Published online on the journal's webpage: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

Journal of Dinda

Data Science, Information Technology, and Data Analytics

Vol. 5 No. 2 (2025) 128 - 128 E-ISSN: 2809-8064

Systematic Literature Review: Population Density Mapping Using Data Mining

Naufal Maftuh[1*], Gunawan Ari Nursanto[2], Muhammad Fahrury Romdendine[3]

[1*]Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

[2]Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

[3]Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

[1*]Naufalmaftuh27@gmail.com, [2]gunawan.arinur@gmail.com, [3]romdendine@poltekim.ac.id

Abstract

Mapping population density plays a crucial role in designing and developing urban policies. Traditional methods are often unable to capture complex spatial patterns, making the application of data mining techniques crucial. In this study, we conducted a Systematic Literature Review (SLR) of various data mining techniques, including K-Means, KDE, DBSCAN, Random Forest, linear regression, Cellular Automata, and Fuzzy C-Means. The findings of this study show that although K-Means proved to be effective, it is quite sensitive to the presence of outliers. On the other hand, DBSCAN successfully detects irregular distributions, while KDE is able to track trends despite being computationally intensive. Random Forest and linear regression can predict growth, but both require large datasets to provide accurate results. Meanwhile, Cellular Automata and Fuzzy C-Means offer flexibility, but also require comprehensive data. For future optimization, we recommend using AI-GIS hybrid models.

Keywords: Data Mining, Population Density Mapping, K-Means Clustering, Kernel Density Estimation, DBSCAN, Random Forest, Linear Regression, Fuzzy C-Means, Cellular Automata

© 2025 Journal of DINDA

1. Introduction

Population density mapping plays an important role in urban planning, infrastructure development, and resource distribution. In the midst of increasing urbanization and population growth, data-driven planning is vital for governments and regional planners.

With the advancement of technology, data mining methods now offer a more accurate approach in analyzing population density patterns compared to conventional methods[1]. Various algorithms, such as K-Means Clustering, Kernel Density Estimation (KDE), Random Forest, Linear Regression, and Cellular Automata, have been applied in previous studies to understand population distribution patterns and population growth trends.

This study aims to:

1. Identify data mining techniques used in population density mapping.
2. Analyze the effectiveness and advantages of each method in the context of population mapping.
3. Develop recommendations for future research to optimize the application of data mining in regional planning.

Using the Systematic Literature Review (SLR) approach, this research explores various previous studies to provide a deeper understanding of the application of data mining in population density mapping.

2. Research Methods

This research adopts a Systematic Literature Review (SLR) approach to explore the use of data mining in population density mapping. The literature selection process follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) protocol which includes the following steps:

1. Determine Selection Criteria Determined by the Inclusion Criteria (IC) :
   - IC1: Articles must be original research written in Indonesian and English
   - IC2: The article has been published between 2015 and 2025.

Received: 23-03-2025 | Accepted: 04-06-2025 | Published: 05-08-2025

- IC3: The article aims to analyze the methods or approaches proposed by other researchers to utilize research in Education with the SLR method used in this research.

1. Data Sources

- Articles were collected from databases such as IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar to find research reference sources.
- In articles that are eligible for IC, a search is also conducted to find other studies related to this research,

2. Literature Selection

- Determining keywords are:"Data Mining on Population Density" and "Mapping Clustering for Population Density"
- Search and select abstract titles, and keyword articles obtained from search results based on predefined criteria.
- Read articles that weren't eliminated from the previous stage, either in their entirety or articles that are eligible for future review.
- Selected articles are Re-rated to find related studies. Articles that are included in the reference list and are related to this research will be re-evaluated by conducting stages 3 to 4

3. Data Collection

This study assessed 884,592 articles based on keywords "Data Mining on Population Density" and "Mapping Clustering for Population Density" From a total of 104 identified journals, through a gradual selection process, 28 of the most relevant journals were finally found. Table 1 shows the data that has been successfully collected.

TABLE I. COLLECTED DATA

Source	Study found (Base on Title and keywords		Candidate	Selected
	Data Mining on Population Density media	Mapping Clustering for Population Density a		
IEE-Xplore	11	426	16	4
ScienceDirect	45.802	220.566	21	5
Springer Link	19.939	50.948	15	3
Google Scholar	559.000	17.900	52	16
Total	624.752	289.840	104	28

3. Results and Discussion

Mapping population density is an important element in spatial planning, infrastructure development, and social and economic policy development. By understanding population distribution, governments can formulate more effective strategies in resource management and mitigate the impacts of urbanization. Various studies have utilized data mining techniques to improve the accuracy and efficiency of population density mapping. Among the frequently used methods are K-Means Clustering, Kernel Density Estimation (KDE), Random Forest, Linear Regression, Fuzzy C-Means, DBSCAN and Celullar Automata.

