Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Citizens needs for smart transportation services in Indonesia: A sentiment analysis approach

Dwi Prabowo *, Andarina Aji Pamurti, Wahjoerini Wahjoerini

Urban and Regional Planning, Semarang University, Semarang, Indonesia

ARTICLE INFO

Article history:
Received 15 July 2023
Received in revised form
11 February 2024
Accepted 6 June 2024

Keywords:
Smart city
Citizen needs
Sentiment analysis
Smart transportation services
Public perception

ABSTRACT

A smart city (SC) uses technology to enhance the social, economic, and environmental quality of urban life. Consequently, addressing citizens' needs is crucial for successfully implementing smart cities. However, much of the focus has been on technological aspects rather than a comprehensive approach that prioritizes people's needs in a SC. This study investigates the needs of citizens for Smart Transportation Services in Indonesia by analyzing public perceptions using sentiment analysis (SA) based on big data from Twitter. While previous studies have applied SA in marketing and health sectors, its application in public services has not been extensively explored. The Naïve Bayes classifier was used to develop a sentiment classifier due to its higher accuracy compared to other methods. SA of tweets containing the keyword 'transportation' revealed that 47.26% were positive, 42.7% were neutral, and 10.04% were negative, with an accuracy rate of 80%. The research identified four main topics related to citizens' needs for smart transportation services in Indonesia: public transportation, motorbikes, challenges, and traffic congestion. These findings highlight the need to address these issues within the context of SC services in Indonesia.

1. Introduction

Many developing countries around the world are faced with major problems of unplanned urbanization and significant regional imbalances. According to United Nations statistics, more than half of the world's population will live in cities by 2050 (Ritchie and Roser, 2023). Compared to other regions of the world, Southeast Asia is among the areas with the highest urbanization rates, with a rate of 49 percent and an annual growth rate of 1.3 percent. The phenomenon of urbanization has led to numerous physical and non-physical issues, which can be observed in many countries worldwide, particularly in developing nations. These issues include traffic congestion, environmental degradation, social fragmentation, insufficient infrastructure, and lifestyle-related diseases (Yang et al., 2021). Furthermore, in the context of global warming, there is a correlation between rising temperatures and increased levels of urbanization (Helbling and Meierrieks, 2023).

Indonesia has experienced an increase in the number of cities from 50 to 94 as a result of the process of urbanization, and there has been an expansion of large urban areas, particularly on the island of Java. According to the United Nations, the urban population of Indonesia was 56 percent in mid-2019, and this is expected to rise to 73 percent by 2050. In 2019, the urban population was 152 million, and this number is projected to reach 244 million in 2050 (Mardiansjah et al., 2021). In 2020, Jakarta, the capital city of Indonesia, became the largest megacity in the country, with a population of over 10 million. This phenomenon has resulted in various serious urban problems faced by cities in Indonesia.

Since 2017, the Indonesian government has implemented a smart city (SC) agenda known as the "100 Smart City Movement." In the early stages of 2017, 25 cities/regencies were targeted, with a target of reaching 100 cities by 2019 (MOCI, 2017). This movement aims to encourage and guide cities in Indonesia to develop a smart cities master plan in their respective regions to maximize the potential of available resources and technology. Because Indonesia's SC initiation is relatively new, these require development based on the Indonesian city's problems and characteristics. Transportation,

* Corresponding Author.
Email Address: dwiprabowo@usm.ac.id (D. Prabowo)
https://doi.org/10.21833/ijaas.2024.06.017
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-6529-8715

particularly public transportation, plays a crucial role in cities to meet the transport needs of the citizens (Bubelíny and Kubina, 2021). However, Indonesia, classified as a low-income middle country, has encountered various challenges in enhancing its public transport services. These challenges encompass inadequate service levels, deficient management practices, financial limitations hindering maintenance efforts, intricate and inflexible regulatory frameworks, competition from paratransit systems, and fragmentation within the bus sector (Irawan et al., 2021).

