



PEMETAAN PENYEBARAN TOPIK PERCAKAPAN PENGGUNA KRL (COMMUTER LINE) PADA MEDIA SOSIAL TWITTER MENGGUNAKAN SNA DAN BERTOPIC

Marsyanda Nur Zahra¹⁾, Monica Cinthya²⁾

¹⁾Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: marsyanda.22119@mhs.unesa.ac.id

²⁾Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: monicacinthya@unesa.ac.id

Abstract

KRL (Commuter Line) is one of the main public transportation services used by the community, particularly in the Greater Jakarta area. The high level of activity among KRL users has generated various conversations on social media X (Twitter) concerning schedules, routes, operational disruptions, complaints, and service facilities. The large volume of unstructured conversations makes it difficult to manually identify the topics discussed and the relationships among them. This study aims to map the topics of KRL Commuter Line user conversations on social media X using BERTopic and analyze the relationships among topics using Social Network Analysis (SNA). The data were collected through web scraping from social media X using the keywords "KRL", "Commuter Line", and "@CommuterLine" during the period from March to May 2026, and were subsequently processed through a preprocessing stage. BERTopic modeling consisted of sentence embedding using SentenceTransformer, dimensionality reduction using UMAP, clustering using HDBSCAN, and topic representation using C-TF IDF. The modeling results identified ten main topics covering schedules and delays, routes and travel directions, operational disruptions, unresolved user complaints, women's carriages, and priority seating, with a Topic Coherence score of 0.4387 and a Topic Diversity score of 0.7897. SNA was then used to map the relationships among topics based on shared keywords, forming a network with 9 nodes and 12 edges, an Average Degree of 1.333, an Average Weighted Degree of 2.778, and three communities with a modularity value of 0.337. The network evaluation showed that Topic 3 (Routes and Stations) and Topic 4 (Unresolved Complaints) had the highest degree value of 4, while Topic 4 had the highest betweenness centrality and served as a bridge between groups. The network of KRL user conversation topics was divided into three communities: schedules and routes, operational disruptions and complaints, and priority facilities, which were interconnected through Topic 3 and Topic 4. Therefore, the combination of BERTopic and SNA provides an overview of KRL user conversation topics and the relationships among topics on social media X.

Keywords: BERTopic, KRL, Social Network Analysis, Topic Modeling, Twitter

Abstrak

KRL (Commuter Line) merupakan salah satu sarana transportasi publik utama yang digunakan masyarakat, khususnya di wilayah Jabodetabek. Tingginya aktivitas pengguna KRL mendorong munculnya berbagai percakapan mengenai jadwal, rute, gangguan operasional, keluhan, hingga fasilitas layanan pada media sosial X (Twitter). Besarnya jumlah percakapan yang tidak terstruktur menyebabkan topik pembahasan dan keterkaitan antar-topik sulit diidentifikasi secara manual. Penelitian ini bertujuan untuk memetakan topik percakapan pengguna KRL Commuter Line pada media sosial X menggunakan BERTopic serta menganalisis keterkaitan antar topik menggunakan Social Network Analysis (SNA). Data penelitian diperoleh melalui teknik web scraping pada media sosial X dengan kata kunci "KRL", "Commuter Line", dan "@CommuterLine" pada periode Maret–Mei 2026, kemudian diproses melalui tahap preprocessing. Pemodelan BERTopic dilakukan melalui tahapan sentence embedding menggunakan SentenceTransformer, reduksi dimensi menggunakan UMAP, clustering menggunakan HDBSCAN, serta pembentukan representasi kata menggunakan C-TF-IDF. Hasil pemodelan menghasilkan sepuluh topik utama yang mencakup jadwal dan keterlambatan, jalur dan rute perjalanan, gangguan operasional, keluhan pengguna yang belum ditangani, gerbong khusus wanita, hingga kursi prioritas penumpang, dengan nilai Topic Coherence sebesar 0,4387 dan Topic Diversity sebesar 0,7897. Selanjutnya, SNA digunakan untuk memetakan keterkaitan antar-topik berdasarkan kesamaan kata kunci, membentuk jaringan dengan 9 node dan 12 edge, Average Degree sebesar 1,333, Average Weighted Degree sebesar 2,778, serta tiga komunitas dengan nilai modularity sebesar 0,337. Hasil evaluasi jaringan menunjukkan Topik 3 (Rute dan Stasiun Perjalanan) dan Topik 4 (Keluhan yang Belum Ditangani) memiliki nilai degree tertinggi sebesar 4, sementara Topik 4 memiliki nilai betweenness centrality tertinggi dan berperan sebagai penghubung antar-kelompok. Jaringan topik percakapan pengguna KRL terbagi menjadi tiga komunitas, yaitu jadwal dan rute perjalanan, gangguan operasional dan keluhan, serta fasilitas prioritas, yang saling terhubung melalui Topik 3 dan Topik 4. Dengan demikian, kombinasi BERTopic dan SNA dapat memberikan gambaran mengenai topik percakapan pengguna KRL sekaligus pola keterkaitan antar-topik pada media sosial X.

