



Twitter Data for Traffic Estimation

Tonny Judiantono¹, Syifa Alia Rahmah^{2*}, Dadan Mukhsin³, Astri Mutia Ekasari⁴

Urban and Regional Planning Program, Faculty of Engineering, Universitas Islam Bandung, Indonesia.

*Corresponding Author

Syifa Alia Rahmah

Urban and Regional Planning
Program, Faculty of
Engineering, Universitas Islam
Bandung, Indonesia.

Article History

Received: 09.02.2025

Accepted: 11.03.2025

Published: 30.03.2025

Abstract: Traffic estimation is a crucial aspect of transportation planning to anticipate increasing mobility and provide transportation infrastructure. Traditional methods, such as household travel surveys and traffic counting, often require significant time and financial resources. Therefore, this study explores the use of geolocation data from Twitter as an alternative for projecting trip generation in East Bandung, a rapidly developing area with new facilities such as the Gelora Bandung Lautan Api Stadium and the Summarecon Bandung commercial area. Data were collected using the Twitter API over one year (November 2023–November 2024) within a 700-meter radius and analyzed through spatial mapping and the development of an origin-destination (OD) matrix. Validation was conducted using the Mean Absolute Percentage Error (MAPE) and the Spatially Weighted Structural Similarity Index (SW-SSIM). The results identified Gedebage, Arcamanik, and Rancasari as the areas with the highest trip generation concentrations. A MAPE of 18.5% and an SW-SSIM of 0.72 indicate that Twitter data is sufficiently representative for modeling mobility patterns. Despite limitations such as population bias, Twitter data offers cost and time efficiency, making it a potential alternative to support more adaptive and real-time data-driven transportation planning in urban areas like East Bandung.

Keywords: traffic estimation, Twitter data, trip generation, big data, MAPE, SW-SSIM.

Cite this article:

Judiantono, T., Rahmah, S. A., Mukhsin, D., Ekasari, A. M., (2025). Twitter Data for Traffic Estimation. *ISAR Journal of Science and Technology*, 3(3), 79-86.

A. Introduction

Trip generation projection is one of the key elements in transportation planning, which aims to understand and anticipate people's mobility needs. However, traditional methods of collecting trip generation data, such as household travel surveys and traffic counting, often require high costs, a long time, and significant labour (Tamin, 2000). In recent years, big data has become an alternative to replace these methods, especially in providing real-time and efficient insights into people's trip generation patterns (Liao et al., 2022).

The East Bandung area is experiencing rapid development, marked by the construction of large-scale public facilities such as the Gelora Bandung Lautan Api Stadium, the Al-Jabbar Mosque, and the Summarecon Bandung commercial area. Based on Bandung City Regional Regulation Number 5 of 2022 concerning the Bandung City Spatial Plan (RTRW) 2022–2042, this area is designated as one of the new activity centres focusing on social, economic, and infrastructure integration. However, this rapid development has significantly increased people's trip generation locally and inter-regionally.

The phenomenon shows pressure on the transportation infrastructure in East Bandung. Based on data from the Bandung City Transportation Agency (2023), more than 60% of trips in this area are still dominated by private vehicles, which causes severe congestion in several main corridors, such as Jalan Gedebage and Jalan Buah Batu. In addition, the high rate of land use change, such as the development of new commercial and residential areas, increases the need for fast and accurate analysis to support sustainable transportation planning.

Big data from social media such as Twitter is a potential alternative. Geotagged data from Twitter offers the advantage of cost efficiency and the ability to provide a real-time picture of trip generation (Verma et al., 2021). However, the use of big data in Indonesia still has limitations, such as user population bias and a lack of exploration of local analysis methods (Liao et al., 2022). Therefore, this study explores how Twitter data can be used to project the community trip generation in East Bandung. With this approach, practical and innovative solutions can be found to support transportation planning that is more responsive to regional development dynamics.

B. Metodologi

The research framework connects the problem of a significant increase in community trip generation with the limitations of traditional methods in understanding the trip generation patterns. East Bandung, which is designated as the second centre of activity in Bandung City, is experiencing massive growth and development. This increases community mobility without being balanced by adequate transportation planning, resulting in traffic congestion and less than optimal land use management. This study

aims to test how Twitter geolocation data can be utilized to estimate trip generation accurately and in real-time. If Twitter data cannot be utilized optimally, alternative strategies such as data integration from GPS surveys or relevant local technology are needed. Conversely, if it can be utilized, this approach enriches transportation literature and supports policies to design a more efficient and connected transportation system in East Bandung. The relationship between these concepts can be seen in the following figure.