In this study, to encourage citizens' need for smart transportation services, we employ sentiment analysis (SA) based on social media Twitter data. The foundation of SA for smart society remains fragmented (Verma, 2022). Earlier reviews were generic, focusing either on mapping analytical processes or on the application of SA in marketing and health. This study exercised a new perspective, combining SA with academic and grey literature that focused on smart transportation in Indonesia.

2. Literature review

2.1. SC and citizens needs

The SC concept, which first appeared in the 1990s, is not the only concept for finding a better place to live (Noori et al., 2021). However, in recent years, the SC concept has had the highest popularity in academic research among other city concepts. The rapid improvement of technology and ICT is the remarkable factor driving massive SC discussion. There are a bunch of SC concepts and definitions (Zhao et al., 2021). This is because a city or urban area can be seen from different perspectives, such as economy, geography, society, etc.

However, there are two main approaches to the concept of SC: The ICT and technology-oriented approach and the people-oriented approach (Kummitha and Crutzen, 2017). The key word remains recurrent in an analysis of 84 definitions of smart cities: quality of life, services, citizens, and ICT (Schiavo and Magalhães, 2022). The citizen-centric approach emphasizes the importance of considering citizens' needs in urban planning without imposing predetermined definitions. The concept of citizen needs is inherently very broad and leaves many questions unanswered for future academic research. However, the primary focus should be on satisfaction surveys and exploring the relationship between citizen needs and urban planning to get an explanation of this relationship (Kopackova, 2019).

2.2. Public sentiment and smart transportation services

Public sentiment refers to the collective attitudes, emotions, and opinions held by the public towards their urban surroundings (Chen and Wei, 2023). The relationship between public sentiment and the urban environment is a complex and growing field of study. While significant progress has been made in understanding how specific urban places influence sentiment, more research is needed to understand the long-term impact of an integrated urban environment. By adopting a multidimensional approach that considers various socio-economic, built environment, and human mobility factors, researchers can gain a holistic understanding of the interaction between public sentiment and the urban environment, which ultimately contributes to the creation of more liveable and sustainable cities. Understanding the correlation between public sentiment and the urban environment is important for policymakers, urban planners, and researchers to create better cities.

2.3. Twitter SA

Nowadays, social media platforms have become a rich source of data for understanding public sentiment and making important decisions about a topic related to urban planning (López-Ornelas et al., 2017). Twitter is considered a valuable resource for SA due to its extensive usage and the availability of real-time data. Twitter SA has emerged as a powerful tool for understanding public sentiment across multiple domains. Despite the challenges posed by the unique characteristics of Twitter data, there has been significant progress in developing methodologies for extracting sentiment from tweets.

SA, also known as opinion mining, is a computational approach that aims to uncover sentiments, opinions, and subjectivity in text data. By utilizing SA, we can computationally identify and evaluate the emotions and attitudes expressed through text (Medhat et al., 2014). There are several methods employed in SA, each with its strengths and limitations. Supervised learning algorithms, such as Naive Bayes, are commonly used in SA to classify text into different sentiment categories. Naïve Bayes classifiers, usually used in machine learning for text classification, depend on the conditional probability of features going to a specific class. These classifiers operate by selecting relevant features through the utilization of feature selection methods (Zhang and Gao, 2011). The Naïve Bayes classifier has a better level of accuracy than other models, and the accuracy appears to be approximately 10% higher than the other model (Xhemali et al., 2009; Medhat et al., 2014). The Naive Bayes classifier method has some advantages, such as efficiency, ease of implementation, scale, good performance with limited data, interpretability, strong baseline, and robustness to irrelevant features.

3. Methods and materials

3.1. Research design

To understand the needs of citizens for smart transportation services, this study analyzes tweets

from Indonesia. The dataset consists of opinions about transportation in Indonesia, which were collected from Twitter. These opinions, shared publicly on social media, often use informal language and slang. We used a script to gather public tweets from Indonesian users containing the keyword "transportation." The research process involved four main steps (Fig. 1):

1. Twitter dataset from 1st January 2023 to 13 June 2023 was utilized to calculate public sentiment to construct citizens' need for smart transportation services in Indonesia.
2. The Twitter dataset was processed to reduce data noise. Indonesian Tweets were extracted for further SA.
3. A Naïve Bayes Classifier method was introduced to quantify public sentiment of smart transportation services in Indonesia.
4. Combined with the data and previous research studies to explore the relationship between Twitter sentiment and the citizen's needs for smart transportation services.
5. To perform data processing and analysis, we employ Python language programming.