Kata Kunci: BERTopic, KRL, Social Network Analysis, Topic Modeling, Twitter.



PENDAHULUAN

KRL atau Commuter Line menjadi salah satu sarana transportasi massal utama yang mendukung mobilitas masyarakat perkotaan, khususnya di wilayah Jabodetabek. Pada tahun 2025, jumlah pengguna KRL Jabodetabek mencapai 349,31 juta orang dengan rata-rata lebih dari 1,1 juta penumpang per hari kerja. Tingginya jumlah pengguna tersebut menunjukkan bahwa KRL memiliki peranan penting dalam mendukung aktivitas masyarakat sehari-hari. Seiring dengan tingginya penggunaan layanan, berbagai pengalaman pengguna, baik berupa informasi, keluhan, maupun apresiasi, turut disampaikan melalui ruang publik digital.

Media sosial menjadi salah satu sarana yang digunakan masyarakat untuk menyampaikan opini tersebut. Pada tahun 2025, jumlah pengguna internet di Indonesia mencapai 229,4 juta jiwa dengan penetrasi sebesar 80,66 persen (APJII, 2025). Salah satu media sosial yang digunakan untuk menyampaikan informasi secara cepat adalah X atau Twitter. Pengguna dapat menyampaikan tweet, retweet, reply, maupun mention secara real-time. Pengguna KRL Jabodetabek juga memanfaatkan platform tersebut untuk menyampaikan pengalaman perjalanan, keluhan, pertanyaan, serta informasi dengan menyebut akun resmi @CommuterLine.

Besarnya volume percakapan di media sosial menghasilkan data tidak terstruktur dalam jumlah besar sehingga informasi yang terkandung di dalamnya sulit dianalisis secara manual. Opini pengguna dapat menjadi sumber informasi untuk memahami kebutuhan maupun permasalahan layanan (Tamba & Kholikin, 2025). Oleh karena itu, diperlukan metode yang dapat mengolah data percakapan secara otomatis agar informasi yang terdapat dalam ribuan tweet dapat dipahami dengan lebih mudah.

Salah satu pendekatan yang dapat digunakan adalah topic modeling. Topic modeling merupakan teknik untuk mengidentifikasi tema tersembunyi dalam sekumpulan dokumen secara otomatis. Beberapa metode yang telah digunakan antara lain Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), Top2Vec, dan BERTopic. BERTopic memanfaatkan representasi semantik berbasis transformer sehingga dapat menangkap konteks teks secara lebih baik. Egger dan Yu (2022) menunjukkan bahwa BERTopic dapat digunakan untuk menghasilkan topik yang koheren dan mudah diinterpretasikan pada data Twitter.

Meskipun pemodelan topik dapat menunjukkan topik yang muncul dalam percakapan, hasil tersebut belum secara langsung menggambarkan keterkaitan antar-topik. Permasalahan seperti keterlambatan, kepadatan, gangguan perjalanan, maupun keluhan pengguna dapat memiliki hubungan satu sama lain. Oleh karena itu, diperlukan pendekatan tambahan untuk mengetahui bagaimana topik-topik tersebut saling terhubung.

