

Analisis Sentimen Masyarakat Terhadap Sekolah Islam Terpadu berbasis Aspek menggunakan *Lexicon based* dan *Machine Learning*

Aspect-based Sentiment Analysis of Public Opinions on Integrated Islamic Schools using Lexicon based and Machine Learning Approaches

¹Fitriani Muttakin*, ²Daffa Takratama Savra

^{1,2}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

^{1,2}Jl. H.R. Soebrantas No. 155, KM. 15, Simpang Baru, Panam, Kecamatan Tampan, Kota Pekanbaru, Riau, Indonesia

e-mail: fitrianimuttakin@uin-suska.ac.id, daffatakratama13@gmail.com

(received: 13 November 2025, revised: 31 January 2026, accepted: 1 February 2026)

Abstrak

Penelitian ini bertujuan untuk menelaah persepsi masyarakat terhadap Sekolah Islam Terpadu melalui analisis sentimen berbasis aspek dengan mengintegrasikan *Latent Dirichlet Allocation*, *Lexicon Based* dan *Deep Neural Network*. LDA digunakan untuk mengekstraksi struktur topik yang merepresentasikan konteks semantik dari ulasan masyarakat *Lexicon Based* untuk analisis sentimennya, sedangkan DNN menginferensi orientasi sentimen berdasarkan representasi tersebut. Pendekatan ini berupaya menggabungkan keunggulan pemodelan probabilistik dan pembelajaran mendalam guna memperoleh pemahaman yang lebih komprehensif terhadap opini publik. Analisis dilakukan terhadap kumpulan ulasan daring sejumlah 2.280 ulasan, kemudian melalui tahap praproses menjadi 1.438 data yang diolah dengan kombinasi LDA–DNN yang terbukti mampu mengidentifikasi dimensi opini secara lebih kontekstual dan meningkatkan interpretabilitas hasil analisis. Evaluasi empiris menunjukkan bahwa model mencapai akurasi 63,89% untuk klasifikasi aspek dan 93,06% untuk klasifikasi sentimen, lebih tinggi dibandingkan pendekatan *K-Means–LSA* dan *K-Means–PCA* yang masing-masing mencatat 45,14% dan 31,94% untuk klasifikasi aspek serta 92,36% untuk klasifikasi sentimen. Temuan ini menegaskan keunggulan pemodelan topik berbasis probabilistik dalam menangkap hubungan semantik yang kompleks, serta memberikan kontribusi metodologis bagi pengembangan analisis sentimen dalam konteks pendidikan Islam terpadu.

Kata kunci: DNN, sekolah islam terpadu, *lexicon based*, LDA, analisis sentimen,

Abstract

This study aims to examine public perceptions of Integrated Islamic Schools through aspect-based sentiment analysis by integrating Latent Dirichlet Allocation (LDA), a lexicon-based approach, and a Deep Neural Network (DNN). LDA was employed to extract topic structures representing the semantic context of public reviews, the lexicon-based method was used to determine sentiment polarity, and the DNN was applied to infer sentiment orientation based on the extracted representations. This approach combines the strengths of probabilistic topic modeling and deep learning to achieve a more comprehensive understanding of public opinion. The analysis was conducted on 2,280 online reviews, which were reduced to 1,438 valid records after preprocessing. The proposed LDA–DNN combination proved effective in identifying opinion dimensions in a more contextual manner and improving the interpretability of sentiment analysis results. Empirical evaluation showed that the model achieved an accuracy of 63.89% for aspect classification and 93.06% for sentiment classification, outperforming the K-Means–LSA and K-Means–PCA approaches, which achieved 45.14% and 31.94% accuracy for aspect classification, respectively, and 92.36% accuracy for sentiment classification. These findings highlight the superiority of probabilistic topic modeling in capturing complex semantic relationships and provide a methodological contribution to the development of sentiment analysis in the context of integrated Islamic education.

Keywords: DNN, Integrated Islamic Schools, Lexicon Based, LDA, Sentiment Analysis

1 Pendahuluan

Perkembangan era disrupsi informasi dan keterbukaan digital telah mengubah cara masyarakat menilai dan menyampaikan opini terhadap layanan publik, termasuk sektor pendidikan. Masyarakat kini dapat secara langsung memberikan ulasan dan penilaian terhadap lembaga pendidikan melalui berbagai platform daring. Salah satu model pendidikan yang banyak menjadi sorotan adalah Sekolah Islam Terpadu (SIT), yaitu lembaga pendidikan yang mengintegrasikan nilai-nilai keislaman dengan kurikulum formal nasional. Platform berbasis lokasi seperti *Google Maps* memberikan ruang bagi masyarakat untuk menuliskan komentar dan memberikan rating terhadap sekolah, sehingga menghasilkan data opini publik yang dapat dimanfaatkan untuk memahami persepsi masyarakat terhadap kualitas layanan pendidikan Islam terpadu.

Analisis terhadap opini publik tersebut dapat dilakukan melalui analisis sentimen seperti pada penelitian Dahlia dkk meneliti pengaruh sentimen orang tua terhadap kualitas pendidikan agama dan umum dalam keputusan memilih Sekolah Islam Terpadu [1]. Penelitian lain menerapkan machine learning untuk mengidentifikasi sentimen masyarakat terhadap sistem pendidikan pesantren melalui media sosial [2][3], Penelitian-penelitian tersebut menelusuri aspek yang menjadi sumber pembentukan sentimen yang menyebabkan hasil analisis namun belum sepenuhnya menggambarkan persepsi masyarakat yang bersifat multidimensi dan kontekstual sehingga melalui *Aspect-Based Sentiment Analysis (ABSA)* dapat dipahami persepsi masyarakat terhadap aspek-aspek spesifik tertentu yang memengaruhi penilaian masyarakat.

