

Analisis Business Intelligence Ulasan Negatif Aplikasi M-Pajak Menggunakan BERTopic untuk Evaluasi Layanan Digital

Henry Pandia

Fakultas Teknologi Informasi, Sistem Informasi, Universitas Advent Indonesia

Bandung Barat, Jawa Barat, Indonesia

Email: pandiahenry@unai.edu

Abstrak—Aplikasi M-Pajak merupakan bagian dari transformasi layanan perpajakan digital pemerintah yang bertujuan untuk meningkatkan efisiensi dan aksesibilitas layanan perpajakan bagi masyarakat. Namun, berbagai keluhan pengguna di Google Play Store menunjukkan bahwa masih terdapat permasalahan teknis dan operasional dalam penggunaan aplikasi ini. Penelitian ini bertujuan untuk menganalisis ulasan negatif pengguna aplikasi M-Pajak menggunakan BERTopic guna menghasilkan insight Business Intelligence terkait kualitas layanan perpajakan digital pemerintah. BERTopic dipilih karena menggunakan contextual embedding berbasis transformer yang sesuai untuk menganalisis ulasan aplikasi mobile yang pendek, tidak terstruktur, dan mengandung bahasa informal, sehingga dapat membantu mengidentifikasi tema keluhan secara lebih kontekstual dibandingkan pendekatan topic modeling berbasis distribusi kata tradisional. Data penelitian diperoleh dari Google Play Store pada periode Mei 2023 hingga Mei 2026 dengan total 6.024 ulasan pengguna. Sebanyak 5.008 ulasan negatif dengan rating 1 dan 2 digunakan sebagai fokus analisis. Tahapan penelitian meliputi scraping data, preprocessing, pemodelan BERTopic, evaluasi topik, serta interpretasi Business Intelligence. Hasil penelitian menunjukkan bahwa permasalahan utama pengguna berkaitan dengan autentikasi dan login, error sistem, verifikasi email dan OTP, registrasi NPWP dan NIK, serta stabilitas layanan digital. Evaluasi model menghasilkan Topic Coherence 0,522, Topic Diversity 0,667, dan Topic Quality 0,348, yang menunjukkan kualitas topik berada pada tingkat moderat tetapi masih dapat diinterpretasikan. Kontribusi penelitian ini terletak pada pemanfaatan ulasan negatif aplikasi layanan perpajakan digital sebagai sumber insight operasional untuk mendukung evaluasi dan peningkatan kualitas layanan publik digital pemerintah.

Kata Kunci: BERTopic, Business Intelligence, App Store Review Mining, Layanan Perpajakan Digital, Analisis Ulasan Negatif, Topic Modeling, M-Pajak, Google Play Store

Abstract—The M-Pajak application is part of the government's digital tax service transformation aimed at improving the efficiency and accessibility of tax services for the public. However, user complaints on the Google Play Store indicate that technical and operational problems persist with the application. This study aims to analyze negative user reviews of the M-Pajak application using BERTopic to generate Business Intelligence insights related to the quality of government digital tax services. BERTopic was selected because it uses transformer-based contextual embeddings, which are well-suited to analyzing short, unstructured, and informal mobile application reviews, thereby helping identify complaint themes in a more context-aware way than traditional word-distribution-based topic modeling approaches. The data were collected from Google Play Store reviews from May 2023 to May 2026, totaling 6,024 user reviews. A total of 5,008 negative reviews with ratings of 1 or 2 were used as the primary focus of the analysis. The research stages included data scraping, preprocessing, BERTopic modeling, topic evaluation, and Business Intelligence interpretation. The results show that the main user problems are related to authentication and login, system errors, email and OTP verification, NPWP and NIK registration, and digital service stability. Model evaluation yielded a Topic Coherence score of 0.522, a Topic Diversity score of 0.667, and a Topic Quality score of 0.348, indicating that the topic quality was moderate yet interpretable. The contribution of this study lies in utilizing negative reviews of a digital tax service application as a source of operational insights to support the evaluation and improvement of government digital public services.

Keywords: BERTopic, Business Intelligence, App Store Review Mining, Digital Tax Services, Negative Ulasan Analysis, Topic Modeling, M-Pajak, Google Play Store

1. PENDAHULUAN

Perkembangan transformasi digital telah mendorong institusi pemerintah untuk mengimplementasikan layanan publik berbasis teknologi guna meningkatkan efisiensi, aksesibilitas, dan kualitas layanan. Salah satu bentuk implementasi layanan digital tersebut adalah pengembangan aplikasi M-Pajak oleh Direktorat Jenderal Pajak (DJP). Aplikasi ini digunakan sebagai layanan administrasi pajak yang dapat diakses dengan lebih cepat, fleksibel dan transparan. Meskipun memberikan banyak keuntungan, implementasi teknologi digital sering kali harus dievaluasi, apakah memberikan manfaat yang sesuai dengan harapan. Karena itu, pengembang aplikasi perlu mendapatkan umpan balik dari pengguna.

Salah satu cara untuk mendapatkan umpan balik dari pengguna adalah melalui ulasan aplikasi di platform distribusi aplikasi seperti Google Play Store. Ulasan pengguna pada platform tersebut merepresentasikan pengalaman aktual masyarakat dalam menggunakan layanan digital. Data ulasan tersebut dapat digunakan sebagai sumber informasi untuk mengevaluasi kualitas layanan, performa aplikasi, kemudahan penggunaan, serta persepsi pengguna terhadap layanan yang diberikan. Rachmiani et al. [1] menjelaskan bahwa ulasan online berperan penting dalam membentuk persepsi dan kepercayaan pengguna terhadap layanan digital. Selain itu, ulasan pengguna dapat memberikan insight strategis bagi organisasi untuk memahami kebutuhan pengguna dan mengidentifikasi area layanan yang perlu ditingkatkan. Dalam konteks layanan publik digital, pemanfaatan ulasan pengguna menjadi semakin penting karena memungkinkan institusi pemerintah memperoleh evaluasi layanan secara langsung berdasarkan pengalaman masyarakat.

Teknologi yang ada saat ini memungkinkan pendekatan analitis terhadap teks ulasan yang tidak terstruktur dilakukan secara efektif. Pendekatan business intelligence dan Natural Language Processing (NLP) dapat digunakan untuk

mengeksktraksi informasi penting dari data teks dalam jumlah besar, termasuk ulasan pengguna aplikasi mobile. Business Intelligence memungkinkan organisasi memperoleh insight berbasis data untuk mendukung evaluasi layanan dan pengambilan keputusan secara lebih responsif, sedangkan NLP memungkinkan proses analisis teks dilakukan secara otomatis untuk menemukan pola informasi, kecenderungan pembahasan, maupun tema tertentu dari kumpulan data teks yang tidak terstruktur. Pendekatan ini menjadi relevan mengingat data ulasan pengguna yang sangat besar, dalam bentuk teks yang tidak terstruktur, yang mustahil dianalisis menggunakan pendekatan statistik konvensional.

Salah satu pendekatan NLP yang banyak digunakan untuk menganalisis data teks tidak terstruktur adalah *topic modeling*. *Topic modeling* merupakan teknik yang digunakan untuk mengidentifikasi tema atau pola pembahasan dominan dari kumpulan dokumen teks secara otomatis. Pendekatan ini memungkinkan peneliti menemukan struktur informasi tersembunyi dalam data teks tanpa memerlukan proses pelabelan manual. Dalam konteks ulasan aplikasi mobile, *topic modeling* dapat membantu mengidentifikasi tema pembahasan utama, pola keluhan pengguna, serta isu-isu yang paling sering dibahas oleh pengguna aplikasi. Dengan demikian, *topic modeling* menjadi pendekatan yang relevan untuk mendukung evaluasi layanan digital pemerintah berbasis data pengguna.

Metode *topic modeling* tradisional seperti Latent Dirichlet Allocation (LDA) telah banyak digunakan dalam penelitian terkait analisis media sosial, ulasan pengguna, dan data teks online lainnya. LDA bekerja menggunakan pendekatan probabilistik berbasis *bag-of-words* untuk menemukan distribusi topik pada kumpulan dokumen teks [2], [3], [4]. Akan tetapi, model probabilistik seperti LDA memiliki keterbatasan dalam menganalisis teks pendek dan *noisy* karena hanya berfokus pada hubungan statistik antar kata tanpa memahami konteks semantik secara mendalam [5], [6]. Permasalahan tersebut menjadi semakin kompleks pada ulasan aplikasi mobile yang umumnya mengandung bahasa informal, singkatan, *typo*, campuran bahasa, serta struktur kalimat yang tidak konsisten. Kondisi ini menyebabkan interpretasi topik yang dihasilkan metode tradisional sering kali kurang optimal dalam merepresentasikan makna kontekstual dari ulasan pengguna.

Perkembangan terbaru dalam NLP menghadirkan pendekatan *topic modeling* berbasis *contextual embedding* yang mampu menghasilkan representasi semantik yang lebih baik dibandingkan dengan metode probabilistik tradisional. Salah satu metode yang berkembang adalah BERTopic, yaitu pendekatan *topic modeling* yang mengintegrasikan *transformer-based contextual embeddings*, *dimensionality reduction*, *density-based clustering*, dan *class-based TF-IDF* untuk menghasilkan topik yang lebih interpretatif dan kontekstual. BERTopic mampu menghasilkan representasi topik yang lebih semantik dibandingkan pendekatan *topic modeling* tradisional [5], [7]. Selain itu, pendekatan *contextual topic modeling* mampu meningkatkan kualitas interpretasi topik pada data teks yang pendek dan tidak terstruktur [5], [6]. Dengan demikian, BERTopic dinilai lebih sesuai untuk menganalisis ulasan aplikasi mobile yang mengandung variasi bahasa informal dan tingkat noise yang tinggi [6], [7], [8], [9].