Social Network Analysis (SNA) dapat digunakan untuk menganalisis hubungan antarentitas dalam suatu jaringan. Dalam penelitian ini, setiap topik diperlakukan sebagai node, sedangkan hubungan antar-topik dibentuk berdasarkan kesamaan kata kunci. SNA memungkinkan

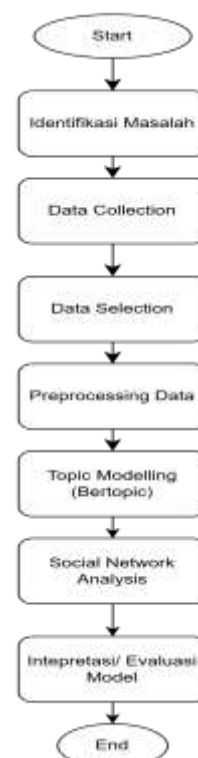
hubungan tersebut dianalisis menggunakan beberapa ukuran sentralitas, seperti degree centrality, betweenness centrality, dan closeness centrality.

Penelitian terdahulu telah menerapkan BERTopic pada berbagai jenis data media sosial, tetapi sebagian penelitian masih berfokus pada pemodelan topik tanpa memetakan hubungan antar-topik. Penelitian lain yang menggabungkan topic modeling dan SNA juga menunjukkan bahwa kombinasi kedua metode dapat memberikan gambaran yang lebih luas mengenai struktur percakapan digital.

Berdasarkan kondisi tersebut, penelitian ini berjudul “Pemetaan Penyebaran Topik Percakapan Pengguna KRL (Commuter Line) pada Media Sosial Twitter menggunakan SNA dan BERTopic”. Penelitian ini menerapkan BERTopic untuk mengidentifikasi topik percakapan pengguna KRL dan menggunakan SNA untuk memetakan keterkaitan antar-topik berdasarkan kesamaan kata kunci.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dalam bentuk text mining. Kerangka penelitian mengikuti tahapan Knowledge Discovery in Databases (KDD), yang meliputi proses pengumpulan dan pemilihan data, preprocessing, data mining, serta interpretasi dan evaluasi hasil.



Gambar 1. Metode Penelitian

3.1. Data Collection

Data berupa percakapan pengguna KRL Commuter Line Jabodetabek di media sosial X (Twitter), dikumpulkan melalui web scraping dengan Twitter/X Data Tweet Scraper pada platform Apify. Kata kunci yang



digunakan adalah "KRL", "Commuter Line", dan "@CommuterLine" untuk menjaring informasi perjalanan, pertanyaan, keluhan, dan laporan gangguan. Pengumpulan dilakukan pada Maret-Mei 2026 dan menghasilkan 5.000 tweet sebagai data mentah.

3.2. Data Selection

Atribut yang digunakan untuk BERTopic adalah id, text, createdAt, dan lang. Atribut text menjadi data utama pemodelan, sedangkan createdAt dan lang menjadi pendukung seleksi. Untuk SNA, data hubungan antar-topik terdiri atas source (topik asal), target (topik tujuan), dan weight (jumlah kata kunci yang sama).

3.3. Data Preprocessing

Preprocessing meliputi penghapusan data kosong dan duplikat, case folding, penghapusan URL, mention, hashtag, angka, dan tanda baca dengan regular expression, normalization kata tidak baku (misalnya "bgt" menjadi "banget", "gk" menjadi "tidak"), serta custom stopwords removal untuk membuang kata kurang informatif. Jumlah data turun dari 5.000 tweet mentah menjadi 4.654 (filter bahasa dan seleksi kolom), 4.652 (cleaning dan hapus duplikat), lalu 4.644 tweet (normalisasi dan stopwords removal).

