



Evaluasi Model Latent Dirichlet Allocation untuk Topic Modelling pada Ulasan SIMKOPDES Menggunakan Coherence Score dan Perplexity

Nuri Cahyono*, Kamarudin

Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Amikom Yogyakarta, Sleman, Indonesia

Email: ^{1,*}nuricahyono@amikom.ac.id, ²kamarudin@amikom.ac.id

Email Penulis Korespondensi: nuricahyono@amikom.ac.id

Abstrak—Transformasi digital pada Koperasi Desa/Kelurahan Merah Putih didukung melalui pengembangan Aplikasi SIMKOPDES sebagai platform layanan digital bagi anggota koperasi. Seiring meningkatnya jumlah pengguna, ulasan pada Google Play Store menjadi sumber informasi yang penting untuk mengevaluasi kualitas layanan aplikasi. Karakteristik ulasan yang berupa data teks tidak terstruktur dalam jumlah besar menyebabkan proses analisis secara manual menjadi kurang efektif. Penelitian ini bertujuan mengidentifikasi topik-topik dominan pada ulasan pengguna Aplikasi SIMKOPDES menggunakan metode Latent Dirichlet Allocation (LDA) serta menentukan model topik terbaik berdasarkan evaluasi statistik dan interpretabilitas topik. Penelitian diawali dengan proses scraping ulasan Google Play Store, dilanjutkan text preprocessing yang meliputi case folding, tokenizing, normalization, stopword removal, dan stemming. Selanjutnya dilakukan pemodelan LDA menggunakan tiga variasi jumlah topik yang dievaluasi menggunakan Coherence Score dan Perplexity, serta divisualisasikan melalui Intertopic Distance Map dan Word Cloud. Hasil penelitian menunjukkan bahwa model satu topik memperoleh Coherence Score tertinggi sebesar 0,419677, sedangkan model tiga topik menghasilkan nilai Perplexity terendah sebesar -6,2917. Meskipun kedua metrik menunjukkan hasil yang berbeda, model tiga topik dipilih karena menghasilkan pemisahan isu yang lebih jelas dan interpretasi topik yang lebih representatif, meliputi keanggotaan koperasi dan manfaat aplikasi, proses registrasi dan verifikasi akun, serta pengalaman penggunaan aplikasi. Penelitian ini menunjukkan bahwa pemilihan model LDA sebaiknya tidak hanya didasarkan pada metrik statistik tunggal, tetapi juga mempertimbangkan interpretabilitas topik untuk menghasilkan analisis yang lebih komprehensif dan mendukung peningkatan kualitas layanan Aplikasi SIMKOPDES.

Kata Kunci: Latent Dirichlet Allocation; Topic Modelling; SIMKOPDES; Coherence Score; Perplexity; Ulasan Pengguna

Abstract—The digital transformation of Koperasi Desa/Kelurahan Merah Putih is supported through the development of the SIMKOPDES Mobile application as a digital service platform for cooperative members. As the number of users continues to grow, reviews posted on Google Play Store have become an important source of information for evaluating application quality and user satisfaction. The large volume of unstructured textual reviews makes manual analysis inefficient. This study aims to identify the dominant topics in SIMKOPDES Mobile user reviews using Latent Dirichlet Allocation (LDA) and evaluate the quality of the generated topic models based on Coherence Score and Perplexity. User reviews were collected through web scraping, followed by text preprocessing consisting of case folding, tokenization, normalization, stopword removal, and stemming. LDA models were then developed using three different numbers of topics and evaluated using Coherence Score and Perplexity, while Intertopic Distance Map and Word Cloud visualizations were employed to support topic interpretation. The results indicate that the one-topic model achieved the highest Coherence Score of 0.419677, whereas the three-topic model produced the lowest Perplexity value of -6.2917. Although these evaluation metrics yielded different optimal results, the three-topic model was selected because it generated more distinct, representative, and semantically interpretable topics related to cooperative membership and application benefits, account registration and verification, and user experience. This study demonstrates that selecting the optimal LDA model should not rely solely on statistical evaluation metrics but should also consider semantic interpretability, providing a more comprehensive approach for topic modelling of digital application reviews.

Keywords: Latent Dirichlet Allocation; Topic Modelling; SIMKOPDES; Coherence Score; Perplexity; User Review

1. PENDAHULUAN

Sebagai bagian dari transformasi digital layanan publik, Kementerian Koperasi Republik Indonesia mengembangkan Aplikasi SIMKOPDES sebagai platform digital yang mendukung operasional Koperasi Desa/Kelurahan Merah Putih. Aplikasi ini menyediakan berbagai layanan, seperti pengelolaan data anggota, transaksi koperasi, layanan simpan pinjam, serta penyampaian informasi dalam satu platform digital [1],[2],[3],[4]. Seiring meningkatnya jumlah pengguna, ulasan yang diberikan melalui Google Play Store menjadi sumber informasi penting untuk mengevaluasi kualitas layanan aplikasi karena memuat pengalaman, kritik, saran, dan berbagai kendala yang dialami pengguna [5]. Namun, volume ulasan yang terus bertambah dalam bentuk data teks tidak terstruktur menyebabkan proses analisis secara manual menjadi tidak efisien, membutuhkan waktu yang lama, serta berpotensi menghasilkan interpretasi yang subjektif [6]. Karakteristik ulasan pengguna yang menggunakan bahasa informal, singkatan, kesalahan penulisan, dan variasi kosakata juga menambah kompleksitas proses analisis sehingga diperlukan pendekatan komputasional yang mampu mengekstraksi informasi penting secara otomatis.

Salah satu pendekatan yang banyak digunakan untuk menganalisis kumpulan dokumen teks adalah Topic Modelling, yaitu teknik unsupervised machine learning yang bertujuan mengidentifikasi struktur semantik tersembunyi berupa sekumpulan topik berdasarkan pola kemunculan kata pada suatu korpus dokumen [7]. Berbeda dengan analisis sentimen yang hanya mengelompokkan opini ke dalam kategori positif, negatif, atau netral, Topic Modelling mampu mengidentifikasi isu, tema, dan permasalahan yang dominan dibahas pengguna tanpa memerlukan proses pelabelan data. Pendekatan ini telah banyak diterapkan pada analisis media sosial, ulasan produk, layanan publik, dokumen ilmiah, maupun berita daring untuk mendukung pengambilan keputusan, evaluasi layanan, identifikasi kebutuhan pengguna, dan



penyusunan strategi pengembangan sistem berbasis data [8]. Dalam konteks penelitian ini, Topic Modelling digunakan untuk mengidentifikasi isu-isu dominan yang muncul pada ulasan pengguna Aplikasi SIMKOPDES sebagai dasar evaluasi kualitas layanan aplikasi.

Salah satu algoritma Topic Modelling yang paling banyak digunakan adalah Latent Dirichlet Allocation (LDA). LDA merupakan model probabilistik generatif yang mengasumsikan bahwa setiap dokumen terdiri atas kombinasi beberapa topik, sedangkan setiap topik direpresentasikan oleh distribusi probabilitas kata-kata tertentu. Keunggulan LDA terletak pada kemampuannya mengidentifikasi hubungan laten antar kata sehingga mampu menghasilkan kelompok topik yang representatif terhadap keseluruhan dokumen [9]. Berbagai penelitian menunjukkan bahwa LDA memiliki performa yang baik dalam mengolah data teks berukuran besar dan mampu menghasilkan topik yang mudah diinterpretasikan. Namun demikian, kualitas topik yang dihasilkan sangat dipengaruhi oleh pemilihan parameter model, terutama jumlah topik (number of topics). Penentuan jumlah topik yang tidak tepat dapat menghasilkan topik yang saling tumpang tindih atau justru terlalu umum sehingga sulit diinterpretasikan. Diperlukan proses evaluasi model untuk memperoleh konfigurasi LDA yang optimal. Dua metrik evaluasi yang banyak digunakan adalah Coherence Score dan Perplexity. Coherence Score digunakan untuk mengukur tingkat keterkaitan semantik antar kata dalam suatu topik sehingga semakin tinggi nilainya menunjukkan topik semakin mudah dipahami oleh manusia [10]. Sebaliknya, Perplexity digunakan untuk mengukur kemampuan model dalam merepresentasikan kumpulan dokumen, di mana nilai Perplexity yang lebih rendah menunjukkan kemampuan prediksi model yang lebih baik [11].