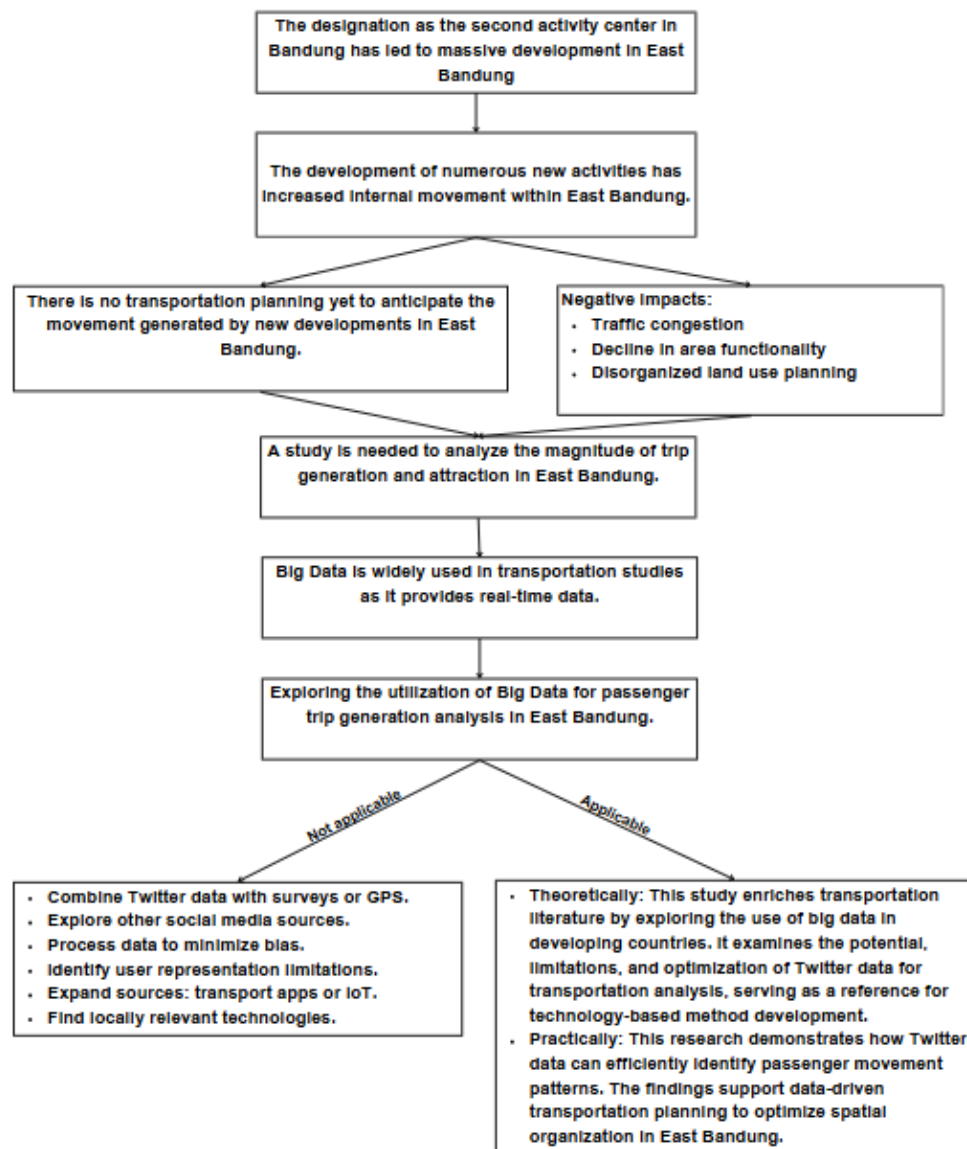


Figure 1. Framework

This study uses a quantitative approach based on big data. The research method is designed to explore how Twitter data can be used to describe the rise of community trip generations in the East Bandung area.

This study includes data collection, data processing, and data analysis with the following stages:

1. Data Collection and Data Processing Methods

This study uses primary data in the form of Twitter geotags, which are obtained through a crawling process using the Application Programming Interface (API).