Tabel 1. Hasil Preprocessing Data

Tahap	Jumlah Data
Data mentah	5.000
Setelah filter bahasa dan seleksi kolom	4.654
Setelah cleaning dan hapus duplikat	4.652
Setelah normalisasi dan stopwords removal	4.644

3.4. Topic Modeling menggunakan BERTopic

Tahapan BERTopic terdiri atas empat langkah. Pertama, setiap tweet diubah menjadi embedding dengan SentenceTransformer paraphrase-multilingual-MiniLM-L12-v2, menghasilkan matriks 4.644×384 . Kedua, dimensi direduksi menjadi 5 dengan UMAP (n_neighbors 15, min_dist 0,0, metric cosine, random_state 42). Ketiga, tweet dikelompokkan dengan HDBSCAN, dan tweet yang tidak masuk kelompok diberi label -1 sebagai outlier. Terakhir, C-TF-IDF menentukan kata yang paling mewakili tiap topik. Sebanyak 2.879 tweet (61,98%) masuk ke topik dan 1.766 tweet (38,02%) menjadi outlier yang tidak dianalisis lebih lanjut.

3.5. Evaluasi Topic Modeling

Model dievaluasi dengan Topic Coherence (metrik C_v) untuk mengukur keterkaitan makna antarkata dalam topik, dan Topic Diversity untuk mengukur keberagaman kata kunci antar-topik (semakin tinggi, semakin sedikit tumpang tindih).

3.6. Social Network Analysis

Hasil BERTopic digunakan sebagai dasar pembentukan jaringan. Setiap topik diperlakukan sebagai node, sedangkan edge dibentuk ketika terdapat kata kunci

yang sama antara dua topik. Bobot edge menunjukkan jumlah kata kunci yang sama.

Edge list yang terdiri atas Source, Target, dan Weight kemudian diolah menggunakan Gephi. Evaluasi jaringan dilakukan menggunakan degree centrality, betweenness centrality, closeness centrality, dan modularity class.

HASIL DAN PEMBAHASAN

4.1. Hasil Preprocessing

Hasil preprocessing menunjukkan bahwa dari 5.000 data tweet mentah diperoleh 4.644 tweet bersih yang dapat digunakan dalam proses pemodelan. Pengurangan data terjadi karena adanya data yang tidak sesuai dengan kriteria bahasa, data kosong, data duplikat, serta proses pembersihan teks.

Data hasil preprocessing kemudian digunakan sebagai input BERTopic. Tahapan preprocessing menjadi penting karena data Twitter memiliki karakteristik berupa bahasa informal, singkatan, mention, tanda baca, dan unsur lain yang dapat menjadi noise dalam pemodelan topik.

4.2. Hasil Pemodelan BERTopic

Pemodelan BERTopic menghasilkan sepuluh topik utama dari tweet yang berhasil dikelompokkan. Setiap topik diinterpretasikan berdasarkan kata kunci representatif yang dihasilkan oleh C-TF-IDF.

Tabel 2. Hasil Pemodelan dan Interpretasi Topik

Topik	Kata Kunci Representatif	Inti Pembicaraan
0	jadwal, jam, pagi, telat, berapa, menit, krl, brp, hari	Jadwal dan Keterlambatan KRL
1	tanah, jalur, jalan, arah, kebayoran, rangkas, serpong, parung, sudimara	Jalur dan Arah Perjalanan KRL
2	greenline, green, line, normal, gangguan, pagi, beroperasi, info, belum	Gangguan Operasional KRL
3	stasiun, rute, besar, senen, cisauk, mangga, stasiunnya, info, bekasi, lrt	Rute dan Stasiun Perjalanan
4	belum, tidak, ada, masih, kebayoran, sudimara, info, tindakan, pernah	Keluhan yang Belum Ditangani
5	cewek, perempuan, gerbong, woman, wanita, sesama, support, laki, campur	Gerbong Khusus Wanita
6	manggarai, via, jam, krl, berapa, bogor, senen, stasiun, paling, menit	Rute dan Waktu Perjalanan melalui Manggarai



7	kereta, jam, jadwal, stasiun, menit, berangkat, telat, bekasi, lama, cikarang	Jadwal dan Keterlambatan KRL Bekasi–Cikarang
8	penumpang, prioritas, kursi, mereka, transportasi, lain, duduk, naik	Kursi Prioritas Penumpang

Hasil tersebut menunjukkan bahwa percakapan pengguna KRL tidak hanya membahas satu jenis permasalahan. Topik yang muncul mencakup informasi perjalanan seperti jadwal, waktu, rute, dan stasiun, serta isu layanan seperti gangguan operasional dan keluhan pengguna.