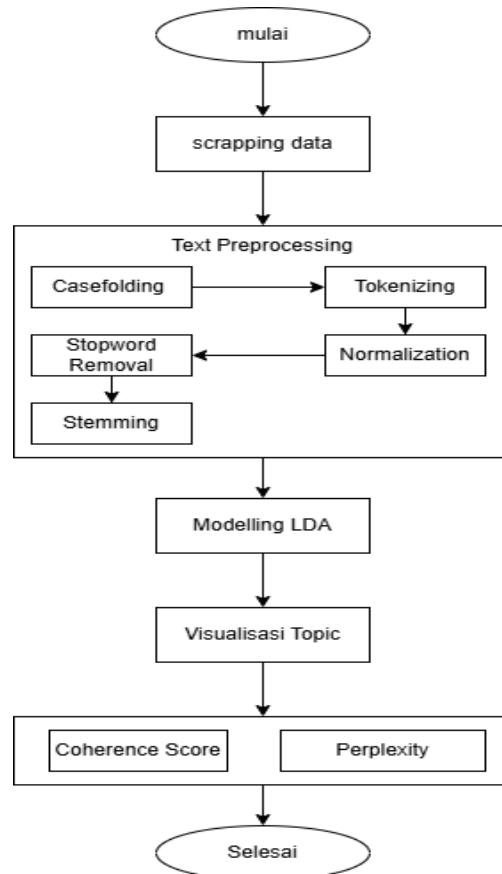
Berbagai penelitian terdahulu menunjukkan bahwa metode Latent Dirichlet Allocation telah diterapkan pada berbagai domain analisis teks. Lee dan Ostwald (2024) mengombinasikan metode PRISMA dengan LDA untuk mengidentifikasi enam topik utama pada penelitian Space Syntax [12]. Geeganage, Xu, dan Li (2024) mengembangkan semantic-LDA dengan memanfaatkan ontologi sehingga mampu meningkatkan kualitas topik dibandingkan LDA konvensional [13]. Ozyurt, Özköse, dan Ayaz (2024) menggunakan LDA untuk menganalisis 8.584 artikel ilmiah mengenai Industry 4.0 dan berhasil mengidentifikasi dua belas tren penelitian utama [14]. Pada lingkup nasional, Khadijah dan Cahyono (2024) menerapkan LDA untuk menganalisis topik pariwisata Yogyakarta berdasarkan data Twitter dengan mengevaluasi model menggunakan Coherence Score dan Perplexity [15], sedangkan Utami (2022) menggunakan LDA untuk menganalisis ulasan hotel sebagai dasar evaluasi kualitas layanan [16]. Meskipun demikian, sebagian besar penelitian masih berfokus pada media sosial, literatur ilmiah, maupun sektor pariwisata. Penelitian mengenai evaluasi model LDA menggunakan kombinasi Coherence Score dan Perplexity pada ulasan aplikasi digital pemerintah, khususnya Aplikasi SIMKOPDES, masih sangat terbatas. Selain itu, sebagian besar penelitian lebih menitikberatkan pada identifikasi topik tanpa membahas keseimbangan antara kualitas statistik model dan interpretabilitas semantik dalam menentukan konfigurasi jumlah topik yang optimal.

Berdasarkan uraian tersebut, penelitian ini bertujuan mengevaluasi model Latent Dirichlet Allocation (LDA) untuk Topic Modelling pada ulasan pengguna Aplikasi SIMKOPDES menggunakan Coherence Score dan Perplexity sebagai metrik evaluasi. Penelitian dilakukan dengan membangun beberapa model LDA menggunakan variasi jumlah topik, kemudian menentukan model terbaik berdasarkan hasil evaluasi kedua metrik tersebut. Model terbaik selanjutnya digunakan untuk mengidentifikasi topik-topik dominan yang merepresentasikan kebutuhan, pengalaman, kendala, serta harapan pengguna terhadap aplikasi Aplikasi SIMKOPDES. Kebaruan (novelty) penelitian ini terletak pada penerapan evaluasi ganda menggunakan Coherence Score dan Perplexity untuk memperoleh konfigurasi model LDA yang optimal pada domain ulasan aplikasi digital koperasi pemerintah, yang hingga saat ini belum banyak dilaporkan pada penelitian sebelumnya. Hasil penelitian diharapkan tidak hanya memberikan kontribusi terhadap pengembangan metode Topic Modelling, tetapi juga menyediakan informasi berbasis data yang dapat dimanfaatkan oleh Kementerian Koperasi Republik Indonesia sebagai bahan evaluasi dan penyempurnaan layanan Aplikasi SIMKOPDES sehingga mampu mendukung keberhasilan transformasi digital Koperasi Desa/Kelurahan Merah Putih secara berkelanjutan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode text mining untuk mengidentifikasi topik-topik dominan pada ulasan pengguna aplikasi SIMKOPDES. Proses penelitian terdiri atas lima tahapan utama, yaitu pengumpulan data ulasan, text preprocessing, pemodelan topik menggunakan Latent Dirichlet Allocation (LDA), visualisasi topik, serta evaluasi model menggunakan Coherence Score dan Perplexity. Tahapan tersebut disusun secara berurutan untuk memastikan data yang digunakan telah melalui proses pembersihan dan transformasi sehingga siap diproses oleh model LDA. Selanjutnya, hasil pemodelan divisualisasikan menggunakan Intertopic Distance Map dan Word Cloud untuk mempermudah interpretasi hubungan antartopik serta mengidentifikasi kata-kata yang paling representatif pada setiap topik. Kualitas model kemudian dievaluasi menggunakan Coherence Score dan Perplexity untuk menentukan jumlah topik yang paling optimal berdasarkan keseimbangan antara kualitas statistik dan interpretabilitas semantik. Melalui rangkaian tahapan tersebut, penelitian ini tidak hanya bertujuan mengidentifikasi isu-isu dominan yang dibahas oleh pengguna SIMKOPDES, tetapi juga mengevaluasi performa model LDA dalam menghasilkan topik yang representatif sebagai dasar pengambilan keputusan untuk pengembangan layanan aplikasi. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1.1 Scrapping Data Ulasan Aplikasi SIMKOPDES

Tahap pertama penelitian adalah pengumpulan data ulasan pengguna Aplikasi SIMKOPDES yang diperoleh dari Google Play Store menggunakan teknik web scraping. Proses pengambilan data dilakukan menggunakan pustaka Google Play Scraper pada bahasa pemrograman Python [17]. Pengambilan data dilakukan pada bulan Juli 2026 dengan mengambil seluruh ulasan yang tersedia pada saat proses scraping dilakukan. Informasi yang dikumpulkan meliputi nama pengguna, tanggal ulasan, rating, isi ulasan (review), versi aplikasi, serta metadata lain yang disediakan oleh Google Play Store. Selanjutnya dilakukan proses seleksi data berdasarkan kriteria inklusi dan eksklusi. Kriteria inklusi meliputi ulasan yang ditulis dalam bahasa Indonesia dan memiliki isi teks sehingga dapat dianalisis menggunakan metode Topic Modelling. Kriteria eksklusi meliputi ulasan kosong, ulasan duplikat, ulasan yang hanya berisi simbol atau karakter tanpa makna, serta data yang tidak relevan terhadap penggunaan aplikasi. Setelah proses pembersihan data, diperoleh dataset akhir sebanyak 694 ulasan yang digunakan sebagai masukan pada proses text preprocessing.