Kesenjangan tersebut menunjukkan perlunya pendekatan yang mampu memodelkan distribusi topik untuk memperoleh representasi semantik yang lebih kaya, yaitu dengan kombinasi metode *Latent Dirichlet Allocation (LDA)*, *Lexicon-Based* dan *Deep Neural Network (DNN)*. LDA untuk menghasilkan representasi topik dan sentimen yang lebih komprehensif [4] kemudian *Lexicon based* memanfaatkan kamus kata dengan bobot polaritas positif, negatif, atau netral cukup efektif digunakan untuk mengidentifikasi kata emosional [5], dan DNN dalam sebagai model untuk klasifikasi sentimen yang lebih akurat [6] [7]. Dibandingkan dengan *Latent Semantic Analysis (LSA)* [8] dan *Principal Component Analysis (PCA)* [9] [10], LDA menghasilkan representasi topik yang lebih semantik dan interpretatif sehingga lebih sesuai untuk analisis teks yang kompleks dan kontekstual [11].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Sekolah Islam Terpadu berbasis aspek dengan mengombinasikan metode LDA, *Lexicon Based* dan DNN dengan tujuan untuk mengekstraksi aspek, menentukan sentimen, dan menggabungkannya secara lebih akurat dibandingkan metode konvensional sehingga memberikan kontribusi terhadap pengembangan pendekatan *Aspect-Based Sentiment Analysis*. Hasil penelitian diharapkan dapat memberikan gambaran persepsi sentimen masyarakat yang dikelompokkan dalam aspek terhadap layanan pendidikan Islam terpadu yang berguna bagi pengambil kebijakan dalam peningkatan mutu dan pelayanan pendidikan berbasis nilai keislaman.

2 Tinjauan Literatur

Penelitian mengenai analisis sentimen berbasis aspek telah berkembang pesat seiring dengan meningkatnya volume data opini publik di media sosial dan platform daring. Beberapa penelitian terkini mencoba mengombinasikan LDA dengan pendekatan Lexicon-based guna memperkaya analisis sentimen dengan informasi polaritas kata. Pendekatan *Lexicon-based* memanfaatkan kamus sentimen untuk menentukan nilai positif, negatif, atau netral pada setiap kata [12]. Integrasi LDA dan Lexicon memungkinkan pengenalan aspek sekaligus estimasi polaritas yang muncul dalam setiap topik [13]. Meskipun demikian, pendekatan berbasis lexicon masih menghadapi keterbatasan pada konteks domain karena makna suatu kata dapat berubah tergantung pada aspek yang dibahas. Misalnya, kata “murah” dianggap positif dalam konteks biaya, tetapi negatif dalam konteks kualitas layanan [14]. Hal ini menunjukkan bahwa metode berbasis lexicon membutuhkan dukungan model pembelajaran yang mampu menangkap hubungan semantik dan konteks kalimat secara lebih mendalam. LDA digunakan sebagai model distribusi topik secara probabilistik. LDA tidak hanya mengidentifikasi kata-kata dominan dalam suatu kumpulan dokumen, tetapi juga menyajikan keterhubungan antar kata pada tingkat konteks. Lebih lanjut, keluaran LDA berupa representasi topik digunakan sebagai fitur yang lebih bermakna bagi algoritma klasifikasi lanjutan [15]. Sehingga Integrasi LDA dan Lexicon based memungkinkan pengenalan aspek sekaligus estimasi polaritas yang

<http://sistemasi.ftik.unisi.ac.id>

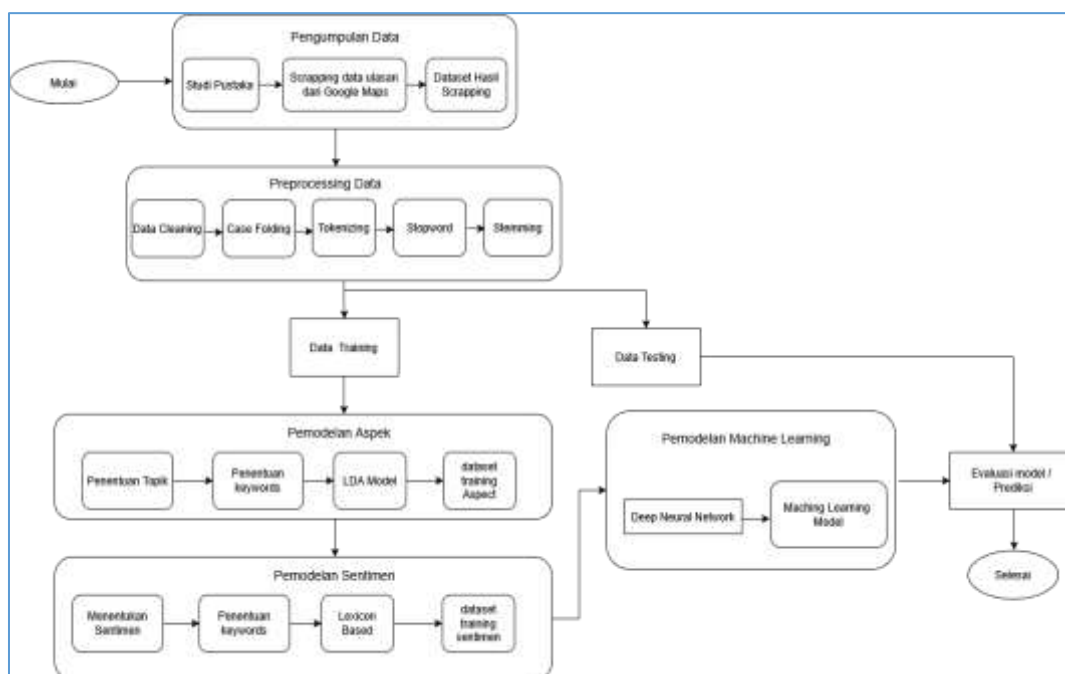
muncul dalam setiap topik. Pendekatan ini memungkinkan peneliti untuk mengidentifikasi aspek atau tema spesifik yang menjadi fokus penilaian masyarakat dan menentukan orientasi sentimennya dengan menilai kata dan frasa dalam teks sesuai dengan makna semantiknya untuk menentukan orientasi dokumen teks [16].