Topik 0 dan Topik 7 menunjukkan adanya pembahasan mengenai jadwal dan keterlambatan KRL. Topik 1, Topik 3, dan Topik 6 berkaitan dengan rute, jalur, stasiun, serta perjalanan melalui Manggarai. Sementara itu, Topik 2 dan Topik 4 menggambarkan percakapan mengenai gangguan operasional dan keluhan pengguna yang belum mendapatkan penanganan.

Topik 5, Topik 8, dan Topik 9 berkaitan dengan fasilitas serta layanan prioritas penumpang. Hal ini menunjukkan bahwa selain aspek perjalanan, pengguna juga membahas penggunaan fasilitas khusus di dalam layanan KRL.

4.3. Evaluasi BERTopic

Evaluasi Topic Coherence menghasilkan nilai 0,4387. Nilai tersebut menunjukkan bahwa kata-kata kunci yang terdapat dalam masing-masing topik memiliki keterkaitan makna yang cukup sehingga topik masih dapat diinterpretasikan. Karakteristik data Twitter yang singkat dan menggunakan bahasa tidak baku menjadi salah satu kondisi yang dapat memengaruhi nilai coherence.

Sementara itu, Topic Diversity menghasilkan nilai 0,7897. Nilai tersebut menunjukkan bahwa kata kunci yang digunakan pada masing-masing topik cukup beragam dan tidak terlalu banyak mengalami pengulangan antartopik.

Tabel 3. Hasil Evaluasi BERTopic

Metrik	Nilai
Topic Coherence	0,4387
Topic Diversity	0,7897

Hasil evaluasi menunjukkan bahwa BERTopic mampu menghasilkan topik yang dapat diinterpretasikan dengan tingkat keberagaman kata kunci yang cukup baik. Sepuluh topik yang dihasilkan selanjutnya digunakan sebagai dasar untuk membangun jaringan antar-topik.

4.4. Pembentukan Jaringan Antar-Topik

Pembentukan jaringan dilakukan berdasarkan kesamaan kata kunci dari sepuluh topik yang dihasilkan BERTopic. Jika dua topik memiliki kata kunci yang sama, maka kedua topik tersebut dihubungkan menggunakan

edge. Jumlah kata kunci yang sama digunakan sebagai nilai weight.

Hasil pembentukan edge list menunjukkan bahwa hubungan terkuat terdapat pada Topik 0 dengan Topik 6 dan Topik 0 dengan Topik 7, masing-masing dengan weight sebesar 4. Sementara hubungan Topik 6 dan Topik 7 memiliki weight sebesar 3.

Hubungan tersebut menunjukkan bahwa topik-topik yang membahas jadwal, waktu, dan perjalanan memiliki sejumlah kata kunci yang sama sehingga membentuk keterkaitan dalam jaringan.

4.5. Evaluasi Social Network Analysis

Hasil pembentukan jaringan menggunakan Gephi menghasilkan jaringan yang terdiri atas 9 node dan 12 edge. Sembilan topik masuk ke dalam jaringan karena terdapat hubungan dengan topik lainnya.

Metrik yang digunakan terdiri atas weighted degree, closeness centrality, betweenness centrality, dan modularity class.

