



Penerapan BERTopic dan Analisis Sentimen Leksikal Pada Ulasan Relevan di Google Maps Mengenai Universitas Pamulang

Nilovar Asyiah¹, Widodo Aktavia^{2,*}

¹Fakultas Ilmu Komputer, Prodi Teknik Informatika, Universitas Pamulang, Tangerang Selatan
Jl. Raya Puspittek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten, Indonesia

²Program Pascasarjana, Prodi Teknik Informatika S-2, Universitas Pamulang, Tangerang Selatan
Jl. Raya Puspittek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten, Indonesia

Email: ¹dosen02835@unpam.ac.id, ^{2,*}widodoaktavia95@gmail.com

Email Penulis Korespondensi: widodoaktavia95@gmail.com

Submitted: 14/07/2025; Accepted: 30/07/2025; Published: 31/07/2025

Abstrak—Perkembangan teknologi informasi telah mendorong masyarakat untuk secara aktif memberikan ulasan secara daring melalui berbagai platform digital, salah satunya Google Maps. Ulasan ini tidak hanya bersifat informatif, tetapi juga merepresentasikan opini dan pengalaman nyata pengguna terhadap suatu tempat atau institusi, termasuk perguruan tinggi. Penelitian ini bertujuan untuk menganalisis topik-topik utama serta klasifikasi sentimen dalam ulasan Google Maps yang ditujukan kepada Universitas Pamulang. Pendekatan yang digunakan mencakup dua metode utama. Pertama, pemodelan topik menggunakan BERTopic, yaitu teknik modern berbasis transformer embedding dan algoritma clustering HDBSCAN, yang mampu menangkap konteks semantik dalam teks secara mendalam. Kedua, analisis sentimen dilakukan dengan pendekatan lexicon-based menggunakan kamus sentimen Bahasa Indonesia, yang memungkinkan identifikasi polaritas opini pengguna secara efisien tanpa proses pelatihan model. Data dalam penelitian ini diperoleh melalui proses web scraping dari ulasan publik Google Maps yang relevan pada empat lokasi Universitas Pamulang: Kampus Pusat, Kampus Viktor, Kampus Witanaharja, dan Unpam Serang. Hasil analisis menunjukkan beberapa topik dominan, seperti pelayanan akademik, fasilitas kampus, dan birokrasi. Sentimen yang ditemukan mayoritas bersifat netral hingga positif, namun tetap terdapat ulasan negatif pada aspek-aspek tertentu. Temuan ini diharapkan memberikan kontribusi strategis bagi institusi dalam meningkatkan mutu layanan dan citra digital.

Kata Kunci: BERTopic; Analisis Sentimen; Google Maps; Universitas Pamulang; Pemodelan Topik; Ulasan Daring

Abstract—The rapid advancement of information technology has encouraged the public to actively share reviews through digital platforms such as Google Maps. These reviews are not only informative but also reflect real user opinions and experiences regarding places or institutions, including higher education institutions. This study aims to analyze the main topics and sentiment classification contained in Google Maps reviews related to Universitas Pamulang. The approach used in this research combines two main methods. First, topic modeling is conducted using BERTopic, a modern technique based on transformer embeddings and HDBSCAN clustering algorithms, which can capture the semantic context of text more deeply. Second, sentiment analysis is performed using a lexicon-based approach, applying an Indonesian sentiment lexicon to efficiently identify the polarity of opinions without requiring model training. The data analyzed were collected through web scraping of relevant public reviews on Google Maps across four Universitas Pamulang locations: Central Campus, Viktor Campus, Witanaharja Campus, and Unpam Serang. The analysis revealed several dominant topics such as academic services, campus facilities, and bureaucracy. The majority of sentiments identified were neutral to positive, although negative opinions were also found in certain aspects. These findings are expected to serve as strategic input for the university to enhance service quality and strengthen its institutional image in the digital landscape.

Keywords: BERTopic; Sentiment Analysis; Google Maps; Universitas Pamulang; Topic Modeling; Online Reviews.

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah membawa dampak besar terhadap berbagai aspek kehidupan manusia, termasuk cara individu menyampaikan opini, pengalaman, dan kritik terhadap suatu layanan atau institusi. Salah satu fenomena yang mencerminkan transformasi digital tersebut adalah kemunculan berbagai platform ulasan daring (online reviews) yang memungkinkan masyarakat secara terbuka menyuarakan pandangannya melalui rating serta komentar. Platform-platform ini telah menjadi saluran komunikasi informal yang efektif untuk menjangkau publik luas, salah satunya adalah Google Maps. Google Maps yang awalnya hanya dikenal sebagai layanan navigasi, kini telah berevolusi menjadi ruang sosial tempat pengguna dapat meninggalkan ulasan terkait pengalaman mereka terhadap tempat-tempat yang dikunjungi, termasuk rumah makan, tempat wisata, fasilitas umum, hingga institusi pendidikan tinggi seperti universitas. Ulasan tersebut tidak hanya bersifat informatif, tetapi juga sering kali mencerminkan persepsi jujur masyarakat terhadap kualitas layanan yang diterima.

Sifat ulasan di Google Maps yang terbuka dan dapat diakses publik menjadikannya sebagai sumber data alternatif yang sangat kaya, tidak seperti survei konvensional yang memerlukan biaya, waktu, dan proses administrasi yang panjang, ulasan daring bersifat spontan, real-time, dan dapat mencerminkan emosi atau kepuasan pengguna secara langsung. Dalam konteks pendidikan tinggi, data ini dapat digunakan untuk memahami pandangan mahasiswa dan masyarakat terhadap pelayanan akademik, fasilitas kampus, interaksi dengan staf, maupun aspek administratif lainnya [1]. Selain itu, studi lintas sektor menunjukkan bahwa ulasan Google Maps sebagai bagian dari user-generated content (UGC) semakin banyak dimanfaatkan dalam evaluasi layanan institusi



publik dan pendidikan. Contohnya, penelitian oleh Khan et al. (2024) menggunakan analisis ulasan Google Maps untuk menilai kualitas layanan perpustakaan umum di Taiwan, dan berhasil mengidentifikasi faktor pelayanan dan lingkungan yang mempengaruhi kepuasan pengguna [2]. Di bidang pariwisata dan perhotelan, Kitsios et al. (2022) menunjukkan melalui tinjauan bibliometrik bahwa UGC dari platform seperti Google Maps dan TripAdvisor sangat berpengaruh dalam pengambilan keputusan dan persepsi layanan [3]. Sebagai tambahan, studi terbaru oleh [4] yang berjudul "Public attitudes toward higher education using sentiment analysis and topic modeling" menerapkan teknik analisis sentimen dan pemodelan topik terhadap data media sosial untuk mengukur sikap publik terhadap pendidikan tinggi. Penelitian ini menegaskan bahwa analisis data digital dapat memberikan wawasan mendalam mengenai persepsi dan kekhawatiran masyarakat terhadap institusi pendidikan, sehingga memperkuat argumen bahwa platform daring seperti Google Maps merupakan sumber data alternatif yang valid dan kaya untuk evaluasi layanan pendidikan tinggi secara real-time. Hal ini memperkuat argumen bahwa ulasan daring dapat menjadi data alternatif yang valid dan kaya untuk memetakan persepsi publik secara real-time, termasuk pada institusi pendidikan tinggi seperti Universitas Pamulang.