Dalam beberapa tahun terakhir, pendekatan *Deep Learning* seperti *Convolutional Neural Network (CNN)*, *Long Short-Term Memory (LSTM)*, dan *Deep Neural Network (DNN)* menunjukkan hasil yang signifikan dalam analisis sentimen karena kemampuannya mempelajari representasi fitur nonlinier dari data teks [12]. Model DNN mampu menangkap pola semantik yang kompleks tanpa memerlukan ekstraksi fitur manual, sehingga menghasilkan akurasi yang lebih tinggi dalam klasifikasi sentimen. Namun demikian, pendekatan ini bersifat *black-box*, sehingga interpretabilitasnya rendah dan sulit digunakan untuk memahami alasan di balik keputusan model [17]. Selain itu, model *Deep Learning* umumnya membutuhkan jumlah data latih yang besar agar mencapai performa optimal, yang menjadi kendala pada penelitian berbasis domain tertentu seperti pendidikan Islam.

Berdasarkan studi literatur tersebut, maka penelitian ini berfokus pada pengembangan model hibrid yang menggabungkan LDA, *Lexicon-Based*, dan DNN untuk menganalisis sentimen masyarakat terhadap Sekolah Islam Terpadu berdasarkan aspek-aspek utama seperti kurikulum, fasilitas, pelayanan, dan manajemen. Pendekatan ini diharapkan dapat menghasilkan model yang tidak hanya akurat secara prediktif, tetapi juga informatif secara semantik dalam menggambarkan persepsi publik terhadap kualitas pendidikan Islam di Indonesia.

3 Metode Penelitian

Secara umum, metodologi pada penelitian ini meliputi beberapa tahapan yaitu Pengumpulan data, Preprocessing data, Pemodelan Aspek, Pemodelan sentimen dan Evaluasi model, pada Gambar 1.



Gambar 1 Metode penelitian

Secara terinci, penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis sentimen masyarakat terhadap Sekolah Islam Terpadu berdasarkan aspek-aspek tertentu. Proses penelitian dilakukan melalui beberapa tahapan utama yang meliputi pengumpulan data, praproses data, pemodelan aspek menggunakan LDA, analisis sentimen berbasis Lexicon, pemodelan klasifikasi menggunakan DNN, serta evaluasi model untuk mengukur performa sistem yang diusulkan.

Tahapan pemodelan dimulai dengan proses ekstraksi aspek menggunakan LDA untuk mengidentifikasi topik-topik dominan dalam ulasan publik. Setiap ulasan direpresentasikan sebagai distribusi topik, dan label aspek ditentukan berdasarkan topik dengan probabilitas tertinggi. Ekstraksi

aspek dilakukan menggunakan LDA untuk mengidentifikasi topik-topik dominan pada korpus ulasan. LDA dikonfigurasi menggunakan parameter utama: jumlah topik K dan prior Dirichlet α untuk mengatur kerapatan distribusi topik pada dokumen. pada penelitian ini, jumlah topik ditetapkan $K=4$ untuk merepresentasikan empat aspek utama yaitu aspek kurikulum, fasilitas, pelayanan dan aspek umum. Ulasan yang telah terklasifikasi ke dalam aspek tertentu kemudian dianalisis polaritasnya menggunakan pendekatan *lexicon-based* dengan kamus InSet untuk menentukan kecenderungan sentimen positif dan negatif. Ulasan ini kemudian dikonversi menjadi label sentimen menggunakan aturan threshold yaitu positif jika $S(d)>0$ dan negative jika $S(d) \leq 0$.

Dengan mekanisme ini, keluaran tahap *lexicon-based* berupa skor sentimen sekaligus label sentimen awal yang digunakan sebagai bagian dari target pada proses klasifikasi. Hasil dari tahap LDA dan *lexicon-based* digunakan sebagai fitur masukan bagi model *Deep Neural Network (DNN)* multi-task, yang terdiri dari embedding layer dan beberapa dense layer, dengan dua keluaran softmax untuk klasifikasi aspek dan sentimen. Model dioptimasi menggunakan hyperparameter KerasTuner RandomSearch, kemudian model dikompilasi menggunakan Adam optimizer dan categorical cross-entropy loss function untuk kedua output. Proses pelatihan dilakukan hingga maksimum 100 epoch dengan batch size 32, serta menerapkan EarlyStopping (patience = 15, restore best weights) untuk mencegah overfitting. Selanjutnya dievaluasi menggunakan pembagian data random split 80:20 dengan random seed yang konsisten. Pendekatan berlapis ini diharapkan menghasilkan analisis sentimen berbasis aspek yang lebih akurat dan kontekstual.