Perkembangan *topic modeling* dalam beberapa tahun terakhir menunjukkan pergeseran dari pendekatan tradisional berbasis distribusi kata, seperti LDA, NMF, dan LSI, menuju pendekatan berbasis *contextual embedding* seperti BERTopic. Sejumlah studi sebelumnya telah membandingkan BERTopic dengan metode *topic modeling* tradisional dalam berbagai konteks, seperti tweet [10], [11], headline berita [9], forum MOOC [12], ulasan e-commerce [13], diskursus media sosial [14], dan opini publik terhadap program pemerintah [15]. Temuan dari studi-studi tersebut menunjukkan bahwa BERTopic berpotensi menangkap konteks semantik pada teks pendek, informal, dan tidak terstruktur. Namun, hasil perbandingan juga menunjukkan bahwa performa BERTopic tidak selalu secara mutlak lebih unggul dibandingkan metode lain, karena kualitas model sangat dipengaruhi oleh karakteristik dataset, jumlah topik, representasi teks, serta metrik evaluasi yang digunakan. Oleh karena itu, pemilihan BERTopic dalam penelitian ini tidak dimaksudkan untuk membuktikan superioritas metode, tetapi didasarkan pada kesesuaiannya dengan karakteristik ulasan pengguna aplikasi mobile yang pendek, tidak terstruktur, dan kaya konteks keluhan. Rekapitulasi lengkap penelitian terdahulu disajikan pada Tabel 1.

Meskipun BERTopic telah banyak digunakan dalam berbagai penelitian, sebagian besar studi terdahulu masih berfokus pada evaluasi performa metode, analisis opini publik di media sosial, klasifikasi sentimen, atau pemetaan topik pada domain komersial, pendidikan, dan program publik tertentu. Penelitian yang secara khusus menganalisis ulasan negatif terhadap aplikasi layanan perpajakan digital pemerintah masih relatif terbatas. Padahal, ulasan negatif pengguna aplikasi M-Pajak memuat informasi penting mengenai hambatan operasional yang dialami pengguna secara langsung, seperti kesulitan login, kegagalan autentikasi, keterlambatan verifikasi OTP, kendala integrasi NPWP dan NIK, serta gangguan stabilitas sistem. Dengan demikian, ulasan negatif tidak hanya merepresentasikan ekspresi ketidakpuasan, tetapi juga dapat dipahami sebagai sumber data operasional untuk mengevaluasi kualitas layanan publik digital.

Berdasarkan gap tersebut, penelitian ini bertujuan untuk menganalisis ulasan negatif pengguna aplikasi M-Pajak menggunakan BERTopic dalam kerangka App Store Review Mining dan Business Intelligence. Fokus penelitian diarahkan pada identifikasi tema keluhan dominan serta eksplorasi hambatan operasional yang muncul dalam pengalaman pengguna aktual. BERTopic digunakan untuk mengekstraksi tema-tema utama dari data ulasan yang pendek, tidak terstruktur, dan menggunakan bahasa informal, sedangkan perspektif Business Intelligence digunakan untuk menerjemahkan hasil *topic modeling* menjadi insight operasional yang relevan untuk evaluasi dan peningkatan kualitas layanan perpajakan digital pemerintah.

Kontribusi penelitian ini mencakup tiga aspek utama. Pertama, penelitian ini memperluas penerapan BERTopic dalam konteks layanan publik digital pemerintah, khususnya pada aplikasi perpajakan digital. Kedua, penelitian ini menyajikan pemetaan tema keluhan pengguna berdasarkan ulasan negatif pada aplikasi M-Pajak di Google Play Store sehingga dapat

memberikan gambaran empiris mengenai masalah layanan yang paling sering dialami pengguna. Ketiga, penelitian ini menginterpretasikan hasil topic modeling dalam perspektif Business Intelligence, sehingga temuan yang dihasilkan tidak hanya bersifat deskriptif, tetapi juga dapat digunakan sebagai dasar penyusunan rekomendasi strategis untuk meningkatkan stabilitas, aksesibilitas, dan kualitas layanan digital pemerintah.

Tabel 1. Pengelompokan Penelitian Terdahulu Berdasarkan Bidang Penerapan Topic Modeling dan Relevansinya terhadap Penelitian M-Pajak

No.	Kelompok Penelitian	Bidang/Kasus Penerapan	Fokus Umum Penelitian
1	Studi komparatif metode topic modeling	Korpus Bahasa Arab [16], abstrak PubMed [5], headline berita [9], dan Twitter posts [17].	Membandingkan BERTopic dengan metode tradisional seperti LDA, NMF, LSI, Top2Vec, atau pendekatan lain untuk menilai kualitas topik pada teks pendek dan tidak terstruktur
2	Topic modeling pada media sosial dan opini publik	X/Twitter [11], Instagram [18], TikTok [15], dan YouTube [19].	Mengidentifikasi isu, persepsi, dan opini publik terhadap kendaraan listrik, kesehatan masyarakat, lembaga pemerintah, dan program pemerintah
3	Topic modeling pada ulasan digital dan platform komersial	Ulasan Shopee [13], Google Maps [6], marketplace digital, dan produk template desain [20].	Menganalisis ulasan atau data digital untuk menemukan topik dominan, klasifikasi topik, sentimen, atau tren produk
4	Topic modeling pada pendidikan dan forum pembelajaran	Program akses pendidikan tinggi [21], platform pembelajaran kolaboratif [22], dan forum MOOC [12].	Mengidentifikasi isu mahasiswa, diskusi pembelajaran, dan urgent posts menggunakan BERTopic serta metode pembandingan lain
5	Topic modeling untuk pengembangan sistem, knowledge management, dan idea management	Aktivitas pengembangan perangkat lunak [23] dan sesi brainstorming [24].	Mengelompokkan aktivitas pengembangan perangkat lunak atau ide brainstorming untuk mendukung pengambilan keputusan dan pengelolaan pengetahuan
6	Topic modeling untuk literatur, tren riset, dan domain khusus	Medical AI [25], riset [26], dan tren riset Peatland [27].	Memetakan tren topik, perkembangan riset, dan hubungan antartopik dalam literatur akademik atau domain ilmiah tertentu

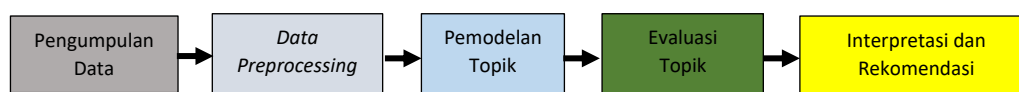
2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian

Penelitian ini menggunakan pendekatan App Store Review Mining untuk menganalisis ulasan negatif pengguna aplikasi M-Pajak yang diperoleh dari Google Play Store. Pendekatan ini banyak digunakan untuk mengekstraksi informasi dari *user-generated content* guna mengidentifikasi permasalahan aplikasi, kebutuhan pengguna, serta peluang perbaikan layanan digital. Martin et al. [28] menjelaskan bahwa ulasan aplikasi mobile dapat dimanfaatkan untuk mendukung pemeliharaan perangkat lunak, identifikasi bug, analisis kebutuhan pengguna, serta evaluasi berkelanjutan terhadap kualitas aplikasi. Selain itu, ulasan pengguna juga dapat digunakan sebagai sumber informasi strategis untuk mendukung pengambilan keputusan dan pengembangan layanan digital berbasis pengalaman pengguna.

Pendekatan ini digunakan untuk mentransformasikan ulasan pengguna yang tidak terstruktur menjadi informasi Business Intelligence yang dapat digunakan untuk mengevaluasi kualitas layanan perpajakan digital pemerintah. Fokus penelitian diarahkan pada identifikasi permasalahan layanan digital, hambatan operasional, kualitas pengalaman pengguna, serta isu tata kelola layanan berdasarkan ulasan negatif pengguna aplikasi.

Kerangka penelitian terdiri dari lima tahapan utama, yaitu pengumpulan dan penyaringan data ulasan, data preprocessing, modeling BERTopic, evaluasi topik, serta interpretasi *Business Intelligence* dan rekomendasi strategis. Tahapan-tahapan tersebut dirancang untuk menghasilkan insight operasional dan rekomendasi strategis berdasarkan ulasan pengguna. Diagram alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram alur penelitian

Pada tahap pengumpulan data, ulasan pengguna aplikasi dikumpulkan dari Google Play Store menggunakan pendekatan *scraping* berbasis Python dengan bantuan *library* Google Play Scraper. Data yang diperoleh meliputi isi ulasan, *rating*

pengguna, tanggal ulasan, versi aplikasi, serta metadata lainnya yang berkaitan dengan pengalaman penggunaan aplikasi. Data ulasan kemudian disaring untuk mendapatkan ulasan negatif yang ditandai dengan rating 1 atau 2. Rating ulasan 1 dan 2 dinilai merepresentasikan ketidakpuasan pengguna akibat kegagalan sistem, masalah autentikasi, serta berbagai hambatan operasional dalam penggunaan aplikasi.

Tahap *data preprocessing* bertujuan untuk meningkatkan kualitas data tekstual sebelum dilakukan analisis topik. Tahap ini meliputi pembersihan teks, penghapusan simbol dan URL, normalisasi kata, penghapusan stopword, serta penghapusan data yang terduplikat. Proses preprocessing diperlukan untuk mengurangi *noise* pada data ulasan sehingga representasi teks menjadi lebih konsisten dan siap digunakan dalam proses analisis topik [10].

Tahap ketiga adalah BERTopic Modeling. Pada tahap ini dilakukan proses identifikasi tema-tema utama yang muncul dalam ulasan negatif dari pengguna. Hasil *topic modeling* kemudian digunakan untuk menggambarkan pola permasalahan dan keluhan pengguna berdasarkan tema-tema yang terbentuk dari data ulasan.

Tahap evaluasi topik digunakan untuk mengevaluasi kualitas topik yang dihasilkan oleh model BERTopic. Evaluasi dilakukan menggunakan parameter Topic Coherence, Topic Diversity, dan Topic Quality. Evaluasi dilakukan untuk memastikan bahwa topik yang dihasilkan memiliki kualitas semantik yang baik dan mampu merepresentasikan tema permasalahan ulasan pengguna dengan jelas.

Tahap terakhir adalah tahap interpretasi Business Intelligence dan pemberian rekomendasi strategis. Pada tahap ini, topik-topik yang dihasilkan dianalisis dari perspektif operasional dan tata kelola layanan digital pemerintah untuk menghasilkan insight terkait kualitas layanan, hambatan operasional, serta pengalaman pengguna terhadap aplikasi M-Pajak. Selain itu, hasil interpretasi digunakan untuk menyusun rekomendasi strategis yang mendukung peningkatan kualitas layanan perpajakan digital pemerintah.