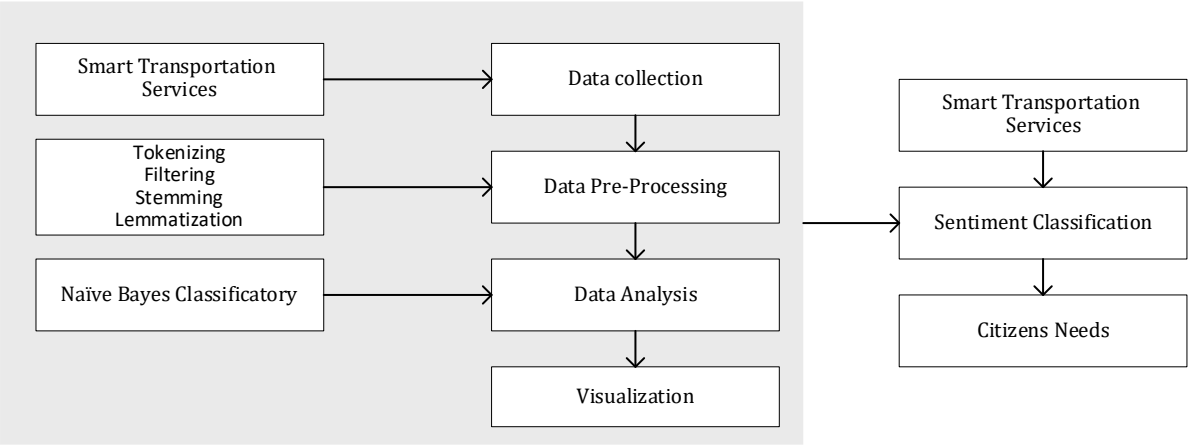

Fig. 1: Research design

3.2. Data collection and processing

1. Data collection: The data set comes from Twitter's social media tweets about "transportation." The process of extracting tweet data from Twitter is conducted by utilizing a Python language programming script. The collected data is subsequently stored in the .csv format, facilitating further analysis and processing stages.

A total of 2093 tweets were collected between January 01, 2023, and July 13, 2023, using the keyword "Transportation" (Table 1). Following the data collection process, the acquired data was transformed into a tabular format to facilitate subsequent processing stages. The resulting data frame comprises three attributes, namely:

a. Username: This attribute stores the username associated with each tweet obtained from the tweet creator's account.
b. Date: This attribute represents the timestamp indicating when the user posted the tweet.
c. Text: The text attribute contains the actual content of the tweet.

Table 1: Twitter data scraping

Username	Tweet	Date
User 0	Development of a mass public transportation system ...	June 2023
User 1	Tickets are ready, food and drinks are ready, transportation is...	June 2023
User 2	Public transportation in the city is according to ...	June 2023
User 3	Continuing on to the concert location, certainly...	June 2023
User 4	I do not support emission reduction through ...	June 2023
...	...	...
User 2898	Some things we can do to improve ...	June 2023
User 2899	Asking about government policies on ...	June 2023
User 2900	Discussion on transportation policies ...	June 2023
User 2901	Further discussion on transportation policies ...	June 2023
User 2902	When traveling using various modes of transportation ...	June 2023
2903 rows × 3 columns		

Note: Usernames and specific tweet details have been anonymized, and exact dates have been generalized to protect privacy while maintaining the necessary context for analysis

2. Data pre-processing: There are several operations in the pre-processing steps, including data cleaning, case folding, tokenization, filtering or stopping word removal, and stemming. These steps are undertaken to remove unnecessary elements from the data, ensuring its readiness for subsequent analysis.

a. Cleaning: The cleaning process involves removing punctuation marks such as commas, URL links,

periods, exclamation points, and question marks, as well as eliminating emojis, hashtags, mentions, and other irrelevant symbols (Table 2).