Tahapan terakhir adalah evaluasi model yang dilakukan menggunakan metrik standar seperti akurasi, presisi, recall, dan F1-score dari *Confusion Matrix*. Pengujian dilakukan pada data uji untuk mengukur kemampuan model dalam mengklasifikasikan sentimen secara konsisten dan akurat. Sebagai dasar untuk menilai efektivitas kombinasi metode *LDA-Lexicon-DNN* dalam mendeteksi sentimen berbasis aspek secara lebih mendalam dan interpretatif.

Secara keseluruhan, kombinasi metode ini memberikan keseimbangan antara kemampuan eksplanatif dari pendekatan leksikon dan kapasitas prediktif dari model pembelajaran mendalam. Dengan demikian, penelitian ini tidak hanya menghasilkan model klasifikasi yang akurat, tetapi juga menyediakan pemahaman yang lebih kaya terhadap persepsi masyarakat terhadap Sekolah Islam Terpadu melalui pendekatan analisis sentimen berbasis aspek.

4 Hasil dan Pembahasan

Tahapan hasil penelitian ini meliputi pemodelan aspek menggunakan LDA, analisis sentimen berbasis Lexicon, serta klasifikasi sentimen akhir menggunakan DNN. Ketiga pendekatan ini saling melengkapi untuk menghasilkan pemetaan sentimen masyarakat terhadap Sekolah Islam Terpadu berdasarkan aspek-aspek yang relevan.

Tahapan penelitian diawali Pengumpulan data dengan melakukan scrapping komentar pada Google Maps terhadap Sekolah Islam Terpadu. Untuk studi kasus ini, scrapping dibatasi untuk Sekolah Islam Terpadu di wilayah Provinsi Riau. Beberapa *keyword* digunakan untuk melakukan identifikasi seperti “Sekolah Islam Terpadu”, “Islamic Boarding School”, “TK Islam Terpadu”, “SD Islam Terpadu”, “SMP Islam Terpadu”, “SMA Islam Terpadu” maupun “Islamic School” dimanfaatkan dalam proses pencarian dan perolehan data sebanyak 2.280 data ulasan dari 205 sekolah islam terpadu se-Provinsi Riau yang di scrapping dari aplikasi *Google Maps*. Contoh proses perolehan data yang dilakukan dengan menggunakan Scrapping pada *Google Maps* pada tabel 1.

Tabel 1 Contoh perolehan data ulasan pada sekolah islam terpadu

Id Review	Text
1	Sekolah yang cocok untuk anak2 penerus bangsa
2	Sekolah swasta dengan fasilitas belajar yang bagus Insyaallah disini mencetak generasi Qur'ani dan membentuk karakter islami berakhlakul karimah
3	Sekolah mahal, tapi fasilitas nihil tidak sesuai
4	Di al-karim islamic school anak akan mendapat lebih banyak pelajaran agama, sehingga anak bisa lebih mendalami ilmu agama dan insya allah bisa menjadi tahfidz qur'an.

5	Sekolah IT dengan fasilitas olahraga berkuda yg ada di kampar
...	...
2280	Ada ekskul tahfiz jg

Tahapan praproses dalam penelitian ini terdiri dari enam tahapan utama, yaitu data cleaning, case folding, tokenizing, normalisasi, stopword removal, dan stemming. Setiap tahap memiliki peran penting dalam mengurangi noise dan mempertahankan makna semantik dari teks ulasan. Secara rinci, tahapan tersebut ditampilkan pada tabel 2 berikut.

