

Identifikasi Kebutuhan Infrastruktur Transportasi Berbasis Analisis Sentimen Media Sosial di Kabupaten Lamongan

N. Riwiowo^{1*}, Fibia Sentaury Cahyaningrum², Wahyu Tri Prambudi³,
Menik Ariyanti⁴

Institut Teknologi dan Bisnis Ahmad Dahlan, Lamongan

e-mail : n.riwibowo.ad.la@gmail.com¹, fibiasenta22@gmail.com,

wahyuprambudi@ahmaddahlan.ac.id, menikariyanti@ahmaddahlan.ac.id

Abstract

Transportation infrastructure planning requires both technical data and insights into public perception. Social media has emerged as a rich source of public opinions and complaints. This study aims to identify needs and determine priority levels and locations for addressing transportation infrastructure issues in Lamongan Regency based on public sentiment expressed on social media. This quantitative study employs text mining, machine learning, topic modeling, and GIS-based spatial analysis. Social media data were collected using transportation-related keywords and subjected to preprocessing stages (cleansing, case folding, tokenizing, normalization, stopword removal, and stemming). Sentiment analysis was conducted using TF-IDF weighting and the Support Vector Machine (SVM) algorithm. Key issues were identified using Latent Dirichlet Allocation (LDA), then combined with technical data via weighted scoring and visualized using GIS. Analysis of 1,500 data samples revealed a sentiment distribution of 58% negative, 27% neutral, and 15% positive. Five primary issue categories were identified: road damage, traffic congestion, transportation safety, lighting and signage/road markings, and pedestrian facilities. The SVM classification model achieved an accuracy of 87%, precision of 86%, recall of 85%, and an F1-score of 85.5%. These findings demonstrate that processing social media data in conjunction with GIS mapping can effectively provide preliminary indications and map priorities for addressing transportation issues. Consequently, this approach can serve as a supporting tool for infrastructure planning in Lamongan Regency, although field verification remains necessary before final decisions are made.

Keywords: social media, sentiment analysis, transportation infrastructure, Lamongan Regency.

Abstrak

Perencanaan infrastruktur transportasi membutuhkan data teknis sekaligus persepsi masyarakat sebagai pengguna. Media sosial kini menjadi sumber data alternatif yang kaya akan opini dan keluhan publik. Penelitian ini bertujuan mengidentifikasi kebutuhan serta menentukan tingkat dan lokasi prioritas penanganan infrastruktur transportasi di Kabupaten Lamongan berdasarkan sentimen masyarakat di media sosial. Penelitian kuantitatif ini menggunakan metode *text mining*, *machine learning*, *topic modeling*, dan analisis spasial berbasis GIS. Data media sosial dikumpulkan berdasarkan kata kunci terkait transportasi lalu melalui tahapan *preprocessing* (*cleansing*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*). Analisis sentimen dilakukan menggunakan pembobotan TF-IDF dan algoritma *Support Vector Machine* (SVM). Identifikasi isu utama memanfaatkan *Latent Dirichlet Allocation* (LDA), yang kemudian dikombinasikan dengan data teknis melalui skoring berbobot dan divisualisasikan dengan GIS. Hasil analisis dari 1.500 sampel data menunjukkan distribusi sentimen negatif sebesar 58%, netral 27%, dan positif 15%. Lima topik permasalahan utama yang teridentifikasi meliputi kerusakan jalan, kemacetan, keselamatan transportasi, penerangan serta rambu/marka, dan fasilitas pejalan kaki. Model klasifikasi SVM menghasilkan akurasi 87%, *precision* 86%, *recall* 85%, dan *F1-score* 85,5%. Temuan ini membuktikan bahwa pengolahan data media sosial yang dipadukan dengan pemetaan GIS mampu memberikan indikasi awal serta memetakan prioritas penanganan masalah transportasi secara efektif. Dengan demikian, pendekatan ini dapat menjadi instrumen pendukung perencanaan infrastruktur di Kabupaten Lamongan, meski tetap memerlukan verifikasi lapangan sebelum keputusan akhir diambil.

Kata kunci: media sosial, analisis sentimen, infrastruktur transportasi, Kabupaten Lamongan.

PENDAHULUAN

Transportasi merupakan salah satu komponen penting dalam mendukung aktivitas masyarakat dan perkembangan wilayah. Infrastruktur transportasi yang memadai berperan dalam memperlancar mobilitas penduduk, distribusi barang dan jasa, meningkatkan akses terhadap fasilitas pelayanan publik, serta mendukung kegiatan ekonomi daerah. Dalam konteks perencanaan transportasi, kebutuhan masyarakat dan kondisi aktual jaringan transportasi menjadi informasi penting dalam menentukan prioritas pembangunan dan pemeliharaan. Oleh karena itu, proses perencanaan infrastruktur transportasi perlu didukung oleh data yang mampu menggambarkan kondisi lapangan sekaligus kebutuhan pengguna secara aktual.

Kabupaten Lamongan merupakan salah satu wilayah di Jawa Timur yang memiliki aktivitas transportasi cukup tinggi karena adanya kegiatan pertanian, perdagangan, industri, perikanan, jasa, dan aktivitas ekonomi lainnya. Perkembangan berbagai kegiatan tersebut menyebabkan kebutuhan terhadap jaringan transportasi yang aman, nyaman, dan memiliki tingkat pelayanan yang baik semakin meningkat. BPS Kabupaten Lamongan melalui publikasi Kabupaten Lamongan Dalam Angka 2026 menyediakan berbagai data mengenai kondisi geografis, kependudukan, sosial ekonomi, serta data dari berbagai instansi daerah yang dapat menjadi dasar dalam penyusunan perencanaan pembangunan daerah (BPS Kabupaten Lamongan, 2026).

Permasalahan infrastruktur transportasi di daerah tidak hanya berkaitan dengan pembangunan jaringan jalan baru, tetapi juga mencakup kondisi jalan, kemacetan, keselamatan lalu lintas, penerangan jalan, rambu dan marka, fasilitas pejalan kaki, parkir, serta pelayanan transportasi. Dalam kondisi seperti ini, pemerintah membutuhkan informasi yang mampu menunjukkan permasalahan secara lebih cepat dan berorientasi pada kebutuhan pengguna. Data konvensional dari survei dan laporan administratif tetap penting, tetapi pengumpulan data tersebut dapat membutuhkan waktu dan sumber daya yang relatif besar. Oleh karena itu, diperlukan sumber data tambahan yang mampu memberikan gambaran mengenai pengalaman dan persepsi masyarakat secara lebih dinamis.

Perkembangan teknologi digital telah menyebabkan media sosial menjadi salah satu ruang utama bagi masyarakat dalam menyampaikan pendapat, kritik, keluhan, dan pengalaman terhadap pelayanan publik. Dalam bidang transportasi, media sosial berpotensi menjadi sumber informasi mengenai kondisi jaringan transportasi, kemacetan, kecelakaan, kualitas pelayanan, serta permasalahan infrastruktur. Kajian sistematis mengenai pemanfaatan media sosial dalam penelitian transportasi menunjukkan bahwa data media sosial dapat digunakan untuk memperoleh umpan balik pengguna, mengetahui opini terhadap jaringan transportasi, mengidentifikasi kondisi jalan, serta mendeteksi berbagai peristiwa transportasi secara lebih cepat (Zayet et al., 2021).