b. Tokenization: The tokenization process was performed to separate individual words in the tweet sentences. The Python NLTK library was used for this process.
c. Case folding: The case folding process involves converting all words in the tweet data to lowercase, making them easier for the computer to process and read.
d. Filtering or removing stop words: In the filtering process, the NLTK library was utilized to facilitate sentence removal. Additionally, this study incorporates the inclusion of several frequently occurring words in tweet data, such as "yg," "tdk," "utk," and others, to reduce noise and improve data cleanliness.
e. Stemming: The stemming process utilizes the stemmer factory library to streamline the process. The objective of stemming was to derive the base forms of words in the tweet data and eliminate any prefixes, infixes, or suffixes present in the tweet words.

Table 2: Twitter clean data

Tweet_Text
Discussion on public mass transportation system development...
Comments on tickets, food, and transport arrangements...
Positive feedback on public transportation in Jakarta...
Comments on transport options for concerts...
Opinions on reducing emissions through vehicle electrification...
Actions for climate control and reducing electricity use...
...
Questions about government support for transportation in Jambi...
Discussions on transportation safety in Jatim...
Further discussions on Jatim transportation issues...
Comments on using various modes of mass transportation...
2810 rows × 1 column

Note: Specific tweet details have been anonymized have been generalized to protect privacy while maintaining the necessary context for analysis

f. Translate: The next step was translating the tweet data from Indonesian to English using a translator library. This step is required because the SA library used in this study operates on English format data.

4. Results and discussions

4.1. SA

SA using the Naïve Bayes Classifier method of 2083 tweets collected between January 01, 2023, and July 13, 2023, using the keyword "Transportation" from Twitter social media users in Indonesia showed that 47.26% (1329), were classified as positive, indicating a favorable perception of transportation-related topics (Table 3). Additionally, 42.7% (1200) of the tweets were classified as neutral, suggesting a lack of strong positive or negative sentiment expressed by the users. On the other hand, 10.04% (282) of the tweets were identified as negative, highlighting a critical or unfavorable sentiment towards transportation issues (Fig. 2). Moreover, the SA achieved an accuracy rate of 80% (Table 4), demonstrating the effectiveness of the Naïve Bayes Classifier method in accurately classifying sentiments expressed in the collected tweets.

4.2. Discussion between the citizens' needs and sentiments analysis

Public transportation: Previous research indicates that many cities in Indonesia need to improve both the quantity and quality of public transportation. The high number of private vehicle owners suggests that public transportation is underutilized for community mobility. Additionally, the public transportation system in Indonesian cities is not well-integrated (Nuha, 2022). Issues with public transportation in Indonesia include individually owned fleets, substandard fleet quality, unhealthy competition between operators, low service performance, non-integrated fares, and lack of infrastructure. SA shows that public transportation receives the most attention across all sentiments (Fig. 3, Fig. 4, and Fig. 5). This suggests that while there have been some revitalization efforts in the public transportation sector, significant improvements are still needed.

Table 3: Sample of SA output

Tweet_Text	Classification
Discussion on developing public transportation...	Negative
Comments on tickets, meals, and transport...	Negative
Positive feedback on public transportation in Jakarta	Positive
Comments on transport options for concerts...	Neutral
Opinions on vehicle electrification and emissions...	Neutral
Positive thoughts on discussing parental topics...	Positive
Team actions related to public transport masks...	Positive
Comments on transportation class and airplane...	Positive
Positive morning messages about transportation...	Positive
Apologies related to transportation conditions...	Negative

Note: Tweet messages have been generalized to protect user privacy while maintaining the necessary context for analysis