Tabel 2 Tahapan praproses data

Casefolding	Tokenizing	Normalization	Stopword Removal	Stemming
sekolah yang cocok untuk anak2 penerus bangsa	['sekolah', 'yang', 'cocok', 'untuk', 'anak', 'penerus', 'bangsa']	['sekolah', 'yang', 'cocok', 'untuk', 'anak', 'penerus', 'bangsa']	['sekolah', 'cocok', 'anak', 'penerus', 'bangsa']	['sekolah', 'cocok', 'anak', 'terus', 'bangsa']
sekolah mahal, tapi fasilitas nihil tidak sesuai	['sekolah', 'mahal', 'tapi', 'fasilitas', 'nihil', 'tidak', 'sesuai']	['sekolah', 'mahal', 'tapi', 'fasilitas', 'nihil', 'tidak', 'sesuai']	['sekolah', 'mahal', 'fasilitas', 'nihil', 'sesuai']	['sekolah', 'mahal', 'fasilitas', 'nihil', 'sesuai']
sekolah swasta dengan fasilitas belajar yang bagus	['sekolah', 'swasta', 'dengan', 'fasilitas', 'belajar', 'yang', 'bagus', 'insyaallah', 'disini', 'mencetak', 'generasi', 'qurani', 'dan', 'membentuk', 'karakter', 'islami', 'berakhlaqul', 'karimah']	['sekolah', 'swasta', 'dengan', 'fasilitas', 'belajar', 'yang', 'bagus', 'insyaallah', 'disini', 'mencetak', 'generasi', 'qurani', 'dan', 'membentuk', 'karakter', 'islami', 'berakhlaqul', 'karimah']	['sekolah', 'swasta', 'fasilitas', 'belajar', 'bagus', 'insyaallah', 'mencetak', 'generasi', 'qurani', 'membentuk', 'karakter', 'islami', 'berakhlaqul', 'karimah']	['sekolah', 'swasta', 'fasilitas', 'ajar', 'bagus', 'insyaallah', 'cetak', 'generasi', 'quran', 'bentuk', 'karakter', 'islami', 'akhlaqul', 'karimah']
di al-karim islamic school anak akan mendapat lebih banyak pelajaran agama, sehingga anak bisa lebih mendalami ilmu agama dan insya allah bisa menjadi tahfidz qur'an.	['di', 'alkarim', 'islamic', 'school', 'anak', 'akan', 'mendapat', 'lebih', 'banyak', 'pelajaran', 'agama', 'sehingga', 'anak', 'bisa', 'lebih', 'mendalami', 'ilmu', 'agama', 'dan', 'insya', 'allah', 'bisa', 'menjadi', 'tahfidz', 'quran']	['di', 'alkarim', 'islamic', 'school', 'anak', 'akan', 'mendapat', 'lebih', 'banyak', 'pelajaran', 'agama', 'sehingga', 'anak', 'bisa', 'lebih', 'mendalami', 'ilmu', 'agama', 'dan', 'insya', 'allah', 'bisa', 'menjadi', 'tahfidz', 'quran']	['alkarim', 'islamic', 'school', 'anak', 'pelajaran', 'agama', 'anak', 'mendalami', 'ilmu', 'agama', 'insya', 'allah', 'tahfidz', 'quran']	['alkarim', 'islamic', 'school', 'anak', 'ajar', 'agama', 'anak', 'dalam', 'ilmu', 'agama', 'insya', 'allah', 'tahfidz', 'quran']
sekolah it dengan fasilitas olahraga berkuda yg ada di kampar	['sekolah', 'it', 'dengan', 'fasilitas', 'olahraga', 'berkuda', 'yg', 'ada', 'di', 'kampar']	['sekolah', 'it', 'dengan', 'fasilitas', 'olahraga', 'berkuda', 'ada', 'di', 'kampar']	['sekolah', 'it', 'fasilitas', 'olahraga', 'berkuda', 'kampar']	['sekolah', 'it', 'fasilitas', 'olahraga', 'kuda', 'kampar']
...	...	...	...	...
ada ekskul tahfiz jg	['ada', 'ekskul', 'tahfiz', 'jg']	['ada', 'ekskul', 'tahfiz', 'juga']	['ekskul', 'tahfiz']	['ekskul', 'tahfiz']

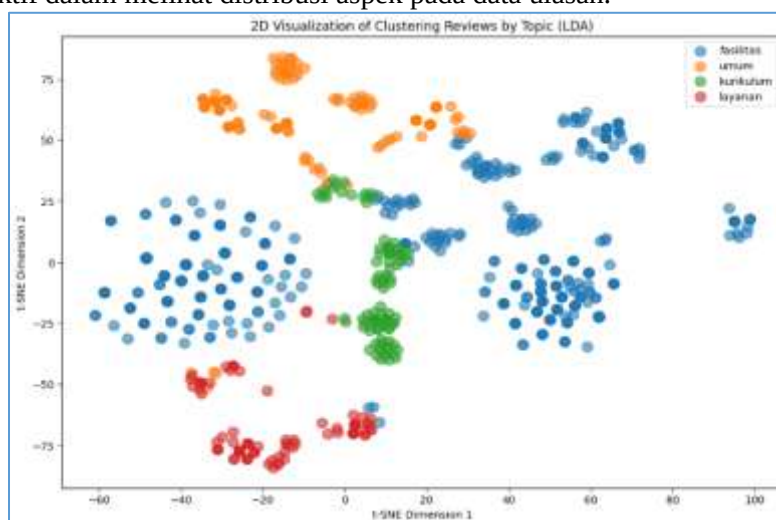
Tahapan-tahapan tersebut menghasilkan korpus teks yang bersih, terstandardisasi, terdiri dari 1.438 data ulasan. Dataset ini kemudian dibagi dengan perbandingan 80:20 untuk keperluan pelatihan (training dataset) dan pengujian (testing dataset). Dengan pembagian ini, terdapat 1.150 data yang digunakan sebagai training dataset dan 288 data yang digunakan sebagai testing dataset, kemudian digunakan sebagai masukan pada tahap pemodelan LDA dan Lexicon based.

Hasil pemodelan LDA menghasilkan empat topik utama yang diinterpretasikan berdasarkan kata-kata dominan pada masing-masing topik. Topik kurikulum dicirikan oleh kata seperti *kurikulum, pembelajaran, guru, agama, akademik*, yang merepresentasikan proses belajar dan integrasi nilai keislaman. Topik fasilitas dicirikan oleh kata *gedung, parkir, ruang kelas, sarana, kebersihan, lingkungan*, yang menggambarkan kondisi fisik sekolah. Topik layanan memiliki kata dominan seperti *pelayanan, administrasi, staf, komunikasi, respon*, yang berkaitan dengan interaksi dan layanan sekolah. Sementara itu, topik umum ditandai oleh kata *sekolah, anak, oke, baik, puas, rekomendasi*, yang mencerminkan penilaian keseluruhan tanpa fokus aspek tertentu. Hal ini terlihat pada hasil klasterisasi LDA sebagaimana dirangkum pada Tabel 3.