Tabel 4. Hasil Evaluasi Social Network Analysis

Label	Weighted Degree	Closeness Centrality	Betweenness Centrality	Modularity Class
Topik 0	9	0,600000	0,000000	0
Topik 6	9	1,000000	0,000000	0
Topik 7	9	0,000000	0,000000	0
Topik 3	6	0,800000	0,035714	0
Topik 4	6	1,000000	0,071429	1
Topik 2	4	0,625000	0,053571	1
Topik 9	3	0,000000	0,000000	2
Topik 1	2	0,666667	0,000000	1
Topik 8	2	1,000000	0,000000	2

Jaringan memiliki nilai Average Degree sebesar 1,333 dan Average Weighted Degree sebesar 2,778. Nilai tersebut menunjukkan bahwa rata-rata topik memiliki sekitar satu hubungan dengan topik lainnya, sedangkan kekuatan hubungan rata-rata berdasarkan jumlah kesamaan kata kunci adalah 2,778.

Hasil modularity menghasilkan nilai 0,337 dan membentuk tiga komunitas. Hal ini menunjukkan adanya pengelompokan topik, tetapi hubungan antarkelompok masih tetap terbentuk.

Berdasarkan degree, Topik 3 dan Topik 4 memiliki jumlah hubungan langsung paling banyak dengan nilai



degree sebesar 4. Sementara itu, Topik 1 dan Topik 8 memiliki jumlah hubungan langsung paling sedikit dengan nilai degree sebesar 1.

Pada betweenness centrality, Topik 4 memiliki nilai tertinggi sebesar 0,071429, diikuti Topik 2 sebesar 0,053571 dan Topik 3 sebesar 0,035714. Ketiga topik tersebut memiliki nilai betweenness lebih dari nol sehingga berada pada jalur penghubung antar-topik dalam jaringan.

Pada closeness centrality, nilai tertinggi sebesar 1,000000 dimiliki oleh Topik 4, Topik 6, dan Topik 8.

4.6. Interpretasi Jaringan Antar-Topik

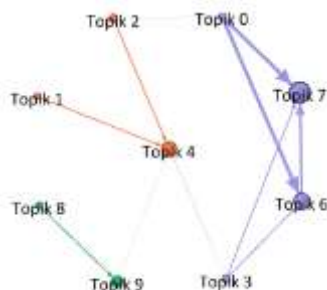
Hasil visualisasi jaringan menunjukkan adanya tiga kelompok topik. Kelompok pertama terdiri atas Topik 0, Topik 3, Topik 6, dan Topik 7 yang berkaitan dengan jadwal, waktu, rute, serta stasiun perjalanan KRL. Hubungan terlihat cukup dekat terutama antara Topik 0, Topik 6, dan Topik 7.

Kelompok kedua terdiri atas Topik 1, Topik 2, dan Topik 4 yang berkaitan dengan jalur perjalanan, gangguan operasional, dan keluhan pengguna. Hubungan tersebut menunjukkan bahwa gangguan perjalanan dapat berkaitan dengan keluhan mengenai informasi maupun penanganan yang belum diberikan.

Kelompok ketiga terdiri atas Topik 8 dan Topik 9 yang berkaitan dengan fasilitas prioritas penumpang. Kedua topik tersebut memiliki hubungan melalui pembahasan mengenai penggunaan fasilitas prioritas di dalam kereta.

Selain hubungan dalam masing-masing kelompok, terdapat pula hubungan antarkelompok. Topik 4 memiliki hubungan dengan topik dari kelompok lain sehingga menjadi salah satu titik yang menghubungkan pembahasan mengenai gangguan dan keluhan dengan topik lainnya. Topik 3 juga memiliki hubungan dengan beberapa topik dalam kelompok jadwal dan rute serta kelompok gangguan dan keluhan.

Topik 1 dan Topik 8 berada lebih luar dari jaringan karena hanya memiliki sedikit hubungan dengan topik lainnya. Secara keseluruhan, hasil visualisasi menunjukkan bahwa topik percakapan pengguna KRL tidak berdiri sendiri, tetapi memiliki keterkaitan melalui kesamaan kata kunci.



Gambar 2. Visualisasi SNA menggunakan Gephi

Dengan demikian, kombinasi BERTopic dan SNA memberikan dua sudut pandang terhadap percakapan pengguna. BERTopic memberikan informasi mengenai topik yang dibahas, sedangkan SNA memberikan gambaran

mengenai hubungan antartopik berdasarkan kesamaan kata kunci.