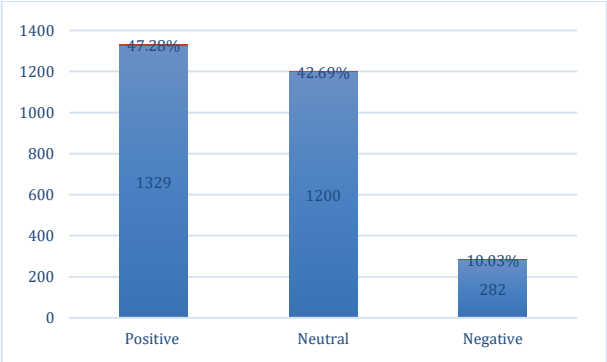

Fig. 2: Composition of SA output

Table 4: The accuracy of Naïve Bayes classifier

	Precision	Recall	F1-score	Support
Negative	0.89	0.47	0.62	529
Neutral	0.74	0.92	0.82	966
Positive	0.84	0.85	0.84	1315
Accuracy			0.80	2810
Macro avg	0.82	0.75	0.76	2810
Weighted avg	0.81	0.80	0.79	2810

A. Motorbike (Motorcycles): Motorcycles cause significant safety and environmental challenges in cities across the Global South (Chiu, 2022). They

also produce high levels of concentrated local pollution, which reduces life expectancy and causes harmful respiratory diseases (Guerra, 2019). The fact also shows that the deaths in Southeast Asia and the Western Pacific mainly involved motorbikes, with a percentage of 43% (WHO, 2018). Empirical data shows the highest proportion of motorcycles in daily traffic on Indonesia's national roads (Wandani et al., 2018). However, Indonesia has been one of the largest motorcycle users in the world. In 2022, the number of motorcycles in circulation in Indonesia reached 125 million. However, Indonesia has steadily ranked among the world's largest motorcycle users. By 2022, the total number of motorcycles in circulation in Indonesia had surged to an astounding 125 million. This is in accordance with the findings of the word cloud analysis, which has both positive and negative sentiments. Yet motorbikes have been the source of some transportation problems in Indonesia. This transportation mode has been the favorite of citizens in Indonesia.

B. Difficult: These difficulties include service coverage and access distance. In Jakarta and its buffer zones, public transportation coverage is limited. It is not surprising that millions of residents have difficulty accessing public transportation, so they eventually choose alternative modes of transportation. In the study conducted by Yumita et al. (2021), the three most considerable barriers for students in using the bus during the morning commute were the limited time of travel, distance to the bus stops, and circuitous routes.

C. Traffic Jam (traffic congestion): The expansion of the city scale came with changes in traffic mode and traffic congestion (Lu et al., 2021). Major cities in Indonesia are grappling with persistent traffic congestion as a consequence of the escalating vehicle population within the country. Jakarta, the capital city of Indonesia, has a traffic index of the top 29 in the world. Approximately 125 million vehicles traversed Indonesian roads, signifying a staggering 300% surge from the 30 million registered vehicles recorded in 2004. Notably, private vehicles constitute a significant proportion, accounting for approximately 85-90% of the vehicles used and owned by Indonesians.

Fig. 3: Word cloud analysis of positive sentiment

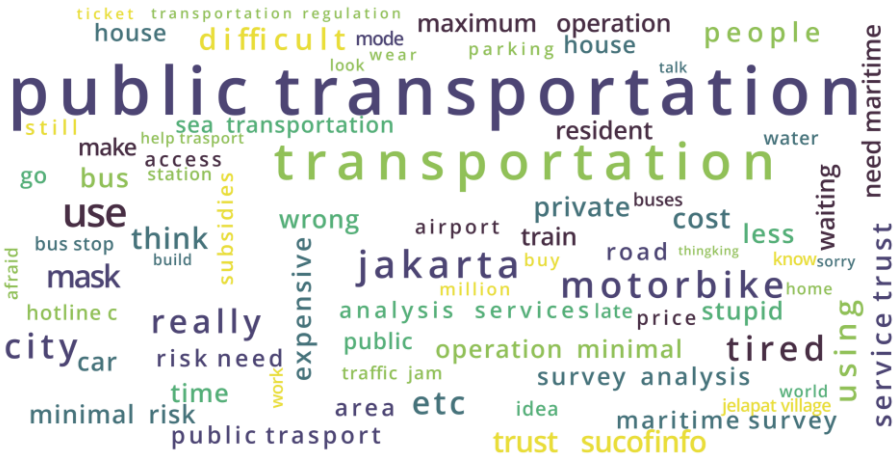

Fig. 4: Word cloud analysis of negative sentiment

5. Conclusions

These findings provide valuable insights into the public opinion regarding transportation in Indonesia, shedding light on the overall sentiment landscape and offering a deeper understanding of how individuals perceive and express their thoughts about transportation through social media Twitter.