Pemanfaatan media sosial dalam perencanaan transportasi juga telah didukung oleh berbagai penelitian terdahulu. Penelitian mengenai potensi media sosial dalam kebijakan transportasi menunjukkan bahwa informasi dari media sosial dapat digunakan untuk menangkap pandangan, kebutuhan, dan pengalaman pengguna transportasi secara langsung dan dinamis. Bahkan, klasifikasi otomatis terhadap informasi transportasi dari media sosial dapat membantu mengidentifikasi informasi yang relevan bagi proses perencanaan dan pengambilan kebijakan (Gal-Tzur et al., 2014). Penelitian lain yang mengembangkan kerangka analisis opini masyarakat terhadap pelayanan transportasi menunjukkan bahwa kombinasi analisis sentimen, analisis waktu, dan analisis semantik dapat menghasilkan informasi yang bermanfaat bagi perencanaan dan pengelolaan transportasi (Qi et al., 2020).

Salah satu metode yang dapat digunakan untuk mengolah data tersebut adalah analisis sentimen. Analisis sentimen merupakan pendekatan dalam *text mining* dan *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi kecenderungan opini dalam suatu teks, misalnya menjadi sentimen positif, negatif, atau netral. Penerapan metode ini pada isu pembangunan infrastruktur jalan di Indonesia telah menunjukkan bahwa media sosial Twitter dapat dimanfaatkan untuk mengidentifikasi opini masyarakat mengenai pembangunan jalan. Penelitian tersebut menggunakan tahapan *crawling*, *preprocessing*, *labeling*, dan klasifikasi berbasis BERT serta memperoleh akurasi klasifikasi sebesar 87% (Fikri et al., 2025).

Penelitian mengenai transportasi di Indonesia juga menunjukkan bahwa analisis sentimen dapat digunakan untuk mengevaluasi persepsi masyarakat terhadap pelayanan transportasi. Penelitian yang membandingkan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) dalam menganalisis opini masyarakat terhadap layanan transportasi menunjukkan bahwa SVM menghasilkan performa klasifikasi yang lebih tinggi dibandingkan *Naive Bayes* pada dataset yang digunakan (Nufus & Gata, 2025). Penelitian lain mengenai pelayanan transportasi publik menggunakan *Naive Bayes* untuk mengidentifikasi sentimen masyarakat terhadap layanan transportasi berbasis media sosial (Aldisa et al., 2021). Temuan tersebut menunjukkan bahwa metode *machine learning* memiliki potensi untuk mengolah data opini masyarakat dalam jumlah besar secara lebih sistematis.

Namun demikian, analisis sentimen saja belum cukup untuk menjawab kebutuhan perencanaan infrastruktur transportasi. Sentimen negatif hanya menunjukkan adanya ketidakpuasan atau permasalahan, tetapi belum menjelaskan secara spesifik jenis infrastruktur yang dibutuhkan, lokasi permasalahan, maupun tingkat prioritasnya. Oleh sebab itu, analisis sentimen perlu dikombinasikan dengan analisis topik. Pendekatan kombinasi analisis sentimen dan *topic modeling* telah digunakan dalam penelitian mengenai pembangunan infrastruktur transportasi, termasuk proyek kereta cepat Jakarta–Bandung. Penggunaan *Naive Bayes* dan *Latent Dirichlet Allocation* (LDA) memungkinkan peneliti tidak hanya mengetahui kecenderungan sentimen masyarakat, tetapi juga mengidentifikasi topik atau isu yang paling banyak dibicarakan (Fatah D.A et al., 2024).

Pendekatan tersebut juga telah diterapkan dalam kajian proyek infrastruktur transportasi berskala besar. Penelitian terhadap proyek kereta api cepat Thailand–China menunjukkan bahwa analisis sentimen media sosial dapat digunakan untuk memahami dinamika opini masyarakat terhadap proyek transportasi dari waktu ke waktu (Nokkaew et al., 2024). Sementara itu, penelitian terhadap proyek infrastruktur kereta api HS2 di Inggris menggabungkan *sentiment analysis* dan *topic modeling* untuk mengevaluasi opini publik dan mengidentifikasi tema-tema utama yang muncul dalam percakapan masyarakat. Hasilnya menunjukkan potensi media sosial sebagai sumber informasi untuk mendukung evaluasi proyek infrastruktur dan pengambilan keputusan (Yao & Gillen, 2023).

Perkembangan penelitian terbaru bahkan menunjukkan bahwa analisis sentimen media sosial dapat dikembangkan untuk kebutuhan perencanaan infrastruktur jalan. Penelitian tahun 2025 mengenai *roadway work zones* menggunakan data media sosial dan beberapa model *machine learning* untuk menganalisis persepsi publik, emosi, serta kaitannya dengan kondisi keselamatan dan kecelakaan. Hasil penelitian menunjukkan adanya hubungan temporal antara percakapan di media sosial dengan kejadian kecelakaan, sehingga data media sosial berpotensi menjadi sumber informasi tambahan dalam pengambilan kebijakan keselamatan transportasi (Sayed et al., 2025). Hal ini memperkuat argumentasi bahwa data media sosial tidak hanya dapat digunakan untuk mengetahui opini, tetapi juga dapat dikaitkan dengan kondisi fisik dan kinerja infrastruktur transportasi.

Dalam konteks Indonesia, penelitian terbaru mengenai pemodelan kebijakan transportasi berbasis *social media data mining* juga menunjukkan bahwa data media sosial dapat digunakan sebagai pelengkap data konvensional dalam mendukung kebijakan transportasi. Penelitian tersebut menggabungkan klasifikasi sentimen menggunakan IndoRoBERTa dan *topic modeling* menggunakan LDA untuk mengidentifikasi preferensi serta permasalahan masyarakat terkait transportasi. Hasilnya menunjukkan bahwa analisis media sosial dapat menghasilkan informasi mengenai kebutuhan dan permasalahan pengguna yang dapat diterjemahkan menjadi rekomendasi kebijakan transportasi (Adi Purnama et al., 2025).

Berdasarkan berbagai penelitian tersebut, terdapat peluang untuk mengembangkan pendekatan yang lebih spesifik pada perencanaan kebutuhan infrastruktur transportasi di tingkat kabupaten, khususnya Kabupaten Lamongan. Penelitian terdahulu banyak berfokus pada pelayanan transportasi, proyek transportasi tertentu, pembangunan jalan secara umum, atau evaluasi opini masyarakat. Sementara itu, pemanfaatan sentimen media sosial yang secara khusus diintegrasikan dengan data kondisi infrastruktur, lokasi permasalahan, analisis spasial, dan model prediksi kebutuhan infrastruktur transportasi pada tingkat wilayah kabupaten masih memiliki ruang

pengembangan yang cukup besar. Dengan demikian, penelitian ini memiliki nilai kebaruan dalam menghubungkan data opini masyarakat dengan kebutuhan infrastruktur secara lebih operasional.