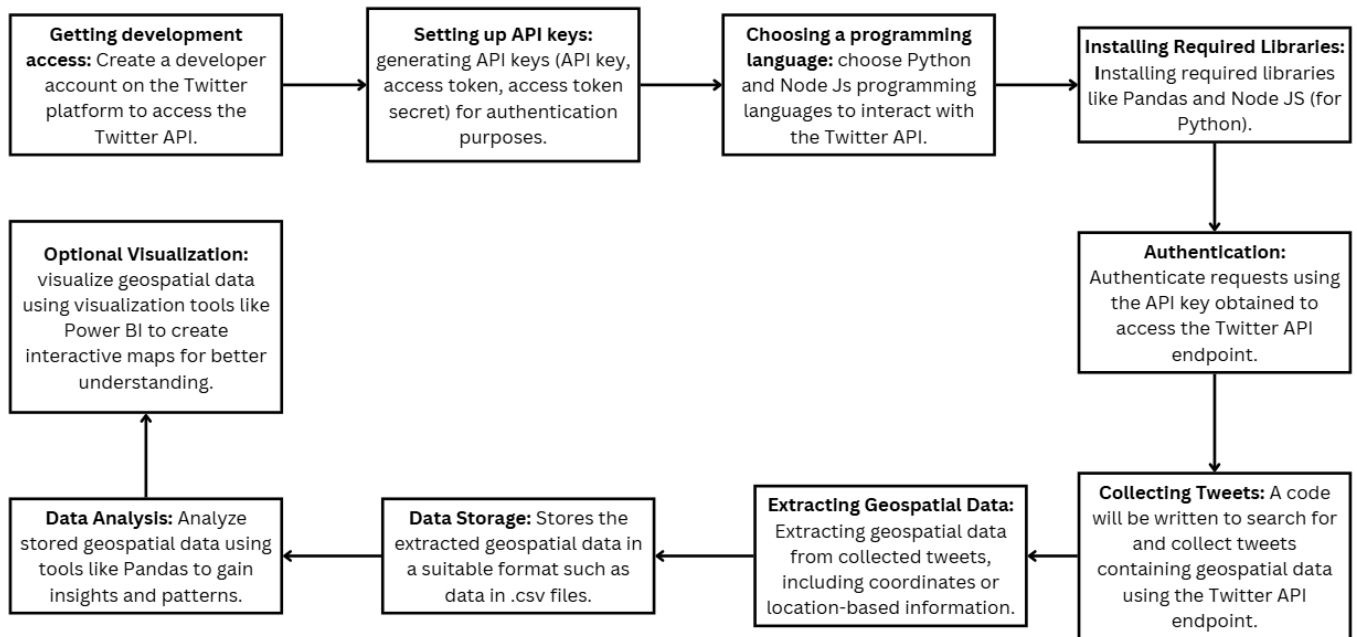


Figure 2. Data Crawling Stages

Data collection is carried out by setting a reference point in the study area and determining the radius as the coverage area. The data collection radius used is 700 meters, which is selected based on the results of previous trials with various radii. This radius is considered the most optimal because it provides maximum data coverage without reducing the accuracy of the information required.

To ensure the relevance of the collected data, the filtering process is carried out using several keywords adjusted to the research objectives. Data is collected for one year, from November 2023 to November 2024, to capture daily and weekly community trip generation patterns. In addition, this study also utilizes additional data in the form of traffic counting statistics in the East Bandung area. This additional data is obtained from direct calculations. The following are points from the traffic counting activities carried out;

Traffic Counting (TC) observation locations in East Bandung were selected based on the characteristics of high mobility in the area. Each point describes the main trip generation pattern between residential areas, commercial centres, and public facilities. The data obtained includes the direction of trip generation from the point of origin to the destination on strategic

routes. This data is used as a reference for validating the results of Twitter data analysis to improve the accuracy and reliability of the representation of the trip generation patterns obtained. This data collection process is designed to ensure that the data collected is relevant and able to describe the trip generation of people in the study area.

Twitter data collected through the crawling process is processed through several stages to ensure quality and relevance. The first stage is data cleaning, which involves removing anomalies such as data duplication, locations outside the scope of the study area, and irrelevant tweets. Furthermore, an assumption is made regarding the difference in posting time of at least one minute, with the assumption that within that time frame, it is possible for trip generation between sub-districts using motorized vehicles. After that, the geotag data is processed into an Origin-Destination Matrix (OD Matrix), which describes the trip generation of people between regions based on the coordinates of the origin and destination locations of each tweet. Finally, the data were grouped based on the temporal dimension, namely weekdays and weekends, to analyze different trip generation patterns according to the characteristics of the time. This approach aims to provide a comprehensive picture of the dynamics of community trip generation in the study area.

