

Comparative Sentiment Analysis and Topic Modeling of Dola AI User Reviews for Knowledge Discovery

Nadia Ramadhani¹, Ken Ditha Tania^{2*}, Naretha Kawadha Pasemah Gumay³

^{1,2,3}Sistem Informasi, Universitas Sriwijaya

nadia.r085@gmail.com¹, kenya.tania@gmail.com², narethakawadha@unsri.ac.id³

Article Info

Article history:

Received 2026-08-06

Revised 2026-09-11

Accepted 2026-10-17

Keywords:

*Sentiment Analysis,
Knowledge Discovery,
Machine Learning,
Latent Dirichlet Allocation,
Dola AI.*

ABSTRACT

As artificial intelligence (AI) technologies become more widely adopted, users increasingly express their experiences through online reviews, providing valuable insights into user satisfaction and opportunities for application improvement. This study compares the performance of Support Vector Machine (SVM), Naïve Bayes (NB), and K-Nearest Neighbor (KNN) for sentiment classification of Dola AI user reviews and applies Latent Dirichlet Allocation (LDA) to identify the dominant topics within each sentiment category. The research followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework using 28,226 Google Play Store reviews collected through web scraping. The reviews were represented using Term Frequency–Inverse Document Frequency (TF-IDF), class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE), and model performance was evaluated through 5-fold stratified cross-validation using accuracy, weighted precision, weighted recall, and weighted F1-score. The results showed that SVM achieved the highest average performance, with an accuracy of 77.68% and a weighted F1-score of 80.02%, followed by Naïve Bayes with a weighted F1-score of 78.60% and K-Nearest Neighbor with 70.42%. Furthermore, LDA identified six topics in positive reviews, three topics in neutral reviews, and two topics in negative reviews. Positive reviews mainly highlighted the application's ease of use, AI capabilities, and conversational features, whereas negative reviews primarily focused on application performance, the quality of AI-based image editing, and advertisements. These findings provide deeper insights into user perceptions and offer data-driven recommendations for improving the quality and user experience of the Dola AI application.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kemajuan teknologi *Artificial Intelligence* (AI) turut mendorong perubahan besar pada berbagai sisi kehidupan manusia, mulai dari cara bekerja, belajar, berkomunikasi, hingga menyelesaikan persoalan sehari-hari secara lebih efisien [1][2]. Kemajuan tersebut didorong oleh hadirnya teknologi *generative AI*, yaitu model pembelajaran mesin yang mampu menghasilkan berbagai bentuk konten serta mendukung proses *brainstorming*, penyusunan ide, dan penyempurnaan solusi dalam pemecahan masalah [3]. Seiring pesatnya perkembangan teknologi tersebut, berbagai aplikasi asisten virtual berbasis *generative AI* seperti *ChatGPT*,

Gemini, *Copilot*, *Claude*, *Perplexity AI* semakin banyak dimanfaatkan oleh masyarakat [4]. Salah satu aplikasi yang turut berkembang adalah *Dola AI*, yaitu asisten manajemen jadwal dan kalender berbasis AI yang memungkinkan pengguna mengatur agenda, pengingat, dan perencanaan kegiatan secara otomatis melalui antarmuka percakapan alami [5].

Seiring dengan meningkatnya adopsi aplikasi berbasis kecerdasan buatan, pengguna semakin aktif menyampaikan ulasan mengenai pengalaman penggunaan melalui platform distribusi aplikasi, seperti *Google Play* dan *App Store*. Ulasan pengguna tersebut menjadi sumber data berharga bagi pengembang dalam memperoleh pemahaman mengenai

persepsi pengguna, mengevaluasi kualitas aplikasi, serta mengidentifikasi aspek-aspek yang perlu ditingkatkan dalam proses pengembangan produk [6].

Salah satu pendekatan yang dapat diterapkan untuk mengolah data ulasan pengguna adalah analisis sentimen, yaitu teknik yang memanfaatkan *text mining* dan *Natural Language Processing* (NLP) untuk mengekstraksi, menganalisis, serta menginterpretasikan opini, pandangan, dan perasaan yang terkandung dalam data tekstual, baik yang bersifat terstruktur maupun tidak terstruktur. Melalui proses tersebut, setiap ulasan dapat diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral [7]. Dalam penelitian analisis sentimen, pemilihan algoritma klasifikasi menjadi faktor penting karena berpengaruh terhadap performa model yang dihasilkan. Algoritma *Support Vector Machine* (SVM) dipilih karena mampu mengklasifikasikan data teks berdimensi tinggi secara efektif melalui penentuan *hyperplane* optimal serta terbukti memberikan akurasi yang tinggi pada berbagai tugas klasifikasi teks [8]. Di sisi lain, *Naïve Bayes* merupakan algoritma pembelajaran terawasi yang menggunakan pendekatan probabilistik dengan asumsi bahwa setiap atribut bersifat independen secara bersyarat dan memiliki kontribusi yang sama dalam menentukan kelas, sehingga memiliki struktur yang sederhana, mudah diimplementasikan, serta tetap efektif meskipun dilatih menggunakan dataset berukuran kecil [9]. Sementara itu, *K-Nearest Neighbor* (KNN) merupakan algoritma pembelajaran terawasi nonparametrik berbasis *instance* yang melakukan klasifikasi berdasarkan tingkat kemiripan dan kedekatan suatu data baru dengan sejumlah data terdekat pada data pelatihan tanpa memerlukan proses pelatihan secara eksplisit [10]. Ketiga algoritma klasifikasi tersebut merupakan metode yang umum diterapkan dalam penelitian analisis sentimen karena masing-masing menawarkan karakteristik dan keunggulan yang berbeda dalam mengolah serta mengklasifikasikan data teks.

Selain melakukan klasifikasi sentimen, penelitian terhadap ulasan pengguna juga perlu dilengkapi dengan identifikasi topik yang sering dibahas pengguna. Salah satu metode yang umum digunakan untuk mengidentifikasi topik dalam kumpulan data teks adalah *topic modeling* berbasis *Latent Dirichlet Allocation* (LDA) [11]. Metode LDA mampu mengidentifikasi topik-topik utama yang tersembunyi dalam kumpulan dokumen sehingga dapat memberikan informasi yang lebih mendalam mengenai aspek-aspek yang menjadi perhatian pengguna [12]. Penelitian sebelumnya mengkombinasikan klasifikasi sentimen menggunakan SVM dan pemodelan topik menggunakan LDA untuk mengetahui interpretasi topik utama yang sering dibahas pada setiap kelompok sentimen, kemudian divisualisasikan menggunakan *word cloud* [13].

Pendekatan *knowledge discovery* yang mengombinasikan analisis sentimen dengan pemodelan topik juga telah diterapkan pada berbagai domain ulasan dan opini pengguna. Penelitian oleh [14] membandingkan persepsi pengguna terhadap dua aplikasi mobile banking, yaitu BCA Mobile dan

myBCA, dengan mengombinasikan klasifikasi sentimen berbasis *deep learning* dan pemodelan topik menggunakan BERTopic, sehingga berhasil mengidentifikasi delapan topik keluhan utama, seperti kegagalan login dan pengurangan saldo yang tidak jelas, pada masing-masing aplikasi. Penelitian lain oleh [15] menerapkan pendekatan *topic mining* dengan membandingkan tujuh metode pemodelan topik untuk mengungkap kebutuhan informasi kesehatan pengguna pada platform konsultasi daring, dan menunjukkan bahwa hasil pemodelan topik dapat diolah lebih lanjut menjadi pengetahuan yang bermanfaat bagi pengembangan layanan digital. Selain itu, penelitian oleh [16] mengintegrasikan lima metode pemodelan topik dengan *Aspect-Based Sentiment Analysis* untuk mengungkap persepsi publik terhadap *generative AI* di media sosial X, dan menemukan bahwa kecenderungan sentimen pengguna bervariasi secara signifikan bergantung pada aspek topik yang dibahas. Ketiga penelitian tersebut menegaskan bahwa pendekatan *knowledge discovery* yang mengombinasikan analisis sentimen dan pemodelan topik mampu menghasilkan pemahaman yang lebih komprehensif mengenai persepsi pengguna dibandingkan penerapan analisis sentimen maupun pemodelan topik secara terpisah, sehingga pendekatan serupa relevan diterapkan pada ulasan aplikasi Dola AI dalam penelitian ini.

Hasil berbagai penelitian sebelumnya menunjukkan bahwa performa algoritma klasifikasi dalam analisis sentimen bervariasi, bergantung pada karakteristik data dan metode yang digunakan. Penelitian oleh [17] membandingkan algoritma *Naïve Bayes* dan SVM, di mana *Naïve Bayes* memperoleh akurasi sebesar 63,02%, sedangkan SVM mencapai akurasi sebesar 70,82%. Penelitian lain melaporkan bahwa model *Support Vector Machine* mampu mencapai tingkat akurasi sebesar 82,56% dalam tugas klasifikasi teks [18]. Selain itu, penelitian oleh [19] menyimpulkan bahwa pengklasifikasi pembelajaran mendalam dan SVM memberikan kinerja yang lebih baik dibandingkan algoritma *Decision Tree*, *K-Nearest Neighbours*, dan *Naïve Bayes* pada dataset yang digunakan. Sementara itu, penelitian oleh [20] melaporkan bahwa algoritma SVM memperoleh tingkat akurasi sebesar 52%, sedangkan KNN mencapai akurasi sebesar 51% berdasarkan pengujian terhadap 189 sampel data uji.

Meskipun SVM, *Naïve Bayes*, KNN, dan LDA telah banyak digunakan dalam penelitian sebelumnya, berdasarkan penelusuran literatur belum ditemukan penelitian yang secara khusus mengombinasikan perbandingan ketiga algoritma klasifikasi tersebut dengan pemodelan topik LDA pada ulasan pengguna Dola AI. Penelitian sebelumnya pada aplikasi AI umumnya berfokus pada perbandingan algoritma klasifikasi atau analisis sentimen dan *topic modeling* secara terpisah maupun dengan pendekatan yang berbeda.

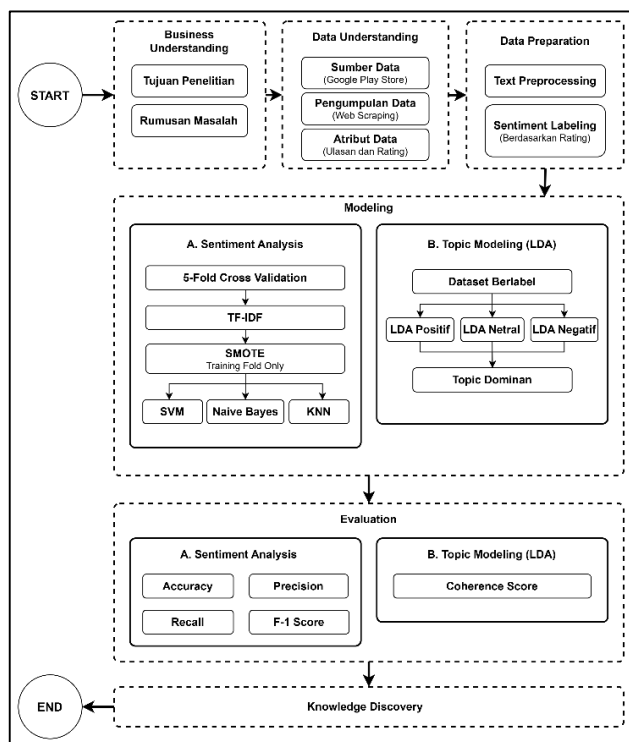
Oleh karena itu, kontribusi penelitian ini tidak terletak pada kebaruan algoritma, tetapi pada penerapannya secara integratif pada ulasan Dola AI. Penelitian ini membandingkan kinerja SVM, *Naïve Bayes*, dan KNN, kemudian

menggunakan hasil klasifikasi sentimen untuk membentuk kelompok positif, netral, dan negatif yang selanjutnya dianalisis menggunakan LDA secara terpisah. Pendekatan tersebut memungkinkan identifikasi topik atau isu yang mendasari masing-masing kelompok sentimen, sehingga memberikan pemahaman yang lebih spesifik mengenai persepsi pengguna terhadap Dola AI.