Universitas Pamulang sebagai salah satu institusi pendidikan tinggi swasta terbesar di Indonesia memiliki banyak mahasiswa dan tersebar di beberapa lokasi. Penelitian ini secara khusus mencakup ulasan publik yang diberikan pada empat lokasi utama kampus Universitas Pamulang, yaitu Kampus Pusat (Jl. Surya Kencana, Pamulang), Kampus Viktor (Jl. Raya Puspitek, Buaran), Kampus Witanaharja (Jl. Witanaharja Raya), dan Unpam Serang (Jl. Raya Serang-Pandeglang). Keempat lokasi ini memiliki karakteristik pelayanan dan persepsi publik yang bisa berbeda, sehingga penting untuk dianalisis secara menyeluruh guna mendapatkan gambaran umum dan perbedaan antar lokasi.

Tantangan utama dari data ulasan yang tersedia di Google Maps adalah sifatnya yang tidak terstruktur (unstructured text data). Ulasan biasanya ditulis dalam bahasa sehari-hari yang bebas, tanpa format baku, dan cenderung bervariasi dalam panjang, gaya, dan isi. Oleh karena itu, pendekatan analisis konvensional tidak cukup efektif untuk menggali informasi berharga dari jenis data ini. Diperlukan teknik analisis teks (text mining) yang mampu memproses, memahami, dan mengekstraksi makna dari kumpulan teks dalam jumlah besar [5]. Tantangan signifikan dalam menganalisis ulasan daring terletak pada keberadaan noise, penggunaan slang, struktur kalimat tidak baku, serta campuran bahasa atau emotikon. Wu et al. (2024) menunjukkan bahwa BERT secara substansial meningkatkan akurasi analisis sentimen pada teks informal dengan memahami konteks semantik tersembunyi [6].

Salah satu teknik text mining yang cukup populer adalah topic modeling. Teknik ini bertujuan untuk menemukan topik-topik tersembunyi secara otomatis dalam kumpulan dokumen teks. Dalam penelitian ini, digunakan metode BERTopic sebagai alat pemodelan topik. BERTopic merupakan pendekatan modern berbasis transformer embeddings dan algoritma klusterisasi HDBSCAN yang memungkinkan identifikasi topik secara kontekstual. Berbeda dengan metode klasik seperti LDA yang hanya mengandalkan distribusi kata, BERTopic mampu memahami konteks kalimat secara lebih dalam karena menggunakan representasi vektor dari model seperti BERT. Hal ini membuat BERTopic sangat cocok untuk menganalisis teks informal dalam bahasa alami, termasuk ulasan pengguna di platform digital [7]. Efektivitas BERTopic dalam mengidentifikasi topik secara efisien juga telah dibuktikan dalam penelitian Khodeir dan Elghannam [8] yang menggunakannya untuk mendeteksi topik penting pada forum diskusi MOOC secara cepat dan akurat. Studi ini menekankan keunggulan BERTopic dalam konteks teks informal dan mendesak, yang sejalan dengan karakteristik ulasan pengguna di platform seperti Google Maps. Sebagai tambahan, penelitian oleh Samsir et al. [9] menegaskan keunggulan metode ini dengan menyoroti bagaimana integrasi BERTopic menggunakan transformer embeddings serta algoritma klusterisasi HDBSCAN tidak hanya meningkatkan akurasi identifikasi topik, tetapi juga memperkaya visualisasi hasil kluster dan memfasilitasi eksplorasi topik yang lebih interaktif.

Penelitian tesis oleh Yaolan Jin (2025) [10] menerapkan BERTopic dengan pendekatan embedding berbasis transformer, dilanjutkan dengan pengurangan dimensi menggunakan UMAP dan pengelompokan data menggunakan HDBSCAN untuk mengkaji dokumen teks. Hasil studi menunjukkan kemampuan BERTopic dalam mengelompokkan dokumen ke dalam topik-topik yang koheren dan bermakna, khususnya untuk teks yang panjang dan terstruktur. Temuan ini menguatkan alasan penggunaan BERTopic sebagai metode analisis topik dalam penelitian ini, terutama dalam konteks rekomendasi berbasis minat pengguna.

Selain memahami topik yang dibahas dalam ulasan, penting juga untuk mengetahui bagaimana sentimen atau sikap penulis ulasan terhadap topik tersebut. Oleh karena itu, penelitian ini juga menggunakan pendekatan lexicon-based sentiment analysis. Pendekatan ini bekerja dengan menggunakan kamus sentimen (sentiment lexicon) yang berisi daftar kata yang diklasifikasikan sebagai positif, negatif, atau netral untuk membantu mengklasifikasikan sentimen teks [11].

Pendekatan lexicon-based terletak pada interpretasinya yang transparan dan proses klasifikasinya yang cepat tanpa perlu pelatihan model yang kompleks. Kamus sentimen berbahasa Indonesia yang digunakan dalam penelitian ini telah divalidasi dalam berbagai studi sebelumnya, sehingga mampu memberikan hasil yang cukup akurat untuk mengkategorikan opini dalam ulasan menjadi sentimen positif, negatif, atau netral [12]. Studi sistematis oleh Fauziah et al. [13] mengidentifikasi bahwa pendekatan lexicon-based merupakan salah satu metode yang paling sering digunakan dalam analisis sentimen berbahasa Indonesia karena efisiensi dan kemudahan penerapannya, meskipun tetap memiliki tantangan dalam penyesuaian domain dan konteks bahasa informal.

Penelitian ini menggabungkan dua pendekatan utama, yaitu BERTopic dan lexicon-based sentiment analysis, untuk memberikan pemahaman komprehensif terhadap persepsi publik mengenai Universitas Pamulang. Analisis dilakukan untuk mengidentifikasi topik-topik utama yang sering dibahas oleh pengguna serta mengevaluasi persebaran sentimen terhadap masing-masing topik. Dengan mengkaji ulasan dari keempat lokasi kampus, hasil penelitian ini diharapkan dapat memberikan masukan strategis berbasis data bagi pihak pengelola universitas dalam meningkatkan mutu layanan, merespons keluhan dan apresiasi publik secara efektif, serta membangun citra positif institusi di ruang digital [14].