2.2 Pengumpulan Data

Data untuk penelitian ini adalah ulasan pengguna aplikasi M-Pajak di Google Play Store. Data ulasan dikumpulkan menggunakan teknik scraping berbasis Python dengan bantuan pustaka Google Play Scraper yang dijalankan pada lingkungan Google Colab. Proses scraping dilakukan pada tanggal 28 Mei 2026 dengan pengaturan parameter pengambilan data ulasan pengguna maksimal 10.000 ulasan atau 1.100 hari (3 tahun), mana yang tercapai lebih dahulu. Pengambilan data dilakukan menggunakan pendekatan *pagination retrieval* untuk memperoleh ulasan terbaru secara bertahap.

Data yang dikumpulkan meliputi ID ulasan, nama pengguna, tanggal ulasan, rating, isi ulasan, versi aplikasi, dan jumlah likes pada ulasan pengguna. Masing-masing atribut ini dijelaskan pada Tabel 2. Dataset yang diperoleh kemudian diperiksa untuk memastikan konsistensi atribut dan kelengkapan data sebelum memasuki tahap preprocessing.

Tabel 2. Atribut Dataset Ulasan

Atribut	Deskripsi
ulasan_id	Identitas unik ulasan
user_name	Nama pengguna
ulasan_date	Tanggal ulasan
rating	Nilai rating pengguna
ulasan	Isi ulasan pengguna
app_version	Versi aplikasi
thumbs up	Jumlah likes ulasan

Setelah proses pengumpulan data selesai dilakukan, dataset difilter berdasarkan nilai rating untuk memperoleh ulasan negatif sebagai fokus penelitian. Penelitian ini menggunakan ulasan dengan rating 1 dan 2 karena kategori tersebut umumnya merepresentasikan ketidakpuasan pengguna, masalah teknis, kegagalan sistem, kesulitan autentikasi, serta hambatan operasional dalam penggunaan aplikasi M-Pajak. Dataset ulasan negatif selanjutnya digunakan sebagai sumber utama pada tahap preprocessing, pemodelan BERTopic, evaluasi topik, dan interpretasi Business Intelligence.

2.3 Data Preprocessing

Data ulasan yang telah dikumpulkan selanjutnya melalui proses *preprocessing* untuk meningkatkan kualitas representasi teks sebelum dilakukan analisis topik menggunakan BERTopic. Tahap *preprocessing* diperlukan karena ulasan pada Google Play Store umumnya mengandung bahasa informal, singkatan, simbol, emoji, URL, dan variasi penulisan yang dapat menghasilkan *noise* dalam proses *topic modeling*. *Preprocessing* merupakan tahapan penting dalam topic modeling karena dapat memengaruhi kualitas representasi teks dan interpretasi topik yang dihasilkan [29].

Proses *preprocessing* meliputi *case folding*, text cleaning, penghapusan emoji dan emotikon, penghapusan stopword, serta normalisasi teks [10]. *Case folding* dilakukan dengan mengubah seluruh teks menjadi huruf kecil untuk menjaga konsistensi token. *Text cleaning* dilakukan dengan menghapus URL, tanda baca, angka, simbol khusus, dan spasi berlebih

yang tidak berkontribusi secara semantik terhadap analisis topik. Selain itu, emoji dan emotikon juga dihapus untuk mengurangi token yang tidak relevan dalam proses pembentukan topik.

Selanjutnya, proses *stopword removal* dilakukan menggunakan pustaka Sastrawi untuk menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap identifikasi topik. Sedangkan tahap normalisasi teks dilakukan untuk mengubah kata tidak baku, singkatan, dan bahasa informal menjadi format bahasa Indonesia standar agar representasi teks lebih konsisten. Contoh normalisasi yang dilakukan meliputi perubahan kata seperti “gk” menjadi “tidak”, “bgt” menjadi “banget”, dan “apk” menjadi “aplikasi”.

Pada penelitian ini, proses stemming tidak dilakukan. Pendekatan ini dipilih karena BERTopic menggunakan contextual embeddings berbasis transformer yang mempertahankan konteks semantik kata dalam suatu kalimat. Proses stemming berpotensi menghilangkan nuansa konteks dan makna kata yang diperlukan untuk membentuk representasi *embedding*. Selain itu, beberapa penelitian menunjukkan bahwa penggunaan contextual embedding tidak selalu memerlukan stemming karena model transformer telah mampu memahami hubungan semantik antarkata berdasarkan konteks penggunaannya [12], [13], [21]. Dengan demikian, teks hasil *preprocessing* digunakan secara langsung sebagai input dalam proses modeling BERTopic untuk mempertahankan kualitas representasi semantik ulasan pengguna.

2.4 Pemodelan BERTopic

Penelitian ini menggunakan BERTopic sebagai metode utama untuk mengidentifikasi tema-tema utama pada ulasan negatif pengguna aplikasi M-Pajak. BERTopic dipilih karena mampu menghasilkan representasi topik yang lebih kontekstual dan semantik dibandingkan metode topic modeling tradisional seperti Latent Dirichlet Allocation (LDA). BERTopic mengintegrasikan *contextual embeddings* berbasis transformer, *dimensionality reduction*, *clustering*, dan *class-based* TF-IDF untuk menghasilkan topik yang lebih mudah diinterpretasikan pada data teks yang tidak terstruktur. Pendekatan ini dinilai sesuai untuk menganalisis ulasan aplikasi mobile yang umumnya pendek, *noisy*, dan mengandung variasi bahasa yang tidak formal [6], [7].

Tabel 3. Perbandingan kinerja Model BERTopic dengan beberapa model lain

No	Kasus	Metode	Kesimpulan
1	Topic modeling pada korpus Bahasa Arab [16].	BERTopic, LDA, NMF	BERTopic menunjukkan hasil keseluruhan yang lebih baik dibandingkan NMF dan LDA berdasarkan pengukuran NPMI.
2	Topic modeling pada Twitter posts [7]	LDA, NMF, Top2Vec, BERTopic	Penulis menyimpulkan bahwa BERTopic dan NMF memiliki potensi yang kuat untuk menganalisis data Twitter.
3	Topic modeling pada 1.837 abstrak PubMed [5]	LDA/MALLET dan BERTopic/BioBERT	BERTopic memiliki keunggulan dalam hal interpretabilitas dan koherensi semantik, terutama karena memanfaatkan embedding kontekstual.
4	Identifikasi topik pada urgent posts di forum MOOC [12]	LDA, LSI, NMF, BERTopic	NMF dan BERTopic menunjukkan performa yang lebih baik dibandingkan LDA dan LSI. NMF lebih unggul ketika jumlah topik sedikit, sedangkan BERTopic lebih baik ketika dibutuhkan jumlah topik yang lebih banyak dan lebih spesifik.
5	Program Makan Bergizi Gratis pada data TikTok [15]	LDA dan BERTopic	LDA memperoleh coherence dan topic diversity lebih tinggi, tetapi BERTopic dinilai mampu menghasilkan topik yang lebih kontekstual dan representatif secara semantik
6	Topic modeling pada news headlines [9]	LDA, NMF, BERTopic	Penulis menekankan bahwa perbedaan performa metode perlu dilihat dalam konteks, interpretabilitas, dan karakteristik teks pendek.

Pemilihan BERTopic dalam penelitian ini didasarkan pada karakteristik data ulasan aplikasi M-Pajak yang pendek, informal, tidak terstruktur, dan mengandung noise. Tabel 3 menunjukkan beberapa penelitian terdahulu yang menunjukkan bahwa BERTopic berpotensi menangkap kedekatan semantik pada teks pendek dan tidak terstruktur, serta telah dibandingkan dengan metode tradisional seperti LDA dan NMF. Abuzayed dan Al-Khalifa [16] menyimpulkan bahwa BERTopic dapat menghasilkan performa yang lebih baik dibandingkan LDA dan NMF dalam topic modeling Bahasa Arab. Egger dan Yu [7] juga menunjukkan bahwa BERTopic dan NMF berpotensi digunakan untuk menganalisis tweet pendek yang bermasalah. Namun, penelitian lain menunjukkan bahwa performa topic modeling dapat bervariasi bergantung pada konteks data, jumlah topik, dan metrik evaluasi yang digunakan. Oleh karena itu, penelitian ini tidak mengklaim bahwa BERTopic merupakan metode terbaik secara umum, tetapi penggunaan BERTopic dinilai sebagai pendekatan eksploratif yang sesuai untuk mengidentifikasi tema keluhan pengguna aplikasi M-Pajak secara kontekstual.

Proses BERTopic dalam penelitian ini terdiri dari 4 komponen utama, yaitu *transformer embeddings*, UMAP (Uniform Manifold Approximation and Projection), HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), dan class-based TF-IDF (c-TF-IDF). Tahap awal dilakukan dengan membentuk *contextual embeddings* menggunakan model transformer untuk menghasilkan representasi semantik dari ulasan pengguna. Representasi *embedding* tersebut kemudian direduksi dimensinya menggunakan UMAP agar proses *clustering* dapat dilakukan dengan lebih efisien. Penelitian sebelumnya menemukan bahwa UMAP mampu mempertahankan struktur semantik data berdimensi tinggi selama proses reduksi dimensi [30].

Pemilihan parameter UMAP dan HDBSCAN dilakukan dengan mempertimbangkan karakteristik data ulasan pengguna aplikasi M-Pajak yang bersifat pendek, tidak terstruktur, informal, dan mengandung noise. Pada tahap reduksi dimensi, UMAP menggunakan parameter $n_neighbors = 5$, $n_components = 2$, dan $metric = cosine$. Nilai $n_neighbors = 5$ digunakan agar model lebih sensitif dalam menangkap struktur lokal antardokumen, sehingga ulasan-ulasan pendek yang memiliki kedekatan makna dapat dikelompokkan dengan lebih jelas. Parameter $n_components = 2$ digunakan untuk mereduksi representasi embedding ke dalam ruang dua dimensi agar proses clustering lebih efisien dan hasilnya dapat divisualisasikan. Sementara itu, $metric = cosine$ dipilih karena sesuai untuk mengukur kemiripan semantik antarrepresentasi teks berbasis embedding.

Selanjutnya, proses *clustering* dilakukan menggunakan HDBSCAN untuk mengelompokkan ulasan yang memiliki kemiripan semantik ke dalam topik tertentu. HDBSCAN dipilih karena mampu menangani distribusi data yang tidak teratur serta mengidentifikasi ulasan yang tidak memiliki kedekatan semantik yang cukup dengan cluster tertentu sebagai outlier topic. Penelitian sebelumnya menemukan bahwa HDBSCAN efektif digunakan untuk *clustering* berbasis densitas pada data yang kompleks dan tidak homogen [31], [32].

