Spektrum Industri
Vol. 23, No. 2, 2025, pp. 192-204
ISSN 1693-6590
https://journal3.uad.ac.id/index.php/spektrum

Modeling Public Transportation Policy Using Macroscopic Social Media Data Mining

Dwi Adi Purnama *, Zahid Anugrah Muzaffar Rana, Distian Pingkan Lumi, Inggil Tahta Haritza, M. Arif Fadhillah

Departement of Industrial Engineering, Universitas Islam Indonesia, Yogyakarta, 55584, Indonesia
* Corresponding Author: dwiadipurnama@uii.ac.id

ARTICLE INFO

Article history
Received December 31, 2025
Revised October 2, 2025
Accepted October 23, 2025

Keywords
Fine-Tuning;
IndoRoBERTa;
Modeling transportation service;
Public policy;
Twitter.

ABSTRACT

Transportation policies must be created by the government, especially in countries with high population expansion, transportation services are used more to meet daily necessities. Conventional surveys to gauge public opinion are costly and slow; social media offers a macro-level proxy that can complement official data. This study employs large-scale online data mining to build decision support for transportation policy. We collected 19,806 Indonesia-based Twitter posts referencing public transport, private transport, sustainable mobility, and electric vehicles. After preprocessing, we fine-tuned IndoRoBERTa for sentiment classification and applied Latent Dirichlet Allocation for topic modeling. The sentiment model achieved 81.17% accuracy, with precision, recall, and F1-scores all above 0.80. Positive discourse concentrated on private vehicles, public transit, multimodal travel, and environmentally responsible practices, with many users endorsing eco-friendly private cars. Negative discourse emphasized severe air pollution, frequently attributing risk to emissions from private automobiles in Jakarta. Translating these insights into policy, we propose expanding electric-vehicle charging infrastructure, implementing vehicle buy-back/retirement programs, establishing low-emission zones, and promoting biofuels. The results demonstrate that macroscopic social media analytics can surface actionable public preferences and pain points, enabling near-real-time monitoring to inform adaptive and equity-oriented transportation policies. This framework provides a scalable approach for governments in rapidly growing contexts to align service provision with community sentiment while advancing sustainability goals.

1. Introduction

Indonesia's rapidly growing population has fueled a surge in transportation and energy demand. Indonesia, with a population exceeding 230 million, is experiencing rapid economic growth (Soehodho, 2017). Between 2009 and 2019, the population density rose from 132 to 144 people per square kilometer (Anugerah et al., 2021). Additionally, Asia Pacific's energy demand is projected to climb from 40 to 55 quadrillion British thermal units between 2020 and 2040 (Anugerah et al., 2021). This population boom has intensified transportation needs and increased gas emissions, contributing to air pollution. Moreover, the rapid population growth has resulted in an increase in the use of transportation services to meet daily needs.

This growth has driven an increase in transportation demand, particularly for personal vehicles. With a limited road network covering less than 7% of the land, combined with high car ownership rates and inadequate public transportation infrastructure, the city's roads are severely congested (Farda & Lubis, 2018; Soehodho, 2017; Sukarno et al., 2016). The reliance on road-based transportation, which accounts for over 90% of all modes, further compounds this problem (Soehodho, 2017).

Vehicles are considered one aspect that can be applied as a potential solution to reduce greenhouse gas emissions and dependence on fossil fuels as the population increases (Bleviss, 2021; Massar et al., 2021; Patil, 2021; Ribeiro et al., 2021). In addition, the concept of green transportation is also gaining attention as a way to support sustainable economic growth without sacrificing the environment. This concept aims to create an environmentally friendly transportation system with more efficient energy use and minimal carbon dioxide production. In some parts of Indonesia, there are still gaps in access to transportation technology, which can lead to inequalities in mobility and economic opportunities (Hidayati et al., 2021). Therefore, it is important to develop policies and regulations that support the production and use of sustainable transportation technologies.

The successful implementation of transportation technologies also depends heavily on the support of effective public policies and regulations. The government needs to play an essential role in formulating policies that encourage the use of environmentally friendly technologies and ensure that all levels of society can access these technologies. Supportive public policies can include incentives for the use of electric vehicles, the development of adequate infrastructure, and public education and awareness about the benefits of green transportation.

Therefore, this research seeks to model public policy regarding transportation services through the application of online big data mining and macroscopic techniques, aiming to enhance the understanding of consumer needs as a reflection of public opinion on transportation services. This study aims to develop optimal policies and comprehend the requirements necessary for fostering more efficient and equitable transportation in countries with high population density and substantial transportation demand, such as Indonesia.

Furthermore, most previous studies on transportation policies only rely on user survey data at the microscopic level. Many of previous studies depend solely on user survey data obtained through conventional methods, including interviews, questionnaires, or surveys, which analyze users at a microscopic level. Qualitative methods, particularly interviews, have been employed to examine developments in electric vehicle technology within the German manufacturing sector (Richter & Anke, 2021), as demonstrated by (Graham, 2020) in their study of new electric vehicle business models across European nations, and by (Richter & Anke, 2021) in their analysis of stakeholders involved in the electric vehicle transition in Germany. Additionally, qualitative data, including interviews, have been utilized to investigate factors influencing electric vehicle adoption in Indonesia (Delmonte et al., 2020; Greene et al., 2020; Natalia et al., 2020; Saputra & Andajani, 2023; Setiawan et al., 2022). The significance of data mining at the macroscopic level regarding public opinion is underscored by the utilization of big data mining techniques. Table 1 presents a comparison between online data mining from social media and traditional methods as referenced in (Balla et al., 2023). The application of online data mining will enhance the design of vehicle technology, aiming to improve transportation facilities and infrastructure in alignment with public adoption patterns.