2.1.2 Text Preprocessing

Data ulasan yang diperoleh masih berupa teks tidak terstruktur sehingga memerlukan tahapan preprocessing untuk meningkatkan kualitas data sebelum dilakukan pemodelan topik [18]. Tahapan preprocessing yang digunakan pada penelitian ini terdiri dari:

- Case Folding**
Seluruh karakter diubah menjadi huruf kecil (lowercase) dan karakter selain huruf seperti angka, tanda baca, URL, emoji, serta simbol dihapus agar diperoleh format teks yang seragam [19].
- Tokenizing**
Kalimat ulasan dipecah menjadi kumpulan kata (token) sehingga setiap kata dapat diproses secara individual pada tahapan berikutnya [20].
- Normalization**
Kata-kata tidak baku, singkatan, bahasa gaul, maupun kesalahan penulisan diubah menjadi bentuk kata baku berdasarkan kamus normalisasi Bahasa Indonesia [21].
- Stopword Removal**
Kata-kata umum yang tidak memiliki makna penting dalam pembentukan topik, seperti dan, yang, di, ke, dan dari, dihilangkan menggunakan daftar stopwords Bahasa Indonesia [22].
- Stemming**
Seluruh kata dikembalikan ke bentuk kata dasar menggunakan algoritma stemming Bahasa Indonesia sehingga kata yang memiliki imbuhan dapat direpresentasikan dalam bentuk yang sama. Hasil akhir dari tahapan preprocessing



berupa kumpulan dokumen yang telah bersih (clean corpus) dan siap digunakan sebagai masukan pada proses Topic Modelling [23].

2.1.3 Topic Modelling Menggunakan Latent Dirichlet Allocation (LDA)

Tahap selanjutnya adalah membangun model Latent Dirichlet Allocation (LDA). Sebelum proses pemodelan dilakukan, corpus hasil preprocessing dikonversi ke dalam representasi Bag of Words (BoW) menggunakan pustaka Gensim. Selanjutnya dibentuk beberapa model LDA dengan variasi jumlah topik (number of topics) untuk memperoleh konfigurasi model yang paling optimal [24]. LDA menghitung distribusi probabilitas antara dokumen terhadap topik (document-topic distribution) dan distribusi probabilitas kata terhadap topik (topic-word distribution) [25]. Hasil proses ini berupa sejumlah topik yang masing-masing direpresentasikan oleh kumpulan kata dengan probabilitas tertinggi sehingga dapat digunakan untuk mengidentifikasi isu-isu utama yang dibahas oleh pengguna Aplikasi SIMKOPDES.

2.1.4 Visualisasi Topik

Topik yang dihasilkan oleh model LDA selanjutnya divisualisasikan menggunakan dua pendekatan, yaitu Intertopic Distance Map dan Word Cloud. Intertopic Distance Map digunakan untuk menggambarkan hubungan antar topik berdasarkan jarak semantik [26]. Visualisasi ini menunjukkan tingkat kemiripan maupun perbedaan antar topik sehingga memudahkan interpretasi kualitas pemisahan topik yang dihasilkan model LDA. Setiap topik divisualisasikan menggunakan Word Cloud untuk menampilkan kata-kata yang memiliki probabilitas tertinggi pada masing-masing topik. Semakin besar ukuran kata pada Word Cloud menunjukkan semakin tinggi kontribusi kata tersebut dalam merepresentasikan suatu topik. Visualisasi ini membantu proses interpretasi dan pemberian label pada setiap topik yang terbentuk [27].

2.1.5 Evaluasi Model

Tahap terakhir adalah mengevaluasi kualitas model LDA menggunakan dua metrik evaluasi, yaitu Coherence Score dan Perplexity. Coherence Score digunakan untuk mengukur tingkat keterkaitan semantik antar kata pada setiap topik. Semakin tinggi nilai Coherence Score, maka semakin baik kualitas topik yang dihasilkan karena lebih mudah dipahami dan diinterpretasikan oleh manusia. Perplexity digunakan untuk mengukur kemampuan model dalam merepresentasikan kumpulan dokumen. Nilai Perplexity yang semakin rendah menunjukkan bahwa model memiliki kemampuan prediksi yang lebih baik terhadap data. Model terbaik pada penelitian ini ditentukan berdasarkan kombinasi kedua metrik tersebut, yaitu model yang memiliki nilai Coherence Score tertinggi dan nilai Perplexity terendah. Selanjutnya model terbaik digunakan untuk mengidentifikasi topik-topik dominan pada ulasan pengguna Aplikasi SIMKOPDES sebagai dasar dalam memberikan rekomendasi pengembangan aplikasi.

3. HASIL DAN PEMBAHASAN

Hasil penerapan metode Latent Dirichlet Allocation (LDA) pada data ulasan pengguna aplikasi Aplikasi SIMKOPDES yang diperoleh dari Google Play Store. Analisis diawali dengan pembentukan model Topic Modelling menggunakan beberapa variasi jumlah topik untuk memperoleh konfigurasi model yang optimal. Selanjutnya dilakukan evaluasi menggunakan Coherence Score dan Perplexity guna menentukan kualitas model yang dihasilkan. Model terbaik kemudian divisualisasikan menggunakan Intertopic Distance Map dan Word Cloud untuk memudahkan interpretasi topik. Hasil analisis selanjutnya dibahas untuk mengidentifikasi isu-isu dominan yang disampaikan pengguna sebagai bahan evaluasi dan rekomendasi pengembangan aplikasi Aplikasi SIMKOPDES.

3.1 Pemodelan LDA

Membangun model Latent Dirichlet Allocation (LDA) menggunakan corpus hasil text preprocessing. Sebelum proses pemodelan dilakukan, dokumen terlebih dahulu dikonversi ke dalam representasi Bag of Words (BoW) menggunakan dictionary yang dibentuk dari seluruh kata unik pada corpus. Selanjutnya dilakukan pembentukan beberapa model LDA dengan variasi jumlah topik (number of topics) untuk mengetahui konfigurasi model yang mampu merepresentasikan kumpulan ulasan pengguna secara optimal. Setiap model menghasilkan distribusi probabilitas antara dokumen terhadap topik (document-topic distribution) serta distribusi probabilitas kata terhadap topik (topic-word distribution). Kata-kata dengan probabilitas tertinggi pada setiap topik digunakan sebagai representasi utama untuk mengidentifikasi tema atau isu yang paling sering dibahas oleh pengguna Aplikasi SIMKOPDES. Proses pemodelan ini menjadi dasar dalam tahap evaluasi menggunakan Coherence Score dan Perplexity yang akan menentukan model terbaik sebelum dilakukan visualisasi dan interpretasi topik.

3.1.1. Pemodelan LDA Topik 1

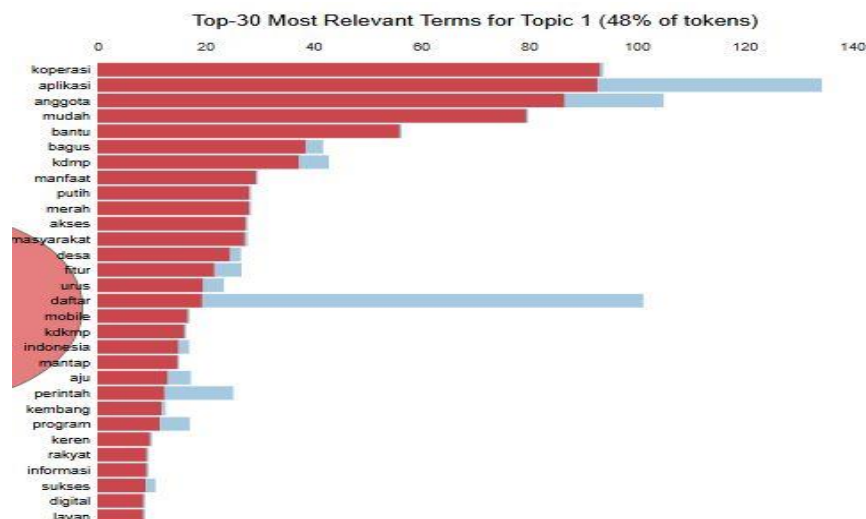
Topik 1 merupakan topik yang memiliki proporsi terbesar dibandingkan topik lainnya, yaitu sebesar 48% dari keseluruhan token hasil pemodelan LDA. Besarnya proporsi tersebut menunjukkan bahwa isu-isu yang terkandung pada topik ini merupakan pembahasan yang paling dominan dalam ulasan pengguna Aplikasi SIMKOPDES. Berdasarkan hasil pemodelan, topik ini berkaitan dengan proses pendaftaran anggota koperasi, kemudahan penggunaan aplikasi, proses verifikasi akun, serta manfaat aplikasi bagi anggota Koperasi Desa/Kelurahan Merah Putih. Dominasi topik ini

mengindikasikan bahwa sebagian besar pengguna memanfaatkan aplikasi sebagai media untuk melakukan registrasi keanggotaan sekaligus memperoleh akses terhadap berbagai layanan digital yang disediakan oleh koperasi.