Pada tahap clustering, HDBSCAN digunakan dengan parameter $min_cluster_size = 10$ dan $metric = euclidean$. Nilai $min_cluster_size = 10$ dipilih untuk menjaga keseimbangan antara jumlah topik yang dihasilkan dan interpretabilitas topik. Nilai yang terlalu kecil dapat menghasilkan terlalu banyak topik yang sempit dan sulit diinterpretasikan, sedangkan nilai yang terlalu besar dapat menggabungkan beberapa isu yang berbeda ke dalam satu topik yang terlalu umum. HDBSCAN dipilih karena mampu mengelompokkan data berdasarkan kepadatan serta mengidentifikasi ulasan yang tidak memiliki kedekatan semantik yang kuat sebagai outlier. Hal ini relevan karena ulasan aplikasi mobile sering mengandung variasi bahasa, kesalahan penulisan, dan konteks keluhan yang beragam.

Tahap akhir menggunakan c-TF-IDF untuk mengidentifikasi kata-kata yang paling representatif pada masing-masing topik berdasarkan distribusi kata dalam cluster ulasan. Kata-kata representatif tersebut kemudian digunakan untuk menginterpretasikan tema utama permasalahan layanan digital yang dialami pengguna aplikasi M-Pajak. Pendekatan ini memungkinkan identifikasi tema keluhan pengguna secara lebih kontekstual sehingga dapat mendukung proses interpretasi Business Intelligence terhadap kualitas layanan perpajakan digital pemerintah.

2.5 Evaluasi Topik

Topik yang dihasilkan dari proses BERTopic dievaluasi menggunakan tiga metrik utama, yaitu *Topic Coherence*, *Topic Diversity*, dan *Topic Quality*. Evaluasi dilakukan untuk mengukur konsistensi semantik, keterpisahan antartopik, serta kualitas representasi tema yang dihasilkan dari dataset ulasan pengguna aplikasi M-Pajak.

Topic Coherence digunakan untuk mengukur tingkat kesamaan semantik antar kata pada setiap topik. Nilai *koherensi* yang lebih tinggi menunjukkan bahwa kata-kata dalam suatu topik memiliki hubungan semantik yang lebih konsisten, sehingga topik tersebut lebih mudah diinterpretasikan [1], [8], [9], [29]. Secara umum, coherence dapat dirumuskan sebagai berikut:

$$TC = \frac{1}{N} \sum_{i=1}^N coherence(T_i)$$

dengan TC merupakan nilai rata-rata topic coherence, N adalah jumlah topik, dan T_i adalah topik ke- i .

Topic Diversity digunakan untuk mengukur tingkat keunikan kata antar topik guna mengurangi redundansi dan kemiripan tema yang dihasilkan model. Topic Diversity menunjukkan keberagaman kata antar topik dengan menghitung proporsi kata unik terhadap seluruh kata representatif yang digunakan dalam topik [1], [8], [9], [29]. Topic Diversity dihitung menggunakan rumus berikut ini:

$$TD = \frac{W_u}{W_t}$$

dimana TD adalah Topic Diversity, W_u adalah jumlah kata unik pada seluruh topik, dan W_t adalah total kata yang merepresentasikan seluruh topik. Nilai Topic Diversity yang semakin tinggi menunjukkan bahwa topik-topik yang dihasilkan memiliki kata-kata representatif yang semakin beragam dan tidak terlalu redundan.

Sementara itu, *Topic Quality* digunakan untuk memberikan evaluasi yang lebih seimbang terhadap performa model berdasarkan kombinasi koherensi dan diversitas. Topic Quality dihitung sebagai kombinasi antara Topic Coherence dan

Topic Diversity untuk memperoleh gambaran keseimbangan antara konsistensi semantik dan keberagaman topik [1], [8], [9], [29]. Rumus yang digunakan adalah menghitung Topic Quality adalah sebagai berikut:

$$TQ = TC \times TD$$

Semakin tinggi nilai Topic Coherence, maka semakin konsisten hubungan semantik antar kata dalam suatu topik. Semakin tinggi nilai Topic Diversity, maka semakin rendah tingkat redundansi antar topik. Sementara itu, Topic Quality menunjukkan kualitas keseluruhan model dengan mempertimbangkan koherensi dan keberagaman topik secara bersamaan.

Ketiga metrik tersebut digunakan untuk memastikan bahwa topik yang dihasilkan BERTopic mampu merepresentasikan permasalahan layanan digital secara semantik, tidak redundan, dan mudah diinterpretasikan, sehingga dapat mendukung proses analisis Business Intelligence terhadap kualitas layanan perpajakan digital pemerintah.

Seluruh proses pengolahan data dilakukan menggunakan Python di lingkungan Google Colab. Library utama yang digunakan meliputi google-play-scraper untuk pengumpulan data, pandas untuk pengelolaan dataset, Sastrawi dan NLTK untuk preprocessing teks, sentence-transformers untuk pembentukan embedding, BERTopic untuk pemodelan topik, UMAP untuk reduksi dimensi, HDBSCAN untuk clustering, serta scikit-learn untuk proses evaluasi dan analisis pendukung. Penggunaan Google Colab dipilih karena menyediakan lingkungan komputasi berbasis cloud yang memadai untuk menjalankan proses embedding dan topic modeling pada dataset ulasan pengguna.

2.6 Interpretasi Business Intelligence dan Rekomendasi Strategis

Topik-topik yang dihasilkan oleh BERTopic diinterpretasikan menggunakan perspektif Business Intelligence untuk menghasilkan insight terkait kualitas layanan perpajakan digital. Interpretasi dilakukan dengan menganalisis tema-tema utama yang muncul dalam ulasan negatif pengguna untuk mengidentifikasi pola permasalahan layanan, hambatan operasional, serta pengalaman pengguna dalam menggunakan aplikasi M-Pajak. Analisis dilakukan untuk memahami berbagai isu layanan digital seperti permasalahan autentikasi, gangguan sistem, stabilitas aplikasi, performa layanan, serta kendala akses yang dialami pengguna. Proses interpretasi dilakukan dengan menghubungkan kata-kata representatif pada setiap topik dengan konteks operasional layanan perpajakan digital.

Hasil interpretasi digunakan untuk menghasilkan insight operasional yang dapat mendukung evaluasi kualitas layanan digital pemerintah. Selain itu, hasil analisis juga digunakan sebagai dasar penyusunan rekomendasi strategis untuk meningkatkan kualitas layanan perpajakan digital pemerintah secara berkelanjutan.

3. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Dataset

Penelitian ini berhasil mengumpulkan 6.024 ulasan pengguna aplikasi M-Pajak dari Google Play Store. Data ulasan merupakan ulasan pengguna selama periode 1.100 hari dari 23 Mei 2023 hingga 26 Mei 2026. Ulasan yang diperoleh mencerminkan berbagai pengalaman pengguna terkait layanan perpajakan digital, termasuk kendala autentikasi, gangguan sistem, error aplikasi, serta berbagai keluhan operasional lainnya.

Distribusi rating ulasan pengguna ditunjukkan pada Tabel 4. Hasil distribusi menunjukkan bahwa rating 1 mendominasi dataset dengan jumlah 4.784 ulasan atau sebesar 79,4% dari total ulasan yang diperoleh. Sementara itu, rating 5 hanya berjumlah 795 ulasan atau sebesar 13,2%. Dominasi rating rendah menunjukkan bahwa sebagian besar pengguna memberikan penilaian negatif terhadap kualitas layanan aplikasi M-Pajak.

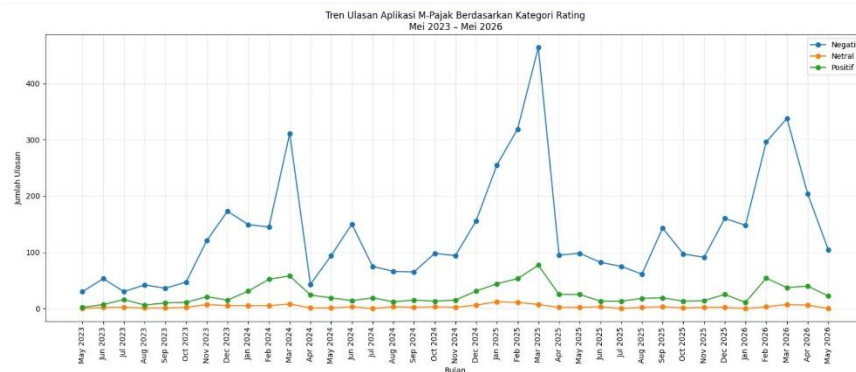
Tabel 4. Distribusi Rating Ulasan Pengguna

Rating	Jumlah Ulasan	Persentase (%)
1	4.784	79,4
2	224	3,72
3	122	2,03
4	99	1,64
5	795	13,20
Total	6.024	100,00

Berdasarkan hasil filtering, diperoleh sebanyak 5.008 ulasan negatif dengan rating 1 dan 2 yang digunakan sebagai fokus utama analisis *topic modeling*. Jumlah tersebut menunjukkan bahwa lebih dari 83% ulasan pengguna berada dalam kategori negatif.

Gambar 2 menunjukkan bahwa ulasan negatif secara konsisten mendominasi ulasan pengguna aplikasi M-Pajak selama periode Mei 2023 hingga Mei 2026. Dominasi ini mengindikasikan bahwa sebagian besar pengguna mengalami kendala dalam menggunakan layanan perpajakan digital, terutama pada aspek akses, autentikasi, verifikasi, dan stabilitas sistem. Pola yang menarik terlihat pada bulan Maret setiap tahun, ketika jumlah ulasan negatif meningkat secara signifikan. Fenomena ini dapat dikaitkan dengan periode pelaporan SPT Tahunan Pajak Orang Pribadi yang memiliki batas waktu pada tanggal 31 Maret. Menjelang tenggat waktu tersebut, intensitas penggunaan aplikasi meningkat sehingga beban

sistem juga meningkat. Peningkatan ulasan negatif yang tidak diikuti oleh peningkatan sebanding pada ulasan positif dan netral menunjukkan bahwa sebagian pengguna mengalami hambatan saat mengakses layanan pada periode puncak. Dari perspektif *Business Intelligence*, tren ini menunjukkan bahwa sentimen negatif pengguna tidak terjadi secara acak, melainkan mengikuti pola musiman yang erat kaitannya dengan siklus pelaporan pajak tahunan. Temuan ini memberikan indikasi bahwa animo pengguna untuk mengadopsi layanan guna menyelesaikan administrasi kewajiban pajak mereka sangat tinggi. Namun, layanan digital tersebut masih menghadapi kendala dan memberikan pengalaman yang kurang memuaskan bagi para pengguna. Dengan memahami pola temporal tersebut, pengelola layanan dapat merencanakan kapasitas sistem, memperkuat sumber daya teknis, dan menerapkan strategi mitigasi risiko sebelum memasuki periode puncak penggunaan, sehingga kualitas layanan digital perpajakan tetap terjaga.