Tabel 3 Distribusi data ulasan berdasarkan aspek hasil pemodelan LDA.

No	Aspek	Jumlah
1.	Fasilitas	488
2.	Umum	433
3.	Layanan	159
4.	Kurikulum	70

Hasil ini dapat divisualisasikan dalam bentuk 2D. Visualisasi ini memperlihatkan persebaran data ulasan berdasarkan empat aspek dominan yang telah diidentifikasi dalam ulasan publik, antara lain aspek fasilitas, umum, kurikulum, dan layanan. Pada gambar 2, visualisasi 2D menampilkan data gambaran yang efektif dalam melihat distribusi aspek pada data ulasan.



Gambar 2 Visualisasi klasterisasi ulasan berdasarkan aspek menggunakan LDA

Selanjutnya dilakukan Pemodelan Sentimen dengan Lexicon Based yang menunjukkan bahwa mayoritas ulasan masyarakat terhadap Sekolah Islam Terpadu memiliki kecenderungan positif. Sebagaimana dirangkum pada Tabel 4. Jumlah ulasan positif yang jauh lebih dominan mengindikasikan bahwa sebagian besar masyarakat memiliki pengalaman yang memuaskan terhadap sekolah Islam terpadu.

Tabel 4 Distribusi data ulasan berdasarkan sentimen hasil pemodelan lexicon based

No	Sentimen	Jumlah
1.	Positif	1074
2.	Negatif	76

Kombinasi kedua metode ini menghasilkan model analisis yang lebih informatif dan kontekstual. Integrasi ini juga memperkuat keandalan analisis karena mampu menghubungkan *apa yang dibicarakan (topik)* dan *bagaimana opini terhadap topik tersebut (sentimen)* dalam satu kerangka analisis yang utuh seperti pada tabel 5.

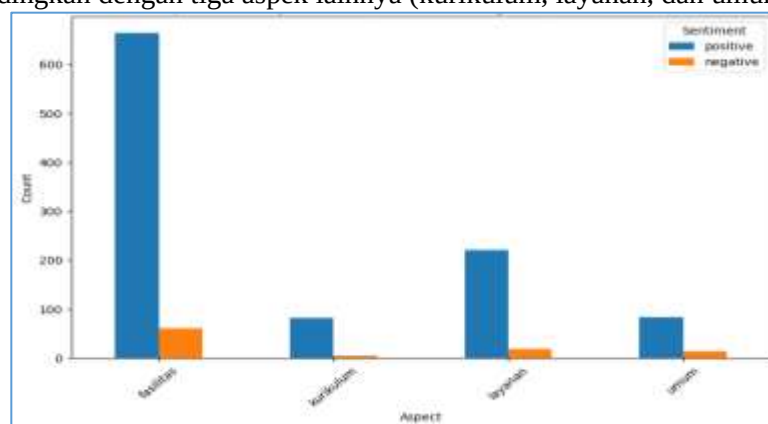
Tabel 5 Sampel data set sentiment dan Aspek hasil pemodelan LDA

No	Text Review	Aspect	Sentiment
1	sekolah mana didik anak anak segi karakter tentu milik unggul luar biasa	kurikulum	positive

<http://sistemasi.ftik.unisi.ac.id>

	speaking english every day tahfizh and computer		
2	sekolah salah satu sekolah dasar swasta baik pekanbaru visi wujud generasi bertaqwa akhlak mulia prestasi maa syaa allah	kurikulum	positive
3	lokasi jauh dari keramaian sangat cocok untuk belajar mendalami isi al quran sebagian pemondokan masih di bangun suasana bersahabat dan nyaman ada musholla dan minimarket khusus	fasilitas	positive
4	bikin macet jalan pas antar jemput this street not yours but for many people tolong securitynya diperbanyak not recommended	fasilitas	negative
5	sekolah sd islam plus didik latih siswa jadi pribadi lebih baik	kurikulum	positive
6	lokasi sulit temu hubung media sosial direspon sama sekali	fasilitas	negative

Seperti yang ditunjukkan pada gambar 3, distribusi jumlah ulasan berdasarkan aspek dan sentimen memperlihatkan ulasan positif memiliki jumlah yang jauh lebih banyak dibandingkan dengan ulasan negatif pada setiap aspek. Selain itu, terlihat bahwa jumlah ulasan pada aspek fasilitas mendominasi dibandingkan dengan tiga aspek lainnya (kurikulum, layanan, dan umum).



Gambar 3 Distribusi jumlah ulasan berdasarkan aspek dan sentimen

Setelah proses identifikasi aspek melalui LDA dan analisis polaritas sentimen menggunakan pendekatan Lexicon-Based, tahap selanjutnya dalam penelitian ini adalah penerapan model DNN untuk meningkatkan kemampuan sistem dalam mengenali pola-pola kompleks pada data teks. Model DNN menghasilkan tingkat akurasi yang signifikan pada aspek, namun menunjukkan tingkat akurasi yang tinggi pada sentimen seperti terlihat pada tabel 6.