The research findings explained that there are four topics that most appear related to citizens' need for smart transportation services in Indonesia based on SA, including public transportation, motorbikes (motorcycles), difficulties, and traffic jams.

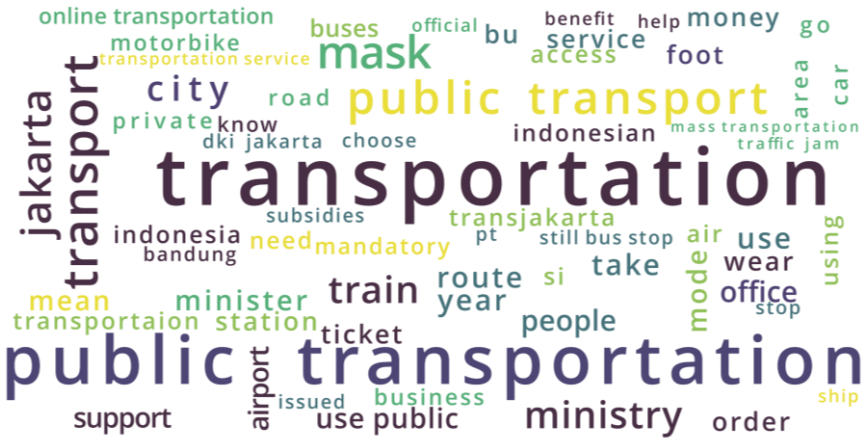

Fig. 5: Word cloud analysis of neutral sentiment

To gain deeper insights and enhance understanding related to these topics, we conducted data analysis and reviewed previous research studies. All of the topics are interconnected or interrelated with each other. Traffic jams or congestion are the most critical problems caused by the massive use of private transport, both motorbikes and private cars, which is in line with the lack of integration of public transportation. People choose private transportation due to difficulties associated with service coverage and access distances. However, this paper has certain limitations; nonetheless, it contributes to future research opportunities. Through SA, we gained insights into citizens' perceptions related to transportation in Indonesia. Future research should focus on addressing the four most apparent topics using appropriate methods and approaches within the context of smart transportation.

Compliance with ethical standards

Ethical considerations

This study used publicly available Twitter data, ensuring no personal or sensitive information was disclosed. All user identities were anonymized to protect privacy. The research complies with ethical standards for social media data use.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Bubelíny O and Kubina M (2021). Impact of the concept smart city on public transport. Transportation Research Procedia, 55: 1361-1367. https://doi.org/10.1016/j.trpro.2021.07.120

Chen K and Wei G (2023). Public sentiment analysis on urban regeneration: A massive data study based on sentiment knowledge enhanced pre-training and latent Dirichlet allocation. PLOS ONE, 18(4): e0285175. https://doi.org/10.1371/journal.pone.0285175 **PMid:37104499 PMCid:PMC10138235**

Chiu BY (2022). Does the bus rapid transit reduce motorcycle use? Evidence from the Jakarta metropolitan area, Indonesia. Case Studies on Transport Policy, 10(3): 1767-1774. https://doi.org/10.1016/j.cstp.2022.07.007

Guerra E (2019). Electric vehicles, air pollution, and the motorcycle city: A stated preference survey of consumers' willingness to adopt electric motorcycles in Solo, Indonesia. Transportation Research Part D: Transport and Environment, 68: 52-64. https://doi.org/10.1016/j.trd.2017.07.027

Helbling M and Meierrieks D (2023). Global warming and urbanization. Journal of Population Economics, 36(3): 1187-1223. https://doi.org/10.1007/s00148-022-00924-y

Irawan MZ, Bastarianto FF, Rizki M, Belgiawan PF, and Joewono TB (2021). Exploring the frequency of public transport use among adolescents: A study in Yogyakarta, Indonesia. International Journal of Sustainable Transportation, 16(11): 978-988. https://doi.org/10.1080/15568318.2021.1959682