Dalam penelitian ini, data media sosial akan dianalisis melalui beberapa tahapan, yaitu pengumpulan data, *preprocessing*, pelabelan sentimen, ekstraksi fitur, klasifikasi sentimen, dan identifikasi topik permasalahan transportasi. Hasil analisis tersebut kemudian dapat dikombinasikan dengan data sekunder mengenai kondisi jalan, volume lalu lintas, kecelakaan, fasilitas transportasi, serta data pembangunan dan pemeliharaan infrastruktur. Selanjutnya, pendekatan *machine learning* dapat digunakan untuk membangun model prediksi tingkat kebutuhan infrastruktur, sedangkan analisis spasial menggunakan GIS dapat digunakan untuk mengetahui persebaran wilayah yang memiliki tingkat kebutuhan rendah, sedang, maupun tinggi.

Dengan pendekatan tersebut, media sosial tidak diposisikan sebagai pengganti data resmi pemerintah, melainkan sebagai sumber data pelengkap yang merepresentasikan pengalaman dan persepsi masyarakat. Hal ini penting karena persepsi pengguna dapat memberikan informasi mengenai permasalahan yang mungkin belum sepenuhnya tercermin dalam data administratif. Penelitian mengenai kepuasan pengguna transportasi juga menunjukkan bahwa integrasi analisis sosial, semantik, dan sentimen dapat membantu memahami kebutuhan serta sensitivitas pengguna terhadap perubahan pelayanan transportasi (El-Diraby et al., 2019).

Berdasarkan uraian tersebut, penelitian “Identifikasi Kebutuhan Infrastruktur Transportasi Berbasis Analisis Sentimen Media Sosial di Kabupaten Lamongan” penting dilakukan sebagai upaya mengembangkan pendekatan perencanaan transportasi yang lebih berbasis data, responsif, dan memanfaatkan perkembangan teknologi informasi. Penelitian ini diharapkan dapat menghasilkan informasi mengenai sentimen masyarakat, jenis permasalahan transportasi yang dominan, lokasi permasalahan, tingkat kebutuhan infrastruktur, serta prioritas penanganan pada masing-masing wilayah di Kabupaten Lamongan. Hasil penelitian selanjutnya dapat menjadi salah satu bahan pendukung bagi pemerintah daerah dalam menentukan prioritas pembangunan, peningkatan, dan pemeliharaan infrastruktur transportasi berdasarkan kombinasi antara kondisi aktual dan kebutuhan yang dirasakan masyarakat.

KAJIAN TEORI

Transportasi merupakan bagian penting dalam mendukung mobilitas penduduk, distribusi barang dan jasa, akses terhadap pelayanan publik, serta aktivitas ekonomi wilayah. Dalam perencanaan transportasi, kebutuhan infrastruktur tidak hanya ditentukan oleh pertumbuhan aktivitas dan volume pergerakan, tetapi juga oleh kondisi aktual jaringan serta kebutuhan pengguna. Infrastruktur transportasi mencakup jalan, jembatan, fasilitas keselamatan, penerangan jalan, rambu dan marka, fasilitas pejalan kaki, parkir, serta fasilitas pelayanan transportasi lainnya. Oleh karena itu, perencanaan infrastruktur membutuhkan data yang mampu menggambarkan kondisi teknis sekaligus pengalaman dan persepsi masyarakat sebagai pengguna transportasi.

Dalam konteks Kabupaten Lamongan, kebutuhan terhadap infrastruktur transportasi berkaitan dengan perkembangan kegiatan pertanian, perdagangan, industri, perikanan, jasa, dan kegiatan ekonomi lainnya. Perkembangan aktivitas tersebut meningkatkan kebutuhan terhadap jaringan transportasi yang aman, nyaman, dan memiliki tingkat pelayanan yang baik. Data mengenai kondisi geografis, kependudukan, sosial ekonomi, serta informasi sektoral yang tersedia melalui publikasi pemerintah dapat digunakan sebagai data pendukung dalam memahami karakteristik wilayah dan perencanaan pembangunan transportasi.

Perencanaan infrastruktur transportasi pada dasarnya membutuhkan informasi yang akurat mengenai kondisi jaringan dan kebutuhan pengguna. Data konvensional seperti survei lapangan, data lalu lintas, laporan kecelakaan, dan data administratif tetap menjadi sumber utama. Namun, perkembangan teknologi informasi memberikan peluang untuk memanfaatkan sumber data alternatif yang lebih dinamis. Media sosial menjadi salah satu sumber tersebut karena mampu merekam pengalaman, keluhan, kritik, dan pendapat masyarakat dalam waktu relatif cepat. Dengan

demikian, pendekatan berbasis data dapat menggabungkan data teknis dengan informasi sosial untuk menghasilkan perencanaan yang lebih responsif terhadap kondisi masyarakat.

Media sosial merupakan platform yang memungkinkan masyarakat menyampaikan informasi, pengalaman, opini, kritik, dan keluhan secara langsung. Dalam bidang transportasi, data media sosial dapat mengandung informasi mengenai kemacetan, kecelakaan, kerusakan jalan, kualitas pelayanan, serta permasalahan fasilitas transportasi. Kajian sistematis yang dilakukan Zayet et al. (2021) menunjukkan bahwa media sosial dapat dimanfaatkan untuk memperoleh umpan balik pengguna, mengetahui opini masyarakat terhadap jaringan transportasi, mengidentifikasi kondisi jalan, dan mendeteksi kejadian transportasi. Gal-Tzur et al. (2014) juga menunjukkan bahwa informasi media sosial memiliki potensi untuk menangkap pandangan, kebutuhan, dan pengalaman pengguna dalam mendukung kebijakan transportasi.

Qi et al. (2020) selanjutnya menunjukkan bahwa pengolahan data media sosial melalui analisis sentimen, analisis waktu, dan analisis semantik dapat menghasilkan informasi yang bermanfaat untuk memahami opini masyarakat terhadap pelayanan transportasi. Dengan demikian, media sosial dapat ditempatkan sebagai sumber data pelengkap, bukan sebagai pengganti data teknis dan administratif pemerintah. Posisi tersebut penting karena data media sosial lebih merepresentasikan pengalaman dan persepsi pengguna, sedangkan data teknis menggambarkan kondisi objektif infrastruktur.

Data media sosial umumnya berbentuk teks tidak terstruktur sehingga memerlukan proses pengolahan sebelum dapat dianalisis. *Text mining* merupakan pendekatan untuk memperoleh informasi dan pola dari kumpulan teks, sedangkan *Natural Language Processing* (NLP) digunakan untuk membantu komputer mengolah dan memahami bahasa manusia. Dalam penelitian ini, *text mining* dan NLP digunakan untuk mengubah unggahan media sosial menjadi data yang dapat dianalisis secara kuantitatif.

Tahapan pengolahan dapat meliputi pengumpulan data, *case folding*, *cleansing*, *tokenizing*, normalisasi kata, *stopword removal*, *stemming*, pelabelan, dan ekstraksi fitur. Tahapan tersebut diperlukan karena data media sosial dapat mengandung bahasa informal, singkatan, kesalahan penulisan, tanda baca, dan istilah yang beragam. Berdasarkan penelitian yang dirangkum dalam sumber penelitian ini, tahapan *crawling*, *preprocessing*, *labeling*, dan klasifikasi telah digunakan dalam penelitian analisis opini masyarakat terhadap pembangunan jalan di Indonesia.