3.1 Data Mining Techniques in Population Density Mapping

Based on the literature review, various data mining techniques have been used in population density mapping. Some of the main methods that are often applied are as follows:

TABEL II. ALGORITHM, MAIN FUNCTION

Algorithm	Main Function
K-Means Clustering	Categorize regions based on population density.
Kernel Density Estimation (KDE)	Analyzing density trends over time.
Random Forest	Predict population growth based on social and economic variables.
Regresi Linear	Identify factors that affect population density.
Cellular Automata	Simulate the dynamics of population density in the long.
DBSCAN	Identifies and categorizes high-density areas, and detects low-density areas or background noise.

3.2 Comparison of Method Effectiveness

Based on the literature review, multiple data mining techniques have been applied to population density mapping. The most frequently used methods include K-Means Clustering, Kernel Density Estimation (KDE), Random Forest, Linear Regression, Fuzzy C-Means, DBSCAN, and Cellular Automata. Each population density mapping method has its own advantages and disadvantages.

1. K-Means Clustering has proven effective in clustering urban and rural areas, although it

lacks flexibility when dealing with data with complex patterns.

2. Kernel Density Estimation (KDE) excels in identifying changes in population density over time, making it an ideal choice for urbanization trend analysis.
3. Random Forest and Linear Regression can be used to forecast future population density with a high degree of accuracy.
4. Cellular Automata could simulate density changes influenced by environmental and spatial factors, although this requires very detailed spatial data.
5. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) could cluster high-density areas without requiring a predetermined number of clusters. In addition, this technique can also recognize low-density areas as noise. Unlike K-Means, DBSCAN offers better flexibility in dealing with irregular distribution patterns. However, it is important to note that the clustering result is highly dependent on the selection of parameters, (epsilon and minPts).

Table III presents a comparative analysis of these methods.

TABEL III. COMPARISON OF DATA MINING ALGORITHM IN POPULATION DENSITY MAPPING

Algorithm	Advantage	Limitation
K-Means Clustering	Effective in categorizing regions based on population density, computationally efficient for large datasets.	Requires predefined cluster numbers, sensitive to outliers.
DBSCAN	Identifies high-density clusters without requiring predefined clusters, effective for irregular distributions.	Highly dependent on parameter selection (epsilon and minPts).
Kernel Density Estimation (KDE)	Captures density trends over time, useful for urbanization trend analysis.	Computationally expensive for large datasets.
Random Forest	Provides high predictive accuracy for population growth based on socio-economic variables.	Computationally intensive, requires large datasets.
Linear Regression	Identifies key factors affecting population density, simple and interpretable.	Assumes linear relationships, may not perform well with complex data.
Fuzzy C-Means	Flexible in handling uncertainty in clustering, useful for mapping social aspects like disease spread.	More computationally expensive than K-Means.
Cellular Automata	Simulates population density dynamics influenced by spatial and environmental factors.	Requires detailed spatial data, sensitive to parameter tuning.

Table IV. shows the heading, Approach, and research results studied by previous researchers for the Use of Data Mining for Population Density Mapping. This classification technique has the potential to grow rapidly in future research, especially if it is further studied to effectively identify variation. With this approach, researchers want to study the Use of Data Mining for Population Density Mapping professionally.

TABEL IV. HEADING, APPROACH AND RESEARCH RESULT

Author	Heading	Approach	Research Results
Jingang Li, Jianwei Li, Yangzi Yuan, Guifang Li	Spatiotemporal Distribution Characteristics and Mechanism Analysis of Urban Population Density: A Case of Xi'an, China	PDI, SCI, OLS, Baidu Heat Map, POI Data	The study shows that the distribution of population density in Xi'an is influenced by land use and population activity, which reflects the characteristics of agglomeration[2].
Katarzyna Kopczewska, Maria Kubara, Mateusz Kopyt	Population density as the attractor of business to the place	DBSCAN clustering	Research shows that population density and business agglomeration have a mutual effect on determining the location of companies in