Gambar 2. Tren sentimen ulasan pengguna aplikasi M-Pajak periode Mei 2023–Mei 2026

3.2 Tahap Preprocessing

Dataset ulasan negatif pengguna aplikasi M-Pajak selanjutnya melalui tahap preprocessing untuk mengurangi noise pada data tekstual sebelum dilakukan proses modeling BERTopic. Tahap ini diperlukan karena ulasan pengguna di Google Play Store umumnya menggunakan bahasa informal, singkatan, simbol, emoji, dan variasi penulisan yang dapat memengaruhi kualitas representasi teks.

Proses preprocessing meliputi *case folding*, *text cleaning*, penghapusan emoji dan karakter khusus, penghapusan stopwords, normalisasi kata tidak baku, serta penghapusan data kosong dan data duplikat. Berbeda dengan modeling topik menggunakan LDA, pada BERTopic proses stemming tidak dilakukan untuk mempertahankan konteks semantik kata dalam proses pembentukan *embedding* kontekstual [12], [13], [21], [33].

Hasil preprocessing menunjukkan bahwa teks ulasan menjadi lebih bersih, lebih konsisten, dan lebih mudah diinterpretasikan. Tabel 5 menunjukkan contoh perubahan ulasan pengguna sebelum dan sesudah proses preprocessing.

Tabel 5. Contoh Hasil Preprocessing Ulasan

Ulasan Asli	Hasil Preprocessing
“Aplikasi nya jelek banget, login terus gagal!!!”	aplikasi jelek banget login terus gagal
“Kode OTP ngak masuk masuk, parah.”	kode otp tidak masuk parah
“Error terus pas buka aplikasi.”	error terus buka aplikasi
“Verifikasi email lama sekali dan gagal.”	verifikasi email lama gagal
“Server down terus, susah login.”	server down terus susah login

Hasil preprocessing membantu mengurangi variasi bahasa informal dan token yang tidak relevan, sehingga representasi teks menjadi lebih konsisten. Dengan demikian, proses BERTopic dapat menghasilkan topik yang lebih jelas, lebih semantik, dan lebih mudah diinterpretasikan.

3.3 Pemodelan BERTopic

Implementasi modeling BERTopic pada penelitian ini menggunakan model embedding *paraphrase-multilingual-MiniLM-L12-v2* dengan kombinasi UMAP, HDBSCAN, dan *class-based TF-IDF* (c-TF-IDF) untuk menghasilkan representasi topik yang lebih kontekstual. Parameter utama yang digunakan pada proses modeling BERTopic ditunjukkan pada Tabel 6.

Hasil pemodelan BERTopic menghasilkan 24 topik utama dan 1 topik *outlier* (Topic -1). Topik *outlier* merepresentasikan kumpulan ulasan yang tidak memiliki kedekatan semantik yang cukup kuat untuk dimasukkan ke dalam cluster topik tertentu. Berdasarkan hasil pemodelan, Topic 0 merupakan topik dengan jumlah ulasan terbesar sebanyak 1.095 ulasan, diikuti oleh *outlier topic* sebanyak 1.093 ulasan. Selain itu, beberapa topik dominan lainnya berkaitan dengan error sistem, verifikasi email dan OTP, autentikasi akun, serta permasalahan NPWP dan NIK. Tabel 7 menunjukkan beberapa topik utama yang dihasilkan oleh BERTopic beserta kata-kata representatif untuk masing-masing topik.

Tabel 6. Parameter BERTopic

Komponen	Parameter
Embedding Model	paraphrase-multilingual-MiniLM-L12-v2
UMAP n_neighbors	5
UMAP n_components	2
UMAP metric	cosine
HDBSCAN min_cluster_size	10
HDBSCAN metric	euclidean

Tabel 7. Contoh Topik Utama BERTopic

Topik	Jumlah Ulasan	Kata Representatif
0	1.095	sulit, buruk, masuk, rumit, buka
1	331	error, terus, gagal, data
2	297	email, verifikasi, kode, sms
3	276	npwp, digit, nik, daftar
-1	1.093	aplikasinya, bayar, masuk, laporan

Hasil *topic modeling* menunjukkan bahwa sebagian besar ulasan negatif berkaitan dengan hambatan akses aplikasi, kegagalan autentikasi, error sistem, serta permasalahan verifikasi akun. Dominasi topik-topik tersebut mengindikasikan bahwa stabilitas sistem dan proses autentikasi menjadi isu utama yang paling sering dikeluhkan oleh pengguna aplikasi M-Pajak.

3.4 Evaluasi BERTopic

Evaluasi dilakukan untuk mengukur kualitas topik yang dihasilkan BERTopic dalam merepresentasikan permasalahan utama pada ulasan negatif pengguna aplikasi M-Pajak. Penelitian ini menggunakan tiga metrik evaluasi, yaitu Topic Coherence (TC), Topic Diversity (TD), dan Topic Quality (TQ). Topic Coherence digunakan untuk mengukur konsistensi semantik antarkata dalam suatu topik, sedangkan Topic Diversity digunakan untuk mengukur tingkat keberagaman kata antartopik sehingga dapat mengurangi redundansi tema. Topic Quality digunakan untuk mengevaluasi keseimbangan antara coherence dan diversity secara keseluruhan [1], [8], [9], [29], [34]. Tabel 8 menunjukkan hasil evaluasi BERTopic pada dataset ulasan negatif aplikasi M-Pajak.

Tabel 8. Hasil Evaluasi BERTopic

Metrik Evaluasi	Nilai
Topic Coherence	0,522
Topic Diversity	0,667
Topic Quality	0,348

Hasil evaluasi menunjukkan bahwa BERTopic menghasilkan topik dengan kualitas moderat. Nilai Topic Coherence sebesar 0,522 menunjukkan bahwa hubungan semantik antarkata dalam topik masih cukup dapat diinterpretasikan, sedangkan Topic Diversity sebesar 0,667 menunjukkan bahwa keberagaman kata antartopik relatif memadai. Kombinasi coherence dan diversity penting untuk menghasilkan topic modeling yang interpretatif dan tidak saling tumpang tindih [29]. Topic Quality sebesar 0,348 menunjukkan bahwa kualitas model secara keseluruhan belum dapat dikategorikan sebagai tinggi, namun masih pada level moderat dan dapat diterima. Oleh karena itu, hasil pemodelan dalam penelitian ini diposisikan sebagai dasar untuk mengeksplorasi pola keluhan pengguna aplikasi M-Pajak dan digunakan untuk melakukan analisis business intelligence guna menemukan topik-topik dominan dari keluhan tersebut.

Selain mengevaluasi keseluruhan model, penelitian ini juga menghitung topic coherence untuk masing-masing topik utama yang dihasilkan oleh BERTopic. Hasil evaluasi ditunjukkan pada Tabel 9. Topik “Verifikasi Email dan SMS” dan “Keterlambatan OTP dan Verifikasi” memiliki nilai koherensi tertinggi, yang menunjukkan bahwa kedua topik tersebut memiliki konsistensi semantik yang kuat. Sementara itu, topik “Kesulitan Autentikasi” memiliki nilai *coherence* terendah sebesar 0,468, yang menunjukkan bahwa topik tersebut mengandung variasi konteks dan kata yang lebih beragam dibandingkan topik lainnya.

Penelitian ini belum melakukan perbandingan empiris dengan metode topic modeling lain seperti LDA atau NMF. Penelitian ini fokus pada penggunaan BERTopic untuk menemukan tema keluhan pengguna M-Pajak dan melakukan analisis Business Intelligence. Perbandingan lintas metode dapat dilakukan pada penelitian selanjutnya untuk memperoleh evaluasi model yang lebih komprehensif.

3.5 Interpretasi Business Intelligence dan Rekomendasi Strategis

Hasil BERTopic menunjukkan bahwa ulasan negatif pengguna aplikasi M-Pajak didominasi oleh permasalahan autentikasi, verifikasi akun, error sistem, dan gangguan akses layanan digital. Beberapa topik dominan yang dihasilkan ditunjukkan pada Tabel 10.

Tabel 9. Topic Coherence per Topik

Topik	Coherence
Verifikasi Email dan SMS	0,782
Keterlambatan OTP dan Verifikasi	0,759
Masalah Password dan Login	0,711
Verifikasi Foto dan eFIN	0,642
Registrasi NPWP dan NIK	0,608
Akses Website dan Layanan Online	0,598
Pembayaran Pajak	0,587
Kritik terhadap Pemerintah	0,547
Error Sistem dan Kegagalan Data	0,529
Kesulitan Autentikasi	0,468

Tabel 10. Topik Dominan Ulasan Negatif Pengguna Aplikasi M-Pajak

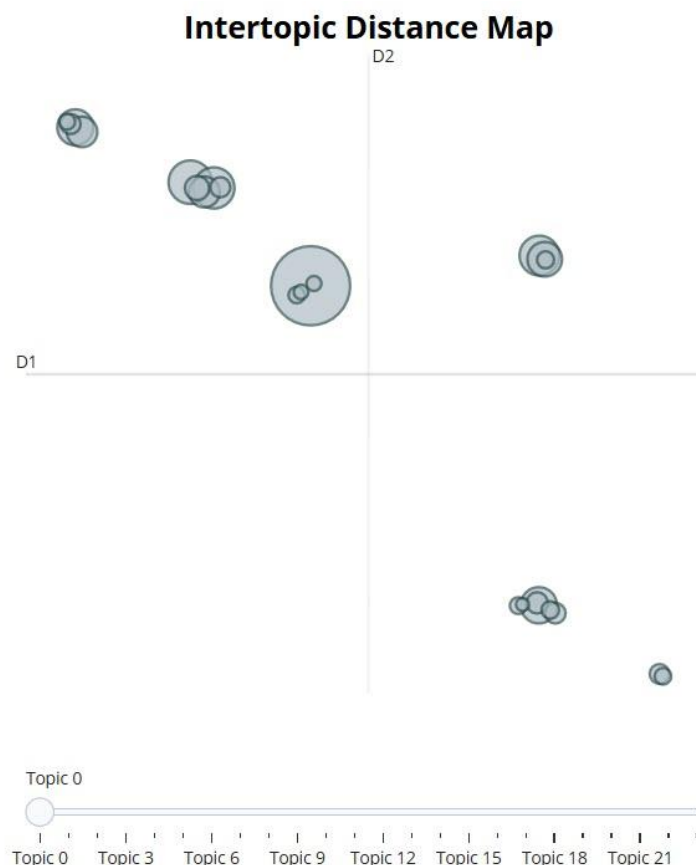
Topik	Kata Representatif	Jumlah Ulasan
Kesulitan Login dan Akses	sulit, buruk, masuk, rumit	1.095
Error Sistem dan Kegagalan Data	error, gagal, data	331
Verifikasi Email dan SMS	email, verifikasi, kode, sms	297
Registrasi NPWP dan NIK	npwp, nik, daftar	276
Website dan Layanan Online	web, online, website	229
Pembayaran Pajak	bayar, uang, negara	205
Password dan Login	password, lupa, salah	194
OTP dan Verifikasi	kode, verifikasi, otp	182