Analisis sentimen merupakan metode dalam *text mining* dan NLP yang digunakan untuk mengidentifikasi kecenderungan opini dalam suatu teks. Pada penelitian ini, sentimen dapat dikategorikan menjadi positif, negatif, dan netral. Sentimen positif menunjukkan penilaian atau pengalaman yang cenderung baik, sentimen negatif menunjukkan keluhan atau ketidakpuasan, sedangkan sentimen netral lebih bersifat informatif tanpa kecenderungan opini yang kuat.

Analisis sentimen relevan digunakan dalam penelitian ini karena keluhan masyarakat mengenai transportasi dapat menjadi indikator awal adanya permasalahan infrastruktur. Penelitian Fikri et al. (2025), misalnya, memanfaatkan data Twitter untuk menganalisis opini masyarakat mengenai pembangunan jalan dengan tahapan *crawling*, *preprocessing*, *labeling*, dan klasifikasi berbasis BERT serta memperoleh akurasi klasifikasi sebesar 87%. Temuan tersebut menunjukkan bahwa data media sosial dapat diolah secara sistematis untuk mengetahui opini masyarakat mengenai infrastruktur jalan.

Machine learning digunakan untuk mempelajari pola dari data dan menghasilkan klasifikasi berdasarkan pola tersebut. Dalam analisis sentimen, algoritma *machine learning* dapat digunakan untuk menentukan apakah suatu teks termasuk sentimen positif, negatif, atau netral. Beberapa algoritma yang relevan adalah *Naive Bayes*, *Support Vector Machine* (SVM), dan BERT/IndoBERT.

Naive Bayes memiliki kelebihan dari sisi kesederhanaan dan efisiensi dalam pengolahan data teks, sedangkan SVM dikenal mampu menangani data berdimensi tinggi seperti data hasil ekstraksi fitur teks. Penelitian Nufus dan Gata (2025) yang membandingkan *Naive Bayes* dan SVM dalam analisis opini pelayanan transportasi menunjukkan bahwa SVM menghasilkan performa

klasifikasi yang lebih tinggi pada dataset yang digunakan. Sementara itu, *Naive Bayes* juga telah digunakan dalam penelitian mengenai sentimen masyarakat terhadap pelayanan transportasi publik.

Analisis sentimen belum dapat menjelaskan secara lengkap apa yang menjadi sumber permasalahan. Sentimen negatif, misalnya, hanya menunjukkan adanya ketidakpuasan, tetapi belum menjelaskan apakah keluhan tersebut berkaitan dengan jalan rusak, kemacetan, kecelakaan, penerangan, rambu, trotoar, atau fasilitas lainnya. Oleh karena itu, analisis sentimen perlu dikombinasikan dengan *topic modeling* untuk mengidentifikasi tema atau topik utama dalam data.

Salah satu metode *topic modeling* yang dapat digunakan adalah *Latent Dirichlet Allocation* (LDA). Metode ini dapat mengelompokkan kata-kata yang sering muncul bersama ke dalam sejumlah topik sehingga peneliti dapat memberikan interpretasi terhadap isu yang sedang dibicarakan masyarakat. Penelitian Fatah D.A. et al. (2024) menunjukkan bahwa kombinasi *Naive Bayes* dan LDA dapat digunakan untuk mengetahui sentimen sekaligus mengidentifikasi topik yang banyak dibicarakan masyarakat dalam konteks pembangunan infrastruktur transportasi.

Kombinasi analisis sentimen dan *topic modeling* menjadi penting karena kedua metode memberikan informasi yang berbeda tetapi saling melengkapi. Analisis sentimen menjawab bagaimana persepsi masyarakat, sedangkan *topic modeling* menjawab permasalahan apa yang menjadi perhatian masyarakat. Dengan menggabungkan keduanya, penelitian dapat mengidentifikasi, misalnya, bahwa sentimen negatif dominan pada topik kerusakan jalan atau bahwa keluhan mengenai kemacetan menjadi permasalahan yang paling sering muncul.

Pendekatan tersebut telah diterapkan pada berbagai proyek infrastruktur transportasi. Penelitian Nokkaew et al. (2024) mengenai proyek kereta api cepat Thailand–China menunjukkan bahwa analisis sentimen dapat digunakan untuk memahami dinamika opini masyarakat dari waktu ke waktu. Sementara itu, penelitian Yao dan Gillen (2023) terhadap proyek HS2 di Inggris menggabungkan *sentiment analysis* dan *topic modeling* untuk mengevaluasi opini publik serta mengidentifikasi tema utama dalam percakapan masyarakat.

Dalam penelitian ini, kebutuhan infrastruktur transportasi dipahami sebagai tingkat kebutuhan terhadap pembangunan, peningkatan, rehabilitasi, pemeliharaan, atau penanganan fasilitas transportasi berdasarkan kombinasi permasalahan yang dirasakan masyarakat dan kondisi aktual infrastruktur. Dengan demikian, prediksi tidak hanya didasarkan pada jumlah sentimen negatif, tetapi juga mempertimbangkan jenis permasalahan, frekuensi keluhan, lokasi, data kondisi infrastruktur, serta data transportasi lainnya.

Pendekatan prediktif dapat menggunakan *machine learning* untuk menghubungkan variabel-variabel tersebut dengan tingkat kebutuhan, misalnya rendah, sedang, dan tinggi. Data teknis seperti kondisi jalan, volume lalu lintas, kecelakaan, fasilitas transportasi, serta data pembangunan dan pemeliharaan dapat digunakan sebagai variabel pendukung. Kerangka tersebut sesuai dengan arah penelitian yang menggabungkan data media sosial dengan data sekunder dan *machine learning* untuk membangun model prediksi kebutuhan infrastruktur.

Permasalahan transportasi selalu memiliki dimensi lokasi. Keluhan mengenai jalan rusak atau kemacetan akan memiliki nilai perencanaan yang lebih tinggi apabila dapat diketahui lokasi terjadinya permasalahan tersebut. Oleh karena itu, hasil analisis sentimen dan topic modeling dapat dikombinasikan dengan *Geographic Information System* (GIS) untuk memetakan persebaran permasalahan dan tingkat kebutuhan infrastruktur.

Analisis spasial dapat menghasilkan informasi mengenai wilayah dengan tingkat kebutuhan rendah, sedang, atau tinggi. Pemetaan tersebut dapat membantu mengubah hasil analisis teks menjadi informasi yang lebih mudah digunakan dalam proses pengambilan keputusan. Dalam rancangan penelitian ini, GIS berfungsi sebagai tahap lanjutan untuk menunjukkan lokasi dan distribusi kebutuhan infrastruktur transportasi di Kabupaten Lamongan.

Teori utama yang mendasari penelitian ini adalah bahwa data sosial dan data teknis perlu digunakan secara komplementer. Data media sosial memberikan informasi mengenai persepsi, pengalaman, dan keluhan masyarakat, sedangkan data teknis memberikan informasi mengenai

kondisi objektif infrastruktur. Integrasi keduanya dapat mengurangi kelemahan apabila masing-masing sumber digunakan secara terpisah.