II. METODE

CRISP-DM merupakan metodologi yang menyediakan proses terstruktur dalam pelaksanaan proyek *data mining* dan bersifat independen terhadap bidang aplikasi. Metodologi ini terdiri atas enam tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Penerapannya bersifat iteratif sehingga memungkinkan peneliti kembali ke tahapan sebelumnya apabila diperlukan. Dalam penelitian ini, CRISP-DM digunakan sebagai pedoman mulai dari pemahaman permasalahan, pengolahan data, pemodelan, hingga evaluasi hasil [21].

Penelitian ini menerapkan tahapan CRISP-DM secara sistematis hingga tahap lanjutan Knowledge Discovery [22], sebagaimana juga diterapkan pada penelitian sejenis yang menggabungkan klasifikasi sentimen dan pemodelan topik menggunakan kerangka CRISP-DM pada domain aplikasi komunitas digital [11], serta penelitian yang mengintegrasikan analisis sentimen dan pemodelan topik ke dalam pendekatan knowledge discovery pada domain aplikasi perbankan digital [14]. Gambar 1 menampilkan alur penelitian dalam bentuk flowchart.



Gambar 1. Alur Penelitian

A. Business Understanding

Fase pertama dalam kerangka CRISP-DM adalah pemahaman bisnis, yang berfokus pada pemahaman terhadap permasalahan penelitian, perumusan tujuan penelitian, serta identifikasi kebutuhan pengguna atau organisasi [23].

B. Data Understanding

Pada fase ini, dilakukan pemahaman terhadap karakteristik data penelitian melalui serangkaian proses, meliputi pengumpulan data tahap awal, eksplorasi data, identifikasi atribut, serta penilaian terhadap kualitas data yang digunakan [24]. Data ulasan pengguna aplikasi Dola AI diperoleh dari platform *Google Play Store* melalui teknik *web scraping* dengan memanfaatkan pustaka *google-play-scraper*. Proses pengumpulan data dilakukan pada periode 1 Maret hingga 25 Juni 2026 dengan mengambil seluruh ulasan yang tersedia dalam rentang periode tersebut dan memenuhi kriteria penelitian, yaitu ulasan yang ditulis dalam bahasa Indonesia. Data yang diperoleh mencakup atribut teks ulasan dan nilai rating bintang pada skala 1–5. Rating tersebut selanjutnya digunakan sebagai dasar pelabelan sentimen, sedangkan teks ulasan digunakan sebagai masukan pada proses *text mining*. Pemanfaatan *Google Play Store* dipilih karena menyediakan ulasan pengguna secara langsung sehingga mampu merepresentasikan pengalaman penggunaan aplikasi Dola AI pada kondisi nyata.

C. Data Preparation

Fase persiapan data merupakan fase terpanjang dan paling krusial dalam kerangka CRISP-DM, meliputi seluruh proses yang diperlukan untuk mentransformasikan data mentah menjadi dataset yang siap digunakan pada tahap pemodelan [25]. Pada penelitian ini, fase ini terbagi menjadi dua sub-tahapan utama, yaitu *praproses teks* dan *pelabelan sentimen*.

1) *Text Preprocessing*: Data yang telah terkumpul pada tahap sebelumnya selanjutnya diolah guna membersihkan dan menyeragamkan teks mentah hasil *scraping* sehingga layak digunakan pada tahap analisis berikutnya. Tahapan pertama, yaitu *cleaning*, berfungsi menghilangkan karakter khusus, angka, URL, emotikon, serta tanda baca yang tidak berkontribusi terhadap analisis sentimen. Tahapan kedua, yaitu *case folding*, berfungsi menyeragamkan seluruh karakter teks ke dalam huruf kecil. Tahapan ketiga, yaitu normalisasi, berfungsi mengonversi kata-kata tidak baku, singkatan, bahasa gaul, maupun kata dengan kesalahan penulisan ke dalam bentuk baku sesuai kaidah bahasa Indonesia. Tahapan keempat, yaitu tokenisasi, berfungsi memisahkan teks menjadi satuan-satuan token tersendiri. Tahapan kelima, yaitu *stopword removal*, berfungsi membuang kata-kata umum yang tidak memberikan kontribusi makna terhadap analisis sentimen, seperti "yang", "dan", "di", dan "dengan", dengan merujuk pada daftar *stopword* berbahasa Indonesia. Tahapan keenam atau terakhir, yaitu *stemming*, berfungsi mengembalikan kata-kata berimbuhan ke bentuk kata dasarnya melalui pustaka *Sastrawi*

yang secara khusus dirancang untuk pemrosesan bahasa Indonesia.

2) *Pelabelan Sentimen*: Pelabelan sentimen dilakukan berdasarkan rating bintang yang diberikan oleh pengguna pada Google Play Store sebagai *proxy* atau pendekatan terhadap sentimen ulasan. Ulasan dengan rating 1–2 bintang dikategorikan sebagai sentimen negatif, rating 3 bintang sebagai sentimen netral, dan rating 4–5 bintang sebagai sentimen positif. Skema pelabelan tersebut digunakan karena rating bintang tersedia secara langsung pada setiap ulasan dan memungkinkan proses pelabelan dilakukan secara konsisten pada jumlah data yang besar. Meskipun demikian, rating bintang tidak selalu merepresentasikan sentimen yang secara eksplisit terdapat dalam teks ulasan. Ulasan dengan rating tinggi dapat tetap mengandung kritik, sedangkan ulasan dengan rating rendah dapat memuat aspek positif terhadap aplikasi.

Validasi manual dilakukan untuk mengevaluasi konsistensi anotasi dan kesesuaian label berbasis rating terhadap sentimen tekstual, dilakukan validasi manual terhadap 1.200 ulasan oleh dua anotator secara independen menggunakan *annotation guidelines*. Ulasan dikategorikan ke dalam sentimen positif, netral, atau negatif berdasarkan sentimen dominan yang terdapat dalam teks. Hasil validasi menunjukkan tingkat *inter-annotator agreement* sebesar 87,08% dengan nilai *Cohen's Kappa* sebesar 0,759, yang menunjukkan substantial agreement antara kedua anotator. Selanjutnya, perbandingan antara label berbasis rating dan hasil anotasi manual menunjukkan tingkat kesesuaian sebesar 61,47% dengan anotator pertama dan 58,30% dengan anotator kedua. Hasil tersebut menunjukkan bahwa label berbasis rating memiliki keterbatasan dalam merepresentasikan sentimen tekstual secara langsung. Oleh karena itu, rating bintang digunakan sebagai *proxy* sentimen dengan tetap mempertimbangkan keterbatasannya.

D. Modeling

Tahap *modeling* berfokus pada pemilihan dan penerapan teknik pemodelan yang sesuai dengan tujuan penelitian yang telah ditetapkan [25]. Pada penelitian ini, proses klasifikasi sentimen dilakukan menggunakan tiga algoritma *machine learning*, yaitu SVM, *Naïve Bayes*, dan KNN. Pemilihan ketiga algoritma tersebut didasarkan pada perbedaan karakteristik dan mekanisme klasifikasi yang berbeda sehingga memungkinkan dilakukan analisis komparatif untuk mengidentifikasi model yang paling optimal untuk dataset penelitian ini.

Proses pembentukan model menerapkan *Stratified 5-Fold Cross Validation* untuk membagi dataset menjadi data pelatihan dan data pengujian dengan tetap mempertahankan proporsi setiap kelas sentimen pada setiap *fold*. Pendekatan ini dipilih agar distribusi data pada setiap proses pelatihan dan pengujian tetap merepresentasikan distribusi dataset asli, terutama pada kondisi *imbalanced dataset* [26]. Pada setiap iterasi, empat *fold* digunakan sebagai data pelatihan dan satu *fold* sebagai data pengujian hingga seluruh *fold* memperoleh

giliran sebagai data uji. Preprocessing dilakukan sebelum proses *cross-validation*, sedangkan TF-IDF dibentuk dengan melakukan *fitting* hanya pada data pelatihan dan kemudian diterapkan pada data pengujian. Selanjutnya, SMOTE diterapkan hanya pada data pelatihan untuk menyeimbangkan kelas sebelum proses pelatihan model, sedangkan data pengujian tetap menggunakan distribusi aslinya. Dengan demikian, informasi dari data pengujian tidak digunakan dalam proses pembentukan fitur maupun pelatihan model.

Pada setiap *fold*, data pelatihan terlebih dahulu direpresentasikan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) melalui *TfidfVectorizer*. Proses pembentukan kosakata (*fit*) dilakukan hanya pada data pelatihan, sedangkan data pengujian ditransformasikan (*transform*) menggunakan kosakata yang telah terbentuk dari data pelatihan.

Selanjutnya, ketidakseimbangan kelas pada data pelatihan ditangani menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). SMOTE dipilih karena memungkinkan pembentukan sampel sintesis pada kelas minoritas sehingga model memperoleh representasi kelas yang lebih seimbang tanpa melakukan duplikasi langsung terhadap data yang telah tersedia [27]. Namun, penerapan SMOTE pada representasi TF-IDF memiliki keterbatasan karena proses interpolasi dilakukan pada ruang fitur berdimensi tinggi dan sparse, sehingga sampel sintesis yang terbentuk tidak selalu merepresentasikan struktur linguistik teks secara aktual. Oleh karena itu, penggunaan SMOTE dalam penelitian ini diposisikan sebagai strategi penanganan ketidakseimbangan kelas yang dievaluasi pada karakteristik dataset Dola AI.

Penerapan SMOTE dilakukan setelah proses TF-IDF dan hanya pada data pelatihan di setiap *fold*, sedangkan data pengujian tidak mengalami proses *oversampling*. Dengan demikian, informasi dari data pengujian tidak digunakan dalam pembentukan sampel sintesis maupun proses pelatihan model. Seluruh tahapan, mulai dari pembentukan fitur TF-IDF, penerapan SMOTE, hingga pelatihan model dilakukan secara independen pada setiap *fold* guna mencegah terjadinya *data leakage* selama proses evaluasi [28].

Model klasifikasi kemudian dibangun menggunakan algoritma SVM, *Naïve Bayes*, dan KNN. SVM menggunakan *LinearSVC* dengan $C=1$, *Naïve Bayes* menggunakan *MultinomialNB* dengan $\alpha=1.0$, sedangkan KNN menggunakan $k=5$.