Tabel 1. LDA Topik 1

LDA Topik 1
(0, '0.054*"koperasi" + 0.054*"aplikasi" + 0.050*"anggota" + 0.046*"mudah" + 0.032*"bantu" + 0.022*"bagus" + 0.022*"kdmp" + 0.017*"manfaat" + 0.016*"putih" + 0.016*"merah"),

Berdasarkan Tabel 1, kata-kata penyusun Topik 1 memiliki distribusi probabilitas yang relatif terkonsentrasi pada beberapa kata utama, yaitu "koperasi" (0,054), "aplikasi" (0,054), "anggota" (0,050), dan "mudah" (0,046). Kedekatan nilai probabilitas pada empat kata tersebut menunjukkan bahwa topik ini dibangun oleh sekumpulan kata yang saling berkaitan dan memiliki kontribusi yang hampir seimbang dalam merepresentasikan tema utama. Penurunan probabilitas secara bertahap pada kata "bantu" (0,032), "bagus" (0,022), "kdmp" (0,022), "manfaat" (0,017), "merah" (0,016), dan "putih" (0,016) menunjukkan bahwa kata-kata tersebut berfungsi sebagai kata pendukung yang memperkuat konteks topik tanpa mendominasi distribusi kata. Pola distribusi probabilitas yang tidak didominasi oleh satu kata tertentu mengindikasikan bahwa Topik 1 memiliki representasi semantik yang cukup stabil. Kombinasi kata-kata tersebut menunjukkan bahwa pembahasan pengguna berfokus pada pengalaman menggunakan SIMKOPDES dalam mendukung keanggotaan koperasi serta manfaat aplikasi, yang tercermin dari kemunculan kata-kata bernuansa positif seperti mudah, bantu, bagus, dan manfaat.



Gambar 2. Diagram Topik 1

Visualisasi menggunakan Intertopic Distance Map yang ditampilkan melalui Gambar 2. Diagram Topik 1 memberikan informasi mengenai distribusi kata-kata yang paling relevan pada topik ini. Kata koperasi, aplikasi, anggota, dan mudah terlihat memiliki frekuensi dan relevansi paling tinggi dibandingkan kata lainnya sehingga menjadi representasi utama Topik 1. Selain itu, terdapat kata-kata lain seperti daftar, verifikasi, akses, bantuan, program, dan digital yang masih memiliki kontribusi terhadap pembentukan topik. Hubungan antar kata tersebut menunjukkan bahwa pembahasan utama pengguna berpusat pada proses pendaftaran anggota, kemudahan verifikasi akun, serta pemanfaatan aplikasi sebagai media digital dalam mendukung layanan Koperasi Desa/Kelurahan Merah Putih. Proporsi token sebesar 48% semakin memperkuat bahwa topik ini merupakan isu yang paling dominan dibicarakan oleh pengguna Aplikasi SIMKOPDES.



Gambar 3. Wordcloud Topik 1

Hasil visualisasi menggunakan Word Cloud pada Gambar 3 semakin memperjelas karakteristik Topik 1. Ukuran huruf yang lebih besar menunjukkan kata dengan tingkat probabilitas kemunculan yang lebih tinggi dalam topik tersebut. Kata koperasi, anggota, dan daftar tampak mendominasi visualisasi, diikuti oleh kata mudah, bantu, verifikasi, aplikasi, manfaat, serta Merah Putih. Dominasi kata-kata tersebut mengonfirmasi bahwa fokus utama ulasan pengguna berkaitan dengan proses registrasi anggota koperasi dan kemudahan penggunaan aplikasi. Selain itu, kemunculan kata-kata bernada positif seperti mudah, bantu, dan manfaat menunjukkan bahwa pengguna merasakan manfaat aplikasi dalam mendukung aktivitas koperasi, meskipun aspek seperti proses verifikasi dan akses layanan masih menjadi bagian penting yang sering dibahas. Dengan demikian, Topik 1 dapat diberi label sebagai "Pendaftaran dan Keanggotaan Koperasi Berbasis Digital", karena sebagian besar kata yang membentuk topik ini menggambarkan pengalaman pengguna dalam melakukan registrasi, verifikasi, dan pemanfaatan layanan Aplikasi SIMKOPDES sebagai platform digital koperasi.

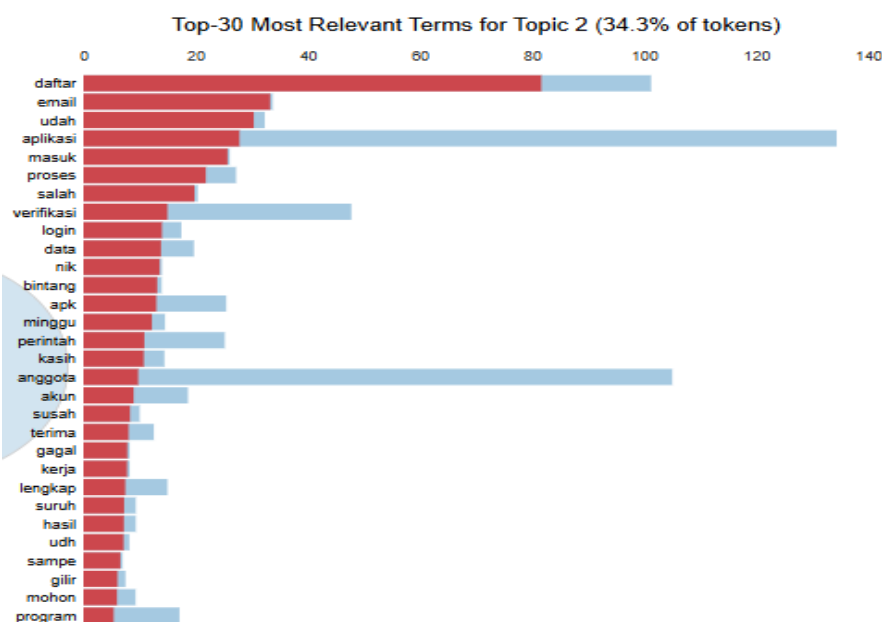
3.1.2. Pemodelan LDA Topik 2

Topik kedua merupakan topik dengan proporsi sebesar 34,3% dari keseluruhan token hasil pemodelan Latent Dirichlet Allocation (LDA). Persentase tersebut menunjukkan bahwa Topik 2 menjadi isu dominan kedua yang banyak dibahas oleh pengguna Aplikasi SIMKOPDES. Berdasarkan kumpulan kata penyusunnya, topik ini berkaitan dengan proses pendaftaran akun, verifikasi identitas, dan kendala yang dialami pengguna selama proses registrasi. Berbeda dengan Topik 1 yang lebih menitikberatkan pada manfaat dan kemudahan penggunaan aplikasi, Topik 2 lebih menggambarkan pengalaman pengguna ketika melakukan proses aktivasi akun, verifikasi data pribadi, serta hambatan teknis yang muncul selama proses pendaftaran.

Tabel 2. LDA Topik 2

LDA Topik 2
(1,0.051*"verifikasi" + 0.042*"wajah" + 0.022*"aplikasi" + 0.015*"akun" + 0.014*"anggota" + 0.009*"foto" + 0.009*"hapus" + 0.009*"tunggu" + 0.008*"apk" + 0.008*"fitur")

Berdasarkan Tabel 2, distribusi probabilitas Topik 2 menunjukkan dominasi yang cukup kuat pada kata "verifikasi" (0,051) dan "wajah" (0,042) dibandingkan kata-kata lainnya. Perbedaan probabilitas yang cukup besar antara kedua kata tersebut dengan kata berikutnya, seperti "aplikasi" (0,022), "akun" (0,015), dan "anggota" (0,014), menunjukkan bahwa proses verifikasi identitas merupakan tema utama yang membentuk topik ini. Sementara itu, kata "foto", "hapus", dan "tunggu" yang masing-masing memiliki probabilitas 0,009, serta "angk" dan "fitur" sebesar 0,008, berperan sebagai kata pendukung yang menjelaskan konteks permasalahan selama proses registrasi dan verifikasi akun. Distribusi probabilitas tersebut memperlihatkan bahwa sebagian besar kata dengan probabilitas tinggi berkaitan dengan satu proses bisnis yang sama, yaitu autentikasi pengguna, sehingga Topik 2 memiliki fokus pembahasan yang relatif spesifik dan tidak bercampur dengan isu lain.



