengaruhi oleh karakteristik data. MNB bekerja baik pada Twitter karena teks singkat dengan kosakata emosional yang relatif konsisten. MNB menghitung probabilitas berdasarkan frekuensi kata, sehingga aspek dengan kata-kata sering muncul

("delay", "refund", "tiket") lebih mudah diklasifikasikan. Namun, aspek dengan kosakata jarang atau variatif lebih sulit, sehingga *smoothing* α berperan penting untuk menjaga performa. XGBoost unggul pada Google Play karena mampu menangani interaksi fitur kompleks dan variasi panjang teks. Algoritma ini membangun pohon keputusan berlapis yang memperkuat prediksi pada kosakata teknis yang konsisten ("akun", "verifikasi", "cair"). AdaBoost cukup adaptif pada Female Daily, tetapi performanya menurun pada aspek dengan kosakata beragam. Hal ini sesuai dengan sifat AdaBoost yang memperkuat *weak learners* pada pola konsisten, namun tetap sensitif terhadap variasi kata yang tinggi dalam ulasan naratif.

Hal ini menegaskan bahwa kompleksitas teks berdampak langsung pada performa algoritma. Model probabilistik seperti MNB lebih cocok untuk teks informal dan pendek, sedangkan model *Boosting* lebih stabil pada teks panjang, teknis, atau naratif. Meskipun kinerja model bervariasi antar platform, integrasi LDA terbukti secara konsisten mampu menyediakan *feature space* yang representatif untuk berbagai algoritma klasifikasi, yang dibuktikan dengan skor *F1-Score* agregat di atas 0.70 di seluruh platform.

Temuan ini sejalan dengan penelitian Chen & Guestrin (2016) yang menunjukkan keunggulan XGBoost dalam menangani data kompleks dengan interaksi fitur yang tinggi, serta studi Zhang et al. (2022) yang menekankan efektivitas algoritma *Boosting* dalam meningkatkan akurasi analisis sentimen pada ulasan e-commerce dengan variasi konteks.

Sintesis Lintas Platform

Sintesis hasil analisis menunjukkan bahwa karakteristik data seperti panjang teks, tingkat formalitas, dan gaya bahasa berkorelasi langsung dengan performa algoritma klasifikasi. Pada Twitter, teks yang pendek, emosional, dan sering menggunakan kosakata berulang membuat MNB bekerja cukup efektif. Distribusi kata yang konsisten seperti "delay", "refund", atau "nyaman" menghasilkan probabilitas yang lebih stabil, sehingga *F1-Score* rata-rata mencapai 0.76. Namun, aspek dengan kosakata jarang atau variatif lebih sulit dipetakan, sehingga *smoothing* α berperan penting untuk menjaga performa.

Sebaliknya, pada Google Play, ulasan yang lebih panjang dan teknis memberikan keuntungan bagi algoritma XGBoost. Model ini unggul karena mampu menangani interaksi fitur kompleks dan variasi panjang teks. Kosakata teknis yang konsisten seperti "akun", "verifikasi", dan "cair" membuat XGBoost mencapai *F1-Score* rata-rata tertinggi (0.82). Hal ini menunjukkan bahwa semakin formal dan teknis data, semakin efektif algoritma *Boosting* berbasis pohon keputusan.

Pada Female Daily, ulasan naratif dan deskriptif menantang algoritma AdaBoost. Meskipun adaptif, performa menurun pada aspek dengan kosakata beragam dan tumpang tindih, seperti deskripsi efek kulit ("jerawat", "breakout"). Namun, aspek dengan kosakata konsisten seperti tekstur ("gel", "busa", "lembut") tetap menghasilkan skor lebih tinggi. Dengan rata-rata *F1-Score* 0.70, AdaBoost terbukti cukup stabil, tetapi tetap sensitif terhadap variasi kata yang tinggi dalam ulasan naratif.

Dari sintesis ini, dapat ditarik *insight* bahwa pemilihan algoritma harus mempertimbangkan karakteristik platform. MNB direkomendasikan untuk platform dengan teks pendek, informal, dan emosional seperti Twitter. XGBoost lebih sesuai untuk platform dengan ulasan teknis dan formal seperti Google Play. AdaBoost dapat digunakan pada platform dengan ulasan naratif seperti Female Daily, meskipun perlu diantisipasi adanya penurunan performa pada aspek dengan kosakata beragam. Dengan demikian, integrasi LDA sebagai tahap ekstraksi aspek terbukti mampu menyediakan *feature space* yang representatif, sementara pemilihan algoritma klasifikasi harus disesuaikan dengan sifat data agar hasil evaluasi optimal.

Temuan ini sejalan dengan penelitian Kowsari et al. (2019) yang menekankan pentingnya pemilihan

algoritma sesuai karakteristik teks untuk meningkatkan akurasi analisis sentimen, serta studi Hidayati (2023) yang menunjukkan bahwa kombinasi LDA dengan algoritma *Boosting* dapat meningkatkan kualitas analisis aspek pada ulasan berbahasa Indonesia.

KESIMPULAN

Penelitian ini menunjukkan bahwa model *hybrid* LDA-ML efektif secara universal dalam mengekstraksi aspek dan mengklasifikasikan sentimen lintas platform, namun performanya tetap sensitif terhadap karakteristik data. Teks pendek dan emosional seperti di Twitter lebih sesuai dengan Multinomial Naïve Bayes, sementara ulasan teknis di Google Play lebih stabil dengan XGBoost, dan ulasan naratif di Female Daily relatif adaptif dengan AdaBoost meskipun menghadapi tantangan kosakata beragam.