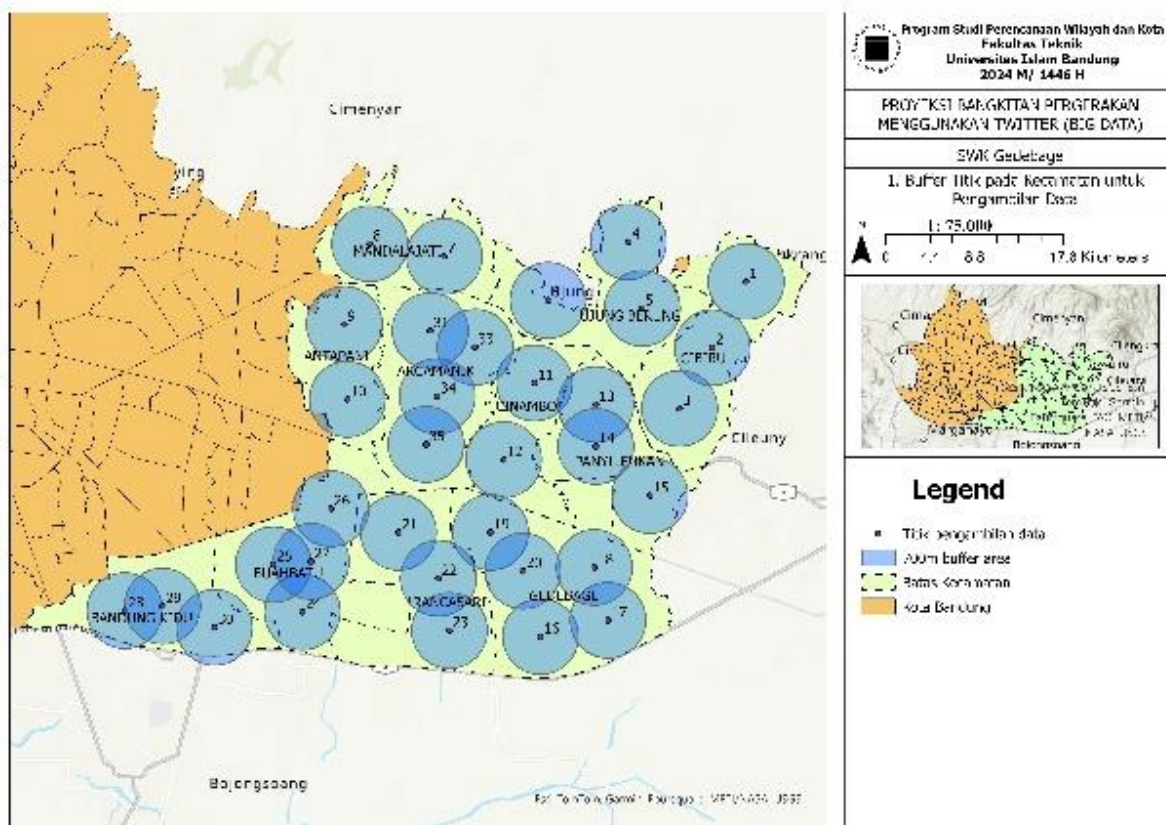


Figure 3. Data Collection Radius

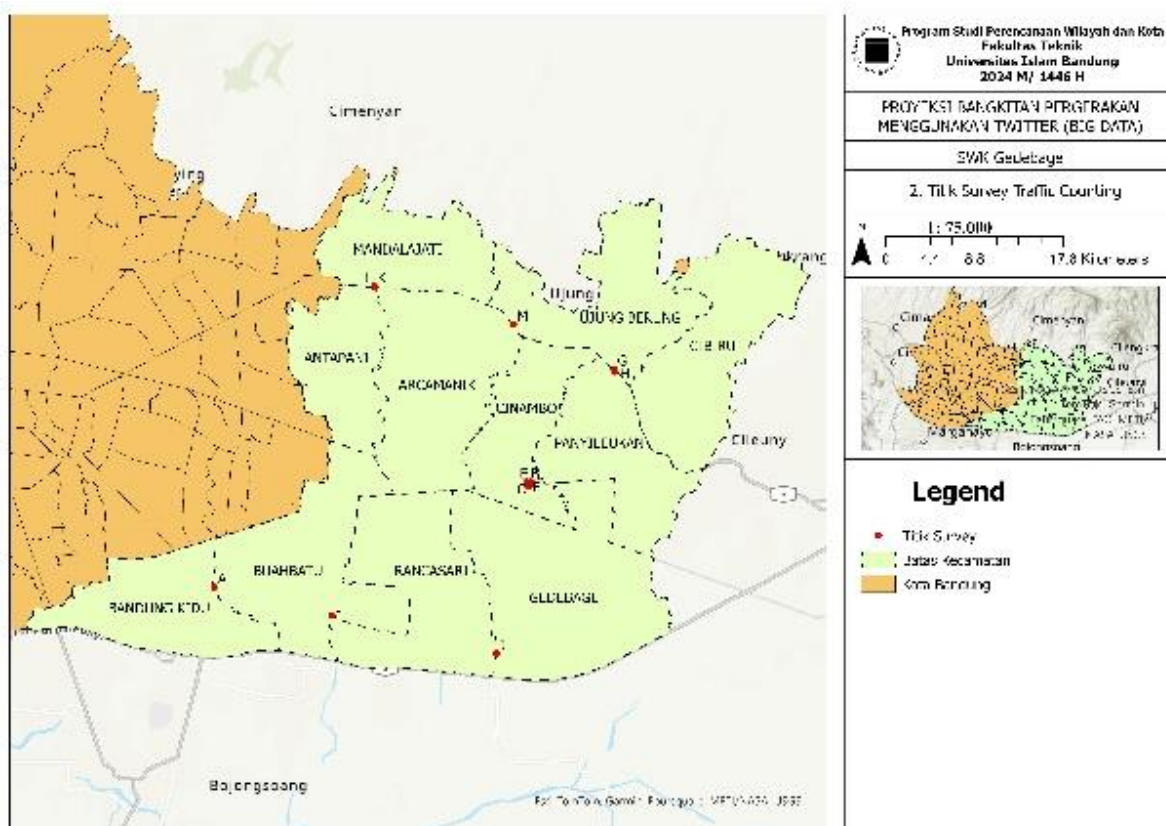


Figure 4. Traffic Counting Survey Location

2. Data Analysis Method

Data analysis was conducted to evaluate the representation and accuracy of Twitter data in projecting the community trip generation. The first stage is spatial analysis, which uses GIS software to map the concentration of community trip generation in the East Bandung area. This mapping provides a visual depiction of the areas with the highest levels of trip generation, which is the basis for further analysis.

The next stage is data validation, which uses two main parameters. The first parameter is the Mean Absolute Percentage Error (MAPE), which measures the accuracy level of Twitter data predictions compared to actual data from traffic counting. The formula used is:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Actual_i - Predicted_i}{Actual_i} \right| \times 100\%$$

Where n is the number of observations, $Actual_i$ is the actual value of the traffic counting data at point i , and $Predicted_i$ is the predicted value of the Twitter data analysis results at point i . This parameter provides an average error value in percentage form, which reflects the extent to which the Twitter data can represent the actual conditions.

The second parameter is the Spatially Weighted Structural Similarity Index (SW-SSIM), which compares spatial patterns between Twitter and traffic counting data. The formula is

$$SW-SSIM = \frac{\sum W_{ij} \cdot SSIM(X_{ij}, Y_{ij})}{\sum W_{ij}}$$

Where $SSIM(X_{ij}, Y_{ij})$ as the structural similarity value between elements X_{ij} (OD matrix of Twitter data) and Y_{ij} (OD matrix of traffic counting data), and W_{ij} as the spatial weight applied to the element, SW-SSIM provides a measure of the degree of similarity of spatial patterns between the two datasets, considering the