Gambar 4. Diagram Topik 2

Visualisasi pada Gambar 4 memperlihatkan distribusi kata-kata yang paling relevan dalam Topik 2 berdasarkan hasil pemodelan LDA. Kata proses, daftar, email, aplikasi, verifikasi, dan akun tampak memiliki tingkat relevansi yang tinggi sehingga menjadi representasi utama topik ini. Selain itu, terdapat beberapa kata lain seperti login, data, NIK,

lengkap, gagal, minggu, dan program yang turut memperkuat karakteristik Topik 2. Hubungan antar kata tersebut menunjukkan bahwa pembahasan pengguna tidak hanya berkaitan dengan proses registrasi akun, tetapi juga mencakup berbagai kendala teknis, seperti kegagalan verifikasi, kesalahan data, proses login, serta lamanya waktu menunggu proses aktivasi akun. Kondisi tersebut menunjukkan bahwa aspek registrasi dan verifikasi masih menjadi perhatian utama pengguna Aplikasi SIMKOPDES.



Gambar 5. Wordcloud Topik 2

Hasil visualisasi menggunakan Word Cloud pada Gambar 5 semakin memperjelas karakteristik Topik 2. Kata daftar dan proses tampil dengan ukuran paling besar, diikuti oleh kata aplikasi, email, salah, verifikasi, data, masuk, dan APK. Dominasi kata-kata tersebut menunjukkan bahwa pengguna banyak membahas proses pendaftaran akun beserta kendala yang terjadi selama proses tersebut. Kemunculan kata email, salah, dan masuk mengindikasikan bahwa sebagian pengguna mengalami masalah pada tahap validasi data, autentikasi akun, maupun akses ke dalam aplikasi. Sementara itu, kata verifikasi tetap menjadi kata kunci utama yang menegaskan bahwa proses verifikasi identitas merupakan tahapan yang paling sering mendapat perhatian dalam ulasan pengguna. Berdasarkan keseluruhan hasil analisis LDA, Diagram Topik, dan Word Cloud, Topik 2 dapat diberi label "Proses Registrasi dan Verifikasi Akun Pengguna", karena sebagian besar kata yang membentuk topik ini menggambarkan pengalaman pengguna dalam melakukan pendaftaran, validasi identitas, serta kendala teknis selama proses aktivasi akun pada Aplikasi SIMKOPDES.

3.1.3 Pemodelan LDA Topik 3

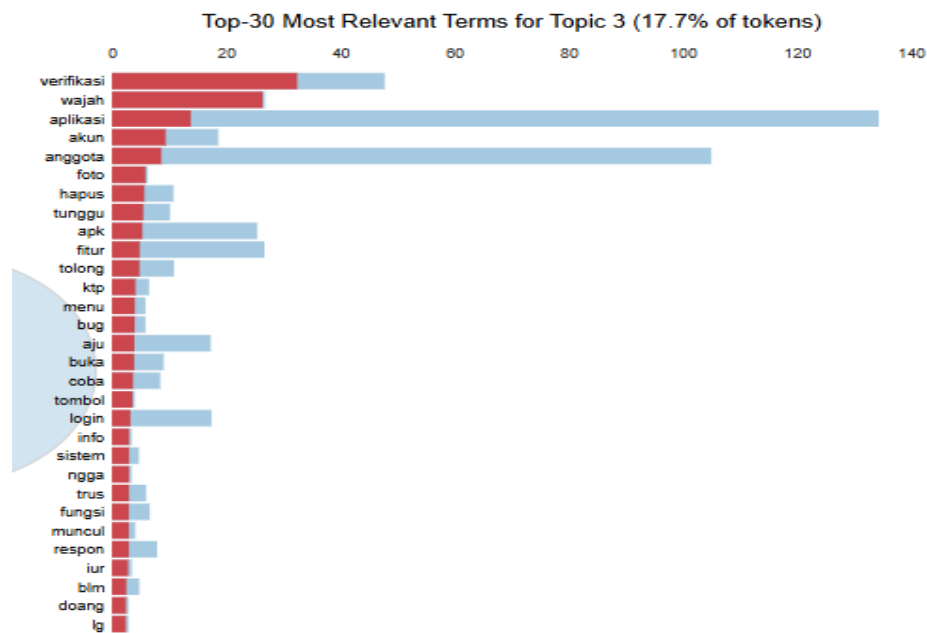
Topik ketiga memiliki proporsi sebesar 17,7% dari keseluruhan token hasil pemodelan Latent Dirichlet Allocation (LDA). Meskipun memiliki proporsi yang lebih kecil dibandingkan Topik 1 dan Topik 2, topik ini tetap memberikan informasi penting mengenai pengalaman pengguna dalam melakukan proses registrasi dan penggunaan awal aplikasi Aplikasi SIMKOPDES. Berdasarkan kumpulan kata penyusunnya, Topik 3 lebih banyak menggambarkan proses pendaftaran akun, validasi data pengguna, serta kendala yang dialami ketika mengakses aplikasi. Hal tersebut menunjukkan bahwa selain membahas manfaat aplikasi, pengguna juga banyak memberikan perhatian terhadap kemudahan proses registrasi, autentikasi akun, serta validasi data yang menjadi tahapan awal penggunaan Aplikasi SIMKOPDES.

Tabel 3. LDA Topik 3

LDA Topik 3
$(2, '0.066*'daftar'' + 0.027*'email'' + 0.025*'udah'' + 0.023*'aplikasi'' + 0.021*'masuk'' + 0.018*'proses'' + 0.016*'salah'' + 0.012*'verifikasi'' + 0.011*'login'' + 0.011*'data'')$

Berdasarkan Tabel 3, distribusi probabilitas pada Topik 3 menunjukkan bahwa kata "daftar" memiliki probabilitas tertinggi sebesar 0,066, lebih tinggi dibandingkan kata-kata lainnya seperti "email" (0,027), "udah" (0,025), "aplikasi" (0,023), "masuk" (0,021), dan "proses" (0,018). Dominasi kata "daftar" menunjukkan bahwa proses pendaftaran akun menjadi tema sentral pada topik ini. Meskipun demikian, keberadaan kata "email", "masuk", "verifikasi", "login", dan "data" dengan probabilitas yang relatif berdekatan menunjukkan bahwa topik ini tidak hanya membahas tahap pendaftaran, tetapi juga mencakup keseluruhan alur penggunaan akun, mulai dari registrasi, autentikasi, hingga proses login. Distribusi probabilitas yang tersebar pada beberapa kata yang saling berhubungan menunjukkan bahwa Topik 3 merepresentasikan satu rangkaian aktivitas pengguna yang utuh. Kehadiran kata "salah" mengindikasikan bahwa sebagian ulasan membahas kendala validasi data atau autentikasi, sehingga topik ini menggambarkan pengalaman pengguna selama proses registrasi dan penggunaan akun secara lebih komprehensif.

Visualisasi pada Gambar 6 memperlihatkan distribusi kata-kata yang paling relevan dalam Topik 3. Kata aplikasi, anggota, verifikasi, wajah, dan akun memiliki tingkat relevansi yang tinggi sehingga menjadi representasi utama topik ini. Selain itu, terdapat kata-kata lain seperti tolong, fungsi, sistem, muncul, login, buku, coba, serta aju yang turut memperkuat karakteristik Topik 3. Hubungan antar kata tersebut menunjukkan bahwa pembahasan pengguna tidak hanya berkaitan dengan proses registrasi, tetapi juga menyangkut fungsi aplikasi setelah proses pendaftaran selesai, seperti proses login, munculnya kendala sistem, serta harapan pengguna agar fitur-fitur aplikasi dapat berjalan dengan baik. Hasil visualisasi ini menunjukkan bahwa aspek registrasi dan akses awal terhadap aplikasi masih menjadi perhatian penting dalam pengalaman pengguna Aplikasi SIMKOPDES.