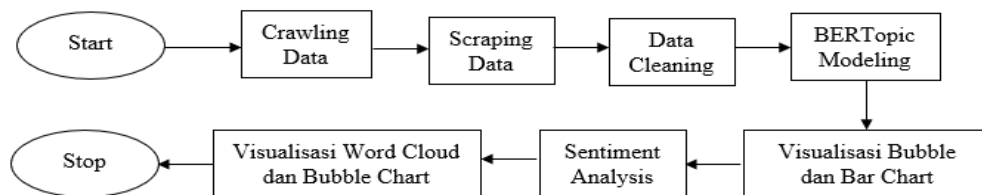
Meskipun sudah banyak studi mengenai analisis sentimen dan topic modeling terhadap ulasan daring, gap penelitian ini terletak pada masih minimnya riset yang secara spesifik mengkaji ulasan institusi pendidikan tinggi berbasis Google Maps dengan mengombinasikan metode BERTopic dan pendekatan leksikal dalam analisis sentimen. Sebagian besar studi sebelumnya hanya menggunakan pendekatan tradisional seperti Latent Dirichlet Allocation (LDA) atau naïve Bayes, dan umumnya berfokus pada sektor komersial, layanan kesehatan, atau transportasi. Beberapa penelitian terkait menunjukkan perkembangan metode analisis ulasan daring. Misalnya, Akbar et al. [1] menggunakan metode leksikal untuk analisis sentimen pada sektor pariwisata berbasis Google Maps, sementara Yahya et al. [5] menerapkan Long Short-Term Memory (LSTM) untuk analisis sentimen rumah sakit. Fatani [14] mengombinasikan pendekatan IndoBERT, Inset lexicon, dan BERTopic untuk mengkaji layanan LRT Jabodebek, yang menunjukkan efektivitas integrasi teknik deep learning dan topic modeling. Di sisi lain, Samsir et al. [15] memanfaatkan BERTopic untuk menganalisis abstrak penelitian NLP, namun tidak mengaitkannya dengan analisis sentimen publik secara langsung. Pada studi terkini yang menggunakan data Twitter terkait kebijakan subsidi bahan bakar di Indonesia, pendekatan lexicon-based sentiment analysis diaplikasikan sebagai pelengkap metode pembelajaran mesin seperti Support Vector Machine (SVM). Meski pendekatan leksikal memiliki keterbatasan karena kamus sentimen yang bersifat statis dan bergantung pada daftar kata yang telah diklasifikasikan, metode ini terbukti efektif dalam memberikan gambaran sentimen publik secara jelas dan akurat [16].

Penelitian ini memperluas pendekatan tersebut dengan menargetkan domain pendidikan tinggi berbasis ulasan publik digital, dan mengombinasikan pemetaan topik tematik dengan persepsi sentimen yang muncul. Selain itu, metode yang digunakan bersifat replikatif dan adaptif, sehingga dapat diterapkan pada institusi lain yang ingin memanfaatkan ulasan daring untuk data-driven evaluation dan peningkatan mutu layanan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan deskriptif kuantitatif berbasis text mining untuk menganalisis ulasan masyarakat terhadap Universitas Pamulang yang diambil dari platform Google Maps dengan mengambil data ulasan relevan. Berikut adalah prosesnya:



Gambar 1. Proses Penelitian

Penjelasan dari Gambar 1 adalah:

a. Crawling Data

Tahapan ini melibatkan proses web crawling, yaitu menelusuri dan mengakses halaman Google Maps yang memuat pengumpulan konten yang menyebutkan atau membahas Universitas Pamulang berfungsi sebagai pengambil data mentah dalam jumlah besar menggunakan script python menggunakan library Selenium dan BeautifulSoup dengan waktu 300 detik. Data yang di ambil merupakan empat kampus universitas pamulang yakni :

- Universitas Pamulang Kampus Pusat
Data yang di peroleh adalah 422 baris dalam format .csv
- Universitas Pamulang Kampus Viktor
Data yang di peroleh adalah 400 baris dalam format .csv
- Universitas Pamulang Kampus Witana Harja
Data yang di peroleh adalah 276 baris dalam format .csv
- Universitas Pamulang Kampus Serang (Unpam PSDKU)
Data yang di peroleh adalah 272 baris dalam format .csv



b. Scraping Data

Setelah data ditemukan, proses web scraping digunakan untuk mengekstrak isi ulasan relevan, elemen penting seperti isi berupa reviewer, rating, dan teks tentang ulasan relevan google maps yang berkaitan langsung dengan Universitas Pamulang.

c. Data Cleaning (Preprocessing Data)

Tahap data cleaning dilakukan setelah data ulasan relevan berhasil dikumpulkan dalam bentuk .csv dengan tahapan nya sebagai berikut:

- a) Penggabungan data dari 4 dataset
- b) Membersihkan data kosong, karakter pendek dan karakter terlalu panjang
- c) Normalisasi kata seluruh teks diubah menjadi huruf kecil untuk menjaga konsistensi
- d) Proses pembersihan ini meliputi penghapusan karakter atau simbol khusus, emoji, angka yang tidak relevan, tanda baca berlebihan, serta spasi ganda.

d. BERTopic Modeling

Data yang telah bersih dianalisis menggunakan metode BERTopic untuk menemukan topik-topik utama yang sering muncul dalam ulasan. Proses ini mencakup embedding teks, reduksi dimensi, dan clustering.

e. Visualisasi Bubble Chart dan Bar Chart

Hasil pemodelan topik divisualisasikan menggunakan diagram bubble dan batang untuk menampilkan persebaran dan kekuatan topik berdasarkan frekuensi dan kedekatan semantik.

f. Sentiment Analysis

Ulasan juga dianalisis menggunakan pendekatan lexicon-based sentiment analysis untuk mengidentifikasi sentimen positif, negatif, atau netral dalam masing-masing ulasan.

g. Visualisasi Word Cloud dan Bubble Chart

Selain visualisasi topik, word cloud digunakan untuk memperlihatkan kata-kata yang paling sering muncul, serta bubble chart tambahan digunakan untuk menyajikan persebaran sentimen secara visual.

2.2 Proses BERTopic

Bidirectional Encoder Representations from Transformers for Topic Modeling adalah metode topic modeling berbasis transformer yang mampu mengidentifikasi topik-topik dominan dalam kumpulan teks secara lebih kontekstual. BERTopic menggabungkan kekuatan representasi semantik dari transformer (khususnya Sentence-BERT), teknik reduksi dimensi (UMAP), dan algoritma klasterisasi berbasis densitas (HDBSCAN), lalu menghasilkan representasi topik dengan metode class-based TF-IDF (c-TF-IDF) [17]. Pendekatan ini juga telah berhasil diterapkan oleh Samsir et al. [18] dalam mengelompokkan abstrak penelitian NLP, menunjukkan bahwa kombinasi Sentence-BERT, UMAP, dan HDBSCAN mampu menghasilkan kluster topik yang bermakna dan informatif dalam konteks teks ilmiah. Studi oleh Babalola et al. [19] menunjukkan bahwa BERTopic menghasilkan topik yang lebih koheren dan representatif dibandingkan metode klasik seperti LDA dan NMF, terutama saat diterapkan pada teks pendek seperti judul berita, menjadikannya cocok pula untuk analisis ulasan pengguna berbasis kalimat pendek seperti di Google Maps.

Setelah proses crawling, scraping, dan pembersihan data dilakukan untuk memperoleh ulasan-ulasan yang relevan tentang Universitas Pamulang dari Google Maps, langkah selanjutnya adalah penerapan algoritma BERTopic. Algoritma ini diawali dengan mengubah setiap ulasan teks menjadi representasi numerik atau embedding menggunakan model bahasa berbasis transformer seperti BERT atau Sentence-BERT [20]. Representasi ini memungkinkan pemahaman semantik yang lebih dalam dari masing-masing ulasan. Embedding yang telah diperoleh kemudian dikelompokkan menjadi beberapa topik menggunakan algoritma clustering HDBSCAN, yang mampu membentuk cluster berdasarkan kepadatan data.