Selain proses klasifikasi sentimen, tahap *modeling* juga mencakup pembangunan model *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik-topik laten yang mendominasi kumpulan dokumen berdasarkan pola kemunculan kata [29]. Pemodelan LDA dilakukan setelah kategori sentimen diperoleh agar topik yang dihasilkan dapat dianalisis berdasarkan kecenderungan sentimen pengguna. Ulasan kemudian dikelompokkan secara terpisah ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Selanjutnya, model LDA dibangun secara independen pada masing-masing kelompok tersebut. Pemisahan data berdasarkan kategori sentimen bertujuan untuk memperoleh

topik yang lebih spesifik dan merepresentasikan isu yang dibahas pada masing-masing kelompok sentimen. Dengan demikian, hasil LDA tidak hanya menunjukkan topik yang dominan secara keseluruhan, tetapi juga menunjukkan topik yang berkaitan dengan kecenderungan opini tertentu.

Penentuan jumlah topik optimal dilakukan secara empiris dengan menguji beberapa kandidat jumlah topik pada masing-masing kelompok sentimen. Setiap kandidat model LDA dievaluasi menggunakan *Coherence Score* dengan metrik C_v . Nilai *Coherence Score* digunakan untuk mengukur tingkat keterkaitan semantik antarkata yang membentuk suatu topik. Jumlah topik yang menghasilkan nilai *Coherence Score* tertinggi pada masing-masing kelompok sentimen dipilih sebagai konfigurasi optimal. Oleh karena itu, jumlah topik tidak ditentukan berdasarkan interpretasi subjektif peneliti, tetapi berdasarkan hasil evaluasi kuantitatif terhadap beberapa kandidat model. Setelah jumlah topik optimal diperoleh, setiap topik diinterpretasikan berdasarkan kata-kata dengan probabilitas kemunculan tertinggi untuk mengidentifikasi tema yang merepresentasikan pembahasan pengguna.

E. Evaluation

Fase *evaluation* berfungsi untuk mengukur kinerja model klasifikasi yang telah dibangun pada tahap *modeling* [25]. Evaluasi dilakukan menggunakan empat metrik, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Nilai *Precision*, *Recall*, dan *F1-score* dihitung menggunakan pendekatan *weighted average* karena distribusi kelas pada dataset penelitian tidak seimbang, sehingga hasil evaluasi lebih merepresentasikan kontribusi setiap kelas dibandingkan *macro average*. Nilai performa akhir masing-masing algoritma diperoleh dari rata-rata hasil evaluasi pada lima *fold*.

Selain itu, *Classification Report* digunakan untuk menyajikan nilai *precision*, *recall*, *F1-score*, *support*, *weighted average* dan *macro average* pada setiap kelas sentimen sehingga kinerja masing-masing algoritma dapat dianalisis secara lebih rinci. Pelaporan *weighted average* dan *macro average* secara bersamaan memungkinkan evaluasi dilakukan dari dua perspektif, yaitu kinerja model terhadap distribusi data secara keseluruhan dan kemampuan model dalam memperlakukan setiap kelas secara seimbang.

Evaluasi model *Latent Dirichlet Allocation* dilakukan secara terpisah pada kelompok sentimen positif, netral, dan negatif. Setiap kelompok sentimen diuji menggunakan beberapa kandidat jumlah topik. Masing-masing konfigurasi LDA kemudian dievaluasi menggunakan *Coherence Score*. Jumlah topik dengan nilai *Coherence Score* tertinggi pada setiap kelompok sentimen ditetapkan sebagai jumlah topik optimal. Pendekatan ini memungkinkan jumlah topik ditentukan berdasarkan kualitas koherensi topik pada masing-masing kelompok, sehingga jumlah topik dapat berbeda antara sentimen positif, netral, dan negatif. Setelah konfigurasi optimal diperoleh, topik diinterpretasikan

berdasarkan kata-kata dominan yang memiliki probabilitas kemunculan tertinggi pada masing-masing topik.

F. Deployment

Fase *deployment* merupakan tahap penerapan model yang telah memenuhi kriteria evaluasi. Pada tahap ini, hasil implementasi model disajikan dan didokumentasikan secara jelas sehingga dapat dipahami oleh pengguna serta dimanfaatkan sebagai dasar dalam mendukung pengambilan keputusan dan meningkatkan pengetahuan [24]. Namun, penelitian ini dibatasi hingga tahap *evaluation* dan *knowledge discovery*, sehingga model klasifikasi sentimen maupun pemodelan topik tidak diimplementasikan ke dalam suatu sistem atau aplikasi. Hasil penelitian berupa model terbaik, hasil evaluasi performa, serta temuan topik utama disajikan sebagai rekomendasi berbasis data bagi pengembang Dola AI dalam menentukan prioritas pengembangan dan peningkatan kualitas aplikasi.

G. Knowledge Discovery

Tahap *knowledge discovery* merupakan tahap lanjutan setelah kerangka CRISP-DM yang bertujuan mengintegrasikan hasil klasifikasi sentimen dan pemodelan topik menjadi suatu pengetahuan yang dapat dimanfaatkan dalam proses pengambilan keputusan. Pada tahap ini, hasil klasifikasi sentimen dimanfaatkan untuk mengetahui proporsi opini positif, netral, dan negatif pengguna terhadap aplikasi Dola AI, sedangkan hasil pemodelan topik menggunakan LDA dimanfaatkan untuk mengidentifikasi isu-isu utama yang mendasari setiap kelompok sentimen tersebut. Kedua hasil analisis tersebut selanjutnya diinterpretasikan secara bersamaan guna mengungkap pola hubungan antara kecenderungan sentimen pengguna dengan topik-topik spesifik yang melatarbelakanginya. Melalui proses ini, penelitian tidak hanya menghasilkan gambaran umum mengenai tingkat kepuasan pengguna, tetapi juga menghasilkan pemahaman yang lebih mendalam mengenai aspek-aspek konkret yang perlu dipertahankan maupun diperbaiki oleh pengembang aplikasi. Pengetahuan yang diperoleh pada tahap ini selanjutnya dirumuskan menjadi rekomendasi berbasis data yang dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam pengembangan dan peningkatan kualitas layanan aplikasi Dola AI. Pendekatan ini sejalan dengan konsep *knowledge discovery*, yaitu proses mengekstraksi pengetahuan yang bermakna dan bermanfaat dari hasil analisis data untuk mendukung pengambilan keputusan yang lebih efektif [30].

III. HASIL DAN PEMBAHASAN

A. Business Understanding

Pada fase *Business Understanding*, telah ditetapkan bahwa tujuan utama penelitian adalah membangun model klasifikasi sentimen yang akurat serta mengidentifikasi topik-topik dominan pada setiap kelompok ulasan pengguna aplikasi Dola AI. Fase ini menghasilkan rumusan pertanyaan analitis

yang menjadi acuan seluruh proses penelitian, yaitu: (a) algoritma manakah di antara SVM, Naïve Bayes, dan KNN yang menghasilkan performa klasifikasi sentimen terbaik pada dataset ulasan Dola AI; dan (b) topik-topik utama pada ulasan pengguna dianalisis menggunakan pemodelan LDA untuk mengungkap aspek-aspek yang paling sering dibahas dalam setiap kategori sentimen. Hasil penelitian diharapkan dapat menjadi dasar rekomendasi berbasis data bagi pengembang dalam menetapkan prioritas pengembangan dan peningkatan kualitas aplikasi.

B. Data Understanding

Pada fase *data understanding*, data ulasan pengguna aplikasi Dola AI berhasil dikumpulkan dari *Google Play Store* melalui proses *web scraping* menggunakan pustaka *google-play-scraper*. Total data yang berhasil dikumpulkan berjumlah 28.226 ulasan yang mencakup periode 1 Maret hingga 25 Juni. Data yang diperoleh terdiri atas dua atribut utama, yakni teks ulasan dan rating. Atribut tersebut berfungsi sebagai fondasi dalam pelaksanaan dua tahapan analisis, yaitu analisis sentimen guna mengukur tingkat kepuasan pengguna, serta pemodelan topik guna mengidentifikasi isu-isu yang paling dominan diperbincangkan dalam ulasan pengguna. Contoh data hasil pengumpulan ditampilkan pada Tabel 1.

TABEL 1
ULASAN DOLA AI

Rating	Ulasan
1	download dulu suaranya cewek aku jadi nggak suka sekarang suaranya cowok Bikin kesal ah
5	oke memang bagus saya suka dengan keterampilanmu☑
5	banyak memberi jalan keluar dalam semua masalah dlm semua bidang
5	ak mengerti saat ak menanyakan Dola tentang hal yg ak tidak tau
3	ak gak tau manfaatnya apaa aku cuman fomoo hahahaha

C. Data Preparation

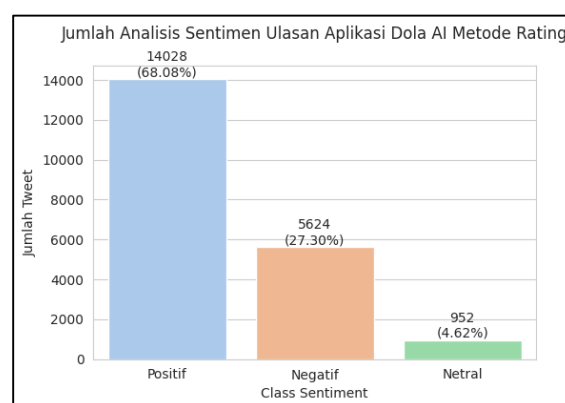
Proses persiapan data mencakup penyaringan data (*data filtering*), praproses teks (*text preprocessing*), dan pelabelan sentimen. Pada tahap penyaringan data, dilakukan penghapusan terhadap ulasan duplikat, ulasan kosong, ulasan yang hanya berisi simbol atau emoji, serta ulasan yang tidak mengandung informasi tekstual yang memadai. Selain itu, beberapa ulasan tidak dapat digunakan setelah melalui tahapan *text preprocessing* karena seluruh token pada ulasan tersebut tereliminasi selama proses *cleaning*, *stopword removal*, dan *stemming*. Dari total 28.226 ulasan hasil *web scraping*, diperoleh 20.604 ulasan yang memenuhi kriteria kualitas data dan selanjutnya digunakan pada proses pelabelan sentimen, pembentukan fitur TF-IDF, klasifikasi sentimen, dan *topic modeling*. Contoh hasil penerapan setiap tahapan *text preprocessing* disajikan pada Tabel 2.