Gambar 6. Diagram Topik 3



Gambar 7. Wordcloud Topik 3

Hasil visualisasi menggunakan Word Cloud pada Gambar 7 semakin memperjelas karakteristik Topik 3. Kata aplikasi dan mudah tampak mendominasi visualisasi dengan ukuran huruf yang paling besar, diikuti oleh kata bagus, bantu, anggota, mantap, verifikasi, koperasi, dan APK. Dominasi kata mudah, bagus, bantu, dan mantap menunjukkan bahwa sebagian besar pengguna memberikan tanggapan positif terhadap kemudahan penggunaan aplikasi setelah berhasil menyelesaikan proses registrasi dan verifikasi akun. Sementara itu, kemunculan kata verifikasi dan anggota menunjukkan bahwa proses validasi identitas tetap menjadi bagian penting yang sering dibahas dalam ulasan pengguna. Berdasarkan keseluruhan hasil analisis LDA, Diagram Topik, dan Word Cloud, Topik 3 dapat diberi label sebagai "Proses Registrasi Akun dan Pengalaman Penggunaan Aplikasi", karena sebagian besar kata yang membentuk topik ini menggambarkan pengalaman pengguna mulai dari proses pendaftaran, login, verifikasi data hingga kemudahan penggunaan Aplikasi SIMKOPDES sebagai platform digital layanan koperasi.

3.2 Evaluasi

Evaluasi model merupakan tahapan penting dalam proses Topic Modelling untuk menentukan kualitas model Latent Dirichlet Allocation (LDA) yang dibangun. Pada penelitian ini, evaluasi dilakukan menggunakan dua metrik, yaitu Coherence Score dan Perplexity. Coherence Score digunakan untuk mengukur tingkat keterkaitan semantik antar kata dalam setiap topik sehingga menunjukkan kemudahan interpretasi hasil Topic Modelling, sedangkan Perplexity digunakan untuk mengukur kemampuan model dalam merepresentasikan kumpulan dokumen berdasarkan distribusi probabilitas kata. Kombinasi kedua metrik tersebut memberikan dasar yang lebih komprehensif dalam menentukan jumlah topik (*number of topics*) yang paling optimal. Pada subbab ini terlebih dahulu dibahas hasil evaluasi menggunakan Coherence Score, kemudian dilanjutkan dengan evaluasi menggunakan Perplexity.

3.2.1 Coherence Score

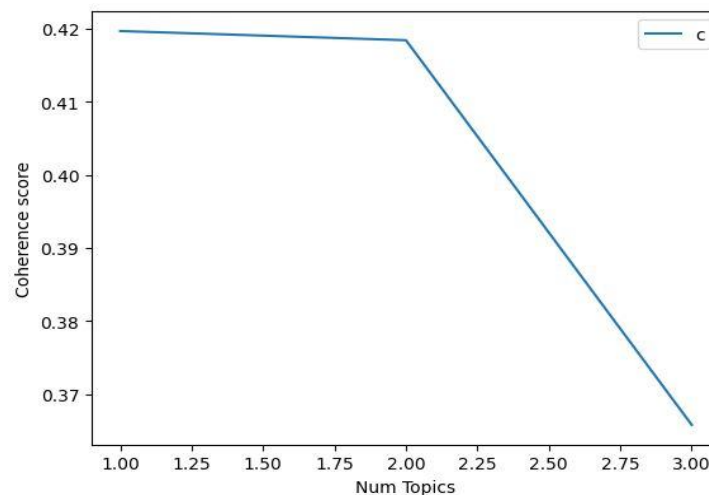
Berdasarkan hasil evaluasi menggunakan Coherence Score, model dengan 1 topik menghasilkan nilai Coherence Score tertinggi sehingga secara statistik memiliki tingkat koherensi terbaik dibandingkan model lainnya. Akan tetapi, dalam konteks penelitian ini, tujuan utama Topic Modelling tidak hanya memperoleh nilai evaluasi yang tinggi, tetapi juga menghasilkan beberapa kelompok topik yang mampu merepresentasikan beragam isu yang terdapat pada ulasan pengguna Aplikasi SIMKOPDES. Meskipun model dengan 2 topik memiliki nilai Coherence Score sedikit lebih rendah dibandingkan model dengan satu topik, perbedaannya sangat kecil sehingga model tersebut masih memiliki kualitas topik

yang baik. Selain itu, model dengan 3 topik dipilih pada penelitian ini karena mampu menghasilkan pemisahan isu yang lebih spesifik dan memberikan informasi yang lebih kaya mengenai berbagai aspek yang dibahas oleh pengguna, seperti keanggotaan koperasi, proses registrasi dan verifikasi akun, serta pengalaman penggunaan aplikasi. Dengan demikian, pemilihan model terbaik tidak hanya didasarkan pada nilai Coherence Score semata, tetapi juga mempertimbangkan kemudahan interpretasi dan kebermanfaatan topik yang dihasilkan dalam menjelaskan karakteristik ulasan pengguna Aplikasi SIMKOPDES.

Tabel 4. Coherence Score

Jumlah Topik	Coherence Score
1	0.419677
2	0.418423
3	0.36582

Berdasarkan Tabel 4, nilai Coherence Score tertinggi diperoleh pada model dengan 1 topik, yaitu sebesar 0,419677. Selanjutnya, model dengan 2 topik memperoleh nilai Coherence Score sebesar 0,418423, sedangkan model dengan 3 topik memperoleh nilai terendah, yaitu 0,365820. Selisih nilai antara model dengan satu dan dua topik relatif kecil, yaitu sekitar 0,001254, yang menunjukkan bahwa kedua model memiliki kualitas topik yang hampir sama. Namun demikian, ketika jumlah topik ditingkatkan menjadi tiga, nilai Coherence Score mengalami penurunan yang cukup signifikan. Penurunan tersebut menunjukkan bahwa penambahan jumlah topik menyebabkan keterkaitan semantik antar kata pada setiap topik menjadi lebih rendah sehingga kualitas interpretasi topik juga menurun.



Gambar 8. Grafik Coherence Score

Hasil evaluasi tersebut diperkuat oleh Gambar 8 yang menunjukkan tren perubahan nilai Coherence Score terhadap variasi jumlah topik. Grafik memperlihatkan bahwa nilai Coherence Score cenderung stabil pada model dengan satu dan dua topik, kemudian mengalami penurunan yang cukup tajam pada model dengan tiga topik. Pola tersebut menunjukkan bahwa peningkatan jumlah topik tidak selalu menghasilkan kualitas Topic Modelling yang lebih baik. Sebaliknya, semakin banyak topik yang dibentuk dapat menyebabkan distribusi kata menjadi lebih tersebar sehingga hubungan semantik antar kata dalam setiap topik menjadi semakin lemah. Oleh karena itu, pemilihan jumlah topik perlu mempertimbangkan keseimbangan antara jumlah topik yang dihasilkan dan tingkat keterkaitan kata pada masing-masing topik.

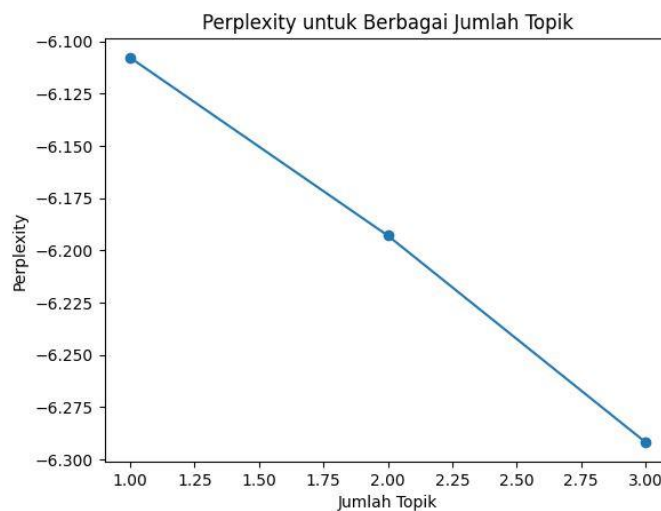
3.2.2 Perplexity

Berdasarkan hasil evaluasi menggunakan Perplexity, model dengan 3 topik menghasilkan nilai Perplexity terendah sehingga memiliki kemampuan representasi data yang paling baik dibandingkan model lainnya. Apabila dibandingkan dengan hasil evaluasi menggunakan Coherence Score, terlihat adanya perbedaan kecenderungan antara kedua metrik. Coherence Score menunjukkan nilai tertinggi pada model dengan satu topik, sedangkan Perplexity menunjukkan performa terbaik pada model dengan tiga topik. Perbedaan tersebut merupakan kondisi yang umum dijumpai pada penelitian Topic Modelling, karena kedua metrik mengevaluasi aspek yang berbeda. Coherence Score berfokus pada keterkaitan semantik antar kata sehingga menghasilkan topik yang mudah diinterpretasikan, sedangkan Perplexity berfokus pada kemampuan statistik model dalam merepresentasikan dokumen. Oleh karena itu, penentuan model terbaik pada penelitian ini tidak hanya mempertimbangkan satu metrik, tetapi menggunakan kombinasi Coherence Score, Perplexity, serta kemampuan model dalam menghasilkan topik yang mudah diinterpretasikan. Dengan mempertimbangkan ketiga aspek tersebut, model dengan 3 topik dipilih sebagai model terbaik karena mampu menghasilkan pemisahan isu yang lebih spesifik dan representatif terhadap berbagai permasalahan yang dibahas oleh pengguna Aplikasi SIMKOPDES.