Sebagai contoh, apabila suatu ruas jalan memperoleh banyak sentimen negatif mengenai kerusakan dan hasil survei teknis juga menunjukkan kondisi jalan yang buruk, maka indikasi kebutuhan penanganan menjadi lebih kuat. Sebaliknya, apabila sentimen negatif tinggi tetapi kondisi fisik relatif baik, diperlukan pemeriksaan lebih lanjut untuk mengetahui apakah permasalahan berkaitan dengan kemacetan, pelayanan, keselamatan, atau faktor lainnya. Dengan pendekatan tersebut, media sosial berfungsi sebagai indikator kebutuhan masyarakat, bukan sebagai satu-satunya dasar penentuan pembangunan.

Berdasarkan kajian teori tersebut, penelitian ini dapat dirumuskan dalam suatu alur konseptual: media sosial → *text mining* dan NLP → *preprocessing* → analisis sentimen → *topic modeling* → integrasi data teknis → *machine learning* → prediksi tingkat kebutuhan → analisis spasial/GIS → prioritas infrastruktur transportasi. Alur tersebut memungkinkan opini masyarakat yang awalnya berbentuk teks tidak terstruktur diubah menjadi informasi yang dapat mendukung perencanaan.

Dengan demikian, penelitian “Identifikasi Kebutuhan Infrastruktur Transportasi Berbasis Analisis Sentimen Media Sosial di Kabupaten Lamongan” memiliki landasan teori pada konsep transportasi, perencanaan infrastruktur berbasis data, media sosial sebagai sumber data alternatif, *text mining*, NLP, analisis sentimen, *machine learning*, *topic modeling*, dan analisis spasial. Pendekatan tersebut diharapkan mampu menghasilkan informasi mengenai sentimen masyarakat, jenis permasalahan, lokasi permasalahan, tingkat kebutuhan, dan prioritas penanganan infrastruktur transportasi di Kabupaten Lamongan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *text mining*, analisis sentimen, dan *machine learning* yang didukung oleh analisis spasial. Pendekatan ini dipilih karena penelitian bertujuan mengolah opini masyarakat di media sosial menjadi informasi terukur mengenai permasalahan dan kebutuhan infrastruktur transportasi di Kabupaten Lamongan. Data media sosial digunakan sebagai sumber informasi pelengkap yang selanjutnya dikombinasikan dengan data teknis transportasi dan kondisi infrastruktur.

Data penelitian berupa unggahan media sosial yang berkaitan dengan transportasi di Kabupaten Lamongan, seperti kerusakan jalan, kemacetan, kecelakaan, penerangan jalan, rambu dan marka, fasilitas pejalan kaki, serta pelayanan transportasi. Data dikumpulkan berdasarkan kata kunci tertentu, kemudian diseleksi berdasarkan relevansi lokasi, topik, dan periode penelitian. Data yang diperoleh selanjutnya melalui proses *preprocessing* yang meliputi *cleansing*, *case folding*, *tokenizing*, *normalisasi*, *stopword removal*, dan *stemming* sehingga diperoleh data teks yang siap dianalisis.

Analisis sentimen dilakukan dengan mengklasifikasikan opini masyarakat ke dalam sentimen positif, netral, dan negatif. Metode TF-IDF dan *Support Vector Machine* (SVM) dipilih sebagai metode utama karena sesuai untuk klasifikasi data teks dan relatif efektif dalam menangani data berdimensi tinggi. Kinerja model dievaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Selanjutnya, *Latent Dirichlet Allocation* (LDA) digunakan untuk mengidentifikasi topik utama yang muncul dalam percakapan masyarakat sehingga dapat diketahui jenis permasalahan transportasi yang paling banyak mendapat perhatian.

Hasil analisis sentimen dan topik kemudian diintegrasikan dengan data teknis infrastruktur, seperti kondisi jalan, volume lalu lintas, kecelakaan, serta ketersediaan fasilitas transportasi. Berdasarkan integrasi tersebut, tingkat kebutuhan infrastruktur ditentukan menggunakan scoring berbobot dengan mempertimbangkan intensitas sentimen negatif, frekuensi keluhan, jenis permasalahan, kondisi infrastruktur, dan karakteristik lalu lintas. Hasilnya diklasifikasikan menjadi tingkat kebutuhan rendah, sedang, dan tinggi.

Tahap akhir dilakukan menggunakan analisis spasial berbasis GIS untuk memetakan persebaran permasalahan dan tingkat kebutuhan infrastruktur transportasi. Hasil pemetaan kemudian divalidasi dengan kondisi lapangan dan/atau data dari instansi terkait. Dengan metode tersebut, penelitian diharapkan menghasilkan identifikasi permasalahan, tingkat kebutuhan, lokasi prioritas, dan rekomendasi penanganan infrastruktur transportasi di Kabupaten Lamongan secara lebih objektif dan berbasis data.

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan mengidentifikasi kebutuhan infrastruktur transportasi di Kabupaten Lamongan dengan memanfaatkan informasi yang diperoleh dari media sosial. Data media sosial digunakan untuk menangkap persepsi masyarakat terhadap kondisi transportasi, sedangkan data teknis digunakan untuk memberikan gambaran kondisi objektif infrastruktur. Pendekatan tersebut memungkinkan permasalahan transportasi dianalisis tidak hanya berdasarkan kondisi fisik, tetapi juga berdasarkan pengalaman dan keluhan masyarakat sebagai pengguna transportasi.

Tahapan analisis dimulai dari pengumpulan data media sosial, seleksi data, *preprocessing*, klasifikasi sentimen, identifikasi topik permasalahan, integrasi dengan data teknis, penentuan tingkat kebutuhan, hingga pemetaan menggunakan GIS. Hasil akhir penelitian diarahkan untuk menghasilkan informasi mengenai jenis permasalahan transportasi, kecenderungan sentimen masyarakat, lokasi permasalahan, tingkat kebutuhan, dan prioritas penanganan infrastruktur.

Berdasarkan proses pengumpulan data, diperoleh 2.000 unggahan yang berkaitan dengan transportasi di Kabupaten Lamongan. Setelah dilakukan seleksi terhadap data duplikat, iklan, spam, unggahan yang tidak berkaitan dengan transportasi, serta data yang tidak memiliki relevansi lokasi penelitian, diperoleh 1.500 data yang dapat digunakan untuk analisis.

Data tersebut kemudian melalui tahapan *preprocessing* yang meliputi *case folding*, *cleansing*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*. Tahapan ini menghasilkan data teks yang lebih seragam dan mengurangi kata-kata yang tidak memiliki kontribusi terhadap proses klasifikasi. Seperti kalimat “Jalan di daerah ini rusak banget!!!” setelah *preprocessing* dapat diubah menjadi bentuk yang lebih sederhana seperti “jalan daerah rusak”.

Hasil *preprocessing* menunjukkan bahwa kualitas data media sosial sangat dipengaruhi oleh karakteristik bahasa yang digunakan masyarakat. Penggunaan bahasa informal, singkatan, istilah lokal, kesalahan pengetikan, dan penggunaan emoji menjadi tantangan dalam proses klasifikasi. Oleh karena itu, normalisasi kata menjadi tahap penting karena kata-kata yang memiliki makna sama tetapi ditulis dengan bentuk berbeda perlu diseragamkan agar dapat dikenali oleh model.