Langkah penting berikutnya adalah melakukan ekstraksi kata kunci representatif dari masing-masing topik menggunakan pendekatan class-based Term Frequency-Inverse Document Frequency atau **c-TF-IDF**. Dalam tahap ini, seluruh dokumen dalam satu cluster digabung menjadi satu dokumen besar, lalu dilakukan perhitungan c-TF-IDF untuk mengukur pentingnya kata tertentu dalam sebuah topik dibandingkan dengan seluruh topik lainnya dalam korpus. Dengan Rumus c-TF-IDF:

$$c - TF - IDF_{t,c} = \frac{ft,c}{\sum_{t'} ft',c} \cdot \log \left(\frac{N}{n_t} \right) \quad (1)$$

Perhitungan representasi kata untuk setiap topik dilakukan dengan mempertimbangkan beberapa parameter, yaitu ft,c yang merupakan frekuensi kemunculan kata t dalam kumpulan dokumen pada kelas atau topik c , N sebagai jumlah total dokumen dalam seluruh korpus, dan n_t yang menyatakan jumlah kelas yang mengandung kata t . Melalui proses ini, diperoleh kata-kata yang paling representatif untuk masing-masing topik, yang kemudian digunakan sebagai label deskriptif yang menggambarkan karakteristik utama dari topik tersebut [21].

2.3 Proses Analisis Sentimen Leksikal

Analisis sentimen berbasis leksikal dilakukan dengan membandingkan kata-kata dalam teks ulasan dengan daftar kata yang telah diberi label sebagai positif atau negatif. Setiap ulasan dianalisis dengan menghitung jumlah kata positif (P) dan jumlah kata negatif (N), kemudian dihitung skor sentimen (S) menggunakan rumus sederhana $S =$



P – N. Hasil dari skor ini digunakan untuk mengklasifikasikan sentimen: jika $S > 0$, maka ulasan dianggap positif; jika $S = 0$ maka netral; dan jika $S < 0$ maka negatif [22]. Metode lexicon-based polarity detection secara efektif digunakan dalam pemrosesan data media sosial, seperti yang dilakukan dalam penelitian analisis sentimen terhadap tweet terkait rebranding Meta menggunakan alat RapidMiner. Dalam studi tersebut, 2997 tweet dianalisis dan hasil klasifikasi sentimen menunjukkan distribusi data yang signifikan dengan 36% sentimen positif, 29% negatif, dan 35% netral [23].

Dalam konteks analisis sentimen terhadap opini vaksin COVID-19 di Twitter Italia, metode lexicon-based dikembangkan lebih jauh dengan menilai koeksistensi antara frasa nomina dan kata-kata bermuatan sentimen dalam struktur kalimat evaluatif. Pendekatan ini memperhitungkan tidak hanya keberadaan kata positif/negatif tetapi juga intensitas polaritasnya melalui skor standar (Z-score), yang memungkinkan analisis sentimen berbasis fitur (feature-based sentiment analysis) menjadi lebih detail dan kontekstual. Model ini menunjukkan keefektifan dalam mengidentifikasi opini dan sikap publik terhadap isu kesehatan yang kompleks seperti vaksin, dengan potensi penyesuaian lexicon sesuai domain aplikasi sebagai kunci peningkatan akurasi analisis [24]. Pendekatan ini bersifat transparan dan dapat diterapkan secara luas, termasuk pada ulasan relevan Google Maps terhadap Universitas Pamulang.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data dan Interpretasi Hasil BERTopic dan Sentimen Leksikal

Analisis topik dilakukan menggunakan pendekatan BERTopic, yaitu model NLP (Natural Language processing) berbasis transformasi embedding dan pengelompokan topik secara dinamis. Hasil pemodelan menghasilkan beberapa kluster topik dominan yang berulang dalam ulasan pengguna. Di sisi lain, analisis sentimen dilakukan dengan metode leksikal (lexicon-based), yaitu mencocokkan kata-kata dalam ulasan dengan kamus sentimen yang telah dikategorikan sebagai positif atau negatif. Setiap kata yang cocok dengan daftar kata positif diberi skor +1, sedangkan kata negatif diberi skor -1 sesuai dengan rumus yang ada di point 2.3. Kombinasi BERTopic dan analisis sentimen leksikal memberikan pemetaan tematik dan emosional yang saling melengkapi. Penerapan metode ini memungkinkan pihak pengelola Universitas Pamulang untuk mendapatkan wawasan berbasis data dalam mengidentifikasi masalah prioritas dan peluang peningkatan layanan.

3.2 Dataset

Penelitian ini menggunakan data ulasan dari masyarakat terhadap Universitas Pamulang yang dikumpulkan melalui proses scraping pada Google Maps dengan bantuan Selenium dan BeautifulSoup. Informasi yang diambil meliputi nama pemberi ulasan, rating, isi teks ulasan, dan lokasi kampus. Setelah proses pengumpulan data, tahap selanjutnya adalah pembersihan data, yang bertujuan menghilangkan karakter yang tidak diperlukan, kata-kata umum yang tidak penting (stopwords), serta melakukan normalisasi teks. Hasil akhir dari tahap ini berupa 791 baris ulasan yang sudah bersih dan siap dianalisis. Selanjutnya, data tersebut dianalisis menggunakan metode pemodelan topik BERTopic untuk mengenali tema-tema utama yang muncul, serta dilakukan analisis sentimen dengan pendekatan berbasis leksikon guna menentukan sikap positif atau negatif dalam opini pengguna. Semua data yang telah diproses beserta hasil analisisnya disimpan dalam format CSV dan HTML, yang berfungsi sebagai arsip dokumentasi penelitian. Pendekatan ini memastikan bahwa data yang digunakan memiliki validitas, relevansi, serta sesuai dengan standar dalam analisis ilmiah.

3.3 Data Preprocessing

Pada tahap awal pengolahan data, dilakukan proses pembersihan dan standarisasi teks agar data siap digunakan dalam analisis. Langkah ini meliputi case folding, yaitu mengonversi semua huruf dalam teks menjadi huruf kecil untuk memastikan konsistensi format. Selain itu, dilakukan penghilangan tanda baca, simbol, karakter khusus, dan angka yang dianggap tidak relevan untuk analisis topik dan sentimen. Kata-kata umum yang tidak memberikan makna spesifik atau disebut stopwords seperti "yang", "di", dan "dan" dikeluarkan agar fokus analisis tertuju pada kata kunci penting. Selanjutnya, dilakukan penyempurnaan dengan menghapus spasi yang berlebihan agar teks menjadi lebih rapi dan mudah diproses. Hasil dari rangkaian proses ini adalah kumpulan data ulasan yang telah bersih, disimpan pada kolom clean_review, yang kemudian digunakan sebagai input utama dalam pemodelan topik menggunakan BERTopic dan analisis sentimen dalam penelitian ini.