Kopackova H (2019). Reflexion of citizens' needs in city strategies: The case study of selected cities of Visegrad group countries. Cities, 84: 159-171. https://doi.org/10.1016/j.cities.2018.08.004

Kummitha RKR and Crutzen N (2017). How do we understand smart cities? An evolutionary perspective. Cities, 67: 43-52. https://doi.org/10.1016/j.cities.2017.04.010

López-Ornelas E, Abascal-Mena R, and Zepeda-Hernández S (2017). Social media participation in urban planning: A new way to interact and take decisions. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42(4/W3): 59-64. https://doi.org/10.5194/isprs-archives-XLII-4-W3-59-2017

Lu J, Li B, Li H, and Al-Barakani A (2021). Expansion of city scale, traffic modes, traffic congestion, and air pollution. Cities, 108: 102974. https://doi.org/10.1016/j.cities.2020.102974

Mardiansjah FH, Rahayu P, and Rukmana D (2021). New patterns of urbanization in Indonesia: Emergence of non-statutory towns and new extended urban regions. Environment and Urbanization ASIA, 12(1): 11-26. https://doi.org/10.1177/0975425321990384

Medhat W, Hassan A, and Korashy H (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4): 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

MOCI (2017). Langkah Menuju "100 Smart City." Ministry of Communication and Information of the Republic of Indonesia, Jakarta, Indonesia.

Noori N, de Jong M, Janssen M, Schraven D, and Hoppe T (2021). Input-output modeling for smart city development. Journal of Urban Technology, 28(1-2): 71-92. https://doi.org/10.1080/10630732.2020.1794728

Nuha U (2022). Public transportation transformation towards a smart, efficient and inclusive system. In: Trialih R, Wardiani FE, Anggriawan R, Putra CD, and Said A (Eds.), Indonesia post-pandemic outlook: Environment and technology role for Indonesia development: 255-277. Penerbit BRIN, Jakarta, Indonesia. https://doi.org/10.55981/brin.538.c500 **PMCid:PMC9694243**

Ritchie H and Roser M (2023). Urbanization. Our World in Data, Global Change Data Lab, Oxford, UK.

Schiavo FT and Magalhães CFD (2022). Smart sustainable cities: The essentials for managers' and leaders' initiatives within the complex context of differing definitions and assessments. Smart Cities, 5(3): 994-1024. https://doi.org/10.3390/smartcities5030050

Verma S (2022). Sentiment analysis of public services for smart society: Literature review and future research directions. Government Information Quarterly, 39(3): 101708. https://doi.org/10.1016/j.giq.2022.101708

Wandani FP, Siti M, Yamamoto M, and Yoshida Y (2018). Spatial econometric analysis of automobile and motorcycle traffic on Indonesian national roads and its socio-economic determinants: Is it local or beyond city boundaries? IATSS Research, 42(2): 76-85. https://doi.org/10.1016/j.iatssr.2017.07.001

WHO (2018). Global status report on road safety 2018. Global Report, World Health Organization, Geneva, Switzerland.

Xhemali D, Hinde CJ, and Stone RG (2009). Naïve Bayes vs. decision trees vs. neural networks in the classification of training web pages. International Journal of Computer Science Issues, 4(1): 16-23.

Yang J, Kwon Y, and Kim D (2021). Regional smart city development focus: The South Korean national strategic smart city program. IEEE Access, 9: 7193-7210. https://doi.org/10.1109/ACCESS.2020.3047139

Yumita FR, Irawan MZ, Malkhamah S, and Kamal MIH (2021). School commuting: Barriers, abilities and strategies toward sustainable public transport systems in Yogyakarta, Indonesia. Sustainability, 13(16): 9372. https://doi.org/10.3390/su13169372

Zhang W and Gao F (2011). An improvement to naive Bayes for text classification. Procedia Engineering, 15: 2160-2164. https://doi.org/10.1016/j.proeng.2011.08.404

Zhao F, Fashola OI, Olarewaju TI, and Onwumere I (2021). Smart city research: A holistic and state-of-the-art literature review. Cities, 119: 103406. https://doi.org/10.1016/j.cities.2021.103406