Berdasarkan hasil klasifikasi, data kemudian dikelompokkan menjadi sentimen positif, netral, dan negatif. Dari 1.500 data yang dianalisis diperoleh 870 data atau 58% sentimen negatif, 405 data atau 27% sentimen netral, dan 225 data atau 15% sentimen positif. Dominannya sentimen negatif menunjukkan bahwa media sosial lebih banyak digunakan masyarakat untuk menyampaikan keluhan atau permasalahan dibandingkan memberikan apresiasi. Kondisi ini dapat terjadi karena pengguna cenderung menggunakan media sosial sebagai sarana menyampaikan pengalaman ketika menghadapi gangguan atau ketidaknyamanan dalam perjalanan. Sentimen negatif yang dominan tidak secara otomatis menunjukkan bahwa seluruh infrastruktur transportasi Kabupaten Lamongan berada dalam kondisi buruk. Media sosial memiliki karakteristik *self-selection*, yaitu masyarakat yang mengalami masalah atau memiliki pengalaman tertentu lebih terdorong untuk menyampaikan keluhan. Oleh karena itu, persentase sentimen harus dipahami sebagai indikator persepsi masyarakat, bukan sebagai pengukuran langsung kondisi teknis infrastruktur. Sebaliknya, sentimen positif dapat memberikan informasi mengenai aspek transportasi yang dianggap sudah baik oleh masyarakat. Misalnya, komentar yang memberikan apresiasi terhadap jalan yang baru diperbaiki, kondisi lalu lintas yang lancar, atau fasilitas penerangan yang sudah tersedia dapat digunakan sebagai informasi evaluasi keberhasilan pembangunan. Klasifikasi sentimen dilakukan menggunakan kombinasi TF-IDF dan *Support Vector Machine* (SVM). TF-IDF digunakan untuk

mengubah teks menjadi representasi numerik berdasarkan tingkat kepentingan suatu kata dalam dokumen, sedangkan SVM digunakan untuk menentukan kelas sentimen berdasarkan pola data yang telah dipelajari. Hasil pengujian model menunjukkan *accuracy* sebesar 87%. Nilai tersebut menunjukkan bahwa sebagian besar data pengujian dapat diklasifikasikan dengan benar oleh model. Untuk memberikan evaluasi yang lebih komprehensif, pengujian juga dilakukan menggunakan *precision*, *recall*, dan *F1-score*. Berdasarkan pengujian model diperoleh hasil sebagai berikut: *precision* 86%, *recall* 85%, dan *F1-score* 85,5%. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang cukup baik untuk membedakan sentimen masyarakat. Namun, beberapa kesalahan klasifikasi tetap ditemukan. Kesalahan umumnya terjadi pada kalimat yang mengandung sarkasme, makna ganda, bahasa daerah, atau pernyataan yang tidak secara eksplisit menunjukkan sentimen. Sebagai contoh, kalimat “Mantap sekali jalannya, lubangnya sampai tidak kelihatan karena gelap” secara literal mengandung kata positif “mantap”, tetapi konteks keseluruhannya merupakan kritik terhadap kondisi jalan. Temuan tersebut menunjukkan bahwa peningkatan kualitas model tidak hanya bergantung pada algoritma, tetapi juga pada kualitas dataset, proses pelabelan, normalisasi bahasa, dan jumlah data pelatihan. Untuk penelitian lanjutan, model berbasis transformer seperti IndoBERT dapat dibandingkan dengan SVM untuk mengetahui apakah pemahaman konteks bahasa Indonesia dapat meningkatkan akurasi.

Analisis *Latent Dirichlet Allocation* (LDA) dilakukan untuk mengetahui topik yang paling sering dibicarakan masyarakat. Berdasarkan hasil pemodelan, diperoleh lima kelompok topik utama, yaitu: a) kerusakan jalan dan kondisi perkerasan; b) kemacetan dan kepadatan lalu lintas; c) keselamatan dan kecelakaan lalu lintas; d) penerangan jalan, rambu, dan marka; e) fasilitas pejalan kaki dan fasilitas pendukung transportasi. Topik kerusakan jalan ditandai dengan kemunculan kata seperti jalan, rusak, lubang, aspal, berlubang, dan perbaikan. Hal tersebut menunjukkan bahwa kondisi fisik jalan merupakan salah satu aspek yang secara langsung dirasakan oleh pengguna. Topik kemacetan ditandai oleh kata macet, kendaraan, padat, antre, lalu lintas, dan simpang. Permasalahan tersebut tidak selalu menunjukkan kekurangan kondisi fisik jalan, tetapi dapat berkaitan dengan peningkatan volume lalu lintas, kapasitas jalan, aktivitas samping jalan, pengaturan simpang, maupun manajemen lalu lintas. Sementara itu, topik keselamatan ditandai oleh kata kecelakaan, tabrak, motor, bahaya, lampu, dan rambu. Kemunculan topik tersebut menunjukkan bahwa masyarakat juga memperhatikan aspek keselamatan sebagai bagian dari kebutuhan infrastruktur transportasi.

Analisis yang lebih mendalam dilakukan dengan menghubungkan sentimen dengan topik permasalahan. Hasilnya menunjukkan bahwa topik kerusakan jalan memiliki proporsi sentimen negatif paling tinggi, yaitu sekitar 75%. Topik kemacetan memiliki sentimen negatif sekitar 70%, sedangkan topik keselamatan mencapai 68%. Temuan ini penting karena menunjukkan bahwa tidak semua permasalahan transportasi memiliki tingkat ketidakpuasan yang sama. Topik dengan jumlah pembicaraan tinggi tetapi sentimen negatif rendah dapat menunjukkan bahwa masyarakat banyak membicarakannya tetapi tidak selalu merasa sangat tidak puas. Sebaliknya, topik dengan frekuensi lebih rendah tetapi dominasi sentimen negatif tinggi dapat menjadi indikasi permasalahan yang lebih serius. Dengan demikian, penentuan prioritas tidak sebaiknya hanya berdasarkan jumlah unggahan, tetapi perlu mempertimbangkan kombinasi antara frekuensi, sentimen, topik, lokasi, dan kondisi teknis. Pendekatan ini membuat hasil penelitian lebih relevan untuk kebutuhan perencanaan infrastruktur.

Hasil analisis media sosial selanjutnya dibandingkan dengan data teknis. Variabel yang dapat digunakan antara lain kondisi jalan, volume lalu lintas, tingkat kecelakaan, keberadaan fasilitas keselamatan, dan data pemeliharaan atau pembangunan infrastruktur. Suatu ruas jalan yang memiliki tingkat keluhan tinggi mengenai kerusakan jalan dan data lapangan menunjukkan adanya kerusakan perkerasan yang signifikan, maka kedua sumber informasi tersebut menunjukkan pola yang konsisten. Kondisi tersebut memberikan dasar yang lebih kuat untuk memasukkan ruas tersebut sebagai kandidat prioritas rehabilitasi. Sebaliknya, apabila suatu lokasi memiliki banyak keluhan mengenai kemacetan tetapi kondisi fisik jalan masih baik, maka kebutuhan penanganan mungkin bukan berupa pembangunan jalan baru. Permasalahan dapat lebih tepat ditangani melalui manajemen lalu lintas, pengaturan simpang, pengendalian parkir, atau peningkatan kapasitas pada titik tertentu. Hasil tersebut memperlihatkan bahwa media sosial mempunyai fungsi penting sebagai

sistem deteksi awal permasalahan transportasi, sedangkan survei teknis digunakan untuk memastikan jenis dan tingkat permasalahan yang sebenarnya.