			various sectors. These two factors complement each other[3].
Philia Christi Latue, Susan E. Manakane, Heinrich Rakuasa	Analisis Perkembangan Kepadatan Permukiman di Kota Ambon Tahun 2013 dan 2023 Menggunakan Metode Kernel Density	Kernel Density, Citra Satelit IKONOS dan SPOT	The results of the study show that during the period 2013-2023, there was a significant increase in settlement density in Ambon City, which reflects rapid urban growth[4].
Mangapul Siahaan	Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means	K-Means, CRISP-DM	Population density mapping in Batam aims to identify development priorities in areas with high density and growth rates[5].
Preddy Marpaung, R. Fanry Siahaan	Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan	K-Means Clustering	The mapping results show that there are 121 villages in Medan that have a very high population density, while 30 other villages are recorded with high density. Interestingly, there are no low-density villages in the area[6].
Frisma Handayanna, Sunarti	Penerapan Algoritma K-Means Untuk Mengelompokkan Kepadatan Penduduk Di Provinsi DKI Jakarta	K-Means Clustering, RapidMiner	The results of the cluster analysis show that South, East, and West Jakarta have high levels of density and continue to increase. These findings provide the basis for developing solutions for managing congestion in the region[7].
Basri, Andani Achamad, Hazriani Cita St. Munthakhabah	Pemetaan dan Prediksi Pertumbuhan Penduduk Menggunakan Regresi Linear	Regresi Linear, SIG	Population growth predictions in Tutar District show that there is a need to build Pustu health facilities between 2022 and 2027[8].
Preddy Marpaung, Ibnu Pebrian, Widia Putri	Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means	K-Means Clustering	The mapping results show that in Deli Serdang there are 3 sub-districts with very high population density, 4 sub-districts with high density, and 15 sub-districts with moderate density[9].
Rizki Fadillah, Sarjon Defit, Sumijan	Implementasi Data Mining untuk Pemetaan Persebaran Infeksi HIV/AIDS di Provinsi Riau	Fuzzy C-Means, Clustering, Sistem Informasi Geografis (GIS)	The mapping results revealed the existence of three clusters, namely a safe zone that includes five districts/cities, a alert zone with five districts/cities, and a dangerous zone consisting of two districts/cities[10].

Gilang Yudistira Hilman, Bandi Sasmito, Arwan Putra	Pemetaan Daerah Rawan Kriminalitas di Wilayah Hukum Poltabes Semarang Tahun 2013 dengan Metode Clustering	K-Means, SIG	The results of the clustering analysis showed that urban areas had a higher level of crime vulnerability, while suburban areas tended to be safer[11].
Nana Suarna, Willy Prihartono	Analisis Jumlah Penduduk Menggunakan Algoritma K-Means Berdasarkan Kabupaten/Kota di Indonesia	K-Means, Data Mining, Knowledge Discovery and Data Mining (KDD)	The results of the grouping process show that Indonesia can be divided into four clusters based on population density, namely very dense, dense, medium, and low[12].
Abdul Karim, Adeni, Fitri, Alifa Nur Fitri	Pemetaan untuk Strategi Dakwah di Kota Semarang Menggunakan Pendekatan Data Mining	Fuzzy C-Means, Clustering	The results of the FCM analysis revealed that there are two clusters: clusters with high da'wah potential and clusters with low da'wah potential. These findings can help in planning more effective da'wah strategies[13].
Martin Lnenicka, Jan Hovad, Jitka Komarkova	A Proposal of Web Data Mining Application for Mapping Crime Areas in the Czech Republic	Web Data Mining, GIS, Python	This web application successfully detects criminal hotspots by utilizing data from various mass media sources and mapping existing crime trends[14].
Hasi Bagan, Yoshiki Yamagata	Analysis of Urban Growth and Estimating Population Density Using Satellite Images of Nighttime Lights	DMSP Nighttime Lights, OLS, GWR, Data Penginderaan Jauh	Analysis of the relationship between population density and urban areas utilizing night lighting data (DMSP) provides valuable insights for urban planning. [15].
Tianjun Wu, Jiancheng Luo, Wen Dong, Lijing Gao	Disaggregating County-Level Census Data for Population Mapping Using Residential Geo-Objects	Random Forests, XGBoost, Geospatial Data, Residential Geo-Objects	This method successfully maps the distribution of the population with a higher level of detail, utilizing geo-residential objects and more accurate spatial data[16].
Kevan Edinborough, Rémi Martineau, Alexa Dufraisse, Stephen Shennan, Marie Imbeaux, Anthony Dumontet, Peter Schauer, Gordon Cook	A Neolithic population model based on new radiocarbon dates from mining, funerary and population scaled activity in the Saint-Gond Marshes region of North East France	Radiocarbon dating, Summed Probability Distribution (SPD), Simulation-based modeling	This study revealed that there was a peak population density in certain regions associated with the construction of the hypogeum in the late Neolithic period, which was between 3650 and 2900 BC[17].
Forrest R. Stevens, Andrea E. Gaughan, Catherine Linard, Andrew J. Tatem	Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data	Random Forest, Remote sensing, Ancillary data	By utilizing census data and spatial data, this study succeeded in producing population density predictions with a resolution of about 100 meters and