differences in spatial structure in the study area.

The combination of spatial analysis and data validation ensures that Twitter data can accurately represent people's trip generation patterns in East Bandung. The results are visualized in a desired line map, which shows the relationship between sub-districts, volume, and direction of travel. This map helps identify sub-districts with high trip generation concentrations and key connections between regions as a basis for transportation planning.

The method used in this study refers to the approach applied by Liao et al. (2022). However, this study has been modified to suit the context of the East Bandung area, which has specific urbanization and community mobility characteristics. These modifications were made to increase the relevance and accuracy of the analysis results and support the development of a more efficient and connected transportation system in the study area.

C. Analysis Results and Discussion

Trip generation based on Traffic Counting

In this study, initial findings show the results of vehicle volume measurements using the traffic counting method carried out at several strategic points in the East Bandung area, such as Arcamanik, Rancasari, Ujung Berung, Cibiru, Cinambo, and Bandung Kidul. This data represents vehicle volume based on time (morning, afternoon, evening) and day category (weekdays and weekends), providing a comprehensive picture of community trip generation patterns.

The analysis results show that vehicle volume tends to increase significantly in the morning and evening, especially in the Rancasari and Arcamanik areas on weekends, which indicates high community activity towards recreational locations or public facilities. Conversely, on weekdays, vehicle volume is more evenly distributed across locations, with a dominance in the afternoon in the Gedebage and Rancasari areas, reflecting routine travel patterns to and from work or educational institutions.