Previous studies have demonstrated that the big data approach, through the analysis of public perception and attitudes, is effective in numerous instances of successful product or service adoption (Jeong et al., 2019; Purnama et al., 2023; Subagyo et al., 2024; J. Wang et al., 2023; Zhang et al., 2022). Previous studies have identified customer needs for product design (Jeong et al., 2019; Purnama et al., 2023; Subagyo et al., 2024), but have not accounted for the dynamic nature of public perception and attitude data over time. (Subagyo et al., 2024) employed a model to analyze customer perception dynamics regarding aircraft services. Consequently, the big data method (macroscopic) proves to be more effective and comprehensive in modeling the transportation service decision support for public policy. Most research relies on microscopic survey data, while this study uses macroscopic online big data to inform public policy. Macroscopic, social-media–derived data provide an urgent, population-level lens to identify customer preferences and to model transportation services.

The contribution and novelty of this research is modeling of transportation service decision support through online big data mining, employing social media as a macroscopic survey method to inform public policy, particularly in countries characterized by high population density and significant transportation demand. The success of transportation modernization depends on effective public policy support. Moreover, the examination of public opinion regarding transportation services to inform public decision-making in Indonesia has not been addressed in previous studies.

2. Method

Topics related to optimizing transportation policies through Twitter can be seen from each tweet that is used for analysis in the form of data mining to determine what is believed by the public, and this study will focus on the results of positive and negative sentiments to assess the topics obtained. The following is a flowchart of the method used.

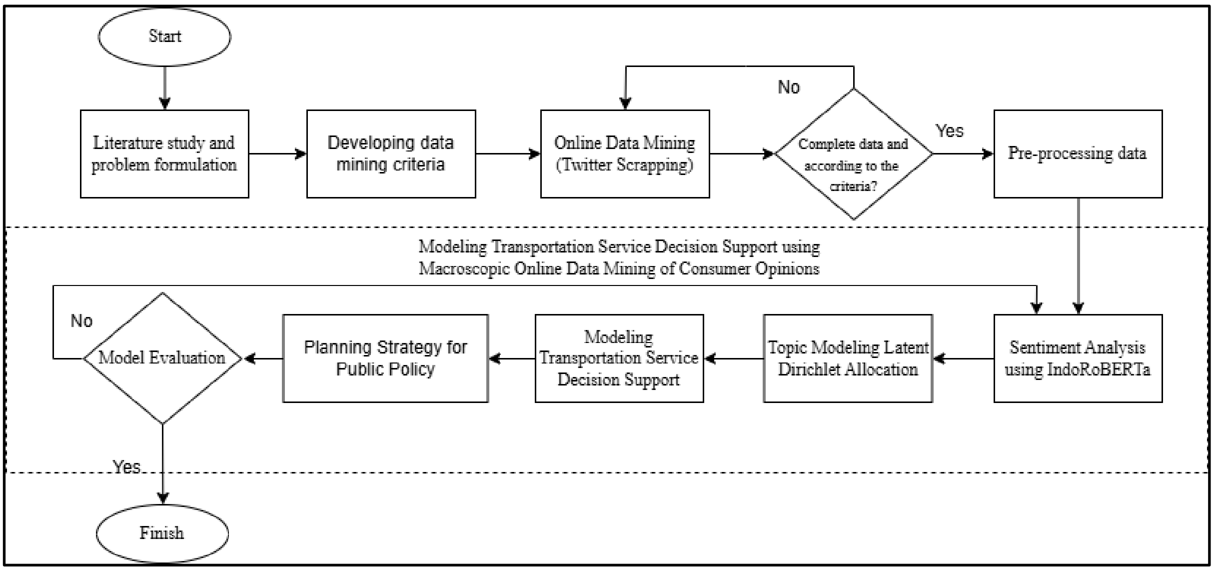

Fig. 1. The proposed method. stages of research carried out according to the flow diagram depicted, starting with Twitter data collection

Fig. 1 shows the stages of research carried out according to the flow diagram depicted, starting with Twitter data collection, data pre-processing, sentiment analysis classification into positive, neutral, and negative forms, model evaluation, and topic modeling. This study was implemented entirely in Python within Jupiter notebooks. Data collection and preprocessing were conducted using standard Python tooling (Pandas, NumPy, regex utilities) and web-scraping scripts executed in Collab. Sentiment classification relied on the Hugging Face Transformers stack with PyTorch to fine-tune IndoRoBERTa, while scikit-learn supported stratified splits, cross-validation, and evaluation (accuracy, precision, recall, F1). Topic modeling was performed with Genism (LDA), and topic interpretability was enhanced through interactive visualizations using PyLDAvis.

2.1. Online Data

This study collected Indonesian public opinions on transportation from Twitter. Using data-scraping in Google Collab, we compiled 19.806 unique tweets posted between January 1, 2023 and July 20, 2024. We adopted this 18-month horizon to balance recency with sufficient temporal coverage to capture seasonality (e.g., holiday travel peaks), policy shifts, and exogenous shocks affecting mobility and emissions, while keeping the insights timely for current decision-making. Queries targeted four themes: public transportation (12.556 tweets), private transportation (4.425), green/sustainable transportation (497), and electric vehicles (2.330). The sum of theme counts slightly exceeds the unique total because some tweets matched multiple themes; after de-duplication, the analytic dataset comprises 19.806 tweets.

2.2. Pre-Processing Data

After collecting data, the data pre-processing stage is carried out to ensure the quality of the data that will be used in model training. The steps taken are eliminating URLs, mentions, hashtags, and special characters such as irrelevant punctuation to reduce noise in the data. Duplicate data removal is also carried out to avoid bias in model training. Finally, all text is converted to lowercase to ensure consistency and facilitate further processing. Thus, the data used will be ready for sentiment analysis so that the sentiment results obtained are more accurate.

2.3. Sentiment Analysis

To create a sentiment classification on all tweets, this study uses the Fine-Tuning method that utilizes the pre-trained Indonesian RoBERTa Base Sentiment Classifier model, which is a further improvement of BERT, which is able to represent words and focuses on producing a language model. This pre-trained model is then Fine-Tuned to solve the problem by modifying the latest dataset so that these parameters can be adjusted to the classification of data that has positive (0), neutral (1), and negative (2) sentiments (Islamey et al., 2024; Padma et al., 2024; Widarmanti et al., 2022).