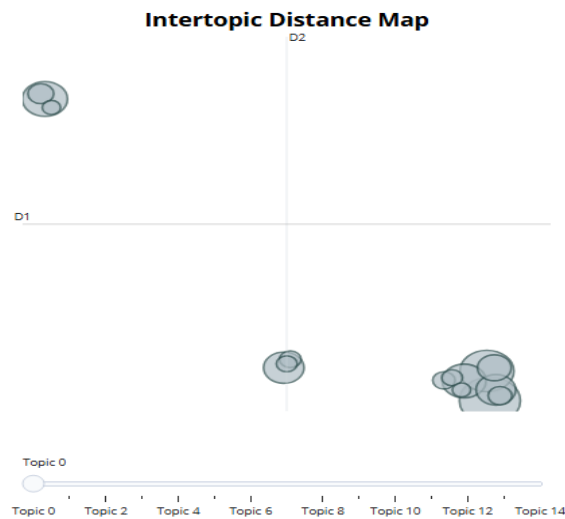
Nama Reviewer	Teks Review	clean_review	Rating	Tanggal Review	Kampus
Piti Cantii	SAYA SANGAT MENYESAL MASIH KULIAH DISINI!Selama saya ada k...	saya sangat menyesal masih kuliah disini selama saya ada kendala dan...	3 bintang	-	Pusat
Rchma Azzahra	Bagus sih tapi saya gak suka warna kampus nya Oren gitu 🤔👎👎 ...	bagus sih tapi saya gak suka warna kampus nya oren gitu	4 bintang	-	Pusat
Natalia Paderico	Hai Kampus, tolong perhatikan sekitar mu, kamu pembuat kemaceta...	hai kampus tolong perhatikan sekitar mu kamu pembuat kemacetan t...	1 bintang	-	Pusat
Fiki Songgeng ...	Saya di sini jujur merasa kecewa istri saya yang tadinya kuliah di sini ...	saya di sini jujur merasa kecewa istri saya yang tadinya kuliah di sini d...	1 bintang	-	Pusat
Nama Reviewer	Teks Review	clean_review	Rating	Tanggal Review	Kampus
Febri H	universitas dengan biaya yang terjangkau 🧡 ...	universitas dengan biaya yang terjangkau	4 bintang	-	Serang
Mirza Chan	Universitas Unpam Serang, totalnya mau ada 10 lantai	universitas unpam serang totalnya mau ada lantai	5 bintang	-	Serang
Romadhon Ades	Masih satu naungan dengan universitas pamulang yang lumayan m...	masih satu naungan dengan universitas pamulang yang lumayan mur...	5 bintang	-	Serang
yofi almahmud...	Yuk segera gabung dan daftar di Universitas Sutomo (Unpam 4 Serang)	yuk segera gabung dan daftar di universitas sutomo unpam serang	5 bintang	-	Serang
kiatiawan 13	Universitas Sutomo dan universitas Pamulang, mereka 1 yayasan yait...	universitas sutomo dan universitas pamulang mereka yayasan yaitu sa...	5 bintang	-	Serang

Nama Reviewer	Teks Review	dean_review	Rating	Tanggal Review	Kampus
Bagas Setiawan...	Kampus,Mantap Fasilitas Terbaik,Ndak ada obat mungkin saran saya ...	kampus mantap fasilitas terbaik ndak ada obat mungkin saran saya ka...	5 bintang	-	Viktor
dian novianti	Kampus megah berada di Viktor, masih bagian dari universitas Pamula...	kampus megah berada di viktor masih bagian dari universitas pamula...	5 bintang	-	Viktor
Nur Syifa Fauziah	Memiliki Fasilitas yang sangat baik, nyaman, parkir sampai 7 lantai, la...	memiliki fasilitas yang sangat baik nyaman parkir sampai lantai lab ter...	5 bintang	-	Viktor
Reni Permanasari	Universitas yang sangat terjangkau biayanya atau bahkan dapat dikata...	universitas yang sangat terjangkau biayanya atau bahkan dapat dikata...	5 bintang	-	Viktor
Nama Reviewer	Teks Review	dean_review	Rating	Tanggal Review	Kampus
su harti	Sepi karena pas libur kuliah.... 😊👍 ...	sepi karena pas libur kuliah	4 bintang	-	Witana
Anna Mariana E...	Suka banget praktek fisika n kimia d sini...	suka banget praktek fisika n kimia d sini	5 bintang	-	Witana
Hariri Asidik	Sangat nyaman tempatnya banyak pepohonanya	sangat nyaman tempatnya banyak pepohonanya	5 bintang	-	Witana
Abdul Choliq	Kampus Witanaharja. Kampus asri, sederhana.	kampus witanaharja kampus asri sederhana	3 bintang	-	Witana

Gambar 2. Contoh data hasil scraping dan proses pembersihan teks ulasan

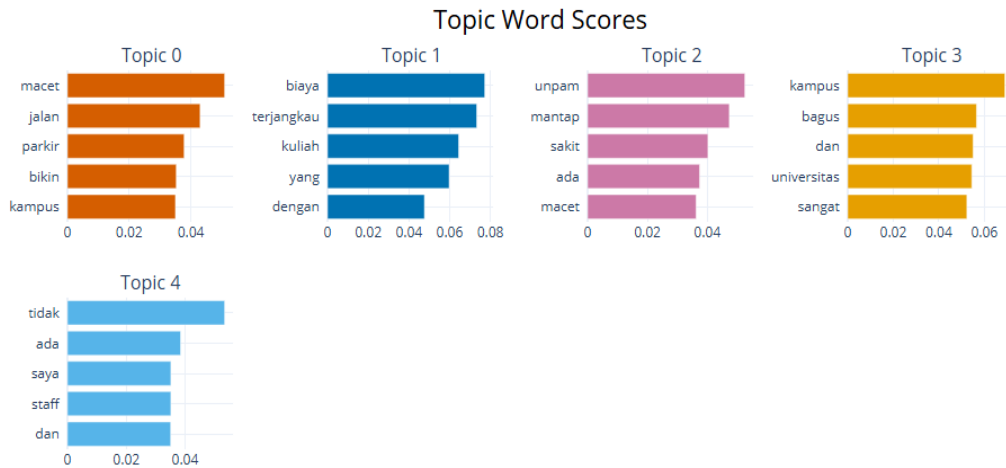
3.4 Topic Modeling (BERTopic)

Dalam tahap analisis topik, penelitian ini mengaplikasikan metode BERTopic yang menggunakan representasi vektor dari dokumen teks untuk menemukan pola dan tema utama dalam kumpulan ulasan. BERTopic bekerja dengan menghasilkan embedding teks lewat model bahasa seperti BERT, kemudian melakukan reduksi dimensi dan pengelompokan dengan algoritma clustering. Hasilnya berupa sejumlah topik yang memiliki kata kunci serupa. Setiap topik diwakili oleh kumpulan kata dengan skor tertinggi yang menunjukkan tema pembahasan dalam ulasan. Dari proses ini, diidentifikasi total 11 topik (termasuk cluster noise) yang meliputi berbagai isu seperti biaya pendidikan, fasilitas kampus, kemacetan, kualitas pelayanan staf, dan pengalaman mahasiswa. Seluruh hasil pemodelan kemudian disimpan dalam format CSV dan divisualisasikan menggunakan bubble chart serta bar chart untuk mempermudah pemahaman terhadap topik dominan dan kata kunci utama pada masing-masing topik.