TABEL 2
HASIL PREPROCESSING DATA

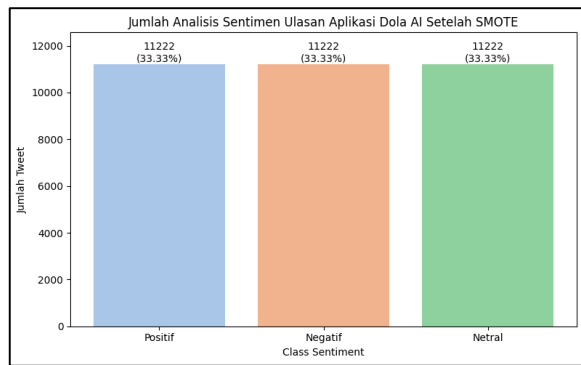
Tahap	Data
Dataset	MEMBANTU BNGT LAH POKOKNYA☐
Cleaning	MEMBANTU BNGT LAH POKOKNYA
Case Folding	membantu bngt lah pokoknya
Normalisasi	membantu banget lah pokoknya
Tokenisasi	['membantu', 'banget', 'lah', 'pokoknya']
Stopword Removal	membantu banget pokoknya
Stemming	bantu banget pokok

Berdasarkan Tabel 2, terlihat bahwa setiap tahapan praproses memberikan transformasi yang signifikan terhadap teks mentah. Proses *cleaning* berhasil menghilangkan karakter non-alfabetik dan emotikon yang tidak relevan, sedangkan *case folding* berhasil menormalkan variasi penulisan huruf kapital. Proses normalisasi kemudian mengubah kata-kata tidak baku, singkatan, atau bahasa gaul menjadi bentuk baku sesuai dengan kaidah bahasa Indonesia, sehingga dapat mengurangi variasi penulisan yang memiliki makna sama. Setelah itu, tokenisasi memecah setiap kalimat menjadi kumpulan token atau kata tunggal yang menjadi dasar dalam proses analisis berikutnya. Proses *stopword removal* mengurangi dimensi fitur dengan membuang kata-kata umum, dan proses *stemming* menggunakan pustaka Sastrawi berhasil mengkonversi kata berimbuhan ke bentuk dasarnya.

Sebelum dilakukan proses analisis sentimen, setiap ulasan terlebih dahulu diberi label sentimen dengan mengacu pada nilai rating yang diberikan pengguna di *Google Play Store*. Pada penelitian ini, Ulasan dengan rating 4–5 ditetapkan sebagai kategori positif, rating 1–2 sebagai kategori negatif, dan rating 3 sebagai kategori netral.



Gambar 2. Distribusi Kelas Sebelum SMOTE



Gambar 3. Distribusi Kelas Setelah SMOTE

Gambar 2-3 menunjukkan perubahan distribusi kelas sentimen sebelum dan sesudah penerapan SMOTE pada data *training* salah satu *fold* sebagai ilustrasi representatif. Sebelum SMOTE, distribusi keseluruhan data menunjukkan ketidakseimbangan kelas dengan kelas positif mendominasi sebanyak 14028 data, diikuti kelas negatif sebesar 5624 data dan kelas netral sebesar 952 data. Setelah penerapan SMOTE, jumlah data pada seluruh kelas menjadi setara sebanyak 11.222 data per kelas, mengikuti jumlah kelas mayoritas pada *fold* tersebut.

D. Modeling

Fase modeling dilakukan setelah seluruh data ulasan melalui tahapan *text preprocessing* dan pelabelan sentimen pada fase *data preparation*. Pada penelitian ini, proses pembangunan model klasifikasi dilakukan menggunakan tiga algoritma machine learning, yaitu SVM, Naïve Bayes, dan KNN. Sebelum proses klasifikasi, data teks diubah menjadi representasi numerik menggunakan metode TF-IDF. Penanganan ketidakseimbangan kelas dan pembentukan model dilakukan mengikuti prosedur *cross-validation* yang telah dijelaskan pada bagian metode.

Seluruh model dievaluasi menggunakan skema *Stratified 5-Fold Cross Validation* untuk memperoleh estimasi performa yang lebih stabil sekaligus mempertahankan proporsi setiap kelas sentimen pada setiap *fold*. Seluruh proses tersebut diulangi sebanyak lima kali hingga setiap *fold* pernah berperan sebagai data pengujian.

Selain proses klasifikasi sentimen, fase *modeling* juga mencakup pembangunan model *topic modeling* menggunakan metode LDA. Berbeda dengan proses klasifikasi yang dilakukan terhadap keseluruhan dataset, pemodelan LDA dilakukan secara terpisah pada masing-masing kelompok hasil pelabelan sentimen, yaitu positif, netral, dan negatif, sehingga topik yang dihasilkan pada setiap kelompok dapat merepresentasikan isu yang lebih spesifik sesuai kecenderungan sentimennya. Proses pembentukan model LDA dilakukan dengan menguji sejumlah kandidat jumlah topik pada setiap kelompok sentimen, kemudian jumlah topik terbaik ditentukan berdasarkan nilai *Coherence Score* tertinggi, sebagaimana dijelaskan lebih lanjut pada bagian *evaluation*.

E. Evaluation

Fase *evaluation* merupakan tahapan untuk menilai kinerja model yang telah dibangun pada fase *modeling*. Pada penelitian ini, evaluasi dilakukan terhadap tiga algoritma klasifikasi, yaitu SVM, Naïve Bayes, dan KNN. Evaluasi dilakukan menggunakan skema *Stratified 5-Fold Cross Validation* dengan mengukur nilai *Accuracy*, *Precision*, *Recall*, dan *F1-score* untuk setiap *fold*. Mengingat distribusi kelas pada dataset tidak seimbang, dengan kelas positif sebagai kelas mayoritas dan kelas netral sebagai kelas minoritas, *Precision*, *Recall*, dan *F1-score* utama dihitung menggunakan *weighted average*. Pendekatan *weighted* digunakan agar kontribusi setiap kelas terhadap nilai agregat mempertimbangkan jumlah sampelnya, sehingga metrik keseluruhan tetap merepresentasikan performa model pada distribusi data aktual. Namun, penggunaan *weighted average* dapat memberikan bobot yang lebih besar kepada kelas mayoritas sehingga performa pada kelas minoritas berpotensi tidak terlihat secara optimal. Oleh karena itu, *Macro-F1* dan metrik per kelas juga dilaporkan melalui *classification report* sebagai evaluasi tambahan untuk memastikan kemampuan model dalam mengenali setiap kelas sentimen, khususnya kelas minoritas. Dengan demikian, evaluasi tidak hanya mempertimbangkan performa agregat berdasarkan distribusi data, tetapi juga memberikan gambaran mengenai keseimbangan kinerja antar kelas.

Evaluasi model Latent Dirichlet Allocation (LDA) dilakukan menggunakan Coherence Score untuk menentukan model dengan jumlah topik yang paling optimal pada setiap kelompok sentimen. Nilai coherence yang lebih tinggi menunjukkan keterkaitan semantik antarkata yang lebih baik sehingga topik yang dihasilkan lebih koheren dan mudah diinterpretasikan.

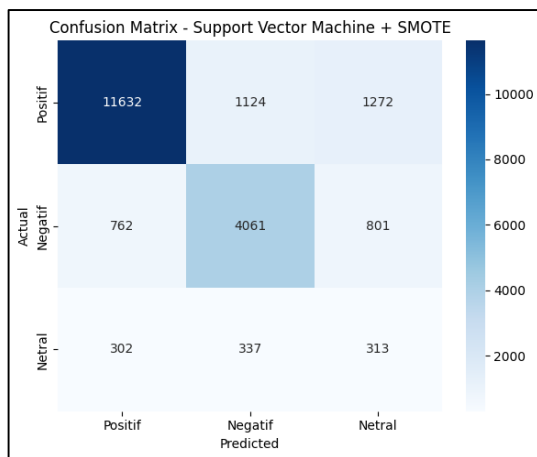
1) SVM: Merujuk Tabel 3, algoritma SVM tercatat stabil pada kelima *fold* pengujian, dengan nilai *accuracy* berkisar antara 76,87% hingga 78,26% dan nilai *F1-score* berkisar antara 79,16% hingga 80,63%. Rata-rata *accuracy* yang diperoleh sebesar 77,68%, *precision* sebesar 83,06%, *recall* sebesar 77,68%, dan *F1-score* sebesar 80,02%. Rentang nilai antarfold yang relatif kecil menunjukkan bahwa performa SVM cukup konsisten pada berbagai pembagian data. Untuk melihat performa klasifikasi secara lebih rinci pada masing-masing kelas sentimen, Gambar 4 menyajikan *classification report* dari model SVM.

TABEL 3
HASIL STRATIFIED 5-FOLD CROSS VALIDATION SVM

Fold	Accuracy	Precision	Recall	F1-score
1	77,55%	83,31%	77,55%	80,05%
2	78,02%	83,31%	78,02%	80,36%
3	78,26%	83,79%	78,26%	80,63%
4	77,72%	82,81%	77,72%	79,93%
5	76,87%	82,09%	76,87%	79,16%
Rata-rata	77,68%	83,06%	77,68%	80,02%

Classification Report				
	precision	recall	f1-score	support
Negatif	0.74	0.72	0.73	5624
Netral	0.13	0.33	0.19	952
Positif	0.92	0.83	0.87	14028
accuracy			0.78	20604
macro avg	0.59	0.63	0.60	20604
weighted avg	0.83	0.78	0.80	20604

Gambar 4. Classification Report SVM



Gambar 5. Confusion Matrix SVM

Berdasarkan Gambar 4-5, model SVM memperoleh *weighted F1-score* sebesar 80,02%. Nilai tersebut menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik secara keseluruhan. Namun, performa model tidak merata pada seluruh kelas sentimen. Berdasarkan *classification report* pada Gambar 4, kelas positif memperoleh *precision* sebesar 0,92, *recall* sebesar 0,83, dan *F1-score* sebesar 0,87. Kelas negatif memperoleh *precision* sebesar 0,74, *recall* sebesar 0,72, dan *F1-score* sebesar 0,73. Sementara itu, kelas netral memperoleh *precision* sebesar 0,13, *recall* sebesar 0,33, dan *F1-score* sebesar 0,19. Hasil tersebut menunjukkan bahwa kelas netral merupakan kelas yang paling sulit diklasifikasikan oleh SVM.

Rendahnya *recall* kelas netral menunjukkan bahwa hanya sebagian data yang sebenarnya termasuk kelas netral berhasil dikenali dengan benar oleh model, sedangkan sebagian lainnya diklasifikasikan ke kelas positif atau negatif. Pola kesalahan tersebut dapat diamati lebih lanjut melalui *confusion matrix* pada Gambar 5, yang menunjukkan distribusi prediksi benar dan salah pada masing-masing kelas. Dengan demikian, *confusion matrix* memberikan informasi mengenai kecenderungan model dalam mengalihkan data dari kelas netral ke kelas sentimen lainnya. Kondisi tersebut mengindikasikan bahwa karakteristik linguistik ulasan netral memiliki tumpang tindih dengan karakteristik ulasan positif maupun negatif, sehingga pemisahan antarkelas menjadi lebih sulit.

Keunggulan SVM secara teoritis dapat dikaitkan dengan kemampuannya dalam menangani ruang fitur berdimensi

tinggi dan sparse yang dihasilkan oleh TF-IDF. Berbeda dengan pendekatan berbasis jarak, SVM membangun *hyperplane* optimal untuk memaksimalkan margin antarkelas sehingga mampu melakukan klasifikasi secara efektif pada karakteristik data tersebut [31]. Temuan ini sejalan dengan Nashir *et al.* yang menyatakan bahwa SVM memiliki kestabilan yang baik dalam mengolah representasi teks berbasis TF-IDF serta mampu mempertahankan performa yang tinggi pada ruang fitur berdimensi tinggi [31].