We fine-tune an Indonesian RoBERTa (IndoRoBERTa) model, a RoBERTa-style Transformer pre-trained on large Indonesian corpora with byte-level BPE tokenization and a masked-language-modeling (MLM) objective (no next-sentence prediction). Compared with vanilla BERT, RoBERTa's training regime (dynamic masking, longer training, larger batches) yields stronger contextual representations for short, informal Indonesian text such as tweets, where clitics (-ku, -mu, -nya), reduplication, slang, and code-mixing are common.

We selected IndoRoBERTa because it is pre-trained on large Indonesian corpora with byte-level BPE tokenization, which better handles Indonesian morphology (clitics -ku/-mu/-nya), reduplication, slang, and common ID–EN code-mixing in tweets, while RoBERTa's training regime (dynamic masking, longer training, no next-sentence prediction) yields more robust contextual representations for short, noisy text. In our preliminary benchmarks (not shown), it outperformed mBERT/XLM-R and TF-IDF+SVM baselines in macro-F1 across three sentiment classes. The base model fine-tunes efficiently on mid-range GPUs and provides low inference latency, enabling near-real-time monitoring. Mature Indonesian checkpoints and tokenizers facilitate reproducibility and deployment. It also supports continued pre-training on unlabeled in-domain tweets before supervised fine-tuning, reducing subword fragmentation and preserving Indonesian lexical nuance, advantages that make it well suited for Twitter-based sentiment analysis to inform transportation policy.

(1) Preprocessing. We de-duplicate tweets; filter for Indonesian via language ID; lower-case; strip URLs/mentions; normalize elongations (e.g., "maaaacet" → "macet") while preserving emojis/hashtags as signal tokens; and map labels to {positive=0, neutral=1, negative=2}. To reduce temporal leakage, splits are stratified and, when feasible, time-aware (train:validation:test = 80:10:10).

(2) Tokenization. Tweets are tokenized with the model's byte-level BPE; max sequence length 128–160; truncation/padding applied.

(3) Head & Loss. We add a classification head over the [CLS]/pooled representation: Drop out→ Dense→ GELU→ Dropout→ Softmax (3 classes). Cross-entropy is used with class weights (or focal loss) when class imbalance is non-trivial; optional label smoothing ($\varepsilon \approx 0.05$) stabilizes training.

(4) Optimization. Adam with linear warm-up (≈10% of total steps) and linear decay; learning rate 2e-5–3e-5; batch size 16–32; 3–5 epochs; gradient clipping (1.0); early stopping on validation macro-F1.

(5) Domain Adaptation (optional). To better match Twitter style, we perform continued pre-training (MLM) on unlabeled in-domain tweets for 1–3 epochs before supervised fine-tuning.

(6) Evaluation & Analysis. We report accuracy and macro-averaged precision/recall/F1 to reflect class balance; we include per-class metrics, confusion matrices, calibration checks, and targeted error analysis (sarcasm, code-mixing, named entities). Thresholds can be tuned if deploying probabilities. Robustness checks include ablations (no emoji/hashtag), alternative token normalization, and baselines (mBERT/XLM-R or classical TF-IDF+SVM).

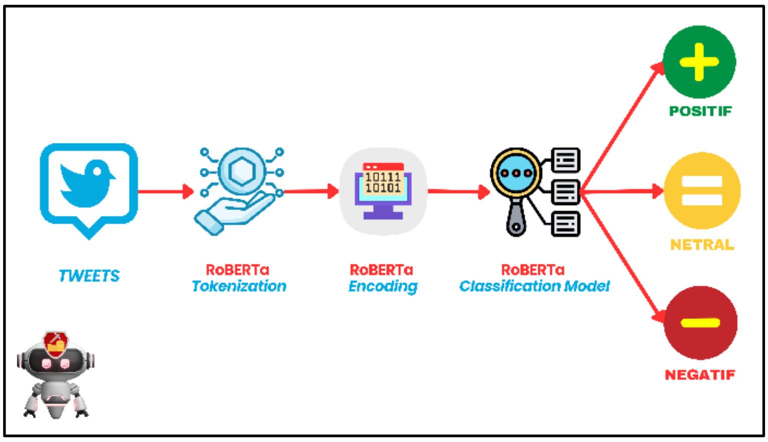

Fig. 2. Sentiment Analysis. process that starts from collecting relevant datasets containing tweets for tokenization and encoding to facilitate model classification

As seen in Fig. 2, shows the process that starts from collecting relevant datasets containing tweets for tokenization and encoding to facilitate model classification. Then, model classification is carried out using the IndoRoBERTa approach, which can adjust model weights based on various prediction errors and perform in-depth specifications of data that may have never been seen before. After classification, three sentiment categories were obtained, namely positive, neutral, and negative.

2.4. Topic Modeling

After doing sentiment analysis and model evaluation, topic modeling can be used to determine which sectors are satisfactory and which want improvement. The method utilized is latent Dirichlet Allocation (LDA), which ensures that the resultant topic model accurately represents the topic and its words. To complete this stage, the data will be turned into a library in the form of a gensim corpus, which will aid in calculating the number of themes. This library will be used as two inputs for Topic Modeling LDA, which will then generate topics that may be visualized using a coherence diagram or PyLDAvis (Purnama et al., 2023; Subagyo et al., 2024).

As seen in Fig. 3, shows the process that starts from the dataset that has been subjected to sentiment analysis in the form of positive, neutral, and negative categories. Then, Topic Modeling LDA will be carried out by determining texts that have similarities, and the similarities of the texts are grouped into one topic. From these topics, the highest probability value is in which topic, so that related words are obtained that will be used as a topic reference for formulating models and updating transportation policies in Indonesia.

2.5. Model Evaluation

The stage will continue with model evaluation, which is carried out to generalize the entire text. The performance of the model will be evaluated using metric parameters in the form of accuracy, which measures the extent to which the model can classify correctly, precision which measures the extent to which the model provides correct predictions, recall which measures the extent to which the model can correctly detect the positive class, and F1-score which combines precision and recall

(Ahmad et al., 2022; Yacouby & Axman, 2020). If the results of the evaluation are satisfactory, the model can be used to modeling.