Tabel 5. Perplexity

Jumlah Topik	Perplexity
1	-6.1079
2	-6.1930
3	-6.2917

Berdasarkan Tabel 5, nilai Perplexity mengalami penurunan seiring bertambahnya jumlah topik yang digunakan pada model LDA. Model dengan 1 topik menghasilkan nilai Perplexity sebesar -6,1079, kemudian menurun menjadi -6,1930 pada model dengan 2 topik, dan mencapai nilai terendah sebesar -6,2917 pada model dengan 3 topik. Hasil tersebut menunjukkan bahwa peningkatan jumlah topik mampu meningkatkan kemampuan model dalam merepresentasikan distribusi kata pada corpus ulasan Aplikasi SIMKOPDES. Dengan kata lain, model dengan tiga topik memiliki kemampuan prediksi yang lebih baik dibandingkan model dengan satu maupun dua topik berdasarkan metrik Perplexity.

**Gambar 9.** Grafik Perplexity

Hasil evaluasi tersebut diperlihatkan secara visual pada Gambar 9, yang menunjukkan tren penurunan nilai Perplexity terhadap variasi jumlah topik. Grafik memperlihatkan pola yang konsisten, yaitu semakin banyak jumlah topik yang dibentuk, maka nilai Perplexity semakin rendah. Penurunan tersebut menunjukkan bahwa model LDA semakin mampu menangkap distribusi probabilitas kata pada dokumen ketika jumlah topik ditambah. Tidak terlihat adanya kenaikan nilai Perplexity pada rentang pengujian satu hingga tiga topik, sehingga model dengan tiga topik memberikan performa terbaik berdasarkan metrik ini. Hasil tersebut sejalan dengan karakteristik Perplexity, di mana model yang memiliki nilai lebih rendah dianggap lebih baik dalam memodelkan kumpulan dokumen.

3.2.3 Pemilihan Model

Meskipun model satu topik menghasilkan nilai Coherence Score tertinggi, model tersebut hanya menghasilkan satu kelompok topik yang bersifat umum sehingga kurang mampu memisahkan isu-isu yang berbeda pada ulasan pengguna. Sebaliknya, model tiga topik menghasilkan nilai Perplexity terendah serta memperlihatkan pemisahan topik yang lebih jelas pada Intertopic Distance Map. Setiap topik merepresentasikan aspek layanan yang berbeda sehingga memberikan interpretasi yang lebih bermakna dibandingkan model satu topik. Oleh karena itu, pemilihan model dilakukan berdasarkan kombinasi evaluasi statistik dan interpretabilitas semantik, bukan hanya berdasarkan satu metrik evaluasi.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa metode Latent Dirichlet Allocation (LDA) mampu mengidentifikasi tiga topik dominan pada ulasan pengguna Aplikasi SIMKOPDES, yaitu keanggotaan koperasi dan manfaat aplikasi, proses registrasi serta verifikasi akun, dan pengalaman penggunaan aplikasi. Temuan ini sejalan dengan penelitian Utami (2022) yang menunjukkan bahwa LDA efektif merangkul ribuan ulasan hotel menjadi beberapa topik utama sebagai dasar evaluasi kualitas layanan [16]. Kesamaan juga ditunjukkan pada penelitian Khadijah dan Cahyono (2024) yang menerapkan LDA pada data Twitter dengan mengevaluasi model menggunakan Coherence Score dan Perplexity, di mana kedua metrik tersebut mampu memberikan evaluasi yang saling melengkapi dalam menentukan kualitas model [15]. Pada lingkup yang lebih luas, penelitian Lee dan Ostwald (2024) membuktikan bahwa LDA mampu mengidentifikasi struktur topik pada kajian Space Syntax [12], sedangkan Ozyurt et al. (2024) berhasil memanfaatkan LDA untuk mengidentifikasi tren penelitian Industry 4.0 berdasarkan ribuan artikel ilmiah [14]. Sementara itu, Geeganage et al. (2024) menunjukkan bahwa pengembangan semantic-LDA mampu meningkatkan kualitas topik melalui pemanfaatan hubungan semantik antar kata [13]. Meskipun memiliki kesamaan dalam penggunaan metode LDA, penelitian ini berbeda dari penelitian-penelitian sebelumnya karena berfokus pada ulasan aplikasi digital pemerintah yang mendukung transformasi koperasi di Indonesia.



Selain itu, penelitian ini tidak hanya mengidentifikasi topik dominan, tetapi juga mengevaluasi konfigurasi model menggunakan kombinasi Coherence Score dan Perplexity untuk menentukan jumlah topik yang paling representatif. Hasil evaluasi menunjukkan bahwa model dengan satu topik menghasilkan Coherence Score tertinggi, sedangkan model tiga topik memperoleh Perplexity terendah. Oleh karena itu, pemilihan model terbaik tidak hanya didasarkan pada satu metrik statistik, tetapi juga mempertimbangkan interpretabilitas topik yang didukung oleh visualisasi Intertopic Distance Map dan Word Cloud, sehingga menghasilkan pemisahan isu yang lebih jelas dan relevan dengan konteks penggunaan Aplikasi SIMKOPDES. Temuan ini memperkuat bahwa evaluasi model Topic Modelling sebaiknya dilakukan menggunakan pendekatan multi-kriteria, yaitu mengombinasikan kualitas statistik model dengan interpretabilitas semantik, sehingga hasil yang diperoleh tidak hanya optimal secara komputasional tetapi juga memberikan informasi yang lebih bermakna bagi pengembang aplikasi dalam menentukan prioritas pengembangan layanan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan Latent Dirichlet Allocation (LDA) mampu mengidentifikasi tiga topik dominan pada ulasan pengguna SIMKOPDES Mobile, yaitu keanggotaan koperasi dan manfaat aplikasi, proses registrasi serta verifikasi akun, serta pengalaman penggunaan aplikasi. Evaluasi model menunjukkan bahwa model satu topik memperoleh Coherence Score tertinggi sebesar 0,419677, sedangkan model tiga topik menghasilkan nilai Perplexity terendah sebesar -6,2917. Perbedaan hasil kedua metrik tersebut menunjukkan bahwa pemilihan model topic modelling tidak cukup dilakukan berdasarkan satu indikator statistik, tetapi perlu mempertimbangkan interpretabilitas semantik agar topik yang dihasilkan lebih representatif. Berdasarkan kombinasi Coherence Score, Perplexity, serta hasil visualisasi menggunakan Intertopic Distance Map dan Word Cloud, model tiga topik dipilih karena mampu memisahkan isu-isu utama pengguna secara lebih jelas dan memberikan interpretasi yang lebih bermakna. Temuan ini menjawab tujuan penelitian dalam menentukan konfigurasi model LDA yang paling sesuai untuk menganalisis ulasan aplikasi digital berbasis teks tidak terstruktur. Secara praktis, hasil penelitian memberikan informasi bagi pengembang SIMKOPDES Mobile mengenai prioritas pengembangan aplikasi, khususnya pada penyempurnaan proses registrasi, verifikasi identitas, serta peningkatan pengalaman pengguna agar kualitas layanan digital koperasi semakin optimal. Penelitian ini juga menunjukkan bahwa kombinasi Coherence Score dan Perplexity dapat digunakan sebagai pendekatan evaluasi yang saling melengkapi dalam pemilihan model LDA pada analisis ulasan aplikasi.