Meskipun teknik SMOTE telah diterapkan pada data latih pada setiap *fold* untuk mengurangi ketimpangan distribusi kelas, performa kelas netral tetap lebih rendah dibandingkan kelas positif dan negatif. Hal ini menunjukkan bahwa penyeimbangan jumlah sampel belum tentu secara langsung menyelesaikan permasalahan klasifikasi apabila karakteristik fitur antarkelas masih memiliki kemiripan. Oleh karena itu, hasil evaluasi SVM tidak hanya menunjukkan performa agregat yang relatif baik, tetapi juga mengidentifikasi keterbatasan model dalam mengenali kelas netral sebagai kelas minoritas.

2) *Naïve Bayes*: Tabel 4 menunjukkan bahwa model *Naïve Bayes* unggul dari sisi *precision* di antara ketiga algoritma, yaitu rata-rata sebesar 84,82%, dengan nilai *precision* per *fold* yang konsisten berada di atas 84%. Namun demikian, nilai *accuracy* dan *recall* yang diperoleh lebih rendah dibandingkan SVM, yaitu rata-rata sebesar 74,43%, dengan *F1-score* rata-rata sebesar 78,60%. Pola ini memperlihatkan kecenderungan konservatif model dalam memprediksi kelas, sehingga prediksi yang dihasilkan cenderung tepat, tetapi model kurang mampu mengenali seluruh data pada masing-masing kelas sentimen, terutama pada kelas dengan proporsi data yang lebih kecil seperti kelas netral. Gambar 6 menyajikan *classification report* dari model *Naïve Bayes* untuk melihat performa klasifikasi pada tiap kelas sentimen secara lebih rinci.

TABEL 4
HASIL STRATIFIED 5-FOLD CROSS VALIDATION NAÏVE BAYES

Fold	Accuracy	Precision	Recall	F1-score
1	74,79%	85,05%	74,79%	78,90%
2	74,64%	84,96%	74,64%	78,87%
3	74,45%	85,34%	74,45%	78,76%
4	74,25%	84,70%	74,25%	78,42%
5	74,00%	84,06%	74,00%	78,07%
Rata-rata	74,43%	84,82%	74,43%	78,60%

Classification Report				
	precision	recall	f1-score	support
Negatif	0.74	0.68	0.71	5624
Netral	0.13	0.50	0.21	952
Positif	0.94	0.79	0.86	14028
accuracy			0.74	20604
macro avg	0.60	0.66	0.59	20604
weighted avg	0.85	0.74	0.79	20604

Gambar 6. Classification Report *Naïve Bayes*

Berdasarkan Gambar 6, model Naïve Bayes menghasilkan nilai *precision* rata-rata sebesar 84,82%, namun nilai *recall* dan *F1-score* masih berada di bawah SVM. mengindikasikan bahwa model memiliki ketepatan tinggi dalam menentukan suatu kelas, tetapi masih belum mampu mengenali seluruh data pada masing-masing kelas sentimen secara optimal.

Performa tersebut berkaitan dengan mekanisme Naïve Bayes yang mengasumsikan independensi antarfitur dalam satu dokumen. Pada representasi TF-IDF, hubungan antarkata dalam suatu dokumen tidak selalu memenuhi asumsi tersebut sehingga informasi kontekstual antarfitur tidak dapat dimanfaatkan secara optimal. Akibatnya, kemampuan model dalam membedakan kelas sentimen menjadi lebih terbatas dibandingkan SVM, terutama pada data teks yang memiliki ruang fitur berdimensi tinggi dan bersifat *sparse* [31].

3) **KNN**: Berdasarkan hasil pada Tabel 5, model ini menghasilkan performa terendah di antara ketiga algoritma, dengan rata-rata *accuracy* 67,39%, *precision* 75,19%, *recall* 67,39%, dan *F1-score* 70,42%. Nilai *F1-score* antar *fold* juga menunjukkan variasi yang lebih besar dibandingkan SVM dan Naïve Bayes, berkisar antara 69,86% hingga 71,24%. Rendahnya performa KNN ini konsisten dengan karakteristik algoritma yang mengandalkan pengukuran jarak antar data, sehingga kurang optimal ketika diterapkan pada representasi fitur TF-IDF yang berdimensi tinggi dan bersifat *sparse* pada kondisi tersebut, jarak antar titik data menjadi kurang informatif untuk membedakan kelas sentimen secara akurat. Gambar 7 menyajikan *classification report* dari model KNN untuk melihat performa klasifikasi pada tiap kelas sentimen secara lebih rinci.

TABEL 5
HASIL STRATIFIED 5-FOLD CROSS VALIDATION KNN

Fold	Accuracy	Precision	Recall	F1-score
1	67,85%	72,72%	67,85%	69,99%
2	65,15%	79,43%	65,15%	69,86%
3	68,41%	72,84%	68,41%	70,42%
4	67,17%	77,65%	67,17%	71,24%
5	68,37%	73,29%	68,37%	70,57%
Rata-rata	67,39%	75,19%	67,39%	70,42%

Classification Report				
	precision	recall	f1-score	support
Negatif	0.59	0.60	0.59	5624
Netral	0.09	0.28	0.14	952
Positif	0.86	0.73	0.79	14028
accuracy			0.67	20604
macro avg	0.51	0.54	0.51	20604
weighted avg	0.75	0.67	0.71	20604

Gambar 7. Classification Report KNN

Berdasarkan *classification report* tersebut, nilai *precision*, *recall*, maupun *F1-score* KNN konsisten lebih rendah dibandingkan SVM dan Naïve Bayes, menandakan model ini masih kesulitan membedakan karakteristik tiap kelas sentimen secara optimal. Secara teoritis, kondisi tersebut

dapat berkaitan dengan karakteristik KNN yang menggunakan kedekatan antar data sebagai dasar klasifikasi. Pada representasi TF-IDF yang berdimensi tinggi dan bersifat *sparse*, pengukuran jarak antardokumen dapat menjadi kurang informatif sehingga proses penentuan tetangga terdekat menjadi lebih menantang. Namun, penelitian ini tidak melakukan eksperimen tambahan untuk mengisolasi pengaruh dimensi fitur atau metrik jarak terhadap performa KNN. Oleh karena itu, penjelasan tersebut diposisikan sebagai interpretasi teoritis terhadap hasil eksperimen.

4) **Perbandingan Performa Model**: Ketiga algoritma yang telah diuji, yaitu SVM, Naïve Bayes, dan KNN, selanjutnya dibandingkan performanya melalui analisis komparatif guna menentukan algoritma yang paling efektif dalam mengklasifikasikan sentimen pada dataset penelitian ini. Hasil perbandingan ketiga algoritma disajikan tabel 6.

TABEL 6
PERBANDINGAN PERFORMA MODEL

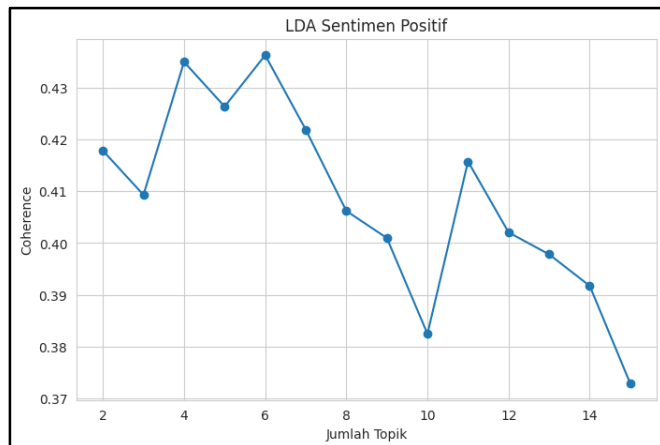
Model	Accuracy	Precision	Recall	F1-score
SVM	77,68%	83,06%	77,68%	80,02%
Naïve Bayes	74,43%	84,82%	74,43%	78,60%
KNN	67,39%	75,19%	67,39%	70,42%

Berdasarkan Tabel 6, SVM mencatatkan nilai *Accuracy*, *Recall*, dan *F1-score* paling tinggi di antara ketiga algoritma, sehingga dipilih sebagai model terbaik pada penelitian ini. Capaian tersebut mengindikasikan bahwa SVM mampu memberikan keseimbangan terbaik antara ketepatan prediksi dan kemampuan mengenali seluruh kelas sentimen. Sebaliknya, Naïve Bayes menghasilkan *precision* tertinggi, tetapi nilai *recall* dan *F1-score* yang lebih rendah menunjukkan bahwa model ini belum sepenuhnya mampu menjangkau seluruh data pada tiap kelas secara optimal. Sementara itu, KNN memperoleh performa terendah karena pendekatan berbasis jarak kurang efektif diterapkan pada representasi fitur TF-IDF yang berdimensi tinggi.

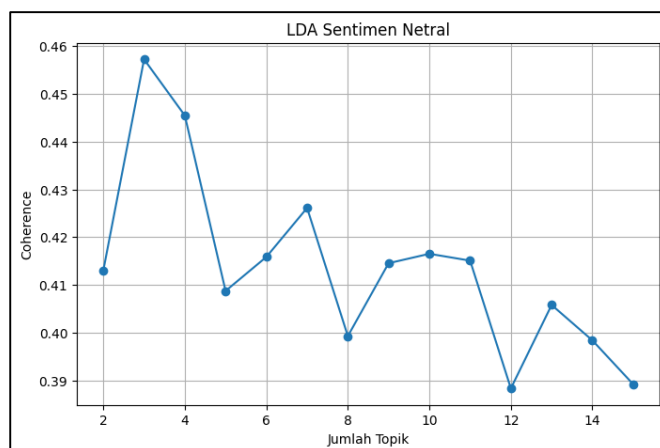
Hasil ini sejalan dengan temuan Wahid *et al.* dan Noor dan Normawati yang melaporkan bahwa SVM memberikan performa lebih unggul dibandingkan Naïve Bayes maupun KNN pada analisis sentimen berbasis TF-IDF LDA [32][33]. Berdasarkan hal tersebut, SVM ditetapkan sebagai model terbaik untuk mendukung tahap knowledge discovery melalui pemodelan topik menggunakan LDA.

5) **Latent Dirichlet Allocation**: Selain melakukan klasifikasi sentimen, penelitian ini menerapkan *Latent Dirichlet Allocation* untuk mengidentifikasi topik utama pada setiap kelompok sentimen. Pemodelan LDA dilakukan setelah kategori sentimen diperoleh dan diterapkan secara terpisah pada kelompok sentimen positif, netral, dan negatif. Pemisahan tersebut bertujuan untuk mengidentifikasi topik yang memiliki karakteristik pembahasan sesuai dengan masing-masing kecenderungan sentimen. Dengan pendekatan tersebut, topik pada kelompok positif dapat merepresentasikan aspek yang diapresiasi pengguna,

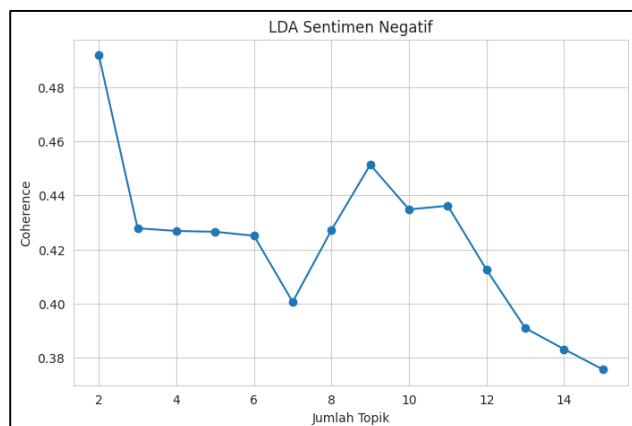
sedangkan topik pada kelompok negatif dapat merepresentasikan isu yang menjadi sumber ketidakpuasan pengguna. Penentuan jumlah topik optimal dilakukan dengan menguji beberapa kandidat jumlah topik pada setiap kelompok sentimen. Setiap model dievaluasi menggunakan *Coherence Score* dengan metrik C_v . Konfigurasi dengan nilai *Coherence Score* tertinggi dipilih sebagai jumlah topik optimal karena menunjukkan tingkat koherensi yang paling baik berdasarkan keterkaitan kata-kata dalam topik.