Gambar 3. Sebaran Topik

Pada gambar 3 Gambar Intertopic Distance Map di atas menampilkan representasi visual dari distribusi topik hasil pemodelan BERTopic terhadap ulasan Google Maps Universitas Pamulang. Setiap bubble mewakili satu topik, di mana ukuran gelembung menunjukkan dominasi atau frekuensi kemunculan topik dalam korpus data, sementara jarak antar topik mencerminkan kedekatan semantik semakin dekat dua bubble, maka semakin mirip konteks atau kata-kata yang digunakan di antara kedua topik tersebut. Terlihat bahwa terdapat sekelompok topics (misalnya Topik 10–14) yang berdekatan secara semantik, yang menunjukkan bahwa kata-kata yang menyusun topik-topik tersebut memiliki kemiripan konteks. Kemungkinan, kelompok topik ini membahas hal-hal yang saling berkaitan seperti administrative services, academic scheduling, atau campus experience, yang meskipun berbeda dari segi fokus, namun memiliki overlapping vocabulary dalam ulasan pengguna. Kedekatan ini menandakan bahwa para pengguna Google Maps mungkin menyampaikan opini atau keluhan mereka dalam satu rangkaian narasi yang mencakup lebih dari satu aspek pelayanan. Sebaliknya, topik yang berada di lokasi terpisah (misalnya Topik 0 atau Topik 6), menunjukkan keunikan semantik, yang bisa mengindikasikan bahwa topik tersebut memiliki konteks yang sangat berbeda, seperti pembahasan spesifik mengenai facilities atau location and accessibility. Salah satu topik tampak memiliki ukuran bubble yang besar, yang mengindikasikan sebagai topik paling dominan dalam korpus ulasan. Topic dominance ini menandakan bahwa banyak pengguna secara eksplisit atau implisit membahas isu yang termasuk dalam topik tersebut. Misalnya, jika topik dominan berisi kata-kata seperti mahasiswa, biaya, kelas, atau pelayanan, maka dapat disimpulkan bahwa persepsi masyarakat terhadap aspek administratif dan akademik merupakan isu utama yang memengaruhi penilaian mereka terhadap institusi. Dengan demikian, topic visualization ini membantu dalam mengidentifikasi klaster pemaknaan, kedekatan tematik, dan bobot perhatian pengguna terhadap berbagai aspek layanan Universitas Pamulang yang tercermin dari ulasan daring mereka



Gambar 4. Topic Word Scores

Setiap topik yang terbentuk dari pemodelan BERTopic direpresentasikan oleh beberapa kata kunci terbaik yang dinilai menggunakan skor c-TF-IDF. Semakin tinggi skor tersebut, semakin kuat keterkaitan kata dengan topik yang dimaksud. Misalnya, pada "topik 1," kata dengan skor tertinggi adalah biaya, diikuti oleh kata terjangkau, kuliah, yang, dan dengan, dengan skor yang menurun secara berurutan. Skor c-TF-IDF ini memudahkan representasi topik, di mana "topik 1" menggambarkan pembahasan mengenai biaya pendidikan yang dinilai terjangkau di Universitas Pamulang. Hal ini menunjukkan bahwa banyak ulasan yang tergabung dalam topik tersebut membahas aspek biaya kuliah, keterjangkauan, serta pandangan masyarakat terhadap hubungan biaya dan kualitas pendidikan di kampus. Dengan bantuan skor c-TF-IDF, peneliti dapat memahami tema utama dan konteks masing-masing topik secara lebih jelas dan terukur. Perlu diketahui bahwa tabel Topic Word Scores hanya menampilkan lima topik utama karena pengaturan standar BERTopic yang memilih topik dengan jumlah dokumen terbanyak atau kata kunci paling signifikan. Hal ini bertujuan agar fokus analisis tetap pada tema dominan tanpa terganggu oleh topik minor yang memiliki jumlah ulasan sedikit. Meski demikian, semua topik lainnya tetap tersedia di output lengkap dari BERTopic dan dapat dianalisis lebih lanjut jika diperlukan.

3.5 Sentiment Analysis (Lexicon-based)

Dalam analisis sentimen penelitian ini, digunakan pendekatan lexicon-based dengan bantuan pustaka Python seperti Sastrawi untuk proses pembersihan teks serta kamus sentimen Bahasa Indonesia yang diperoleh dari sumber eksternal. Setiap kata dalam kolom `clean_review` dicocokkan dengan daftar kata yang diberi label positif atau negatif, lalu dihitung skor polaritasnya. Dokumen yang memiliki skor positif digolongkan sebagai sentimen positif, dokumen dengan skor negatif dikategorikan sebagai sentimen negatif, dan skor yang mendekati nol dianggap netral. Dari total 791 ulasan yang dianalisis, ditemukan distribusi sebanyak 388 ulasan positif, 282 netral, serta 121 negatif. Hasil analisis sentimen ini kemudian diintegrasikan dengan output topik dari BERTopic, sehingga memungkinkan pemetaan distribusi sentimen pada masing-masing topik yang telah diidentifikasi.

```

positive_words = [
    "bagus", "baik", "terjangkau", "Luas", "mantap", "nyaman",
    "murah", "mewah", "sangat", "ramah", "cinta", "terbaik",
    "puas", "senang", "bersih", "rapi", "indah", "megah",
    "unggul", "Lancar", "cepat", "hemat", "mudah", "sukses"
]

negative_words = [
    "macet", "jelek", "buruk", "tidak", "kurang", "keluhan",
    "sakit", "kapok", "menyesal", "rusak", "rotor", "semrawut",
    "Lambat", "antri", "ribet", "berantakan", "penuh",
    "bising", "panas", "ribut", "gangguan", "Lamban", "parah"
]

```

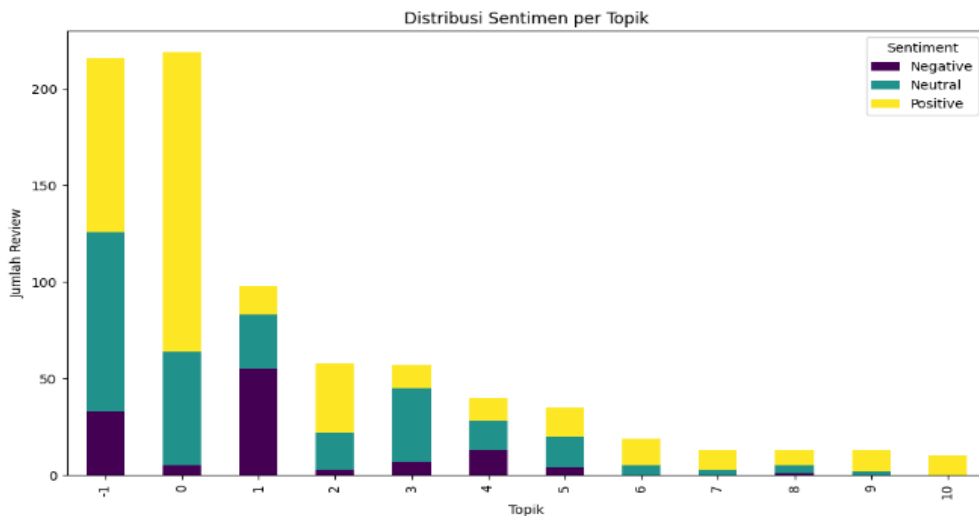
Gambar 5. Kamus kata untuk mengenali nuansa positif/negatif teks (lexicon)

Gambar 5 memperlihatkan dua daftar kata dalam Python, yakni `positive_words` dan `negative_words`, yang berperan sebagai kamus leksikon sentimen dalam metode lexicon-based. Daftar `positive_words` memuat kata-kata bermakna positif seperti bagus, baik, nyaman, terjangkau, dan puas; kehadiran kata-kata ini dalam teks menunjukkan nuansa sentimen positif. Sebaliknya, daftar `negative_words` berisi kata-kata bernada negatif seperti macet, jelek, buruk, menyesal, rusak, dan semrawut, yang jika muncul menandakan sentimen negatif dalam ulasan.

Kumpulan kata-kata ini disusun secara manual berdasarkan kata-kata yang sering muncul dalam konteks ulasan masyarakat terhadap Universitas Pamulang, sehingga kamus ini menjadi pondasi penting dalam proses penilaian sentimen pada tahap analisis selanjutnya.