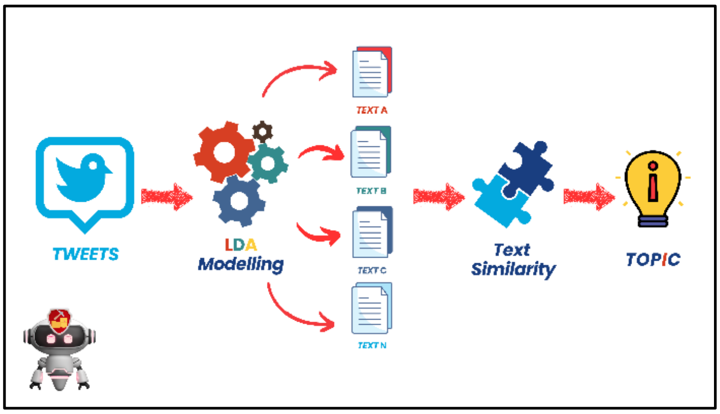

Fig. 3. Topic Modeling LDA. Process that starts from the dataset that has been subjected to sentiment analysis in the form of positive, neutral, and negative categories

3. Results and Discussion

The study presents findings on public sentiment analysis regarding transportation services, topic modeling of these services through latent Dirichlet allocation, public policy strategies related to transportation, and evaluations of high-scoring models. This is beneficial for public policy decision-making, particularly in countries with significant population characteristics and elevated transportation demand, such as Indonesia.

3.1. Fine-Tunning Sentiment Analysis

The following are the results of the polarity of Twitter public opinion regarding transportation in Indonesia:

Table 1. Sentiment Analysis Results

Keyword	Sentiment		
	Positive	Neutral	Negative
Public Transportation	3.968	2.960	5.628
Private Transportation	1.155	985	2.283
Green Transportation	133	282	82
Electric Vehicle	292	1.572	466

Table 1 shows that the total number of cleaned datasets is 19,806 data. The data is divided into four keywords, namely public transportation (12,556 data), private transportation (4,423 data), green transportation (497 data), and electric vehicles (2,330 data). Then, each keyword can be classified into positive sentiment (5,548 data), neutral (5,799 data), and negative (8,459 data). Thus, public transportation is the most discussed keyword, and the most dominant sentiment in this study is negative.

Public transportation refers to shared, scheduled services open to the general public, e.g., buses/BRT, metro/commuter rail, LRT, ferries, and licensed paratransit, typically financed or regulated by government. *Private transportation* denotes individually owned or exclusively hired motorized modes, primarily cars and motorcycles, and includes personal use and ride-hailing trips operated with private vehicles. *Environmentally friendly vehicles* (green transportation) encompass low- or zero-emission modes, including walking and cycling (when mentioned as modal shifts), public transport powered by cleaner technologies, and motor vehicles using clean fuels; *electric vehicles (EVs)* are treated as a prominent subcategory of green transport. Interpreting the cross-tabulation with these definitions, negativity around public/private modes likely reflects congestion, reliability, safety, cost, and emission complaints, whereas neutral EV/green discourse often signals information sharing (policy updates, technology news, incentives) and deliberation rather than strong affect, useful for identifying where policy communication and infrastructure interventions can shift attitudes.

Furthermore, the table analyze the cross-tabulation of tweet sentiment counts by keyword (i.e., a contingency table). Rows are the four transportation themes ("Public Transportation," "Private Transportation," "Green Transportation," "Electric Vehicle"), and the columns are Positive, Neutral, and Negative sentiments. The numerals use dot separators as thousands (e.g., "3.968" = 3,968). Totals implied by the table are:

1. Public: 12,556 (31.6% positive, 23.6% neutral, 44.8% negative)
2. Private: 4,423 (26.1% positive, 22.3% neutral, 51.6% negative)
3. Green: 497 (26.8% positive, 56.7% neutral, 16.5% negative)
4. EV: 2,330 (12.5% positive, 67.5% neutral, 20.0% negative)

Overall (N=19,806), sentiment skews negative (≈42.7%), with neutral ≈29.3% and positive ≈28.0%. Public and private transport are dominated by negative sentiment, while green transport and EVs are dominated by neutral discussion.

3.2. Modeling Transportation Service Decision Support using Topic Modeling Latent Dirichlet Allocation

Next, there are the results of the topic modeling analysis Table 2 related to transportation optimization in Indonesia.

Table 2. Sentiment Analysis Results

		Topic Modeling
Positive	Optimal Number of Topics= 5	Topic 1: Use of environmentally friendly private vehicles. Topic 2: Pedestrians everywhere use public transportation. Topic 3: The increasing use of electric cars and motorbikes in Jakarta. Topic 4: Diversity of transportation modes in Indonesia. Topic 5: The emergence of electric-based public transportation in Indonesia.
Neutral	Optimal Number of Topics= 10	Topic 1: Collaboration of electric vehicle ecosystem in Indonesia. Topic 2: Public transportation for school children. Topic 6: President Jokowi rides an electric vehicle. Topic 9: Going home using environmentally friendly vehicles. Topic 10: Use of battery-based electric vehicles in the Nusantara Capital City (IKN)
Negative	Optimal Number of Topics= 10	Topic 1: Air pollution from private vehicles in Jakarta is major concerns expressed by the public. Topic 4: The small number of electric vehicles has not reduced pollution. Topic 5: Pollution from public and private vehicles is increasing. Topic 8: Transfer of subsidized vehicles in Indonesia. Topic 10: Vehicles are not used through Steam Power Plant (PLTU).

Fig. 4 shows one of the topics taken, topic 1, with 24.3% of words relevant to the overall positive sentiment. This topic has several dominant response models, such as private vehicles, public transportation, modes of transportation, and private. In addition, there is terminology about the use of vehicles and environmentally friendly conditions, so the topic that can be taken is the use of environmentally friendly private vehicles.

Fig. 5 depicts one of the topics extracted from neutral sentiment, namely topic 1, which contains 14.1% of terms relevant to the overall neutral sentiment. This subject has various prominent response models, including electric vehicles, private vehicles, public transportation, air pollution, domestic, and technology. Furthermore, there is terminology connected to the vehicle ecosystem; therefore, one possible issue is the collaboration of the Indonesian electric car ecosystem.