Topik “Kesulitan Login dan Akses” menjadi keluhan paling dominan dengan 1.095 ulasan. Hasil ini mengindikasikan bahwa proses autentikasi masih menjadi hambatan utama dalam penggunaan aplikasi M-Pajak. Selain itu, topik “Error Sistem dan Kegagalan Data” menunjukkan bahwa stabilitas aplikasi dan reliabilitas sistem masih menjadi permasalahan yang sering dihadapi pengguna.

Topik yang berkaitan dengan verifikasi email, OTP, dan sinkronisasi identitas digital juga muncul secara konsisten dalam hasil *topic modeling*. Temuan ini menunjukkan bahwa proses verifikasi akun dan integrasi data identitas digital masih memerlukan perbaikan agar mendukung akses layanan yang lebih stabil dan efisien. Permasalahan tersebut berpotensi memengaruhi kualitas pengalaman pengguna serta persepsi masyarakat terhadap layanan perpajakan digital pemerintah. Direktorat Jenderal Pajak perlu mengingat bahwa keberhasilan layanan publik digital tidak hanya ditentukan oleh ketersediaan sistem, tetapi juga oleh kualitas pengalaman pengguna, reliabilitas layanan, kesiapan sumber daya dan respons institusi [35], [36].

Visualisasi *Intertopic Distance Map* pada Gambar 3 menunjukkan bahwa sebagian besar topik membentuk kluster yang saling berdekatan secara semantik. Kondisi ini mengindikasikan bahwa berbagai keluhan pengguna memiliki keterkaitan yang kuat, khususnya terkait isu autentikasi, verifikasi akun, dan stabilitas sistem. Sementara itu, beberapa topik seperti pembayaran pajak dan layanan website berada relatif terpisah dari kluster utama, yang menunjukkan karakteristik permasalahan yang lebih spesifik.

Gambar 3 menampilkan *Intertopic Distance Map* hasil pemodelan BERTopic yang menggambarkan distribusi, ukuran, dan kedekatan semantik antar topik pada ulasan negatif aplikasi M-Pajak. Setiap lingkaran merepresentasikan satu topik, sedangkan ukuran lingkaran menunjukkan besarnya proporsi ulasan pada topik tersebut. Jarak antarlingkaran menunjukkan tingkat kedekatan makna antartopik: semakin dekat posisi antartopik, semakin besar kemungkinan bahwa topik-topik tersebut memiliki kemiripan konteks keluhan; sebaliknya, topik yang berada lebih jauh menunjukkan isu yang lebih berbeda secara semantik. Berdasarkan visualisasi tersebut, terlihat bahwa sejumlah topik membentuk beberapa kelompok yang saling berdekatan, terutama di area kiri atas, tengah kiri, dan kanan bawah. Pola ini menunjukkan bahwa keluhan pengguna aplikasi M-Pajak tidak sepenuhnya berdiri sendiri, melainkan membentuk kluster permasalahan yang saling berkaitan. Topik-topik yang berdekatan dapat diinterpretasikan sebagai kelompok masalah yang muncul dalam satu rangkaian pengalaman pengguna, misalnya hambatan akses, login, autentikasi, verifikasi email atau OTP, serta validasi data identitas seperti NPWP dan NIK. Dengan kata lain, masalah utama pengguna tidak hanya terjadi pada satu fitur tertentu, tetapi pada keseluruhan alur penggunaan layanan digital, mulai dari proses masuk ke aplikasi, validasi identitas, hingga akses ke layanan perpajakan.



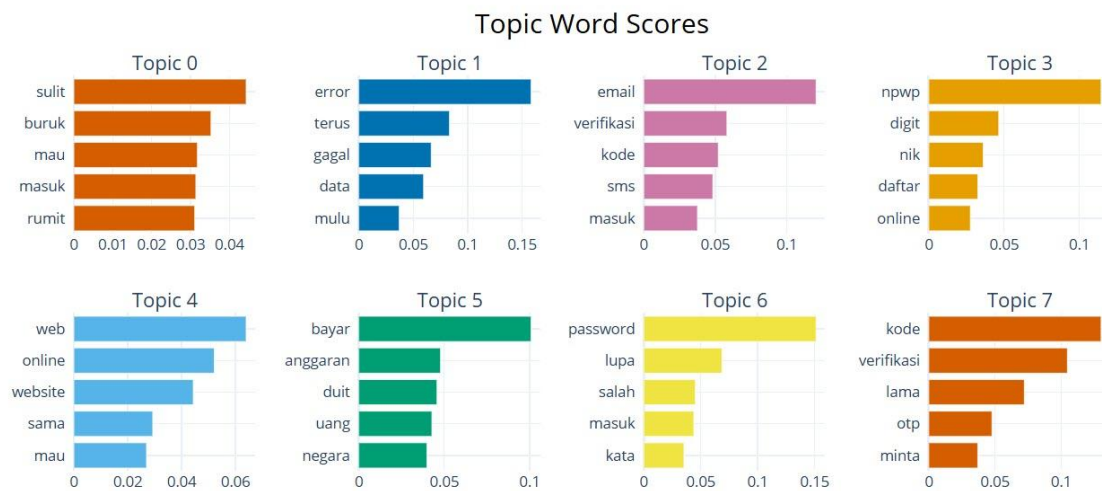
Gambar 3. Intertopic Distance Map Ulasan Negatif Aplikasi M-Pajak

Visualisasi ini juga menunjukkan beberapa topik yang relatif terpisah dari kluster utama. Kondisi tersebut mengindikasikan bahwa sebagian keluhan memiliki karakteristik yang lebih spesifik dan tidak sepenuhnya menyatu dengan kelompok masalah akses atau autentikasi. Topik-topik yang terpisah dapat merepresentasikan isu yang lebih spesifik, seperti pembayaran pajak, layanan website, performa sistem, atau masalah teknis tertentu yang dialami pengguna dalam konteks penggunaan yang berbeda. Keberadaan topik yang terpisah ini penting karena menunjukkan bahwa permasalahan aplikasi M-Pajak tidak bersifat tunggal, melainkan terdiri atas beberapa lapisan masalah: pertama, masalah akses awal seperti login dan autentikasi; kedua, masalah validasi dan verifikasi identitas; ketiga, masalah stabilitas sistem; dan keempat, masalah layanan spesifik seperti pembayaran atau integrasi dengan kanal digital lainnya. Dengan demikian, Intertopic Distance Map tidak hanya berfungsi sebagai visualisasi hasil topic modeling, tetapi juga membantu menunjukkan struktur permasalahan layanan digital yang dihadapi pengguna.

Dari perspektif *Business Intelligence*, temuan pada Gambar 3 menjadi dasar penting dalam menyusun prioritas perbaikan layanan. Kedekatan beberapa topik menunjukkan bahwa penyelesaian masalah sebaiknya tidak dilakukan secara terpisah per fitur, melainkan secara terintegrasi berdasarkan alur perjalanan pengguna. Misalnya, apabila topik login, OTP, verifikasi email, dan validasi NPWP/NIK berada dalam kelompok semantik yang berdekatan, maka perbaikan pada sistem login saja tidak cukup jika proses verifikasi dan validasi data masih bermasalah. Pengelola aplikasi perlu melihat kelompok masalah tersebut sebagai satu rangkaian proses layanan yang saling bergantung. Oleh karena itu, strategi perbaikan aplikasi M-Pajak perlu diarahkan pada penguatan keseluruhan mekanisme akses pengguna, termasuk optimasi autentikasi, peningkatan reliabilitas pengiriman OTP, penyederhanaan proses verifikasi, sinkronisasi data identitas, serta peningkatan stabilitas sistem pada periode penggunaan yang tinggi. Dengan membaca pola kedekatan dan keterpisahan topik pada visualisasi ini, pengelola layanan dapat mengidentifikasi area masalah yang bersifat sistemik dan area masalah yang bersifat spesifik, sehingga keputusan perbaikan dapat dilakukan secara lebih terarah, berbasis data, dan sesuai dengan pengalaman aktual pengguna.

Gambar 4 menampilkan Topic Word Scores yang menunjukkan kata-kata representatif dengan kontribusi tertinggi pada masing-masing topik hasil pemodelan BERTopic. Visualisasi ini penting karena membantu menjelaskan dasar pembentukan setiap topik, yaitu kata-kata yang paling kuat dalam merepresentasikan pola keluhan pengguna aplikasi M-Pajak. Secara umum, kata-kata dominan yang muncul pada beberapa topik menunjukkan bahwa permasalahan utama pengguna terkonsentrasi pada 4 area besar, yaitu akses layanan, kegagalan sistem, verifikasi identitas, dan integrasi data perpajakan. Kata-kata seperti “sulit”, “buruk”, “masuk”, dan “rumit” pada Topic 0 menunjukkan bahwa hambatan akses

merupakan keluhan yang paling menonjol. Hal ini mengindikasikan bahwa sebagian pengguna mengalami kesulitan sejak tahap awal penggunaan aplikasi, bahkan sebelum mereka dapat memanfaatkan fitur layanan perpajakan yang tersedia.