Temuan utama dari penelitian ini adalah bahwa algoritma *Boosting* (XGBoost dan AdaBoost) secara umum lebih stabil dibandingkan MNB, terutama pada platform dengan teks panjang, formal, atau naratif. Namun, MNB tetap kompetitif pada teks pendek dan informal, sehingga tidak ada satu algoritma yang unggul di semua konteks. Jawaban atas pertanyaan penelitian menegaskan bahwa XGBoost adalah algoritma paling stabil dalam menangani variasi konteks dan kosakata, dengan rata-rata *F1-Score* tertinggi di antara platform yang diuji.

Untuk penelitian mendatang, beberapa arah dapat dikembangkan. Pertama, memperluas dataset dengan melibatkan lebih banyak domain dan bahasa agar hasil lebih generalis. Kedua, mengintegrasikan pendekatan *deep learning* seperti BiLSTM, Transformer, atau SetFit untuk membandingkan performa dengan algoritma klasik dan *Boosting*. Ketiga, mengeksplorasi normalisasi *microtext* dan *multi-task learning* (emosi + sentimen) agar analisis lebih kaya dan relevan dengan konteks sosial digital di Indonesia.

DAFTAR PUSTAKA

Akhmedov, F., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using TDS model. *Applied Sciences*, *11*(23), 11091. (https://doi.org/10.3390/app112311091)

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022. (http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf)

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. (https://doi.org/10.1145/2939672.2939785)

Freund, Y., & Schapire, R. E. (2017). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. (https://doi.org/10.1006/jcss.1997.1504)

Hidayati, N. N. (2023). Improving aspect-based sentiment analysis for hotel reviews with Latent Dirichlet Allocation and machine learning algorithms. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, *9*(2), 144–159. (https://doi.org/10.26594/register.v9i2.144-159)

Hossain, S., & Logofătu, D. (2025). Hybrid deep learning and gradient boosting for superior sentiment analysis. Communications in Computer and Information Science (CCIS). Springer. (https://doi.org/10.1007/978-3-031-96196-0_15)

Jayakody, D., Isuranda, K., Malkith, A. V. A., de Silva, N., Ponnamperuma, S. R., Sandamali, G. G. N., &

Sudheera, K. L. K. (2024). Aspect-based sentiment analysis techniques: A comparative study. *arXiv*. (https://doi.org/10.48550/arXiv.2407.02834)

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information, 10*(4), 150. (https://doi.org/10.3390/info10040150)

Lee, S. (2025). The ultimate F1 score guide for imbalanced classes in sentiment analysis. *Number Analytics Journal*. (https://numberanalytics.com/f1-score-guide)

Lghaouch, E. M., Ounacer, S., Ardchir, S., & Azzouazi, M. (2023). Enhancing sentiment analysis through topic modeling: Comprehensive overview. *Springer*. (https://link.springer.com/article/10.1007/springer-absa-2023)

Liu, L., Li, D., Yue, C., Gao, X., & Zhu, Y. (2025). Syntactic denoising and multi-strategy auxiliary enhancement for ABSA. *PLoS One, 20*(8), e0329018. (https://doi.org/10.1371/journal.pone.0329018)

Nazir, A. A., Saiful, I., & Al Sohan, M. F. A. (2025). Real-time review analysis with a sentiment-integrated hybrid LDA approach. In *2025 IEEE 4th International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)* (pp. 274–279). IEEE. (https://doi.org/10.1109/RAAICON69033.2025.11502077)

Negi, G., Sarkar, R., Zayed, O., & Buitelaar, P. (2024). A hybrid approach to aspect-based sentiment analysis using transfer learning. In *Proceedings of LREC-COLING 2024* (pp. 647–658). ELRA and ICCL. (https://aclanthology.org/2024.lrec-main.56/)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2018). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. (http://jmlr.org/papers/v12/pedregosa11a.html)

Said, M., Smairi, N., & Abadlia, H. (2025). LLM powered DeepSeek and BiLSTM pipeline used in aspect-based sentiment analysis. In *2025 IEEE 37th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 813–817). IEEE. (https://doi.org/10.1109/ICTAI66417.2025.00118)

Setiadi, D. R. I. M., Marutho, D., & Setiyanto, N. A. (2024). Comprehensive exploration of machine and deep learning classification methods for aspect-based sentiment analysis with Latent Dirichlet Allocation topic modeling. *Journal of Future Artificial Intelligence and Technologies*. (https://doi.org/10.62411/faith.2024-3)

Shukla, P., Kumar, R., Dwivedi, V. K., & Singh, A. K. (2026). Aspect based sentiment analysis: A systematic review, taxonomy, applications, and future research directions. *Computer Science Review, 61*, 100924. (https://doi.org/10.1016/j.cosrev.2026.100924)

Singh, W. (2025). Aspect-based sentiment analysis for restaurant reviews using Naive Bayes. *arXiv*. (https://doi.org/10.48550/arXiv.2501.01234)

Sojanv, S. (2025). Sentiment analysis using AdaBoost for text classification. *arXiv*. (https://doi.org/10.48550/arXiv.2502.04567)

Symeonidis, S., Effrosynidis, D., & Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis. *Expert Systems with Applications, 110*, 298–310. (https://doi.org/10.1016/j.eswa.2018.06.022)

Wang, J., Li, Y., & Zhang, H. (2023). Evaluating sentiment analysis models on imbalanced datasets: A comparative study. *Expert Systems with Applications, 223*, 119812. (https://doi.org/10.1016/j.eswa.2023.119812)