Table 6 Kinerja model DNN berdasarkan aspek dan sentimen

Metric	Aspect	Sentimen
Accuracy	0,6389	0,9306
Precision	0,5318	0,8659
Recall	0,6389	0,9306

Pada tabel 6 nilai akurasi klasifikasi aspek yaitu 0.63 terlihat lebih rendah dibandingkan klasifikasi sentimen 0.93, hal ini disebabkan oleh identifikasi aspek merupakan tugas yang lebih kompleks. Satu ulasan sering kali memuat lebih dari satu aspek secara bersamaan, sehingga batas antar aspek menjadi tidak tegas dan saling tumpang tindih secara semantik. Sebaliknya, polaritas sentimen cenderung lebih eksplisit dan konsisten secara linguistik, sehingga lebih mudah dipelajari oleh model, terutama dengan menggunakan pendekatan lexicon-based.

Dalam hal pemodelan Aspek, terlihat pada tabel 7 LDA-DNN menunjukkan keunggulan dalam Aspect dengan Accuracy sebesar 0,6389 dan Recall sebesar 0,6389 dibandingkan K-Means LSA dan K-Means PCA-DNN, Hal ini menunjukkan bahwa model LDA-DNN tetap lebih konsisten dalam memberikan prediksi yang akurat untuk aspek-aspek yang ada, sehingga LDA-DNN mampu

<http://sistemasi.ftik.unisi.ac.id>

memberikan hasil klasifikasi aspek yang lebih seimbang dan representatif, yang penting dalam pemodelan berbasis aspek.

Table 7 Perbandingan model yang diusulkan dan eksisting model

Model	Metric					
	Aspect			Sentimen		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
LDA DNN	0,6389	0,5318	0,6389	0,9306	0,8659	0,9306
Kmeans PCA-DNN	0.4514	0.5145	0.4514	0,9236	0,8531	0,9236
Kmeans LSA-DNN	0,3194	0,6321	0,3194	0,9236	0,8531	0,9236

Disisi lain, LDA-DNN terlihat unggul dalam melakukan sentimen yaitu dengan Accuracy sebesar 0,9306, Precision sebesar 0,8659, dan Recall sebesar 0,9306. Hasil ini lebih tinggi dibandingkan dengan K-Means LSA dan K-Means PCA. Keunggulan ini menunjukkan bahwa LDA-DNN lebih efektif dalam mengidentifikasi sentimen positif dan negatif dengan akurasi yang lebih tinggi, sehingga memberikan hasil yang lebih handal dalam pemodelan sentimen.

Meskipun data penelitian masih terbatas pada ulasan yang bersumber dari Google Maps, namun hasil penelitian ini membuktikan bahwa pendekatan gabungan LDA, Lexicon dan DNN efektif dalam menganalisis persepsi masyarakat terhadap Sekolah Islam Terpadu secara terstruktur. Keunggulan ini terutama terlihat dalam pemodelan sentimen, di mana LDA-DNN menghasilkan tingkat akurasi, presisi, dan recall yang lebih tinggi, menjadikannya pilihan model yang lebih baik untuk analisis aspek dan sentimen pada dataset ini dan memberikan implikasi praktis bagi pengelola sekolah dan pembuat kebijakan dalam memahami aspek-aspek utama yang perlu ditingkatkan dalam domain pendidikan Islam.

5 Kesimpulan

Penelitian menunjukkan bahwa integrasi model LDA-Lexicon Based dan DNN untuk melakukan analisis sentimen berbasis aspek terhadap ulasan masyarakat mengenai Sekolah Islam Terpadu secara umum cenderung positif, khususnya pada empat aspek yaitu fasilitas, layanan, kurikulum, dan aspek umum. Hal ini menegaskan bahwa citra Sekolah Islam Terpadu di ruang digital relatif baik dan sesuai dengan persepsi positif pengguna. Kontribusi utama penelitian ini terletak pada pembuktian empiris bahwa pemodelan berbasis aspek membutuhkan pendekatan yang tidak hanya mereduksi dimensi, tetapi juga mempertahankan struktur semantik dalam teks. Model LDA berperan dalam mengekstraksi dan mengidentifikasi topik-aspek secara probabilistik, Lexicon-Based memberikan dasar penilaian polaritas awal berdasarkan makna leksikal, sedangkan DNN memperkuat proses klasifikasi melalui pembelajaran mendalam terhadap fitur semantik. Integrasi ketiganya menghasilkan peningkatan performa model, dengan akurasi mencapai 63,89% untuk pemodelan aspek dan 93,06% untuk klasifikasi sentimen. Dengan demikian, pendekatan ini dapat diposisikan sebagai strategi yang lebih efektif untuk analisis sentimen berbasis aspek pada domain pendidikan, khususnya Sekolah Islam Terpadu. Secara akademis, hasil penelitian ini memperkaya literatur analisis sentimen dengan menawarkan kerangka metodologis hibrid yang menggabungkan kekuatan pemodelan topik, leksikal, dan pembelajaran mendalam. Pendekatan ini dapat menjadi acuan bagi penelitian selanjutnya yang berfokus pada analisis opini publik di berbagai domain. Secara praktis, model yang dikembangkan mampu membantu lembaga pendidikan dalam memahami persepsi masyarakat secara lebih mendalam pada tiap aspek layanan. Informasi tersebut dapat digunakan sebagai dasar evaluasi mutu, peningkatan komunikasi digital, dan perumusan strategi peningkatan kepuasan pengguna. Untuk pengembangan riset selanjutnya berfokus kepada integrasi representasi teks berbasis deep learning yang lebih kompleks untuk meningkatkan akurasi klasifikasi aspek. Selain itu, eksplorasi metode pemodelan topik yang lebih adaptif serta diharapkan dapat memperkaya analisis dan meningkatkan generalisasi model.