REFERENCES

- [1] Setiawan, "Transformasi Digital Berbasis Data di Kabupaten Jember: Strategi Tata Kelola Pemerintahan Daerah Menuju Era 5.0," *Interelasi Humaniora*, vol. 1, no. 4, pp. 396–402, 2025, DOI : 10.24716/sjc8bp94
- [2] Syarifuddin Arief and Haeruddin Haeruddin, "Strategi Pengembangan Koperasi Desa Merah Putih: Tinjauan Literatur atas Praktik Terbaik dan Tantangan Global", *ab*, vol. 21, no. 1, pp. 284–299, Jun. 2025, DOI : 10.30603/ab.v21i1.6469
- [3] Liputan 6, "Puncak Peringatan Hari Koperasi Tahun 2026," YouTube. [Online]. Available: <https://www.youtube.com/watch?v=HwaCgTVCBUw&t=127s>. [Accessed: Jul. 17, 2026].
- [4] Google Play Store, "SIMKOPDES Mobile." [Online]. Available: https://play.google.com/store/apps/details?id=id.kop.merahputih.kdmp_mobile_app&hl=id. [Accessed: Jul. 17, 2026].
- [5] R. K. Wardana, N. Cahyono, U. A. Saputro, A. H. Negoro, and N. Aini, "Implementasi Algoritma SVM dan Optimasi Menggunakan Komparasi N-Gram dan GloVe pada Sentimen Pengguna Aplikasi Access by KAI," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3216–3223, 2025, DOI : 10.36040/jati.v9i2.13284
- [6] M. Y. Assani, A. Rudiyan, and U. Radiyah, "Analisis Sentimen Ulasan Mobile Banking Syariah Berbasis NLP untuk Evaluasi dan Peningkatan Kualitas Layanan Digital NLP-Based Sentiment Analysis of Islamic Mobile Banking Reviews for Digital Service Quality Enhancement," *Jurnal Bumigora Information Technology (BITE)*, vol. 7, no. 2, 2025, DOI : 10.30812/bite.v7i2.5943
- [7] N. Cahyono, *Fundamental Topic Modelling: Memahami Tren & Topik dalam Text Mining*. Yogyakarta, Indonesia: Mega Press Nusantara, 2024.
- [8] M. U. Yanuar and W. Wibowo, "Pemodelan Topik dan Analisis Sentimen pada Ulasan Pengguna Aplikasi Trans Jatim: Topic Modeling and Sentiment Analysis on Trans Jatim Application User Reviews," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 6, no. 1, pp. 156–166, 2026, DOI : 10.57152/malcom.v6i1.2410
- [9] Latifah Nurrohmah Handayani and Lucia Nugraheni Harnaningrum, "Analisis Perbandingan Performa NMF dengan LDA pada Topik Modeling Berita Online Indonesia", *json*, vol. 7, no. 3, pp. 875–886, Mar. 2026, DOI : 10.30865/json.v7i3.9469
- [10] Y. M. Yolanda, D. Swanjaya, and D. P. Pamungkas, "Sistem Pencarian Dongeng Berbasis Topik Menggunakan Latent Dirichlet Allocation (LDA) pada Aplikasi Nusantara Panji Kediri," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 394–407, 2025, DOI : 10.36341/rabit.v10i2.6106
- [11] N. Sofa, F. S. Utomo, and R. E. Saputro, "Eksplorasi Model Hybrid Transformer-Latent Semantic Analysis (LSA) untuk Pemahaman Konteks Teks Berbahasa Indonesia," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 5, pp. 1239–1252, 2025, DOI : 10.52436/1.jpti.662
- [12] J. H. Lee and M. J. Ostwald, "Latent Dirichlet Allocation (LDA) Topic Models for Space Syntax Studies on Spatial Experience," *City, Territory and Architecture*, vol. 11, no. 3, 2024, doi: 10.1186/s40410-023-00223-3.
- [13] D. K. Geeganage, Y. Xu, and Y. Li, "A Semantics-Enhanced Topic Modelling Technique: Semantic-LDA," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 4, pp. 1–27, 2024, DOI : 10.1145/3639409
- [14] O. Ozyurt, H. Özköse, and A. Ayaz, "Evaluating the Latest Trends of Industry 4.0 Based on LDA Topic Model," *Journal of Supercomputing*, vol. 80, pp. 19003–19030, 2024, doi: 10.1007/s11227-024-06247-x, DOI : 10.1007/s11227-024-06247-x



- [15] U. N. Khadijah and N. Cahyono, "Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan Latent Dirichlet Allocation (LDA)," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, 2024, DOI: 10.33022/ijcs.v13i4.3816
- [16] E. Utami, "Pengamatan Tren Ulasan Hotel Menggunakan Pemodelan Topik Berbasis Latent Dirichlet Allocation," *JIKO (Jurnal Informatika dan Komputer)*, vol. 6, no. 2, pp. 169–175, 2022, DOI: 10.26798/jiko.v6i2.579
- [17] G. R. Ramadhan and C. A. Sugianto, "Analisis Sentimen Ulasan Aplikasi Dana di Google Play Store Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 9849–9857, 2024, DOI: 10.36040/jati.v8i5.10732
- [18] A. I. Kamil, O. N. Pratiwi, and D. Witarsyah, "Analisis Sentimen dan Pemodelan Topik terhadap Aplikasi Pembelajaran Online pada Platform Google Play," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 836–849, 2025, DOI: 10.29100/jipi.v10i2.6023
- [19] M. MAHDI FAZA and MOCH. TAUFIK, "PENERAPAN MODEL INDOBERT UNTUK DETEKSI POTENSI SUMBER STRES DALAM TEKS MEDIA SOSIAL," *JRSIT*, vol. 3, no. 1, pp. 271–283, Aug. 2025. DOI: 10.70248/jrsit.v3i1.3046
- [20] A. O. Praneswara and N. Cahyono, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *The Indonesian Journal of Computer Science*, vol. 12, no. 6, 2023, DOI: 10.33022/ijcs.v12i6.3473
- [21] T. S. Haris, D. Alfath, A. R. Marini, H. R. Muhammad, and S. Z. Naya, "Analisis Modelling Pada Reviewes Lazada Indonesia Menggunakan Latent Dirichlet Allocation (LDA) Untuk Optimalisasi Strategi Bisnis," *JEKIN - Jurnal Teknik Informatika*, vol. 5, no. 1, pp. 361–371, 2025, DOI: 10.58794/jekin.v5i1.1325.
- [22] E. L. Utari and S. H. Wibowo, "Analisis Komparatif Algoritma SVM Naive Bayes dan LSTM pada Sentimen Komentar Lagu Labour," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 3, pp. 1276–1286, 2025, DOI: 10.51401/jinteks.v7i3.6213
- [23] M. I. Raif, C. I. Ratnasari, "Pemodelan Topik pada Komunitas Ekspresi Emosi Negatif di Media Sosial X Menggunakan LDA," *Bulletin of Computer Science Research*, vol. 6, no. 1, pp. 224–234, 2025, DOI : 10.47065/bulletincsr.v6i1.877
- [24] D. Hartati, U. Khaira, and R. R. Bintana, "Penerapan Support Vector Machine dan Latent Dirichlet Allocation dalam Analisis Sentimen Terhadap Pengalaman Pengguna Aplikasi Alfagift," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 8, no. 5, Oct. 2025, DOI: 10.32672/jnkti.v8i5.9866.
- [25] I. A. Musdar, H. Angriani, and S. Muriati, "Identifikasi Daya Tarik Wisata Benteng Rotterdam Berdasarkan Ulasan Wisatawan Menggunakan Topic Modeling," *Journal Software, Hardware and Information Technology*, vol. 6, no. 2, pp. 171–184, 2026. , DOI: 10.24252/shift.v6i2.298
- [26] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, "Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis Fitur Layanan Aplikasi Gojek dengan Support Vector Machine," *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 2, no. 1, pp. 61–70, 2024, DOI: 10.24912/jsstk.v2i1.32058