Gambar 4. Topic Word Scores Ulasan Negatif Aplikasi M-Pajak

Topic 1 memperlihatkan kata-kata dominan seperti “error”, “terus”, “gagal”, dan “data”. Kombinasi kata tersebut menunjukkan adanya keluhan yang berkaitan dengan kegagalan sistem yang berulang, bukan hanya gangguan sesaat. Kata “terus” dan “gagal” memperkuat indikasi bahwa sebagian pengguna mengalami masalah yang berulang saat mengakses atau memproses data di dalam aplikasi. Dari sisi kualitas layanan digital, pola ini menunjukkan bahwa stabilitas sistem dan keandalan proses aplikasi masih menjadi isu yang penting. Jika sistem sering mengalami error atau gagal memproses data, kepercayaan pengguna terhadap layanan perpajakan digital dapat menurun, terutama karena layanan pajak berkaitan dengan kewajiban administratif yang memiliki batas waktu tertentu.

Topic 2 dan Topic 7 menunjukkan konsentrasi keluhan pada proses verifikasi. Pada Topic 2, kata-kata seperti “email”, “verifikasi”, “kode”, dan “sms” menunjukkan bahwa pengguna menghadapi kendala dalam proses validasi akun melalui kanal komunikasi digital. Sementara itu, Topic 7 yang memuat kata “kode”, “verifikasi”, “lama”, dan “otp” menunjukkan bahwa keterlambatan atau kegagalan pengiriman kode OTP menjadi salah satu titik kritis dalam pengalaman pengguna. Temuan ini menunjukkan bahwa proses verifikasi identitas, yang seharusnya menjadi mekanisme keamanan, justru dapat menjadi hambatan akses ketika tidak berjalan dengan cepat dan andal. Dalam konteks aplikasi perpajakan digital, kegagalan OTP atau verifikasi email dapat menghambat pengguna untuk masuk ke sistem, memperbarui data, atau menyelesaikan proses administrasi pajak.

Topic 3 memperlihatkan kata-kata dominan seperti “npwp”, “digit”, “nik”, “daftar”, dan “online”. Kata-kata ini menunjukkan bahwa sebagian keluhan pengguna berkaitan dengan registrasi dan validasi identitas perpajakan. Munculnya kata “npwp” dan “nik” mengindikasikan bahwa integrasi data identitas masih menjadi isu penting dalam penggunaan aplikasi M-Pajak. Permasalahan pada tahap ini dapat berdampak serius karena NPWP dan NIK merupakan elemen utama dalam administrasi perpajakan digital. Jika pengguna mengalami kesulitan dalam proses pendaftaran, pencocokan data, atau validasi identitas, akses ke layanan berikutnya juga akan terganggu. Dengan demikian, masalah pada topik ini tidak hanya berkaitan dengan antarmuka aplikasi, tetapi juga dengan kesiapan integrasi data antara sistem perpajakan dan data identitas pengguna.

Topik 4 dan Topik 5 menunjukkan isu yang lebih spesifik. Topik 4 memuat kata “web”, “online”, dan “website”, yang menunjukkan bahwa sebagian pengguna mengaitkan pengalaman mereka dengan akses layanan melalui kanal digital lain atau keterhubungan aplikasi dengan layanan berbasis web. Hal ini mengindikasikan bahwa masalah layanan M-Pajak tidak hanya terbatas pada aplikasi mobile, tetapi juga dapat berkaitan dengan ekosistem layanan digital perpajakan secara lebih luas. Sementara itu, Topic 5 dengan kata-kata seperti “bayar”, “anggaran”, “duit”, “uang”, dan “negara” menunjukkan adanya keluhan atau persepsi pengguna yang berkaitan dengan pembayaran pajak serta dimensi kepercayaan terhadap pengelolaan layanan publik. Topik ini menunjukkan bahwa sebagian ulasan tidak hanya bersifat teknis, tetapi juga menyentuh aspek persepsi publik terhadap kewajiban pajak serta kualitas layanan pemerintah.

Topic 6 memperlihatkan kata-kata seperti “password”, “lupa”, “salah”, “masuk”, dan “kata”. Topik ini menunjukkan bahwa masalah autentikasi tidak hanya terjadi pada sistem login secara umum, tetapi juga pada proses pemulihan atau penggunaan kredensial pengguna. Keluhan terkait password menunjukkan adanya hambatan dalam proses masuk ke aplikasi, baik karena pengguna lupa kata sandi, kesalahan input, maupun kemungkinan alur pemulihan akun yang belum cukup mudah digunakan. Dalam layanan digital pemerintah, proses pemulihan akun perlu dirancang sederhana dan responsif karena pengguna memiliki tingkat literasi digital yang beragam.

Secara keseluruhan, visualisasi Topic Word Scores memperkuat temuan bahwa masalah utama aplikasi M-Pajak terletak pada tahap awal perjalanan pengguna, terutama pada akses, login, autentikasi, verifikasi, dan validasi identitas. Dari

perspektif Business Intelligence, temuan ini dapat digunakan sebagai dasar untuk menentukan prioritas perbaikan layanan. Topik-topik yang berkaitan dengan akses dan autentikasi perlu menjadi prioritas utama karena berfungsi sebagai pintu masuk ke seluruh layanan aplikasi. Jika pengguna gagal melewati tahap awal ini, fitur lain dalam aplikasi menjadi tidak dapat dimanfaatkan, terlepas dari kualitas fitur tersebut. Oleh karena itu, perbaikan aplikasi M-Pajak perlu difokuskan pada peningkatan stabilitas sistem login, percepatan pengiriman OTP, perbaikan proses verifikasi email dan SMS, penyederhanaan proses pemulihan kata sandi, serta penguatan integrasi data NPWP dan NIK. Dengan demikian, visualisasi ini tidak hanya menjelaskan kata-kata dominan dalam topik, tetapi juga memberikan arahan operasional mengenai area layanan yang paling mendesak untuk diperbaiki.

Jika dibandingkan dengan penelitian terdahulu, temuan penelitian ini memiliki karakteristik yang berbeda. Studi-studi sebelumnya yang menggunakan BERTopic maupun metode topic modeling lain umumnya berfokus pada media sosial, opini publik terhadap program pemerintah, ulasan e-commerce, forum pembelajaran daring, berita, atau tren literatur akademik. Pada konteks tersebut, hasil topic modeling lebih banyak digunakan untuk memetakan persepsi publik, mengidentifikasi isu diskursif, membandingkan performa metode, atau mengelompokkan tema umum dalam teks digital. Berbeda dari studi-studi tersebut, penelitian ini menggunakan ulasan negatif terhadap aplikasi M-Pajak sebagai sumber data utama untuk mengidentifikasi hambatan operasional dalam layanan perpajakan digital pemerintah. Dengan demikian, temuan penelitian ini tidak hanya menunjukkan topik yang sering dibicarakan pengguna, tetapi juga mengungkap titik-titik kegagalan layanan yang dialami secara langsung oleh pengguna, terutama pada tahap akses, autentikasi, verifikasi, validasi identitas, serta stabilitas sistem.

Kontribusi utama penelitian ini terletak pada pemanfaatan hasil BERTopic sebagai dasar interpretasi Business Intelligence dalam evaluasi layanan publik digital. Berbeda dari penelitian yang hanya berhenti pada klasifikasi sentimen atau pemetaan topik, penelitian ini menghubungkan setiap topik keluhan dengan implikasi operasional serta prioritas perbaikan layanan. Misalnya, dominasi topik “Kesulitan Login dan Akses” menunjukkan bahwa autentikasi digital merupakan titik kritis dalam perjalanan pengguna, sedangkan topik OTP, verifikasi email, NPWP, dan NIK menunjukkan bahwa masalah akses tidak berdiri sendiri, melainkan berkaitan dengan proses validasi identitas dan integrasi data. Dengan demikian, novelty penelitian ini terletak pada kombinasi antara App Store Review Mining, BERTopic, dan interpretasi Business Intelligence dalam konteks aplikasi perpajakan digital pemerintah. Pendekatan ini memperluas penggunaan topic modeling dari sekadar identifikasi tema menjadi alat pendukung pengambilan keputusan untuk meningkatkan kualitas, stabilitas, dan responsivitas layanan publik digital.

4. KESIMPULAN

Penelitian ini menerapkan pendekatan App Store Review Mining menggunakan BERTopic untuk menganalisis ulasan negatif pengguna aplikasi M-Pajak sebagai sumber Business Intelligence dalam mengevaluasi layanan perpajakan digital pemerintah. Hasil analisis menunjukkan bahwa permasalahan utama pengguna berkaitan dengan kesulitan login dan autentikasi, error sistem, verifikasi email dan OTP, registrasi NPWP dan NIK, serta stabilitas layanan digital. Topik “Kesulitan Login dan Akses” menjadi topik paling dominan, yang menunjukkan bahwa autentikasi digital masih menjadi salah satu hambatan utama dalam penggunaan aplikasi M-Pajak. Hasil evaluasi model memperoleh nilai Topic Coherence sebesar 0,522, Topic Diversity sebesar 0,667, dan Topic Quality sebesar 0,348, yang menunjukkan bahwa topik yang dihasilkan berada pada tingkat kualitas moderat, tetapi masih dapat diinterpretasikan untuk mengidentifikasi pola keluhan pengguna. Dengan demikian, BERTopic dalam penelitian ini diposisikan sebagai pendekatan eksploratif untuk membantu menemukan tema-tema utama dalam ulasan negatif, bukan sebagai metode yang diklaim paling unggul dibandingkan pendekatan topic modeling lainnya. Dari perspektif Business Intelligence, temuan penelitian ini menunjukkan bahwa ulasan pengguna dapat dimanfaatkan sebagai sumber insight operasional untuk mengidentifikasi kelemahan layanan digital pemerintah berdasarkan pengalaman pengguna aktual. Berdasarkan hasil analisis, peningkatan kualitas aplikasi M-Pajak perlu diprioritaskan pada perbaikan mekanisme autentikasi digital, stabilitas sistem, proses verifikasi, integrasi data NPWP dan NIK, serta layanan bantuan pengguna agar layanan perpajakan digital menjadi lebih responsif, stabil, dan berorientasi pada kebutuhan masyarakat. Namun, karena penelitian ini belum melakukan pembandingan empiris dengan metode lain seperti LDA atau NMF, penelitian selanjutnya disarankan untuk melakukan evaluasi lintas metode agar kualitas model dapat dinilai secara lebih komprehensif.