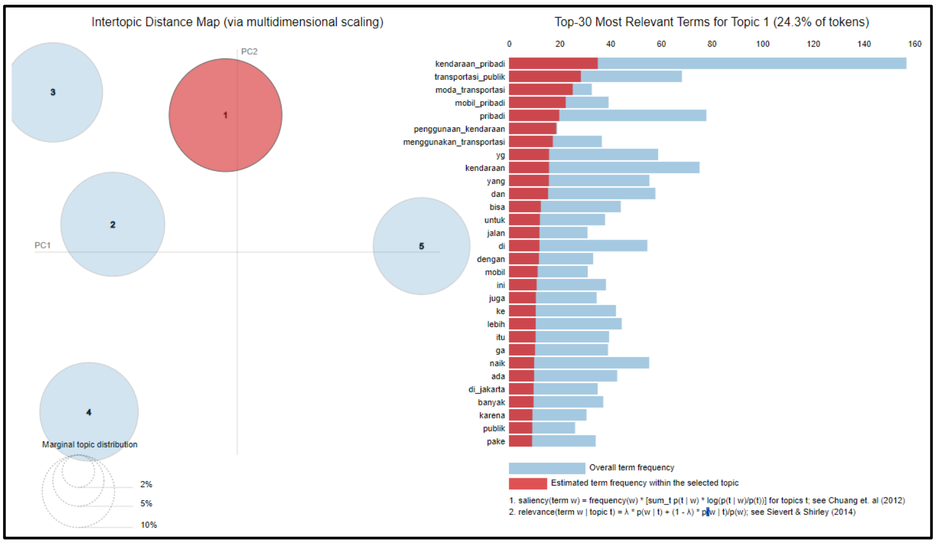

Fig. 4. PyLDAvis Positive Topics in Indonesia

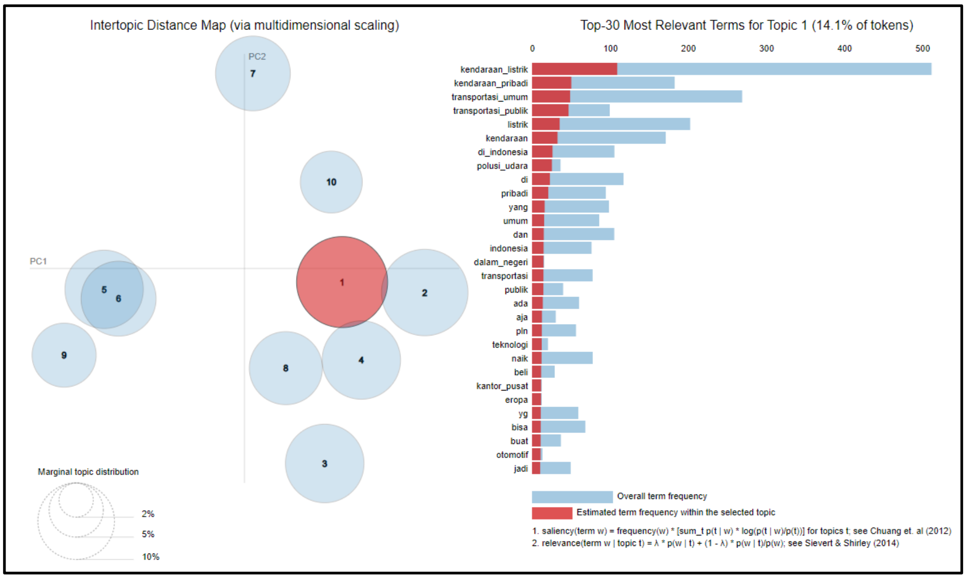

Fig. 5. PyLDAvis Neutral Topics in Indonesia

Fig. 6, depicts one of the topics derived from negative sentiment, namely topic 1, which has 13.4% of terms related to total negative sentiment. This topic is dominated by responses related to models, including private autos, eco-friendliness, electricity, the environment, and Jakarta. Furthermore, there is vocabulary connected to the description of air pollution as a major concern expressed by the public, thus the issue that might be chosen is that private vehicle air pollution in Jakarta is extremely distressing.

According to the subjects covered, the most important one is the adoption of ecologically friendly private automobiles such as electric and emission-free vehicles. Then, the topic used as material for future improvements is related to very upsetting private car air pollution, increasing public vehicle pollution, and the lack of usage of Steam Power Plant (PLTU)-supported vehicles as fuel.

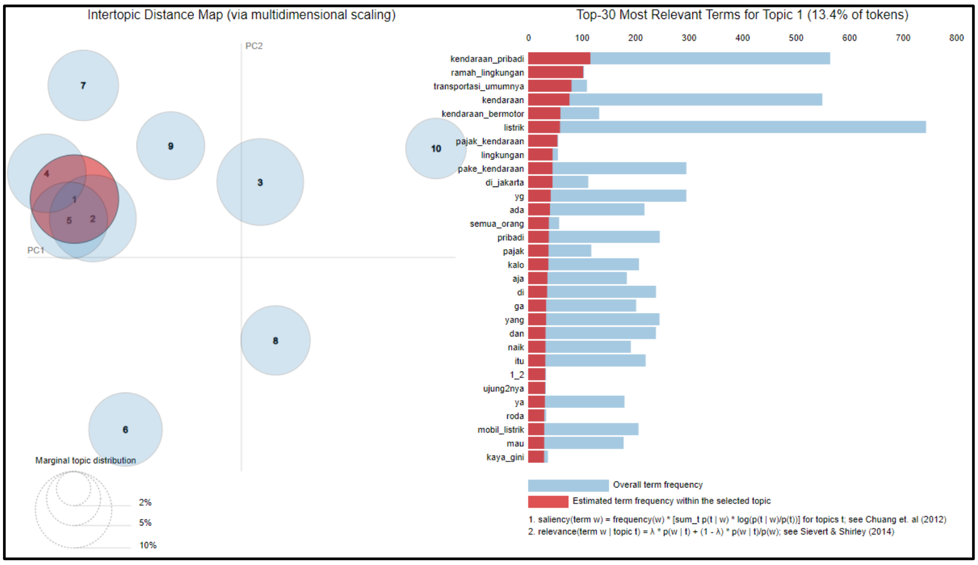

Fig. 6. PyLDAvis Negative Topics in Indonesia

3.3. Research Discussion and Planning Strategy for Public Policy

This policy analysis is essential for ensuring that formulated or implemented policies align with the current requirements of environmentally friendly transportation in Indonesia, thereby effectively achieving objectives and optimally addressing existing challenges.