3.6 Hasil Visualization Sentiment Analysis (Lexicon-based)

Dari 791 ulasan yang dianalisis menggunakan pendekatan lexicon-based, ditemukan 388 ulasan dengan sentimen positif, 282 netral, dan 121 negatif. Sentimen ini kemudian dipetakan ke dalam setiap topik dari hasil BERTopic. Contohnya, Topik 0 yang membahas kemacetan dan parkir terdiri dari 5 ulasan negatif, 59 netral, dan 155 positif. Sementara itu, topik 1 tentang biaya kuliah didominasi ulasan negatif sebanyak 55, diikuti 28 netral, dan hanya 15 positif. Topik 2 memiliki proporsi 3 negatif, 19 netral, dan 36 positif, sedangkan Topik 3 lebih banyak ulasan netral sebanyak 38. Beberapa topik minor seperti topik 10 hanya mengandung 10 ulasan yang semuanya berkonotasi positif. Temuan ini menandakan bahwa terdapat isu-isu tertentu yang memicu sentimen negatif lebih tinggi, sedangkan tema lain seperti biaya yang terjangkau atau pengalaman kuliah cenderung mendapat tanggapan positif.



Gambar 6. Distribusi Sentimen per Topik Hasil Pemodelan BERTopic

Gambar 6 memperlihatkan distribusi sentimen untuk setiap topik yang divisualisasikan dalam bentuk stacked bar chart menggunakan Matplotlib. Pada grafik, sumbu X menunjukkan nomor topik hasil dari BERTopic, sementara sumbu Y menggambarkan jumlah ulasan yang termasuk dalam setiap topik tersebut. Setiap batang pada grafik terbagi menjadi tiga warna yang mewakili jenis sentimen, yaitu ungu untuk sentimen negatif, hijau kebiruan untuk sentimen netral, dan kuning untuk sentimen positif. Dengan visualisasi ini, dapat terlihat proporsi jenis sentimen pada masing-masing topik, sekaligus memudahkan identifikasi topik yang paling dominan dan hubungan antara tema ulasan dengan kecenderungan sentimen masyarakat. Sebagai contoh, Topik -1 dan Topik 0 memiliki jumlah ulasan terbanyak, sedangkan topik 1 cenderung didominasi oleh sentimen negatif.



Gambar 7. Visualisasi dengan Word Cloud



Penjelasan dari Gambar 7 word cloud di atas:

- a. Gambar word cloud topik 1 menggambarkan kata-kata yang berkaitan erat dengan masalah kampus dan parkir. Kata-kata utama yang muncul seperti "kampus," "jalan," "parkir," dan "macet" menjadi indikasi adanya keluhan maupun tantangan yang sering dihadapi oleh mahasiswa dan masyarakat di sekitar kampus, terutama terkait kemacetan dan keterbatasan fasilitas parkir serta akses jalan. Kehadiran kata-kata seperti "tolong," "ada," dan "bikin" menunjukkan adanya aspirasi atau permintaan dari para pengguna untuk perbaikan kondisi tersebut.
- b. Gambar word cloud topik 3 kata-kata yang paling sering muncul meliputi "ada," "yang," "dada," "macet," dan "unpam." Topik ini berfokus pada kehidupan sehari-hari di lingkungan kampus, dengan sorotan pada kondisi di lingkungan sekitar kampus, termasuk interaksi sosial dan situasi sehari-hari. Terdapat kata-kata sapaan dan harapan seperti "bikin," "semoga," dan "amin," yang mencerminkan doa dan harapan pengguna terkait situasi sekitar kampus atau universitas tersebut.
- c. Gambar word cloud topik 5 menyoroti aspek pendidikan dan pengalaman mahasiswa. Kata-kata yang paling menonjol seperti "saya," "kalian," "kuliah," "unpam," "murah," "lulus," dan "biaya" menunjukkan perhatian pada akses pendidikan yang terjangkau, proses kelulusan, serta biaya kuliah. Selain itu, kehadiran kata-kata bernada positif seperti "alhamdulillah" dan "ramah" merefleksikan kepuasan dan apresiasi dari mahasiswa maupun alumni terhadap fasilitas dan layanan di universitas tersebut.
- d. Gambar word cloud topik 7 memperlihatkan kata-kata yang berkaitan erat dengan gedung dan fasilitas kampus. Kata-kata seperti "gedung," "bagus," "mewah," "luas," dan "baru" menonjol, menunjukkan fokus pada aspek fisik kampus yang mencakup bangunan serta fasilitas modern dan terawat yang mungkin menjadi daya tarik atau poin pembahasan utama dalam topik ini.
- e. Gambar word cloud topik 9 menampilkan fokus pada kenyamanan serta mutu suatu tempat, dengan kata-kata utama seperti "tempat," "nyaman," "bagus," dan "sangat." Hal ini mengindikasikan bahwa pembicaraan lebih diarahkan pada suasana dan fasilitas tempat yang dirasakan menyenangkan, nyaman, dan berkualitas oleh para penggunanya.
- f. Gambar word cloud topik 10 point utama terkait dengan fasilitas parkir dan kampus. Kata-kata penting seperti "parkir," "bagus," "luas," "gedung," dan "akses" menandakan diskusi mengenai besarnya kapasitas parkir, kemudahan akses ke bangunan kampus, serta fasilitas pendukung yang menunjang kegiatan mahasiswa dan pengunjung.

Dari keseluruhan word cloud yang dihasilkan mulai dari Topik 0 hingga Topik 10, penelitian ini memilih hanya untuk menyertakan visualisasi word cloud dari Topik 1, 3, 5, 7, 9, dan 10. Keputusan ini diambil karena topik-topik tersebut dianggap paling representatif dan mengandung tema yang informatif, spesifik, serta unik dalam menggambarkan ulasan masyarakat terhadap Universitas Pamulang. Sebaliknya, beberapa topik lain seperti Topik 0, 2, 4, 6, dan 8 menampilkan kata-kata yang terlalu umum, sering berulang, atau tumpang tindih dengan pembahasan di topik lain sehingga kurang memberikan kontribusi berarti bagi bahasan dalam penelitian.

4. KESIMPULAN

Penelitian ini menganalisis ulasan masyarakat tentang Universitas Pamulang dengan pendekatan text mining yang menggabungkan pemodelan topik BERTopic dan analisis sentimen leksikal. Dari 791 ulasan Google Maps yang telah diproses, diperoleh 11 topik berbeda seperti isu biaya kuliah, fasilitas kampus, hingga kemacetan sekitar. Kata kunci tiap topik diidentifikasi dengan c-TF-IDF dan divisualisasikan melalui bubble chart dan bar chart. Analisis sentimen menunjukkan mayoritas ulasan bersifat positif (388 ulasan), meskipun beberapa topik seperti biaya dan kemacetan memicu sentimen negatif. Word cloud digunakan untuk memperjelas kata dominan pada tiap tema. Penelitian ini membuktikan efektivitas integrasi metode BERTopic dan analisis sentimen berbasis leksikon dalam mengeksplorasi opini publik secara lebih mendalam. Temuan yang diperoleh tidak hanya memberikan gambaran mengenai persepsi masyarakat terhadap Universitas Pamulang, tetapi juga berpotensi dapat menjadi dasar evaluasi strategis bagi pihak pengelola Universitas Pamulang dalam meningkatkan kualitas pendidikan dan layanan. Untuk pengembangan ke depan, penelitian dapat diperluas dengan penggunaan model analisis sentimen berbasis machine learning yang lebih kompleks atau melakukan analisis korelasi yang lebih lanjut antara topik ulasan dengan karakteristik pengguna.