Berdasarkan hasil integrasi data, kebutuhan infrastruktur dapat dihitung menggunakan metode skoring berbobot. Indikator yang digunakan terdiri atas intensitas sentimen negatif, frekuensi keluhan, tingkat permasalahan, kondisi fisik infrastruktur, volume lalu lintas, dan tingkat kecelakaan. Setiap indikator diberi skor 1–5. Nilai yang semakin tinggi menunjukkan kondisi yang semakin membutuhkan penanganan. Selanjutnya setiap indikator diberikan bobot sesuai tingkat kepentingannya. Hasil perkalian antara skor dan bobot kemudian dijumlahkan untuk memperoleh nilai indeks kebutuhan. Wilayah dengan nilai indeks tinggi dikategorikan sebagai prioritas tinggi, sedangkan nilai menengah dan rendah dikategorikan sebagai prioritas sedang dan rendah. Pendekatan ini memungkinkan peneliti mengubah data kualitatif berupa opini masyarakat menjadi indikator kuantitatif yang dapat digunakan dalam proses penentuan prioritas.

Hasil analisis selanjutnya divisualisasikan menggunakan GIS. Pemetaan dilakukan dengan menghubungkan lokasi permasalahan dengan nilai sentimen dan indeks kebutuhan infrastruktur. Hasil pemetaan menunjukkan adanya beberapa lokasi yang memiliki konsentrasi keluhan tinggi. Lokasi tersebut kemudian dikategorikan berdasarkan tingkat kebutuhan. Zona prioritas tinggi menunjukkan wilayah yang memiliki kombinasi sentimen negatif tinggi, frekuensi permasalahan tinggi, dan kondisi teknis yang kurang baik. Zona prioritas sedang menunjukkan adanya permasalahan tetapi tingkat urgensinya relatif lebih rendah, sedangkan zona prioritas rendah menunjukkan kondisi relatif baik atau memiliki intensitas keluhan yang rendah. Peta prioritas ini memiliki nilai praktis bagi perencanaan transportasi karena dapat memberikan gambaran lokasi yang perlu diverifikasi dan ditangani terlebih dahulu. Hasil tersebut juga dapat digunakan sebagai dasar penyusunan program pemeliharaan dan peningkatan infrastruktur secara lebih terarah.

Berdasarkan keseluruhan hasil analisis, kerusakan jalan menjadi salah satu permasalahan yang paling relevan untuk dijadikan indikator kebutuhan infrastruktur. Tingginya keluhan mengenai jalan berlubang, permukaan rusak, dan kebutuhan perbaikan menunjukkan bahwa kondisi fisik jalan masih menjadi perhatian masyarakat. Namun, kebutuhan infrastruktur transportasi tidak terbatas pada pembangunan atau perbaikan jalan. Tingginya pembahasan mengenai kemacetan menunjukkan perlunya mempertimbangkan aspek kapasitas dan manajemen lalu lintas. Sementara itu, munculnya keluhan mengenai kecelakaan, rambu, marka, dan penerangan menunjukkan bahwa peningkatan keselamatan juga merupakan bagian penting dari kebutuhan infrastruktur.

Dengan demikian, hasil penelitian menunjukkan bahwa kebutuhan infrastruktur harus dipahami secara multidimensional. Infrastruktur transportasi yang baik tidak hanya ditentukan oleh kondisi perkerasan jalan, tetapi juga oleh kapasitas, keselamatan, kenyamanan, aksesibilitas, dan kualitas fasilitas pendukung. Temuan penelitian menunjukkan adanya peluang untuk mengembangkan proses perencanaan transportasi yang lebih responsif terhadap masyarakat. Data konvensional seperti volume lalu lintas dan kondisi jalan memberikan informasi objektif, sedangkan media sosial memberikan informasi mengenai pengalaman pengguna yang tidak selalu tercatat dalam survei formal. Kombinasi kedua jenis data tersebut dapat menghasilkan sistem informasi yang lebih komprehensif. Laporan teknis menunjukkan bahwa suatu ruas jalan memiliki kondisi sedang, tetapi media sosial menunjukkan peningkatan keluhan yang signifikan. Kondisi tersebut dapat menjadi sinyal bagi pemerintah untuk melakukan pemeriksaan lebih lanjut sebelum masalah berkembang menjadi lebih serius. Dengan demikian, pemanfaatan media sosial dapat diarahkan sebagai data pendukung dalam proses monitoring dan evaluasi, bukan sebagai pengganti metode perencanaan transportasi konvensional.

Meskipun memberikan informasi yang bermanfaat, penggunaan media sosial memiliki beberapa keterbatasan. Pertama, tidak semua masyarakat menggunakan media sosial sehingga data tidak sepenuhnya merepresentasikan seluruh populasi Kabupaten Lamongan. Kedua, pengguna media sosial cenderung lebih aktif menyampaikan pengalaman negatif sehingga dapat menyebabkan dominasi sentimen negatif. Ketiga, tidak semua unggahan mencantumkan lokasi secara jelas sehingga analisis spasial dapat mengalami keterbatasan. Keempat, bahasa informal, sarkasme, singkatan, dan bahasa lokal dapat menurunkan akurasi klasifikasi. Oleh karena itu, hasil analisis media sosial perlu dikombinasikan dengan data teknis dan validasi lapangan.

Secara keseluruhan, hasil penelitian menunjukkan bahwa analisis sentimen media sosial dapat digunakan untuk mengidentifikasi permasalahan dan memberikan indikasi kebutuhan infrastruktur transportasi di Kabupaten Lamongan. Analisis sentimen memberikan informasi mengenai tingkat persepsi masyarakat, sedangkan LDA memberikan informasi mengenai jenis permasalahan. Integrasi dengan data teknis memungkinkan hasil tersebut diverifikasi, sementara GIS memberikan gambaran mengenai lokasi prioritas.

Kontribusi utama penelitian terletak pada perubahan data media sosial yang semula berupa teks tidak terstruktur menjadi informasi yang dapat mendukung pengambilan keputusan. Alur hasil penelitian ini dapat dirumuskan sebagai berikut: Media Sosial → *Preprocessing* → Sentimen → Topik Permasalahan → Integrasi Data Teknis → Indeks Kebutuhan → Pemetaan GIS → Prioritas Infrastruktur. Dengan pendekatan tersebut, pemerintah daerah dapat memperoleh informasi awal mengenai lokasi dan jenis permasalahan transportasi yang paling dirasakan masyarakat. Namun, keputusan akhir mengenai pembangunan atau peningkatan infrastruktur tetap harus mempertimbangkan hasil survei teknis, ketersediaan anggaran, standar teknis, tingkat pelayanan jalan, serta kebijakan pembangunan daerah.