KESIMPULAN

Berdasarkan hasil penelitian, metode BERTopic mampu memetakan percakapan pengguna KRL Commuter Line pada media sosial X ke dalam sepuluh topik utama. Dari 5.000 data tweet mentah, diperoleh 4.644 tweet setelah preprocessing. Sebanyak 2.879 tweet berhasil dikelompokkan ke dalam topik, sedangkan 1.766 tweet menjadi outlier. Sepuluh topik yang dihasilkan mencakup pembahasan mengenai jadwal dan keterlambatan, jalur dan rute perjalanan, gangguan operasional, keluhan pengguna, gerbong khusus wanita, serta fasilitas prioritas.

Hasil evaluasi BERTopic menghasilkan nilai Topic Coherence sebesar 0,4387 dan Topic Diversity sebesar 0,7897. Nilai tersebut menunjukkan bahwa kata kunci pada topik memiliki keterkaitan yang cukup baik dan cukup beragam.

Penerapan SNA digunakan untuk mengetahui keterkaitan antar-topik berdasarkan kesamaan kata kunci. Jaringan yang terbentuk terdiri atas 9 node dan 12 edge serta terbagi menjadi tiga komunitas dengan nilai modularity sebesar 0,337. Topik 3 dan Topik 4 memiliki jumlah hubungan langsung paling banyak, sedangkan Topik 4 memiliki nilai betweenness centrality tertinggi sebesar 0,071429. Nilai closeness centrality tertinggi sebesar 1,000000 dimiliki oleh Topik 4, Topik 6, dan Topik 8.

Berdasarkan hasil tersebut, BERTopic dapat digunakan untuk memetakan topik percakapan pengguna KRL, sedangkan SNA dapat digunakan untuk menggambarkan keterkaitan antar-topik berdasarkan kesamaan kata kunci. Kombinasi kedua metode tersebut memberikan gambaran mengenai topik yang dibicarakan sekaligus pola hubungan antartopik dalam percakapan pengguna KRL di media sosial X.

SARAN

Penelitian selanjutnya dapat menggunakan jumlah data yang lebih besar dan rentang waktu yang lebih panjang agar hasil pemodelan dapat menggambarkan percakapan pengguna KRL secara lebih luas. Penyesuaian parameter UMAP dan HDBSCAN, penggunaan fitur `reduce_outliers` pada BERTopic, serta penggunaan model embedding lain dapat dipertimbangkan untuk meningkatkan kualitas pemodelan.

Kamus normalisasi bahasa informal dan daftar stopword juga dapat diperluas agar proses preprocessing dapat menangani karakteristik bahasa media sosial dengan lebih baik. Selain itu, pembentukan jaringan dapat dikembangkan menggunakan pendekatan lain seperti kemunculan bersama topik, interaksi antarpengguna, atau kemiripan embedding. Analisis jaringan juga dapat dikembangkan secara temporal untuk melihat perubahan hubungan antartopik dari waktu ke waktu.

Penelitian selanjutnya juga dapat menggabungkan BERTopic dan SNA dengan analisis sentimen sehingga pembahasan tidak hanya mencakup topik dan hubungan antartopik, tetapi juga kecenderungan sentimen pengguna



pada setiap topik. Sumber data juga dapat diperluas dengan menggunakan kanal lain seperti ulasan aplikasi C-Access, komentar Instagram resmi, dan kanal pengaduan lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Negeri Surabaya serta pihak-pihak yang telah memberikan dukungan selama proses pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada seluruh pihak yang membantu dalam proses pengumpulan data, pengolahan data, hingga penyelesaian penelitian.