REFERENCES

- [1] F. Akbar, Hadiyanto, and C. E. Widodo, "Sentiment analysis of data on Google Maps reviews regarding tourism on Keraton Kasepuhan Cirebon using the lexicon based method," *Proc. 3rd Int. Conf. Adv. Inf. Sci. Dev.*, pp. 19–24, 2023, doi: 10.5220/0012440100003848.
- [2] C.-C. Chen and C.-C. Chang, "Evaluating public library services through user-generated content: Social network analysis of Google Maps reviews," *Preprints*, May 2024, doi: 10.20944/preprints202405.0619.v1.
- [3] F. Kitsios, E. Mitsopoulou, E. Moustaka, and M. Kamariotou, "User-generated content in social media: A twenty-year bibliometric analysis in hospitality," *Information*, vol. 13, no. 12, p. 574, Dec. 2022, doi: 10.3390/info13120574.
- [4] A. Göçen, M. M. Ibrahim, and A. U. I. Khan, "Public attitudes toward higher education using sentiment analysis and topic modeling," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 83, Nov. 2024, doi: 10.1007/s44163-024-00195-4.



- [5] F. Yahya, D. E. Ratnawati, and B. Rahayudi, “Analisis sentimen ulasan pengguna dari Google Maps menggunakan metode Long Short-Term Memory (studi kasus: Rumah Sakit Gatoel),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 4, pp. 1–11, 2025. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] J. Wu, P. Qu, B. Zhang, and Z. Zhou, “Sentiment analysis in social media: Leveraging BERT for enhanced accuracy,” *Journal of Industrial Engineering and Applied Sciences*, vol. 2, no. 4, 2024. [Online]. Available: <https://doi.org/10.5281/zenodo.13144389>
- [7] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” arXiv preprint, arXiv:2203.05794, Mar. 2022. [Online]. Available: <http://arxiv.org/abs/2203.05794>
- [8] N. Khodeir and F. Elghannam, “Efficient topic identification for urgent MOOC forum posts using BERTopic and traditional topic modeling techniques,” *Education and Information Technologies*, vol. 30, no. 5, pp. 5501–5527, Apr. 2025, doi: 10.1007/s10639-024-13003-4.
- [9] D. Kumar and M. Weissenberger-Eibl, “Artificial intelligence driven trend forecasting: Integrating BERT topic modelling and generative artificial intelligence for semantic insights,” in *R&D Management Conference 2024*, 2024, doi: 10.24406/publica-3456.
- [10] Y. Jin, Using Text Mining and BERTopic, Master’s thesis, University of California, 2022. [Online]. Available: <https://escholarship.org/uc/item/69m2f87d>
- [11] F. T. Saputra, S. H. Wijaya, Y. Nurhadryani, and Defina, “Lexicon addition effect on lexicon-based of Indonesian sentiment analysis on Twitter,” in *Proc. 2020 Int. Conf. Informatics, Multimedia, Cyber and Inf. Syst. (ICIMCIS)*, Nov. 2020, pp. 136–141, doi: 10.1109/ICIMCIS51567.2020.9354269.
- [12] A. Rufaida, A. Permanasari, and N. Setiawan, “Lexicon-based sentiment analysis using Inset dictionary: A systematic literature review,” in *Proc. 7th EAI Int. Conf. Big Data Innov. Smart Cities (EAI BDISC)*, Oct. 2022, doi: 10.4108/eai.5-10-2022.2327474..
- [13] Y. Fauziah, B. Yuwono, and A. S. Aribowo, “Lexicon-based sentiment analysis in Indonesia languages: A systematic literature review,” in *RSF Conf. Ser. Eng. Technol.*, vol. 1, no. 1, Dec. 2021, pp. 363–367, doi: 10.31098/cset.v1i1.397.
- [14] I. Fatan and H. Irawan, “Twitter, Instagram, YouTube speak: Understanding sentiments on LRT Jabodebek services via Inset lexicon, IndoBERT and BERTopic approaches,” *J. Electr. Syst.*, vol. 20, no. 4s, pp. 1028–1035, Apr. 2024, doi: 10.52783/jes.2147..
- [15] S. Samsir, R. S. Saragih, S. Subagio, R. Aditiya, and R. Watrianthos, “BERTopic modeling of natural language processing abstracts: Thematic structure and trajectory,” *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1514, Jul. 2023, doi: 10.30865/mib.v7i3.6426.
- [16] S. Ramadhan, T. Siswanto, and S. Sari, “Sentiment analysis and topic modelling of candidate news in the 2024 general election on Twitter social media using latent Dirichlet allocation (LDA) method,” *Intelmatias*, vol. 5, no. 1, pp. 33–41, Feb. 2025, doi: 10.25105/v5i1.21058.
- [17] H. Lee, S. H. Lee, K. R. Lee, and J. H. Kim, “ESG discourse analysis through BERTopic: Comparing news articles and academic papers,” *Comput. Mater. Contin.*, vol. 75, no. 3, pp. 6023–6037, 2023, doi: 10.32604/cmc.2023.039104..
- [18] L. McInnes, J. Healy, and J. Melville, “UMAP: Uniform manifold approximation and projection for dimension reduction,” arXiv preprint arXiv:1802.03426, Sep. 2020, [Online]. Available: <http://arxiv.org/abs/1802.03426>
- [19] O. Babalola, B. Ojokoh, and O. Boyinbode, “Comprehensive evaluation of LDA, NMF, and BERTopic’s performance on news headline topic modeling,” *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 268–289, Nov. 2024, doi: 10.62411/jcta.11635.
- [20] V. Sharifian-Attar, S. De, S. Jabbari, J. Li, H. Moss, and J. Johnson, “Analysing longitudinal social science questionnaires: Topic modelling with BERT-based embeddings,” in *Proc. 2022 IEEE Int. Conf. on Big Data (Big Data)*, Osaka, Japan, Dec. 2022, pp. 5558–5567, doi: 10.1109/BigData55660.2022.10020678.
- [21] W. Yang, X. Li, J. Zhang, K. Wang, and L. Chen, “A two-stage feature extraction approach for green energy consumers in retail electricity markets using clustering and TF-IDF algorithms,” *Energy Eng.*, vol. 122, no. 5, pp. 1697–1713, 2025, doi: 10.32604/ee.2025.060571.
- [22] C. H. Hill, J. E. Fresneda, and M. Anandarajan, “The wisdom of the lexicon crowds: Leveraging on decades of lexicon-based sentiment analysis for improved results,” *J. Big Data*, vol. 12, no. 1, p. 129, May 2025, doi: 10.1186/s40537-025-01186-7.
- [23] N. H. Md Saad, A. S. Zainul Abidi, Z. Yaacod, M. M. Mohd Ali, and Z. Kun, “Twitter sentiment analysis on Meta: A lexicon-based analysis using RapidMiner,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 6, pp. 259–270, Jul. 2023, doi: 10.17762/ijritcc.v11i6.7561.
- [24] R. Catelli, S. Pelosi, C. Comito, C. Pizzuti, and M. Esposito, “Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy,” *Comput. Biol. Med.*, vol. 158, p. 106876, May 2023, doi: 10.1016/j.combiomed.2023.106876.