Gambar 8. Coherence Score Sentimen Positif



Gambar 9. Coherence Score Sentimen Netral



Gambar 10. Coherence Score Sentimen Negatif

Berdasarkan Gambar 8–10, penentuan jumlah topik pada masing-masing kelompok sentimen dilakukan menggunakan *Coherence Score* dengan memilih model yang menghasilkan nilai coherence tertinggi. Hasil evaluasi menunjukkan bahwa kelompok sentimen positif memperoleh nilai coherence tertinggi sebesar 0,4363 pada enam topik, kelompok sentimen netral sebesar 0,4572 pada tiga topik, sedangkan kelompok sentimen negatif sebesar 0,4917 pada dua topik. Hasil tersebut menunjukkan bahwa konfigurasi jumlah topik yang paling optimal berbeda pada setiap kelompok sentimen. Oleh karena itu, pemodelan LDA dilakukan secara independen pada masing-masing kelompok sentimen menggunakan jumlah topik yang menghasilkan nilai *Coherence Score* tertinggi, sehingga topik yang dihasilkan memiliki keterkaitan semantik yang lebih baik dan lebih mudah diinterpretasikan.

Setelah diperoleh jumlah topik optimal berdasarkan nilai *Coherence Score*, tahap selanjutnya adalah menginterpretasikan topik-topik yang dihasilkan oleh model LDA. Interpretasi dilakukan dengan menganalisis kata-kata yang memiliki probabilitas kemunculan tertinggi pada setiap topik untuk mengidentifikasi tema utama yang mendasari pembahasan pengguna. Setiap topik kemudian diberi nama berdasarkan keterkaitan makna dari kata-kata dominan tersebut sehingga mampu merepresentasikan isu yang dibahas pada masing-masing kelompok sentimen. Hasil interpretasi topik pada kelompok sentimen positif, netral, dan negatif berturut-turut disajikan pada Tabel 7, Tabel 8, dan Tabel 9.

TABEL 7
LDA POSITIF

Topik	Kata Dominan	Aspek
1	sangat, aplikasi, bagus, bantu, kerja, suka, ajar, guna, rumah, tugas	Manfaat aplikasi dalam mendukung produktivitas dan pembelajaran
2	dola, kasih, baik, ya, senang, terima, bintang, makasih, moga, cerdas	Kepuasan dan apresiasi pengguna terhadap aplikasi
3	nya, tidak, ai, bagus, ku, tolong, bikin, tahu, pakai, gambar	Kualitas fitur AI untuk pembuatan dan pengolahan gambar
4	banget, mantap, keren, pintar, pokok, si, bicara, cari, bagus, ngerjain	Performa, kecerdasan, dan responsivitas aplikasi
5	ngobrol, curhat, teman, ajak, seru, ngasih, asik, cerita, bukan, buat	Pengalaman berinteraksi melalui fitur percakapan (chat)
6	foto, edit, update, video, sesuai, tidak, tambah, kirim, layan	Fitur edit foto dan video serta pembaruan (update) aplikasi

Hasil pada Tabel 7 menunjukkan bahwa model LDA menghasilkan enam topik utama pada ulasan dengan sentimen positif. Jumlah topik yang lebih banyak dibandingkan kelompok sentimen lainnya menunjukkan bahwa pengalaman positif pengguna terhadap aplikasi Dola AI mencakup berbagai aspek penggunaan. Keenam topik tersebut tidak hanya menggambarkan kepuasan terhadap fungsi utama aplikasi, tetapi juga menunjukkan adanya apresiasi terhadap

kualitas fitur berbasis AI, pengalaman interaksi, serta pembaruan layanan yang diberikan oleh pengembang.

Topik pertama didominasi oleh kata *sangat*, *bagus*, *bantu*, *kerja*, *ajar*, dan *tugas*, yang menunjukkan bahwa pengguna memanfaatkan Dola AI sebagai asisten untuk mendukung produktivitas, pembelajaran, maupun penyelesaian aktivitas sehari-hari. Dominasi kata-kata tersebut mengindikasikan bahwa manfaat fungsional merupakan salah satu faktor utama yang membentuk persepsi positif pengguna terhadap aplikasi.

Topik kedua memperlihatkan dominasi kata *kasih*, *senang*, *terima*, *bintang*, dan *makasih*, yang merepresentasikan bentuk apresiasi pengguna terhadap pengalaman menggunakan aplikasi. Kemunculan kata-kata tersebut menunjukkan bahwa kepuasan pengguna tidak hanya diwujudkan melalui penilaian positif, tetapi juga melalui ekspresi rasa terima kasih dan pemberian rating yang baik.

Topik ketiga berkaitan dengan kualitas fitur berbasis AI, khususnya dalam pembuatan maupun pengolahan gambar. Sementara itu, topik keempat menggambarkan persepsi pengguna terhadap kecerdasan, responsivitas, dan performa aplikasi melalui dominasi kata seperti *mantap*, *keren*, *pintar*, dan *ngeljain*. Kedua topik tersebut menunjukkan bahwa kualitas teknologi AI menjadi salah satu faktor yang memengaruhi pengalaman positif pengguna selama menggunakan aplikasi.

Topik kelima berfokus pada pengalaman pengguna ketika memanfaatkan fitur percakapan (*chat*), sedangkan topik keenam berkaitan dengan fitur pengeditan foto, video, dan pembaruan aplikasi. Keberadaan kedua topik tersebut menunjukkan bahwa pengguna juga memberikan perhatian terhadap pengalaman interaksi serta perkembangan fitur yang disediakan melalui pembaruan aplikasi.

Secara keseluruhan, hasil pemodelan topik menunjukkan bahwa sentimen positif pengguna terbentuk oleh kombinasi antara manfaat aplikasi dalam mendukung produktivitas, kualitas fitur berbasis AI, pengalaman interaksi yang baik, serta pembaruan layanan yang terus dilakukan oleh pengembang. Temuan ini mengindikasikan bahwa kepuasan pengguna tidak hanya dipengaruhi oleh satu fitur tertentu, melainkan oleh keseluruhan pengalaman penggunaan aplikasi. Hasil tersebut sejalan dengan penelitian Zhai *et al.* yang menunjukkan bahwa analisis topik terhadap ulasan pengguna mampu mengidentifikasi faktor-faktor yang memengaruhi kepuasan pengguna dan memberikan masukan bagi pengembangan kualitas layanan aplikasi berdasarkan pengalaman pengguna yang tercermin dalam ulasan [34]. Dengan demikian, aspek-aspek tersebut perlu dipertahankan sebagai keunggulan kompetitif guna meningkatkan kepuasan dan loyalitas pengguna.

TABEL 8
LDA NETRAL

Topik	Kata Dominan	Aspek
1	tidak, nya, bagus, aplikasi, gambar, dola, foto, baik, tolong, edit	Kualitas fitur pengolahan dan pengeditan gambar/foto

2	salah, kadang, bagus, nya, aplikasi, dola, bantu, kasih, kerja, rumah	Kinerja aplikasi dalam membantu aktivitas sehari-hari
3	dola, tidak, bagus, udah, suara, nya, ya, cerita, bikin, suka	Pengalaman penggunaan fitur percakapan dan suara

Hasil pada Tabel 8 menunjukkan bahwa model LDA menghasilkan tiga topik utama pada ulasan dengan sentimen netral. Jumlah topik yang lebih sedikit dibandingkan kelompok sentimen positif menunjukkan bahwa pembahasan pada kelompok ini cenderung berfokus pada beberapa aspek utama tanpa disertai ekspresi kepuasan maupun ketidakpuasan yang kuat. Karakteristik tersebut mengindikasikan bahwa ulasan netral lebih banyak berisi pengalaman penggunaan, pengamatan terhadap fitur, atau penilaian yang bersifat objektif.

Topik pertama didominasi oleh kata *gambar*, *foto*, *edit*, *aplikasi*, *bagus*, dan *tolong*, yang menunjukkan bahwa pengguna banyak membahas fitur pengolahan serta pengeditan gambar. Meskipun terdapat kata *bagus*, kemunculan kata *tidak* dan *nya* secara bersamaan mengindikasikan bahwa pengguna cenderung memberikan penilaian yang bersifat deskriptif atau membandingkan hasil yang diperoleh tanpa menunjukkan kecenderungan sentimen yang dominan. Hal ini menunjukkan bahwa fitur pengolahan gambar menjadi salah satu fungsi yang paling sering digunakan sekaligus dievaluasi oleh pengguna.

Topik kedua berkaitan dengan pemanfaatan aplikasi dalam membantu aktivitas sehari-hari. Dominasi kata seperti *bantu*, *kerja*, *rumah*, dan *bagus* menunjukkan bahwa pengguna mengakui manfaat aplikasi dalam mendukung berbagai aktivitas, namun kemunculan kata *salah* mengindikasikan bahwa aplikasi masih sesekali menghasilkan respons atau keluaran yang belum sesuai harapan. Dengan demikian, ulasan pada topik ini lebih merefleksikan evaluasi terhadap konsistensi performa aplikasi dibandingkan penilaian positif maupun negatif.

Topik ketiga menggambarkan pengalaman penggunaan fitur percakapan (*chat*) dan suara. Kata-kata dominan seperti *suara*, *cerita*, *bikin*, dan *suka* menunjukkan bahwa pengguna memanfaatkan fitur komunikasi sebagai salah satu fungsi utama aplikasi. Namun, keberadaan kata *tidak* bersama kata-kata tersebut mengindikasikan bahwa pengguna lebih banyak menyampaikan pengalaman penggunaan atau memberikan penilaian terhadap fungsi yang tersedia tanpa menunjukkan kecenderungan emosi yang kuat.

Secara keseluruhan, hasil pemodelan topik menunjukkan bahwa sentimen netral didominasi oleh pembahasan mengenai kualitas fitur pengolahan gambar, konsistensi performa aplikasi dalam membantu aktivitas pengguna, serta pengalaman menggunakan fitur percakapan dan suara. Temuan ini menunjukkan bahwa kelompok sentimen netral lebih banyak berisi evaluasi terhadap fungsi aplikasi daripada ekspresi kepuasan ataupun ketidakpuasan. Informasi tersebut memberikan gambaran mengenai aspek-aspek yang masih

memerlukan penyempurnaan agar pengalaman pengguna dapat bergeser menuju sentimen yang lebih positif. Temuan ini juga sejalan dengan penelitian Ender *et al.* yang melaporkan bahwa ulasan dengan sentimen netral lebih banyak memuat informasi mengenai fungsi dan fitur aplikasi serta cenderung bersifat deskriptif tanpa menunjukkan ekspresi emosional yang kuat [35].