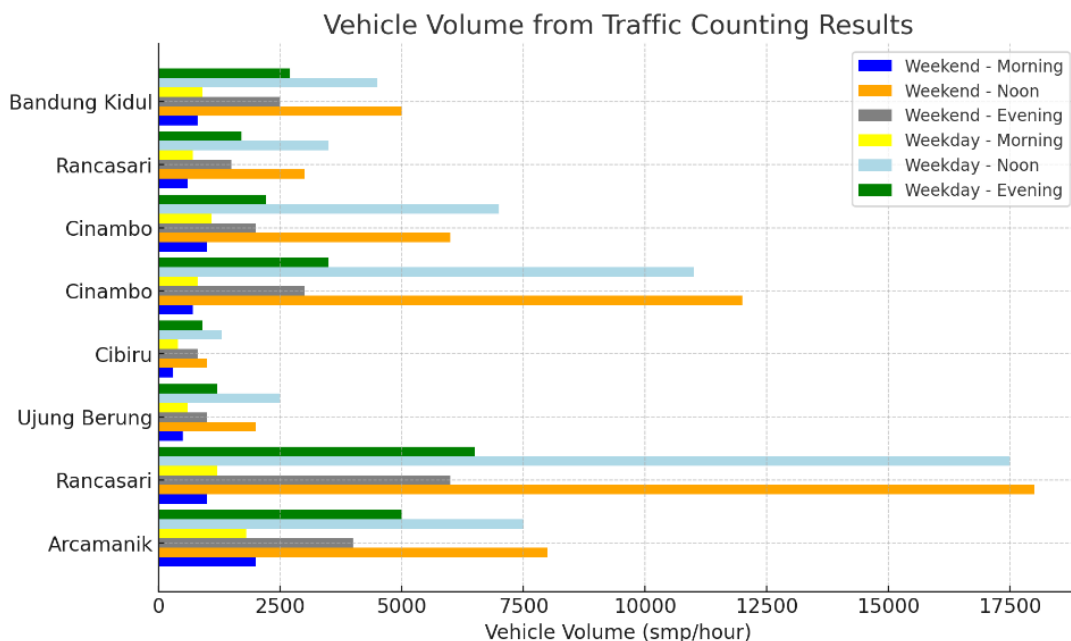


Figure 4. Traffic Volume

These findings reveal the dynamics of community mobility influenced by time and type of activity, such as work on weekdays and social or recreational activities on weekends. This graph provides important insights into the distribution of vehicle volume in certain locations, which is the basis for planning traffic management and developing transportation infrastructure based on the specific needs of each location and time.

Trip Generation Based on Twitter

The community trip generation data based on Twitter geolocation analysis is used to identify mobility patterns in the East Bandung area. The analysis results show that the Gedebage, Arcamanik, and Rancasari areas have the highest concentration of trip generation. This is closely related to the existence of main facilities that are the centre of community activities, such as the Gelora Bandung Lautan Api Stadium, the Al-Jabbar Mosque, and the Summarecon Bandung commercial area, which significantly

affects the intensity of trip generation in the area.

The distribution of trips between sub-districts is visualized through the OD Matrix Table, which provides a quantitative overview of the origin and destination of community trips. For example, Gedebage recorded a high volume of inbound and outbound trips, with the main connections to the Rancasari and Arcamanik sub-districts. Rancasari shows strong connections with other areas, emphasizing its role as a link between the centre of community activities and other sub-districts. In addition, the Arcamanik area, besides having significant connections with Gedebage, also shows moderate interactions with areas such as Cinambo and Panyileukan. This OD Matrix table provides important insights into the main points of community trip generation and how sub-districts in East Bandung are interconnected.

Table 1. OD Matrix

O\D	Antapani	Arcamanik	Bandung Kidul	Buah Batu	Cibiru	Cinambo	Gedebage	Mandalajati	Rancasari	Panyileukan	Ujung Berung	Total generation
Antapani	-	30	6	3	5	2	8	9	2	5	7	77
Arcamanik	45	-	3	2	1	15	7	3	2	10	1	89
Bandung Kidul	10	5	-	15	5	5	5	5	5	5	5	65
Buah Batu	5	4	22	-	7	3	25	5	2	15	0	88
Cibiru	4	7	0	3	-	8	11	2	0	10	3	48
Cinambo	20	18	3	1	0	-	3	4	0	1	2	52
Gedebage	5	5	4	20	6	1	-	2	3	15	0	61
Mandalajati	5	1	3	4	6	0	9	-	0	3	1	32
Rancasari	10	11	2	21	5	5	27	10	-	3	3	97
Panyileukan	2	4	0	5	7	0	6	2	5	-	4	35
Ujung Berung	5	1	0	7	5	3	9	0	4	6	-	40
Total attraction	111	86	43	81	47	42	110	42	23	73	26	684

Sources: Analysis Result, 2024

The Desire Line Map depicts the main trip generation routes between sub-districts to provide a clearer visual interpretation. The thickness of the lines on the map reflects the high volume of travel. The dominant routes are seen between Gedebage, Rancasari, and Gedebage and Arcamanik. This emphasizes the importance of Gedebage as a centre of a community trip generation that functions as the main source and destination of

travel. This map also shows secondary connections between areas such as Cinambo and Panyileukan, indicating medium-level interactions. Thus, the Desire Line Map visualizes the connections between sub-districts and helps identify the main transportation routes that need to be prioritized in managing community mobility.