This policy analysis is essential to ensure that formulated and implemented measures align with Indonesia's current requirements for environmentally friendly transportation and effectively address on-the-ground challenges. Consistent with prior work showing rising public acceptance of electric mobility, our macroscopic sentiment results, highlighting favorable views toward eco-friendly private vehicles, converge with evidence that Presidential Regulation No. 55/2019 advanced a coherent EV program around environmental conservation, energy resilience, and industrial upgrading (Subiantoro & Maharani, 2024). In line with purchase-intention studies that identify charging access as a decisive barrier, our topic modeling similarly surfaces infrastructure gaps; thus, a vehicle replacement (buy-back) scheme coupled with targeted incentives and demand-aligned siting of Electric Vehicle Charging Stations (e.g., along high-use corridors) extends recommendations in (Amilia et al., 2022).

At the same time, the persistent negative sentiment about pollution from public and private vehicles mirrors evaluations of Law No. 32/2009 that underscore implementation and enforcement deficits, despite a comprehensive mix of planning, supervision, and inter-agency instruments (Awewomom et al., 2024; A. Wang et al., 2024). To close this policy-practice gap, our findings support measures already shown effective in the literature: establishing Low Emission Zones (LEZs) to curb inner-city concentrations and accelerate fleet turnover; tightening compliance with emission standards for private vehicles; and promoting clean fuels. Prior studies report that biofuel adoption can materially cut tailpipe emissions and improve urban air quality, reinforcing our recommendation to scale sustainable fuels within the transport mix (Balwan & Kour, 2021; Ebadian et al., 2020; Guddaraddi et al., 2023). Finally, our results align with research advocating public transport modernization, including low-emission or electric fleets, implemented through coordinated central-local policies and operator partnerships, to deliver measurable air-quality gains while improving mobility equity.

This study exploits macroscopic social-media data, 19.806 tweets analyzed via IndoRoBERTa sentiment classification and LDA topic modeling, to generate bottom-up, near-real-time indicators of public preferences. It documents pervasive negative sentiment toward public and private modes (congestion, reliability, pollution) alongside largely neutral, informational discourse surrounding green mobility and electric vehicles (EVs). These signals motivate demand- and emission-management measures, including the introduction of low-emission zones, vehicle retirement/buy-back schemes with EV incentives, targeted biofuel deployment, and demand-responsive siting of EV charging infrastructure (Arifani & Mahfudz, 2024; Hindayanti et al., 2024). By contrast, (Farda & Lubis, 2018) foreground a top-down governance architecture, strong transport authority, an integrated master plan, and sustainable funding/PPPs, paired with capacity expansion across Trans Jakarta BRT, KRL, MRT, and LRT. The bottom-up evidence in this study can operationalize the top-down blueprint of Farda & Lubis (2018): sentiment "hotspots" inform corridor investment priorities and fleet-modernization timelines, guide the phasing and geography of low-emission zones, and reduce siting risk for charging assets. While platform-level biases warrant triangulation with objective indicators (OD flows, ridership, travel speeds, and emissions), the resulting synthesis yields transport governance that is more adaptive, responsive, and outcome-oriented, aligning macro-level capacity planning with continuously monitored citizen preferences.

Unlike traditional travel surveys that are costly, infrequent, and prone to reporting lag, this study leverages near-real-time, macroscopic public opinion from social media to "nowcast" mobility sentiment and topics at population scale (Prawesti et al., 2025). By fine-tuning IndoRoBERTa and applying LDA to continuously harvested tweets, we generate high-frequency indicators that can detect rapid shifts (e.g., policy rollouts, service disruptions, haze events) within days rather than months. This responsiveness enables timely policy design and course-correction, such as calibrating low-emission zones, targeting EV-charging deployment where interest is rising, or prioritizing corridors with acute reliability complaints, while complementing, not replacing, official survey statistics. Methodologically, our pipeline operationalizes public opinion as a decision-support signal, closing the evidence–response gap that often undermines transport reforms. Substantively, it reframes "public participation" from periodic consultation to continuous sentiment monitoring, yielding a more adaptive and citizen-aligned approach to environmentally friendly transport policy.

3.4. Model Evaluation

This policy analysis is essential for ensuring that formulated or implemented policies align with the current requirements of environmentally friendly transportation in Indonesia, thereby effectively achieving objectives and optimally addressing existing challenges. Subsequently, an evaluation can be conducted based on the generated sentiments, yielding the following results.

Table 3. The results of the model evaluation

Sentiment	Precision	Recall	F1-Score	Accuracy
Negative	83,45%	85,17%	84,30%	
Neutral	77,79%	85,12%	81,29%	81,17%
Positive	81,53%	70,98%	75,89%	

Table 3 presents the evaluation results of the Fine-Tuning model across three sentiments. The highest precision value, representing the ratio of correct positive predictions to total predictions, is observed in the negative sentiment category, with a precision of 83.45%. The highest recall value, reflecting the sensitivity of the data comparison with the model's prediction results, is observed in negative sentiment at 85.17%. The F1-score, representing the harmonic mean of precision and recall, is also associated with negative sentiment, recorded at 84.30%. The data indicate an accuracy of 81.17%, reflecting a relatively high score for each sentiment generated.