KESIMPULAN

Berdasarkan hasil dan pembahasan penelitian “Identifikasi Kebutuhan Infrastruktur Transportasi Berbasis Analisis Sentimen Media Sosial di Kabupaten Lamongan”, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Hasil penelitian menunjukkan bahwa integrasi media sosial, analisis sentimen, LDA, *machine learning*, data teknis, dan GIS dapat membentuk pendekatan berbasis data untuk mendukung perencanaan infrastruktur transportasi di Kabupaten Lamongan. Alur penelitian menghasilkan informasi mulai dari 1.500 data terpilih, 58% sentimen negatif, 27% netral, 15% positif, lima topik utama, akurasi model SVM 87%, hingga klasifikasi prioritas kebutuhan rendah, sedang, dan tinggi. Pendekatan tersebut dapat berfungsi sebagai *early indicator* untuk membantu pemerintah menemukan permasalahan yang perlu diverifikasi di lapangan.
2. Topik kerusakan jalan dan kemacetan menjadi permasalahan yang paling menonjol dalam pembahasan masyarakat. Topik kerusakan jalan memiliki sekitar 75% sentimen negatif, sedangkan kemacetan sekitar 70% dan keselamatan sekitar 68%. Angka tersebut menunjukkan bahwa permasalahan fisik jalan, kelancaran pergerakan, dan keselamatan merupakan aspek yang memiliki tingkat ketidakpuasan relatif tinggi. Dengan demikian, kebutuhan infrastruktur transportasi di Kabupaten Lamongan tidak hanya berkaitan dengan pembangunan jalan baru, tetapi juga mencakup rehabilitasi jalan, pemeliharaan perkerasan, penanganan titik kemacetan, peningkatan fasilitas keselamatan, penerangan, rambu, marka, dan fasilitas pejalan kaki.
3. Integrasi antara analisis sentimen dan data teknis memberikan hasil yang lebih kuat dibandingkan penggunaan media sosial secara tunggal. Sentimen negatif yang tinggi belum dapat dijadikan bukti bahwa suatu ruas jalan harus segera diperbaiki. Namun, apabila keluhan masyarakat mengenai kerusakan jalan memiliki frekuensi tinggi dan pada lokasi yang sama terdapat data teknis yang menunjukkan kondisi jalan buruk, maka tingkat kebutuhan penanganan menjadi lebih kuat.
4. Kebutuhan infrastruktur transportasi bersifat multidimensional. Permasalahan tidak hanya berkaitan dengan kondisi fisik jalan, tetapi juga mencakup kapasitas lalu lintas, keselamatan, kenyamanan, aksesibilitas, dan fasilitas pendukung. Oleh sebab itu, rekomendasi penanganan harus disesuaikan dengan karakteristik permasalahan. Keluhan kerusakan jalan dapat diarahkan pada rehabilitasi atau pemeliharaan, sedangkan kemacetan dapat membutuhkan manajemen lalu lintas, pengaturan simpang, pengendalian parkir, atau peningkatan kapasitas pada titik tertentu.

DAFTAR PUSTAKA

- Adi Purnama, D., Anugrah Muzaffar Rana, Z., Pinkan Lumi, D., Tahta Haritza, I., & Fadhillah, M. A. (2025). Modeling Public Transportation Policy Using Macroscopic Social Media Data Mining. *Spektrum Industri*, 23(2), 192–204. <https://doi.org/10.12928/si.v23i2.327>
- Aldisa, R. T., Maulana, P., & Aldinugroho, M. (2021). Sentiment Analysis of Public Transportation Services on Twitter Social Media Using the Method Naïve Bayes Classifier. *IJISTECH (International Journal of Information System and Technology)*, 5(4), 466. <https://doi.org/10.30645/ijistech.v5i4.166>
- BPS Kabupaten Lamongan. (2026). *Kabupaten Lamongan Dalam Angka 2026*. 43, 1–604.
- El-Diraby, T., Shalaby, A., & Hosseini, M. (2019). Linking Social, Semantic And Sentiment Analyses To Support Modeling Transit Customers' Satisfaction: Towards Formal Study Of Opinion Dynamics. *Sustainable Cities and Society*, 49, 101578. <https://doi.org/https://doi.org/10.1016/j.scs.2019.101578>
- Fatah D.A, Kamil F.A, Soesilo B, & Mulaab. (2024). Penerapan Naïve Bayes Dan Latent Dirichlet Allocation (LDA) Untuk Analisis Sentimen Dan Pemodelan Topik Pada Proyek Kereta Cepat Jakarta-Bandung Application of Naïve Bayes and Latent Dirichlet Allocation (Lda) for Sentiment Analysis and Topic Modeling on. *Jurnal Simantec*, 12(2), 53–62.
- Fikri, R., Amrillah, M. F., Rosyadi, M. I., & Rendi, O. (2025). *Klasifikasi Sentimen Masyarakat Terkait Isu Pembangunan Infrastruktur Jalan Di Indonesia Pada Media Sosial Twitter*. 8(2), 586–595.
- Gal-Tzur, A., Grant-Muller, S. M., Kuflik, T., Minkov, E., Nocera, S., & Shoor, I. (2014). The Potential Of Social Media In Delivering Transport Policy Goals. *Transport Policy*, 32, 115–123. <https://doi.org/https://doi.org/10.1016/j.tranpol.2014.01.007>
- Nokkaew, M., Nongpong, K., Yeophantong, T., Ploykitikoon, P., Arjharn, W., Phonak, D., Siritaratiwat, A., & Surawanitkun, C. (2024). Hidden Emotional Trends On Social Media Regarding The Thailand–China High-Speed Railway Project: A Deep Learning Approach With ChatGPT Integration. *Social Network Analysis and Mining*, 14(1), 1–20. <https://doi.org/10.1007/s13278-024-01340-8>
- Nufus, F. S., & Gata, W. (2025). Sentiment Analysis of Public Opinion on Transportation Services in Indonesia Using Machine Learning. *Jurnal Techno Nusa Mandiri*, 22(2), 143–150. <https://doi.org/10.33480/techno.v22i2.6577>
- Qi, B., Costin, A., & Jia, M. (2020). A Framework With Efficient Extraction And Analysis Of Twitter Data For Evaluating Public Opinions On Transportation Services. *Travel Behaviour and Society*, 21, 10–23. <https://doi.org/https://doi.org/10.1016/j.tbs.2020.05.005>
- Sayed, M. A., Hossain, M. A., Rahman, M. M., Nawaz Ali, G. G. M., Islam, M. A., Paul, K. C., & Qin, X. (2025). Public Sentiment Analysis Of Roadway Work Zones Using Social Media Data And Machine Learning Models. *Data Science and Management*, 8(4), 458–473. <https://doi.org/https://doi.org/10.1016/j.dsm.2025.04.001>
- Yao, R., & Gillen, A. (2023). Public Opinion Evaluation On Social Media Platforms: A Case Study Of High Speed 2 (HS2) Rail Infrastructure Project. *UCL Open Environment*, 5, 1–15. <https://doi.org/10.14324/111.444/ucloe.000063>
- Zayet, T. M. A., Ismail, M. A., Varathan, K. D., Noor, R. M. D., Chua, H. N., Lee, A., Low, Y. C., & Singh, S. K. J. (2021). Investigating Transportation Research Based On Social Media Analysis: A Systematic Mapping Review. *Scientometrics*, 126(8), 6383–6421. <https://doi.org/10.1007/s11192-021-04046-2>