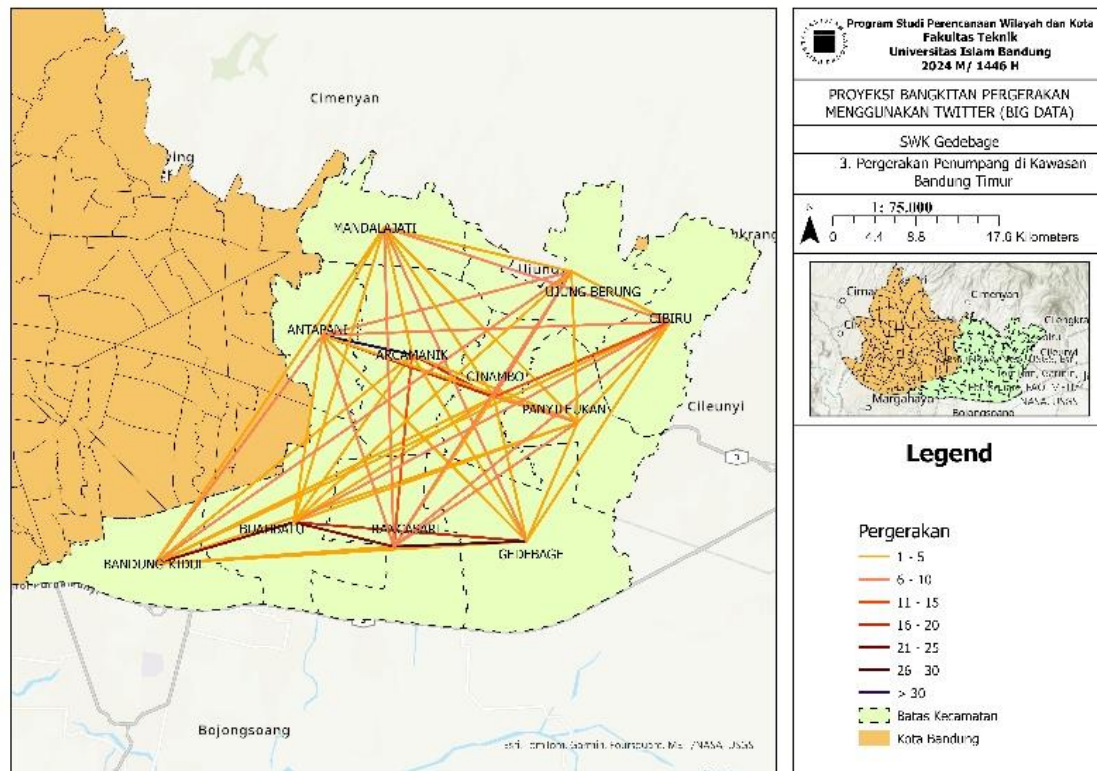


Figure 6. Desire line map of East Bandung

Furthermore, temporal analysis reveals differences in travel patterns between weekdays and weekends. The Weekday and Weekend Trip generation Graphs show that on weekdays, travel is dominated by routine activities to and from work or educational institutions. Travel patterns change significantly on weekends, increasing travel to recreational locations and shopping centres. Areas such as Gedebage, Rancasari, and Arcamanik have become the main destinations for people's travel on weekends, reflecting a change in focus from productivity to recreation and entertainment.

The interpretation of these findings shows that the dynamics of people's mobility are influenced by time, strategic location, and type of activity. The OD Matrix table provides a detailed quantitative description, while the Desire Line Map provides an intuitive and in-depth visualization of people's trip generation patterns. The combination of these two analytical tools provides a strong basis for efficient transportation planning, including managing main travel routes and developing public transportation infrastructure that suits the needs of people in areas with high concentrations of trip generation.

D. Result and Discussion

This study proves that Twitter data has significant potential to project the rise of community trip generation in dynamic urban areas such as East Bandung. Spatial analysis shows that areas with the highest concentration of trip generation, such as Gedebage, Arcamanik, and Rancasari, are the main activity centres, influenced by commercial, sports, educational, and recreational facilities. This data provides real-time and efficient insights compared to traditional methods. However, there are constraints in the form of a bias in the Twitter user population, which young

age groups with high technology access dominate.