REFERENCES

- [1] R. Rachmiani, N. K. Oktadinna, and T. R. Fauzan, “The impact of online reviews and ratings on consumer purchasing decisions on e-commerce platforms,” *International Journal of Management Science and Information Technology*, vol. 4, no. 2, pp. 504–515, 2024, doi: <https://doi.org/10.35870/ijmsit.v4i2.3373>
- [2] A. I. Alfanzar, Khalid, and I. S. Rozas, “Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation,” *JSiI (Jurnal Sistem Informasi)*, vol. 7, no. 1, pp. 7–13, 2020, doi: <https://doi.org/10.30656/jsii.v7i1.2036>
- [3] M. Jelita, “Text Mining dengan Topic Modelling LDA dari Pertanyaan Gelar Wicara Literasi Perpustakaan Nasional RI,” *Media Pustakawan*, vol. 31, no. 3, pp. 253–265, 2024, doi: <https://doi.org/10.37014/medpus.v31i3.5237>

- [4] A. Syaifuddin, R. A. Harianto, and J. Santoso, "Analisis trending topik untuk percakapan media sosial dengan menggunakan topic modelling berbasis algoritme LDA," *Journal of Intelligent System and Computation*, vol. 2, no. 1, pp. 12–19, 2021, doi: <https://doi.org/10.52985/insyst.v2i1.150>
- [5] L. Ma, R. Chen, W. Ge, P. Rogers, B. Lyn-Cook, H. Hong, W. Tong, N. Wu, and W. Zou, "AI-powered topic modeling: Comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women," *Experimental Biology and Medicine* (Maywood), vol. 250, Art. no. 10389, Feb. 2025, doi: 10.3389/ebm.2025.10389.
- [6] N. Asyiah and W. Aktavia, "Penerapan BERTopic dan Analisis Sentimen Leksikal pada Ulasan Relevan di Google Maps Mengenai Universitas Pamulang," *Journal of Information System Research (JOSH)*, vol. 6, no. 4, pp. 2129–2138, Jul. 2025, doi: 10.47065/josh.v6i4.8007.
- [7] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Frontiers in Sociology*, vol. 7, Art. no. 886498, May 2022, doi: 10.3389/fsoc.2022.886498.
- [8] R. Karthick, "Context-aware topic modeling and intelligent text extraction using transformer-based architectures," *Journal of Science Technology and Research (JSTAR)*, vol. 6, no. 1, pp. 1–12, 2025, doi: <https://doi.org/10.2139/ssrn.5275391>
- [9] O. Babalola, B. Ojokoh, and O. Boyinbode, "Comprehensive evaluation of LDA, NMF, and BERTopic's performance on news headline topic modeling," *Journal of Computing Theories and Applications*, vol. 2, no. 2, pp. 268–289, 2024. [Online]. Available: publikasi.dinus.ac.id/index.php/jcta/article/view/11635
- [10] D. Rifaldi, A. Fadlil, and Herman, "Teknik preprocessing pada text mining menggunakan data tweet 'Mental Health'," *DECODE: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, 2023. doi: <https://doi.org/10.51454/decode.v3i2.131>
- [11] M. Rizal and C. Suhaeni, "Topic modeling of public discourse on electric vehicles in Indonesia using BERTopic," *COMPETENTIE Journal*, vol. 2, no. 2, 2025, doi: 10.65049/b44ktr94.
- [12] N. Khodeir and F. Elghannam, "Efficient topic identification for urgent MOOC Forum posts using BERTopic and traditional topic modeling techniques," *Educ. Inf. Technol.*, vol. 30, no. 5, pp. 5501–5527, Apr. 2025, doi: 10.1007/s10639-024-13003-4.
- [13] K. Alifviansyah, A. Saefuddin, and S. Rahardianto, "BERTopic-based multi-class topic classification on Indonesian Shopee e-commerce reviews using ensemble learning," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 11, no. 1, pp. 181–193, May 2026, doi: 10.18860/cauchy.v11i1.37941.
- [14] B. Warsito, J. E. Suseno, and A. Arifudin, "Embedding and topic modeling techniques for short text analysis on social media: A systematic literature review," *Data and Metadata*, vol. 4, Art. no. 1168, 2025, doi: 10.56294/dm20251168.
- [15] N. Hayati, Saikin, and H. Fahmi, "Comparison of LDA and BERTopic in identifying public issues in the MBG program," *JISA (Jurnal Informatika dan Sains)*, vol. 9, no. 1, Jun. 2026. [Online]. Available: <https://ojs.trilogi.ac.id/index.php/JISA/article/view/2782>
- [16] A. Abuzayed and H. Al-Khalifa, "BERT for Arabic topic modeling: An experimental study on BERTopic technique," *Procedia Computer Science*, vol. 189, pp. 191–194, 2021, doi: 10.1016/j.procs.2021.05.096.
- [17] R. Egger and J. Yu, "A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts," *Frontiers in Sociology*, vol. 7, Art. no. 886498, May 2022, doi: 10.3389/fsoc.2022.886498.
- [18] A. F. Anugrah and R. D. Agatha, "Utilizing IndoBERT and BERTopic to explore public opinion on BPS Instagram posts," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, Oct. 2025, doi: 10.30871/jaic.v9i5.10327.
- [19] W. Wahyuni, T. P. Lestari, M. Apriliana, and R. Gumelta, "Implementation of BERTopic for topic modeling analysis of the Free Nutritious Meal Program based on YouTube comments," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, Aug. 2025, doi: 10.30871/jaic.v9i4.9754
- [20] A. Gaviota and F. Ardiani, "Pemodelan Topik Menggunakan BERTopic Pada Nama Produk Template Desain di Peterdraw Studio," *CESS (Journal of Computer Engineering, System and Science)*, vol. 11, no. 1, pp. 27–40, Jan. 2026, doi: 10.24114/cess.v11i1.70335.
- [21] C. Y. Sy, L. L. Maceda, N. M. Flores, and M. B. Abisado, "Unsupervised machine learning approaches in NLP: A comparative study of topic modeling with BERTopic and LDA," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 21s, pp. 3276–3283, 2024. [Online]. Available: <https://ijisae.org/index.php/IJISAE/article/view/6018>
- [22] K. Zhu, J. Weng, and B. Tu, "A comparative study on short-text topic modeling in online collaborative learning platforms: Application and performance analysis of BTM and BERTopic," in *Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education (ICAIE '24)*, 2024, pp. 415–421, doi: 10.1145/3722237.3722309.
- [23] B. R. N. Listyawan, N. Y. Setiawan, and M. C. Saputra, "Topic modelling pada aktivitas pengembangan perangkat lunak menggunakan BERTopic," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 8, Aug. 2024. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14051>
- [24] A. Cheddak, T. Ait Baha, Y. Es-Saady, M. El Hajji, and M. Baslam, "BERTopic for enhanced idea management and topic generation in brainstorming sessions," *Information*, vol. 15, no. 6, Art. no. 365, Jun. 2024, doi: 10.3390/info15060365
- [25] C. Li and X. Hu, "Medical artificial intelligence in scholarly and public perspective: BERTopic-based analysis of topic-sentiment collaborative mining," *Data Science and Informetrics*, vol. 5, no. 1, pp. 33–42, Mar. 2025, doi: 10.1016/j.dsim.2025.05.001
- [26] C. B. Pavithra and J. Savitha, "Topic modeling for evolving textual data using LDA, HDP, NMF, BERTopic, and DTM with a focus on research papers," *Journal of Technology and Informatics*, vol. 5, no. 2, 2024, doi: 10.37802/joti.v5i2.618
- [27] A.-R. Yang, J. Chae, and E. Choi, "Analysis of peatland research trends based on BERTopic," *Land*, vol. 13, no. 5, Art. no. 628, May 2024, doi: 10.3390/land13050628.
- [28] W. Martin, F. Sarro, Y. Jia, Y. Zhang, and M. Harman, "A survey of app store analysis for software engineering," *IEEE Transactions on Software Engineering*, vol. 43, no. 9, pp. 817–847, 2017, doi: <https://doi.org/10.1109/TSE.2016.2630689>.
- [29] S. Terragni, E. Fersini, B. G. Galuzzi, P. Tropeano, and A. Candelieri, "OCTIS: Comparing and optimizing topic models is simple!," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, Online, pp. 263–270, 2021, doi: <https://doi.org/10.18653/v1/2021.eacl-demos.31>



- [30] A. Upadhye, “Enhancing semantic understanding by visualizing sentence-level embeddings,” *International Journal of Computer Applications*, vol. 185, no. 46, pp. 20–24, 2023, [Online]. Available: <https://www.ijcaonline.org/archives/volume185/number46/33000-2023923275/>
- [31] D. Vijayan and I. Aziz, “Adaptive hierarchical density-based spatial clustering algorithm for streaming applications,” *Telecom*, vol. 4, no. 1, pp. 1–14, 2023, doi: <https://doi.org/10.3390/telecom4010001>.
- [32] D. N. Amalina and A. Fauzan, “A Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) Approach for Identifying Potential Villages in Buleleng Regency,” *Knowledge Engineering and Data Science*, vol. 7, no. 2, pp. 187–199, 2024, doi: <https://doi.org/10.17977/um018v7i22024p187-199>
- [33] D. Karabašević, A. Vujko, V. Mirčetić, G. Popović, and D. Stanujkić, “A transformer-based semantic encoding framework for quantitative analysis of large-scale textual reviews,” *Axioms*, vol. 15, no. 3, p. 175, 2026, doi: <https://doi.org/10.3390/axioms15030175>
- [34] M. Röder, A. Both, and A. Hinneburg, “Exploring the space of topic coherence measures,” in *Proceedings of the 8th ACM International Conference on Web Search and Data Mining (WSDM '15), Shanghai, China*, pp. 399–408, 2015, doi: <https://doi.org/10.1145/2684822.2685324>
- [35] I. K. Mensah, “Impact of government capacity and e-government performance on the adoption of e-government services,” *International Journal of Public Administration*, vol. 43, no. 4, pp. 303–311, 2020, doi: <https://doi.org/10.1080/01900692.2019.1628059>
- [36] Rijal, Jusman, Nasrullah, and Sukmawati, “Efektivitas implementasi aplikasi layanan publik berbasis digital (Makassar Smart City) dalam meningkatkan kepuasan masyarakat,” *Moderat: Jurnal Ilmiah Ilmu Pemerintahan*, vol. 11, no. 4, pp. 1352–1375, 2025, doi: <https://doi.org/10.25157/moderat.v11i4.5223>