DAFTAR PUSTAKA

- Albarrak, M., & Pergola, G. (n.d.). "U-BERTopic: An urgency-aware BERT-Topic modeling approach for detecting cyberSecurity issues via social media."
- Anggiyanti, J. N., & Papilaya, F. S. (2026). "Implementasi pemodelan topik BERTopic dan analisis sentimen BERT terhadap ulasan aplikasi SATUSEHAT berbasis natural language processing lanjutan." 11(1), 165–177.
- APJII. (2025). Asosiasi Penyelenggara Jasa Internet Indonesia - Survei.
- Aryani, D. (2024). "Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic pada Artikel Berita Detik.com." Skripsi.
- boyd, D. M., & Ellison, N. B. (2007). "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Cindy Mutia Annur. (2026). "Lampaui Sebelum Pandemi, Jumlah Penumpang KRL di Jabodetabek Tembus 349 Juta Sepanjang 2025."
- DataReportal. (2025). Digital 2025: Indonesia.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). "Indexing by latent semantic analysis." *Journal of the American Society for Information Science*, 41(6), 391–407.
- Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). "Topic Modeling in Embedding Spaces." *Transactions of the Association for Computational Linguistics*, 8, 439–453.
- Egger, R., & Yu, J. (2022). "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts." *Frontiers in Sociology*, 7, 1–16.
- Fayyad, U., & Uthurusamy, R. (1996). "Data Mining and Knowledge Discovery in Databases." *Communications of the ACM*, 39(11), 24–26.
- Freeman, L. C. (1977). "A set of measures of centrality based on betweenness." *Sociometry*, 40(1), 35–41.
- Ghosh, S., Chen, Y., & Dou, W. (2024). "Railroad safety: A systematic analysis of Twitter data." *Case Studies on Transport Policy*, 15, 101154.
- Grootendorst, M. (2022). "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." arXiv:2203.05794.
- Hofmann, T. (1999). "Probabilistic Latent Semantic Analysis." *Advances in Intelligent Systems and Computing*, 410, 149–157.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436–444.
- Lee, D. D., & Seung, H. S. (2000). "Algorithms for Non-negative Matrix Factorization." *Advances in Neural Information Processing Systems*, 11.
- Lim, H., Lee, A., & Eun, S. (2021). "Exploring Nurses' Experience and Grievance: Network Analysis and Topic Modeling using a Social Networking Service." 27(3), 169–180.
- Logan, A. P., LaCasse, P. M., & Lunday, B. J. (2023). "Social network analysis of Twitter interactions: a directed multilayer network approach." *Social Network Analysis and Mining*, 13(1), 65.
- Loilatu, M. J., Irawan, B., Salahudin, S., & Sihidi, I. T. (2021). "Analysis of Twitter's Function as a Media Communication of Public Transportation." *Jurnal Komunikasi*, 13(1), 54.
- Newman, M. E. J., & Girvan, M. (2004). "Finding and evaluating community structure in networks." *Physical Review E*, 69(2), 026113.
- Pullanikkat, R., Poddar, S., Das, A., Jaiswal, T., Singh, V. K., Basu, M., & Ghosh, S. (2024). "Utilizing the Twitter social media to identify transportation-related grievances in Indian cities." *Social Network Analysis and Mining*, 14(1), 118.
- Ramamoorthy, T., Kulothungan, V., & Mappillairaju, B. (2024). "Topic modeling and social network analysis approach to explore diabetes discourse on Twitter in India." *Frontiers in Artificial Intelligence*, 7.
- Reimers, N., & Gurevych, I. (2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." *Proceedings of EMNLP-IJCNLP 2019*.
- Röder, M., Both, A., & Hinneburg, A. (2015). "Exploring the Space of Topic Coherence Measures." *WSDM'15*, 399–408.
- Scott, J. (2017). *Social Network Analysis* (4th ed.). SAGE Publications.
- Tamba, D., & Kholikin, R. A. (2025). "Media Sosial dalam Membangun Opini Publik atas Layanan Pemerintah: Kajian Literatur Interaksi Sosial Digital." *Jejaring Administrasi Publik*, 17(2), 113–126.
- Vaswani, A., et al. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems*, 30.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Zulqarnain, M. I., & W. C. P. (2023). "Topic Modeling dan Social Network Analysis digunakan untuk mencari keterkaitan topik pada Tweet." *Indonesian Journal on Data Science*, 1(1), 47–55.