Ucapan Terima Kasih

Terimakasih yang sebesar-besarnya kepada LPPM UIN Sultan Syarif Kasim Riau atas dukungan pendanaan melalui Program Hibah Penelitian Tahun 2024, sehingga penelitian ini dapat terlaksana dengan baik.

Referensi

- [1] R. Dahlia, Z. Khairi, A. Diniaty, K. Anwar, A. Ahmad Tohar, and V. Shofiah, "Peran Citra Sekolah dalam Memotivasi Orang Tua Memasukkan Anaknya ke Sekolah Dasar Islam Terpadu (SDIT)," *J. Ilmu Psikol. dan Kesehat.*, Vol. 2, No. 1 SE-Articles, pp. 135–144, Jul. 2023, DOI: 10.47353/sikontan.v2i1.1270.
- [2] H. Kiromi, "Analisis Sentimen Pendidikan Pesantren pada Media Sosial Twitter dengan Metode *Naive Bayes*," Universitas Islam Negeri Maulana Malik Ibrahim., 2023. [Online]. Available: <http://etheses.uin-malang.ac.id/id/eprint/52270>
- [3] Z. Drus and H. Khalid, "Sentiment Analysis in Social Media and its Application: Systematic Literature Review," *Procedia Comput. SCI.*, Vol. 161, pp. 707–714, 2019, DOI: 10.1016/j.procs.2019.11.174.
- [4] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing What, How and Why: A Near Complete Solution for Aspect-based Sentiment Analysis," *Proc. AAAI Conf. Artif. Intell.*, Vol. 34, No. 05, pp. 8600–8607, 2020, DOI: 10.1609/aaai.v34i05.6383.
- [5] D. Musfiroh, U. Khaira, P. Eko, P. Utomo, and T. Suratno, "Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. SCI.*, Vol. 1, No. April, pp. 24–33, 2021.
- [6] S. Aisy and B. Prasetyo, "Sentiment Analysist of the TPKE Law on Twitter using InSet Lexicon with Multinomial Naïve Bayes and Support Vector Machine based on Soft Voting," *Recursive J. Informatics*, Vol. 1, No. 2 SE-Articles, Sep. 2023, DOI: 10.15294/rji.v1i2.68324.
- [7] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 391–394. DOI: 10.1109/IALP.2017.8300625.
- [8] I. Journal and O. F. Science, "A Comparative Analysis of Latent Semantic Analysis and Latent Dirichlet Allocation Topic Modeling Methods using Bible Data," pp. 4474–4482, 2020.
- [9] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid Sentiment Classification on Twitter Aspect-based Sentiment Analysis," *Appl. Intell.*, Vol. 48, No. 5, pp. 1218–1232, 2018, DOI: 10.1007/s10489-017-1098-6.
- [10] T. Hecking and L. Leydesdorff, "Topic Modelling of Empirical Text Corpora : Validity , Reliability , and Reproducibility in Comparison to Semantic Maps," pp. 1–17.
- [11] S. Ali, G. Wang, and S. Riaz, "Aspect based Sentiment Analysis of Ridesharing Platform Reviews for Kansei Engineering," *IEEE Access*, Vol. 8, pp. 173186–173196, 2020, DOI: 10.1109/ACCESS.2020.3025823.
- [12] D. Zhang, T. Luo, and D. Wang, "Learning from LDA using Deep Neural Networks," in *Natural Language Understanding and Intelligent Applications*, C.-Y. Lin, N. Xue, D. Zhao, X. Huang, and Y. Feng, Eds., Cham: Springer International Publishing, 2016, pp. 657–664.
- [13] E. Wahyudi and R. Kusumaningrum, "Aspect based Sentiment Analysis in E-Commerce User Reviews using Latent Dirichlet Allocation (LDA) and Sentiment Lexicon," 2019, pp. 1–6. DOI: 10.1109/ICICoS48119.2019.8982522.
- [14] S. Mukherjee, "Sentiment Analysis," in *ML.NET Revealed: Simple Tools for Applying*

<http://sistemasi.ftik.unisi.ac.id>

- Machine Learning to Your Applications*, Berkeley, CA: Apress, 2021, pp. 113–127. DOI: 10.1007/978-1-4842-6543-7_7.
- [15] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare, G. Yogapriya, and P. Pandey, “*Neuroscience Informatics an Efficient Way of Text-based Emotion Analysis from Social Media using LRA-DNN*,” *Neurosci. Informatics*, Vol. 2, No. 3, p. 100048, 2022, DOI: 10.1016/j.neuri.2022.100048.
- [16] S. Ying, “*Guests’ Aesthetic Experience with Lifestyle Hotels: An Application of LDA Topic Modelling Analysis*,” *Heliyon*, Vol. 10, No. 16, p. e35894, 2024, DOI: <https://doi.org/10.1016/j.heliyon.2024.e35894>.
- [17] D. M.D. and A. Ganesh, “*Sentiment Analysis: A Comparative Study on Different Approaches*,” *Procedia Comput. SCI.*, Vol. Volume 87, pp. 44–49, 2016, DOI: <https://doi.org/10.1016/j.procs.2016.05.124>.