TABEL 9
LDA NEGATIF

Topik	Kata Dominan	Aspek
1	aplikasi, dola, salah, nya, iklan, banget, jelek, bagus, ku, ai	Keluhan terhadap kualitas aplikasi dan keberadaan iklan
2	tidak, nya, ai, foto, sesuai, gambar, edit, bikin, dola, kasih	Ketidakpuasan terhadap hasil edit foto dan gambar berbasis AI

Hasil pada Tabel 9 menunjukkan bahwa model LDA menghasilkan dua topik utama pada ulasan dengan sentimen negatif. Hasil LDA mengidentifikasi dua topik pada kelompok negatif, yang secara interpretatif menunjukkan dua kelompok isu utama berdasarkan kata-kata dominan. Kondisi ini mengindikasikan bahwa pengalaman negatif pengguna tidak tersebar pada berbagai aspek aplikasi, melainkan didominasi oleh permasalahan tertentu yang secara konsisten muncul dalam ulasan.

Topik pertama didominasi oleh kata *aplikasi*, *dola*, *salah*, *iklan*, *jelek*, dan *AI*, yang menunjukkan bahwa sebagian besar keluhan pengguna berkaitan dengan kualitas aplikasi secara keseluruhan. Kemunculan kata *salah* mengindikasikan adanya kesalahan atau respons aplikasi yang belum sesuai harapan pengguna, sedangkan kata *iklan* menunjukkan bahwa keberadaan iklan masih dianggap mengganggu kenyamanan selama menggunakan aplikasi. Dominasi kata-kata tersebut menunjukkan bahwa stabilitas aplikasi dan pengalaman penggunaan (*user experience*) masih menjadi perhatian utama pengguna.

Topik kedua didominasi oleh kata *AI*, *foto*, *gambar*, *edit*, *sesuai*, dan *bikin*, yang menggambarkan ketidakpuasan pengguna terhadap fitur berbasis kecerdasan buatan, khususnya pada proses pengeditan foto maupun pembuatan gambar. Kemunculan kata *tidak* dan *sesuai* secara bersamaan menunjukkan bahwa hasil yang dihasilkan oleh fitur AI masih belum mampu memenuhi ekspektasi pengguna. Temuan ini mengindikasikan bahwa kualitas keluaran (*output quality*) dari fitur AI menjadi salah satu faktor yang memengaruhi munculnya sentimen negatif.

Secara keseluruhan, hasil pemodelan topik menunjukkan bahwa sentimen negatif pengguna terutama dipengaruhi oleh dua aspek utama, yaitu kualitas aplikasi dan performa fitur berbasis AI. Berbeda dengan kelompok sentimen positif yang membahas berbagai manfaat aplikasi, kelompok sentimen negatif lebih terfokus pada isu-isu tertentu yang dianggap paling mengganggu pengalaman pengguna. Kondisi ini menunjukkan bahwa peningkatan stabilitas aplikasi, pengurangan gangguan akibat iklan, serta peningkatan

kualitas hasil pengolahan gambar berbasis AI berpotensi memberikan dampak terbesar dalam meningkatkan kepuasan pengguna.

F. Knowledge Discovery

Berdasarkan hasil analisis sentimen menggunakan model SVM sebagai model terbaik serta hasil *topic modeling* menggunakan LDA, diperoleh sejumlah pengetahuan (*knowledge*) mengenai persepsi pengguna terhadap aplikasi Dola AI. Analisis sentimen menunjukkan kecenderungan polaritas ulasan pengguna, sedangkan LDA mengidentifikasi topik-topik yang mendasari munculnya setiap kelompok sentimen. Integrasi kedua hasil tersebut menghasilkan informasi yang lebih komprehensif dibandingkan apabila hanya menggunakan salah satu metode, karena tidak hanya menunjukkan apakah pengguna memberikan respons positif, netral, atau negatif, tetapi juga mengungkap faktor-faktor yang memengaruhi terbentuknya sentimen tersebut.

Hasil analisis sentimen dan topic modeling menggunakan LDA terhadap ulasan pengguna Dola AI menunjukkan tiga pola utama. Sentimen positif didominasi oleh apresiasi pengguna terhadap manfaat aplikasi dalam mendukung produktivitas dan pembelajaran, serta pengalaman berinteraksi melalui fitur percakapan dan pengolahan gambar maupun video berbasis AI. Sebaliknya, sentimen negatif didominasi oleh keluhan pengguna terhadap tiga hal, yaitu kestabilan performa aplikasi, hasil fitur pengeditan gambar berbasis AI yang belum optimal, dan keberadaan iklan yang mengganggu kenyamanan penggunaan. Sementara itu, sentimen netral mengindikasikan bahwa sebagian pengguna masih berada pada tahap mengevaluasi konsistensi kualitas layanan aplikasi. Temuan ini sejalan dengan [13], yang menunjukkan bahwa integrasi analisis sentimen dan topic modeling mampu mengungkap aspek-aspek utama yang menjadi perhatian pengguna secara lebih komprehensif dibandingkan analisis sentimen yang diterapkan secara terpisah.

Temuan ini memiliki relevansi khusus dalam konteks industri aplikasi *generative AI* di Indonesia. Persepsi pengguna terhadap keandalan dan relevansi keluaran AI terbukti berkontribusi pada terbentuknya niat untuk terus menggunakan layanan berbasis AI [36]. Oleh karena itu, tingginya keluhan pada sentimen negatif, terutama pada aspek performa dan hasil fitur pengeditan gambar berbasis AI, perlu menjadi perhatian serius bagi pengembang, karena masalah tersebut berisiko menurunkan niat pengguna untuk bertahan menggunakan aplikasi. Hal ini sejalan dengan temuan bahwa aspek *usability* mencakup kemudahan pakai, fungsionalitas, dan keandalan secara konsisten memengaruhi tingkat kepuasan pengguna aplikasi *generative AI* [2]. Selain itu, keberadaan iklan yang dikeluhkan pengguna mengindikasikan bahwa strategi monetisasi Dola AI belum sepenuhnya diimbangi dengan kenyamanan penggunaan. Studi skala besar terhadap ulasan aplikasi menunjukkan bahwa frekuensi dan gangguan iklan menjadi keluhan yang paling sering memicu penilaian rendah dari pengguna [37],

sehingga kondisi ini berisiko mengurangi daya saing Dola AI di tengah persaingan aplikasi generative AI yang semakin bertumpu pada pengalaman pengguna tanpa gangguan.

Pada sentimen positif, topik mengenai produktivitas, pembelajaran, serta interaksi dengan fitur pengolahan gambar dan video berbasis AI menunjukkan aspek yang telah diapresiasi pengguna. Faktor kegunaan ini, bersama dengan kemudahan pakai dan keandalan, terbukti secara empiris menjadi penentu utama penerimaan pengguna terhadap teknologi generative AI [38], sehingga sebaiknya dipertahankan dan dikembangkan lebih lanjut agar manfaat yang telah dirasakan pengguna tetap terjaga. Sementara itu, kemunculan sentimen netral terkait evaluasi konsistensi kualitas layanan mengindikasikan perlunya peningkatan stabilitas keluaran AI secara berkelanjutan, mengingat keandalan layanan telah terbukti menjadi salah satu faktor yang membentuk niat keberlanjutan penggunaan aplikasi berbasis AI [36].

Dengan demikian, hasil penelitian memberikan rekomendasi pengembangan yang diturunkan dari hubungan antara topik, polaritas sentimen, dan masalah yang ditemukan dalam ulasan pengguna. Aspek performa aplikasi, kualitas pengeditan gambar berbasis AI, dan gangguan iklan menjadi prioritas perbaikan karena muncul pada kelompok sentimen negatif, sedangkan fitur produktivitas, pembelajaran, dan kemampuan AI yang memperoleh respons positif perlu dipertahankan dan dikembangkan. Temuan ini juga dapat menjadi masukan strategis bagi daya saing Dola AI dalam ekosistem aplikasi generative AI di Indonesia yang terus berkembang.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan kerangka kerja CRISP-DM untuk melakukan analisis sentimen dan *knowledge discovery* pada 20.604 ulasan pengguna aplikasi Dola AI. Hasil evaluasi menunjukkan bahwa algoritma SVM merupakan model klasifikasi dengan performa terbaik dibandingkan Naïve Bayes dan KNN, dengan rata-rata *accuracy* sebesar 77,68%, *weighted precision* sebesar 83,06%, *weighted recall* sebesar 77,68%, dan *weighted F1-score* sebesar 80,02%. Capaian tersebut mengindikasikan bahwa SVM lebih unggul dalam mengklasifikasikan sentimen ulasan pengguna yang direpresentasikan menggunakan fitur TF-IDF dibandingkan dua algoritma pembandingan.

Selain menghasilkan klasifikasi sentimen, penelitian ini juga menerapkan LDA guna mengidentifikasi topik-topik dominan yang mendasari setiap kelompok sentimen. Hasil *topic modeling* menunjukkan bahwa sentimen positif didominasi oleh pembahasan mengenai kemanfaatan aplikasi, kualitas fitur berbasis AI, pengalaman berinteraksi melalui fitur percakapan, serta pembaruan fitur aplikasi. Sementara itu, sentimen negatif didominasi oleh keluhan terhadap kualitas aplikasi, keberadaan iklan, serta hasil fitur pengeditan foto berbasis AI yang belum memenuhi harapan pengguna. Temuan tersebut menyajikan wawasan yang lebih

komprehensif mengenai faktor pendorong kepuasan maupun ketidakpuasan pengguna, sehingga dapat menjadi dasar bagi pengembang dalam menentukan prioritas pengembangan dan peningkatan kualitas layanan Dola AI.

Secara akademis, penelitian ini memberikan kontribusi melalui perbandingan performa tiga algoritma klasifikasi pada domain ulasan aplikasi *generative AI* serta mengintegrasikan analisis sentimen dengan *topic modeling* untuk menghasilkan informasi yang lebih komprehensif dibandingkan klasifikasi sentimen saja. Dari sisi praktis, temuan penelitian ini dapat menjadi dasar rekomendasi berbasis data bagi pengembang Dola AI dalam mengevaluasi kualitas layanan, mempertahankan aspek yang telah memberikan pengalaman positif, serta mengutamakan peningkatan kualitas pada aspek-aspek yang paling banyak dikeluhkan oleh pengguna.

Penelitian ini tidak terlepas dari sejumlah keterbatasan. Pertama, data penelitian hanya bersumber dari ulasan Google Play Store sehingga belum sepenuhnya merepresentasikan persepsi pengguna yang berasal dari platform lain, misalnya App Store atau berbagai media sosial. Kedua, distribusi kelas sentimen yang mengalami ketidakseimbangan, khususnya pada kelas netral, masih memengaruhi kemampuan model dalam mengenali kelas tersebut secara optimal. Selain itu, proses *knowledge discovery* masih menggunakan metode LDA, sehingga interpretasi topik yang dihasilkan masih bergantung pada analisis peneliti terhadap kata-kata dominan pada setiap topik. Berdasarkan keterbatasan tersebut, penelitian selanjutnya dapat mempertimbangkan penggunaan data dari berbagai platform, menerapkan optimasi parameter pada model klasifikasi, memanfaatkan representasi teks berbasis *embedding* atau *pre-trained language model* seperti IndoBERT, serta mengeksplorasi metode pemodelan topik yang lebih mutakhir, seperti BERTopic atau Top2Vec, guna memperoleh topik yang lebih representatif dan mudah dipahami.