Data validation through Mean Absolute Percentage Error (MAPE) and Spatially Weighted Structural Similarity Index (SW-SSIM) showed promising results, with an average error rate of 18.5% and a spatial pattern similarity of 0.72. This confirms that Twitter data can be a reliable tool to represent community trip generation patterns spatially and temporally.

The main advantage of Twitter data is the efficiency of time and cost in collecting information, making it a potential alternative to support transportation planning. However, to overcome the limitations in population representation, integration with other data such as household travel surveys, traffic counting, or GPS data is needed to improve the accuracy and scope of the analysis. This research provides an important contribution to modern transportation planning, especially in developing technology-based methods that are responsive to the needs of people in urban areas. Twitter data can identify priority locations for developing transportation infrastructure, designing more efficient and integrated transportation systems, and supporting sustainable policies. This approach enriches the academic literature while providing practical solutions for policymakers to face mobility challenges in rapidly developing areas, such as East Bandung.

E. Conclusion

This study shows that Twitter data has the potential to project the rise of community trip generation in dynamic urban areas such as East Bandung. Spatial analysis identifies areas with high trip generation activity, such as Gedebage and Arcamanik, which record the largest concentration of trip generation due to main facilities such as office, educational, recreational, and

commercial areas. Although this data offers real-time insights, the bias of the Twitter user population dominated by young age groups and users with high technology access remains an obstacle that needs to be considered.

Validation using parameters such as Mean Absolute Percentage Error (MAPE) and Spatially Weighted Structural Similarity Index (SW-SSIM) shows that, despite its limitations, Twitter data can represent community trip generation patterns quite well. The average error rate of 18.5% is still within the tolerance limit. At the same time, the spatial pattern similarity of 0.72 confirms that this data can be a relevant tool for analyzing community trip generation in urban areas.

The main advantage of Twitter data lies in the efficiency of time and cost in data collection, especially when compared to conventional methods such as field surveys or manual traffic counts. However, these benefits need to be balanced with strategies to overcome data limitations, such as integrating additional data sources, such as household travel surveys, traffic counting, or data from navigation applications that are more inclusive of all levels of society.

This research significantly contributes to developing technology-based transportation planning methods in rapidly growing urban areas. Twitter data can determine public transportation needs, identify priority locations for infrastructure development, and design more efficient and connected transportation systems.

Thus, these findings open up opportunities for further exploration in integrating big data with modern transportation planning. This technology-based approach not only enriches academic literature but also offers practical solutions that policymakers can adopt to create a transportation system responsive to regional dynamics, such as East Bandung.

Acknowledgements

The author would like to thank the Islamic University of Bandung for supporting research facilities and fellow researchers who provided input and support during the research process. The support of all parties plays a major role in the success of this research.

References

1. Abu-Eisheh, S. A., & Fayyad, F. A. (2024). Modeling trip generation rates. *Journal of Transportation Engineering*.
2. Ahmadi, T. (2024). *Metodologi penelitian sosial*. Indonesian Social Science Press.
3. Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*.
4. Liao, Y., Yeh, S., & Gil, J. (2022). Feasibility of estimating travel demand using geolocations of social media data. *Transportation*, 49(1), 137–161.
5. McGuckin, N., & Fucci, A. (2018). Summary of travel trends. Federal Highway Administration.
6. Mukherjee, J., & Kadali, B. R. (2022). A comprehensive review of trip generation models based on land use characteristics. *Transportation Research Part D: Transport and Environment*, 109, 103340.
7. Nurrahman, A. (2024). Studi literatur: Panduan praktis untuk peneliti. *Journal of Research Methodology*.
8. Pemerintah Kota Bandung. (2022). Peraturan Daerah Kota Bandung Nomor 5 Tahun 2022 tentang Rencana Tata Ruang Wilayah Kota Bandung Tahun 2022–2042. Bandung: Pemerintah Kota Bandung.
9. Rodrigue, J. (2020). *The Geography of Transport Systems*. Routledge.
10. Satriani, T., Pratama, M., & Nugraha, R. (2022). Fenomena sosial dalam konteks urbanisasi. *Indonesian Urban Studies Journal*, 15(3), 205–220.
11. Tamin, O. Z. (2000). *Perencanaan transportasi dan analisis jaringan jalan*. ITB Press.
12. Wennerberg, E. (2020). Estimating travel demand from Twitter using an individual mobility model. Chalmers University of Technology.