4. Conclusion

The government plays a crucial role in formulating transportation policies, particularly in countries with rapid population growth where the demand for daily mobility continues to rise. This study modeled decision support for transportation services using macroscopic social media data mining as an alternative to traditional surveys. Focusing on Indonesia as a case study, the research employed Twitter data analyzed through Latent Dirichlet Allocation (LDA) and Fine-Tuned

IndoRoBERTa to capture public sentiment and policy-relevant insights. The findings indicate that positive discussions largely revolve around private vehicles, multimodal transport, and environmentally friendly practices, while negative sentiments are dominated by concerns about air pollution caused by private vehicle emissions, especially in Jakarta. Based on these results, the study proposes several policy directions, including the establishment of Low Emission Zones (LEZs), the promotion of biofuels and clean energy, modernization of sustainable transportation, and continuous public education to support the implementation of Presidential Regulation No. 55 of 2019 on emission-free vehicles. The sentiment model achieved a high accuracy of 81.17%, indicating strong reliability and analytical robustness. However, this study is limited by its reliance on Twitter data, which may not fully represent the broader population due to platform and demographic bias. Future research should expand the analysis to multi-platform data and cross-country comparisons, and integrate objective indicators such as ridership, emissions, and mobility data to improve external validity and policy applicability.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research was funded by the Direktorat Penelitian dan Pengabdian Masyarakat (DPPM), Universitas Islam Indonesia (UII), under Grant No. 005/Dir/DPPM/70/Pen.Pemula/XII/2024.

Acknowledgment: The authors would like to express their sincere gratitude to the Direktorat Penelitian dan Pengabdian Masyarakat (DPPM), Universitas Islam Indonesia (UII) for administrative and technical support throughout this research. Appreciation is also extended to the research assistants and data analysts who contributed to the data collection and preprocessing process, as well as to the individuals and communities on social media whose publicly shared opinions made this study possible.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ahmad, G. N., Fatima, H., Abbas, M., Rahman, O., & Imdadullah. (2022). Mixed machine learning approach for efficient prediction of human heart disease by identifying the numerical and categorical features. In *Applied Sciences*. mdpi.com. https://www.mdpi.com/2076-3417/12/15/7449

Amilia, N., Palinrungi, Z., Vanany, I., & Arief, M. (2022). Designing an optimized electric vehicle charging station infrastructure for urban area: A case study from indonesia. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. https://ieeexplore.ieee.org/abstract/document/9922278/

Anugerah, A. R., Muttaqin, P. S., & Purnama, D. A. (2021). Effect of large-scale social restriction (PSBB) during COVID-19 on outdoor air quality: Evidence from five cities in DKI Jakarta Province, Indonesia. *Environmental Research*, *197*, 111164. https://doi.org/10.1016/j.envres.2021.111164

Arifani, E. N., & Mahfudz, M. S. (2024). Corn Agroindustry Supply Chain Management in Indonesia: Increasing Added Value and Competitiveness through the Hayami Method. *Spektrum Industri*. https://journal3.uad.ac.id/index.php/spektrum/article/view/241

Awewomom, J., Dzeble, F., Takyi, Y. D., Ashie, W. B., Ettey, E. N. Y. O., Afua, P. E., Sackey, L. N. A., Opoku, F., & Akoto, O. (2024). Addressing global environmental pollution using environmental control techniques: a focus on environmental policy and preventive environmental management. *Discover Environment*, *2*(1), 8. https://doi.org/10.1007/s44274-024-00033-5

Balla, S. N., Pani, A., Sahu, P. K., & Gonzalez-Feliu, J. (2023). Examining shifts in public discourse on electric mobility adoption through Twitter data. *Transportation Research Part D: Transport and Environment*, *121*, 103843. https://doi.org/10.1016/j.trd.2023.103843

Balwan, W. K., & Kour, S. (2021). A Systematic review of biofuels: the cleaner energy for cleaner environment. In *Indian Journal of Scientific Research*. researchgate.net. https://www.researchgate.net/profile/Wahied-Balwan-

2/publication/354322529_A_Systematic_Review_of_Biofuels_The_Cleaner_Energy_for_Cleaner_Environment/links/6130efa638818c2eaf7a303c/A-Systematic-Review-of-Biofuels-The-Cleaner-Energy-for-Cleaner-Environment.pdf

Bleviss, D. L. (2021). Transportation is critical to reducing greenhouse gas emissions in the United States. *WIREs Energy and Environment*, *10*(2). https://doi.org/10.1002/wene.390

Delmonte, E., Kinnear, N., Jenkins, B., & Skippon, S. (2020). What do consumers think of smart charging? Perceptions among actual and potential plug-in electric vehicle adopters in the United Kingdom. *Energy Research & Social Science*, *60*, 101318. https://doi.org/10.1016/j.erss.2019.101318

Ebadian, M., Dyk, S. van, McMillan, J. D., & Saddler, J. (2020). Biofuels policies that have encouraged their production and use: An international perspective. *Energy Policy*. https://www.sciencedirect.com/science/article/pii/S0301421520306182

Farda, M., & Lubis, H. R. (2018). Transportation system development and challenge in jakarta metropolitan area, indonesia. *International Journal of Sustainable Transportation Technology*. https://unijourn.com/article/5cd3cf6cb2f6bd305742ba42/5ae99ad07348a8567766abe2/2.html

Graham, J. (2020). Twists and turns: New business models. Electric vehicles 101: A series of transport notes on electric vehicle trends and opportunities from IFC. In *World Bank group. Twists-and-Turns-New-Business-Models*. https://www.ifc.org/content/dam/ifc/doc/mgrt/ifc-transportnotes-twist-turns-final.pdf

Greene, D. L., Kontou, E., Borlaug, B., Brooker, A., & Muratori, M. (2020). Public charging infrastructure for plug-in electric vehicles: What is it worth? *Transportation Research Part D: Transport and Environment*, *78*, 102182. https://doi.org/10.1016/j.trd.2019.11.011

Guddaraddi, A., Singh, A., Amrutha, G., Saikanth, D. R. K., Kurmi, R., Singh, G., Chowdhury, M., & Singh, B. V. (2023). Sustainable biofuel production from agricultural residues an eco-friendly approach: a review. In *Int J Environ Clim Change*. https://journalijecc.com/index.php/IJECC/article/view/2956

Hidayati, I., Tan, W., & Yamu, C. (2021). Conceptualizing Mobility Inequality: Mobility and Accessibility for the Marginalized. *Journal of Planning Literature*, *36*(4), 492–507. https://doi.org/10.1177/08854122211012898