DAFTAR PUSTAKA

- [1] F. F.-H. Nah, R. Zheng, J. Cai, K. Siau, and L. Chen, "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration," *J. Inf. Technol. Case Appl. Res.*, vol. 25, no. 3, pp. 277–304, 2023, doi: 10.1080/15228053.2023.2233814.
- [2] R. Alabduljabbar, "User-centric AI: evaluating the usability of generative AI applications through user reviews on app stores," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/peerj-cs.2421.
- [3] D. H. Hagos, R. Battle, and D. B. Rawat, "Recent Advances in Generative AI and Large Language Models: Current Status, Challenges, and Perspectives," *IEEE Trans. Artif. Intell.*, vol. 5, no. 12, pp. 5873–5893, Dec. 2024, doi: 10.1109/TAI.2024.3444742.
- [4] A. Iorliam and J. A. Ingio, "A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing," *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 311–325, 2024, doi: 10.62411/jcta.9447.
- [5] O. N. Salsabila, *Navigasi Dola AI - Smart Assistant: Menjelajahi Arsitektur Kecerdasan Buatan Dalam Ruang Komunikasi Digital*, 1st ed. Karawang: Afdan Rojabi Publisher, 2026.
- [6] M. Alghareeb, A. S. Albeshir, and A. Asif, "Studying Users' Perceptions of COVID-19 Mobile Applications in Saudi Arabia," *Sustainability*, vol. 15, no. 2, pp. 1–17, 2023, doi: 10.3390/su15020956.
- [7] T. N. Prakash and A. Aloysius, "Textual Sentiment Analysis using

- Lexicon Based Approaches,” *Ann. Rom. Soc. Cell Biol.*, vol. 25, no. 4, pp. 9878–9885, 2021.
- [8] A. R. Makhtum and M. Muhajir, “Sentiment Analysis of Omnibus Law Using Support Vector Machine (Svm) With Linear Kernel,” *Barekeng*, vol. 17, no. 4, pp. 2197–2206, 2023, doi: 10.30598/barekengvol17iss4pp2197-2206.
- [9] T. Kim and J. S. Lee, “Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers,” *IEEE Access*, vol. 10, pp. 22724–22736, 2022, doi: 10.1109/ACCESS.2022.3155231.
- [10] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, “Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications,” *J. Big Data*, vol. 11, no. 1, pp. 1–55, 2024, doi: 10.1186/s40537-024-00973-y.
- [11] M. M. A. Pratama, D. Kurniawan, A. Rifai, and K. D. Tania, “Analisis Sentimen dan Topik Ulasan Aplikasi Komunitas Digital Gamer dengan SVM - LDA dan CRISP - DM,” vol. 15, no. 1, pp. 181–193, 2026, doi: 10.32520/stmsi.v15i1.5746.
- [12] N. A. S. ER, “Implementasi Latent Dirichlet Allocation (Lda) Untuk Klasterisasi Cerita Berbahasa Bali,” *J. Teknol. Inf. Ilmu Komput.*, vol. 8, no. 1, pp. 127–134, 2021, doi: 10.25126/jtiik.202183556.
- [13] S. Nurmallasari, A. N. Hidayanto, L. A. Huwaida, and H. Wulandari, “Sentiment Analysis and Topic Modeling of Citizen Satisfaction with the Indonesian Government in Handling a Pandemic,” *OPSearch Am. J. Open Res.*, vol. 2, no. 7, pp. 246–256, 2023, doi: 10.58811/opsearch.v2i6.61.
- [14] S. A. Putri, K. D. Tania, and Naretha Kawadha Pasemah Gumay, “Knowledge Discovery Through Sentiment Analysis and Topic Modeling of BCA Mobile and MyBCA,” *J. Math. Comput. Stat.*, vol. 8, no. 2, pp. 669–682, 2025, doi: 10.35580/jmathcos.v8i2.9782.
- [15] D. K. Harahap, K. D. Tania, and P. E. Sevtiyuni, “Topic Mining-Based Knowledge Discovery of User Health Information Needs,” *Indones. J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 4, pp. 641–653, 2025, doi: https://doi.org/10.35882/ijeemi.v7i4.270.
- [16] M. R. Maulizidan and K. D. Tania, “Knowledge Discovery Based on Sentiment Analysis of Public Perceptions About Generative AI on X,” *Indones. J. Informatics Educ.*, vol. 9, no. 2, pp. 92–107, 2025, doi: http://dx.doi.org/10.20961/ijie.v9i2.107758.
- [17] W. Ningsih, B. Alfianza, R. Rahmadden, and D. Wulandari, “Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [18] R. K. Das, M. Islam, M. M. Hasan, S. Razia, M. Hassan, and S. A. Khushbu, “Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models,” *Heliyon*, vol. 9, no. 9, pp. 1–22, 2023, doi: 10.1016/j.heliyon.2023.e20281.
- [19] M. E. M. Abo *et al.*, “A multi-criteria approach for arabic dialect sentiment analysis for online reviews: Exploiting optimal machine learning algorithm selection,” *Sustainability*, vol. 13, no. 18, pp. 1–20, 2021, doi: 10.3390/su131810018.
- [20] M. Hanafi and M. Furqan, “Perbandingan Analisis Sentimen Presiden 2024 Menggunakan Algoritma Support Vector Machine dan K-Nearest Neighbor,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 10, no. 1, pp. 275–285, 2025, doi: 10.24114/cess.v10i1.67747.
- [21] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [22] M. R. Maulana, K. D. Tania, and P. Sari, “Comparative Analysis of Machine Learning Models for Daily Urban Waste Volume’s Knowledge Discovery,” *Int. Conf. Artif. Intell. Technol. Solut.*, pp. 131–136, 2025, doi: 10.1109/ICAITEch66481.2025.11387653.
- [23] Rudianto, T. I. Solihati, and R. Kania, *Data Mining Untuk Analisis Prilaku dan Kecerdasan Bisnis*, 1st ed. Banten: PT Literapedia Utama, 2025.
- [24] A. Sánchez, C. Vidal-Silva, G. Mancilla, M. Tupac-Yupanqui, and J. M. Rubio, “Sustainable e-Learning by Data Mining—Successful Results in a Chilean University,” *Sustainability*, vol. 15, no. 2, pp. 1–16, 2023, doi: 10.3390/su15020895.
- [25] A. Nursikuwagus, Suherman, H. Purwanto, and T. Hartono, “SUPPORT VECTOR MACHINE TO CLASSIFY SENTIMENT REVIEWS ON GOOGLE PLAY STORE,” *JITK*, vol. 11, no. 3, pp. 724–732, 2022, doi: 10.33480/jitk.v11i3.7282.kemampuan.
- [26] S. Szeghalmy and A. Fazekas, “A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning,” *Sensors*, vol. 23, no. 4, pp. 1–27, 2023, doi: 10.3390/s23042333.
- [27] A. Setiawan, R. Yanti, E. Ali, and H. Yenni, “Optimasi Klasifikasi Tingkat Obesitas Pada Remaja Berdasarkan Pola Hidup Menggunakan SVM Dengan Teknik Smote,” *J. Algoritma*, vol. 22, no. 2, pp. 224–235, 2025, doi: 10.33364/algoritma/v.22-2.2509.
- [28] R. Andriyanto and K. Kusri, “Enhancing Sentiment Analysis Effectiveness with LSTM Variants, and Stratified K-Fold on Imbalanced Dataset,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 2, pp. 981–990, 2024, doi: 10.47065/bits.v6i2.5721.
- [29] M. Rudiger, D. Antons, A. M. Joshi, and T.-O. Salge, “Topic modeling revisited: New evidence on algorithm performance and quality metrics,” *PLoS One*, vol. 17, no. 4, pp. 1–25, 2022, doi: 10.1371/journal.pone.0266325.
- [30] U. Fayyad and R. Uthrusamy, “Data Mining and Knowledge Discovery in Databases,” *Commun. ACM*, vol. 39, no. 11, pp. 24–26, 1996, doi: 10.1145/240455.240463.
- [31] M. Nashir, Dian Ade Kurnia, Yudhistira Arie Wijaya, Ade Irma Purnama Sari, and Nisa Dienwati Nuris, “Perbandingan Kinerja Svm Dan Naïve Bayes Pada Analisis Sentimen Komentar Demonstrasi Dpr 25 Agustus 2025,” *J. Mhs. Sist. Inf.*, vol. 7, no. 1, pp. 392–401, 2025, doi: 10.24127/jmsi.v7i1.10650.
- [32] Y. A. Wahid, Sanatang, and D. D. Andayani, “Performance Comparison of Svm and Naïve Bayes For Indonesian-Language Sentiment Analysis On Free Fire Reviews Using Tf-Idf And Smote,” *J. Embed. Syst. Secur. Intell. Syst.*, vol. 6, no. 4, pp. 674–689, 2025, doi: 10.59562/jessi.v6i4.10818.
- [33] A. R. Noor and D. Normawati, “Perbandingan Support Vector Machine Dan K-Nearest Neighbor Pada Analisis Sentimen Pelaksanaan Pemilihan Presiden Tahun 2024,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 11, no. 2, pp. 1021–1030, 2026, doi: 10.29100/jupi.v11i2.8120.
- [34] Y. Zhai, X. Song, Y. Chen, and W. Lu, “A Study of Mobile Medical App User Satisfaction Incorporating Theme Analysis and Review Sentiment Tendencies,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 12, 2022, doi: 10.3390/ijerph19127466.
- [35] Endar Nirmala and Andri Fahmi, “Comparison of LSTM and Naïve Bayes in Google Play Store App Review Sentiment Analysis,” *J. Inotera*, vol. 11, no. 1, pp. 171–181, 2026, doi: 10.31572/inotera.vol11.iss1.2026.id653.
- [36] E. Ernest Mfumbilwa, C. Mnong’one, E. Chao, and D. Amani, “Invigorating continuance intention among users of AI chatbots in the banking industry: an empirical study from Tanzania,” *Cogent Bus. Manag.*, vol. 11, no. 1, pp. 1–17, 2024, doi: 10.1080/23311975.2024.2419482.
- [37] C. Gao, J. Zeng, D. Lo, X. Xia, I. King, and M. R. Lyu, “Understanding in-app advertising issues based on large scale app review analysis,” *Inf. Softw. Technol.*, vol. 142, pp. 1–13, 2022, doi: 10.1016/j.infsof.2021.106741.
- [38] P. Shrivastava, “Understanding acceptance and resistance toward generative AI technologies: a multi-theoretical framework integrating functional, risk, and sociolegal factors,” *Front. Artif. Intell.*, vol. 8, no. April, pp. 1–10, 2025, doi: 10.3389/frai.2025.1565927.