Hindayanti, H., Septiani, W., & Maulidya, R. (2024). Employee Performance Assessment Model in Survey Service Companies Using Analytical Hierarchy Process and Naive Bayes. *Spektrum Industri*. https://journal3.uad.ac.id/index.php/spektrum/article/view/150

Islamey, J., Jonathan, V., Nurzaki, M., & Lucky, H. (2024). Comparative Analysis of Encoder-Based Pretrained Models: Investigating the Performance of BERT Variants in Indonesian Question-Answering. *International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics*. https://ieeexplore.ieee.org/abstract/document/10704260/

Jeong, B., Yoon, J., & Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*, *48*, 280–290. https://doi.org/10.1016/j.ijinfomgt.2017.09.009

Massar, M., Reza, I., Rahman, S. M., Abdullah, S. M. H., & Jamal, A. (2021). Impacts of autonomous vehicles on greenhouse gas emissions positive or negative? In *International Journal of Environ Science and Engineering*. mdpi.com. https://www.mdpi.com/1660-4601/18/11/5567

Natalia, Y., Rahman, I., & Hidayatno, A. (2020). Conceptual Model for Understanding the Policy Challenges of Electric Vehicle Adoption in Indonesia. *2020 The 6th International Conference on Industrial and Business Engineerin*, 113–117. https://doi.org/10.1145/3429551.3429557

Padma, T., Visweshvar, R., Tamilarasan, K., & Bhadrinath, C. J. (2024). Dynamic YouTube Comment Sentiment Analysis with Supervised Fine-Tuned BERT. *2024 International Conference on Cognitive Robotics and Intelligent Systems*. https://ieeexplore.ieee.org/abstract/document/10533926/

Patil, P. (2021). Sustainable transportation planning: Strategies for reducing greenhouse gas emissions in urban areas. In *Empirical Quests for Management Essences*.

Prawesti, G., Supriatna, A., Isniah, S., & Khoirunisa, A. (2025). Determination of Cultural Tourism Development Strategy in the Special Capital Region of Jakarta Using the Dynamic Model Method. *Spektrum Industri*. https://journal3.uad.ac.id/index.php/spektrum/article/view/284

Purnama, D. A., Subagyo, & Masruroh, N. A. (2023). Online data-driven concurrent product-process-supply chain design in the early stage of new product development. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(3), 100093. https://doi.org/10.1016/j.joitmc.2023.100093

Ribeiro, F. N. D., Umezaki, A. S., Chiquetto, J. B., & Ueno, H. M. (2021). Impact of different transportation planning scenarios on air pollutants, greenhouse gases and heat emission abatement. *Science of The Total Environment*. https://www.sciencedirect.com/science/article/pii/S0048969721017769

Richter, D., & Anke, J. (2021). Exploring Potential Impacts of Self-Sovereign Identity on Smart Service Systems. *Business Information Systems*, 105–116. https://doi.org/10.52825/bis.v1i.68

Saputra, M. C., & Andajani, E. (2023). Analysis of Factors Influencing Intention to Adopt Battery Electric Vehicle in Indonesia. *ADI Journal on Recent Innovation (AJRI)*, *5*(2), 100–109. https://doi.org/10.34306/ajri.v5i2.993

Setiawan, A. D., Zahari, T. N., Purba, F. J., Moeis, A. O., & Hidayatno, A. (2022). Investigating policies on increasing the adoption of electric vehicles in Indonesia. *Journal of Cleaner Production*, *380*, 135097. https://doi.org/10.1016/j.jclepro.2022.135097

Soehodho, S. (2017). Public transportation development and traffic accident prevention in Indonesia. In *IATSS research*. Elsevier. https://www.sciencedirect.com/science/article/pii/S0386111216300085

Subagyo, Purnama, D. A., Masruroh, N. A., & Pratama, R. R. (2024). Modeling Dynamic Consumer Preferences in Product Attributes for Social Media-Based Product Improvement Planning. *Malaysian Journal of Consumer and Family Economics*, *32*(1), 104–140. https://doi.org/10.60016/majcafe.v32.05

Subiantoro, H., & Maharani, A. E. P. (2024). Analisis Perpres Nomor 55 tahun 2019 Terkait Program Kendaraan Listrik Dalam Rangka Mewujudkan Transportasi Ramah Lingkungan. *Jurist-Diction*. https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=27218392&AN=176577758&h=N1VqkOdA%2Flc7GbNYnIuUGCpIf3w7eTSBg4mN%2FBTnbpavRdqXpH60oPjmBxu8PLDl4QBTIyrxLxrU2%2FazUeI9lg%3D%3D&crl=c

Sukarno, I., Matsumoto, H., & Susanti, L. (2016). Transportation energy consumption and emissions-a view from city of Indonesia. In *Future Cities and Environment*. Springer. https://doi.org/10.1186/s40984-016-0019-x

Wang, A., Wang, J., Zhang, R., & Cao, S. J. (2024). Mitigating urban heat and air pollution considering green and transportation infrastructure. *Transportation Research Part A: Policy and Practice*. https://www.sciencedirect.com/science/article/pii/S0965856424001277

Wang, J., Lai, J.-Y., & Lin, Y.-H. (2023). Social media analytics for mining customer complaints to explore product opportunities. *Computers & Industrial Engineering*, *178*, 109104. https://doi.org/10.1016/j.cie.2023.109104

Widarmanti, T., Widodo, M. P., Ramadhani, D. P., & Danlami, M. (2022). Text emotion detection: Discover the meaning behind YouTube comments using indo RoBERTa. *2022 International Conference on Advanced Creative Network and Intelligent Systems*. https://ieeexplore.ieee.org/abstract/document/10055265/

Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*. https://aclanthology.org/2020.eval4nlp-1.9/

Zhang, Y., Chen, K., Weng, Y., Chen, Z., Zhang, J., & Hubbard, R. (2022). An intelligent early warning system of analyzing Twitter data using machine learning on COVID-19 surveillance in the US. *Expert Systems with Applications*, *198*, 116882. https://doi.org/10.1016/j.eswa.2022.116882