			comparing them with other methods[18].
Lionel Gueguen, Jan Koenig, Carl Reeder, Tim Barksdale, Jon Saints, Kostas Stamatiou, Jeffery Collins, Carolyn Johnston	Mapping Human Settlements and Population at Country Scale From VHR Images	Ultra-high-resolution satellite imagery (VHR) processing, Crowdsourced validation, Automated mosaic engineering	The results of the study show a very high level of accuracy in mapping human settlements. With the validation carried out in a crowdsourced manner, we managed to achieve 100% precision at the village boundaries. [19].
Litao Wang, Shixin Wang, Yi Zhou, Wenliang Liu, Yanfang Hou, Jinfeng Zhu, Futao Wang	Mapping population density in China between 1990 and 2010 using remote sensing	Remote sensing, Partial correlation analysis, Geo-weighted regression	This method results in a more accurate mapping of population density compared to previous methods, with a correlation coefficient that reaches more than 0.90[20].
Nirav N. Patel, Emanuele Angiuli, Paolo Gamba, Andrea Gaughan, Gianni Lisini, Forrest R. Stevens, Andrew J. Tatem, Giovanna Trianni	Multitemporal settlement and population mapping from Landsat using Google Earth Engine	Google Earth Engine, Landsat, Klasifikasi Normalized Difference Spectral Vector (NDSV)	By utilizing Google Earth Engine (GEE) and Landsat imagery, the mapping results show a high level of accuracy in the classification of urban areas. GEE-based population mapping has also increased significantly[21].
Edith Darin, Ahmadou Hamady Dicko, Hisham Galal, Rebeca Moreno Jimenez, Hyunju Park, Andrew J. Tatem, Sarchil Qader	Mapping refugee populations at high resolution by unlocking humanitarian administrative data	Random Forest, Satellite-based settlement maps, Refugee administration data	This method maps the refugee population with a high degree of accuracy, utilizing administrative data and satellite imagery. This provides a more detailed picture to support humanitarian relief efforts. [22].
Yaping Liu, Juanle Wang, Keming Yang, Altansukh Ochir	Mapping livestock density distribution in the Selenge River Basin of Mongolia using random forest	Random Forest, Livestock density mapping, Environmental data	The results of the study show that the density of livestock in the southwestern part of Selenge is very high. Further analysis revealed that human factors are the main drivers in the distribution of livestock in the region[23].
John R. Ellis, Natalia B. Petrovskaya	A computational study of density-dependent individual movement and the formation of population clusters in two-dimensional spatial domains	Individual-based models, simulations of Brownian and non-Brownian movements	The results of the study show that the density of livestock is quite high in the southwestern part of Selenge. The analysis revealed that human factors are the main drivers for this condition. In addition, the simulations performed showed that the movement of individuals affected by density contributes to the formation of stable

			population clusters in two-dimensional space. [24].
Yimeng Song, Yong Xu, Bin Chen, Qingqing He, Ying Tu, Fei Wang, Jixuan Cai	Dynamic population mapping with AutoGluon	AutoGluon, Geospatial data, Machine learning	This framework produces a highly accurate population density map, depicting population dynamics in China with an R2 value of 0.974 and an RMSD of 427.61[25].
Marita Prasetyani, R. Rizal Isnanto, Catur Edi Widodo	Exploring Visitor Density Trends in Rest Areas Through Google Maps Data and Data Mining	Data Mining, Web Scraping, Google Maps API	Identify visitor density trends in rest areas to support more optimal planning[26].
Jinchao Song, Xiaoye Tong, Lizhe Wang, Chunli Zhao, Alexander V. Prishchepov	Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach	Geographically Weighted Regression (GWR), Remote Sensing, Night-Time Light Data	GWR-based models that utilize fusion data can improve accuracy in mapping population densities in urban areas[27].
Yao Yao, Xiaoping Liu, Xia Li, Jinbao Zhang, Zhaotang Liang, Ke Mai, Yatao Zhang	Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data	Random Forest Algorithm, Gravity Model, GIS, Big Data	Integrative models can improve accuracy in population mapping down to the building level[28].
Miao Wang, Meizi Yang, Xu-dong Yang, Juan Chen, Bogang Yang	Data Mining of National Geographical Census for Decision-making in Urban Planning: A Geo-simulation of Urban Size in Beijing, China	Cellular Automaton, Machine Learning, Data Mining	EL-CA model improves urban development prediction in [29].

Based on the data that has been obtained from the table above, the methods used in the study can be classified into:

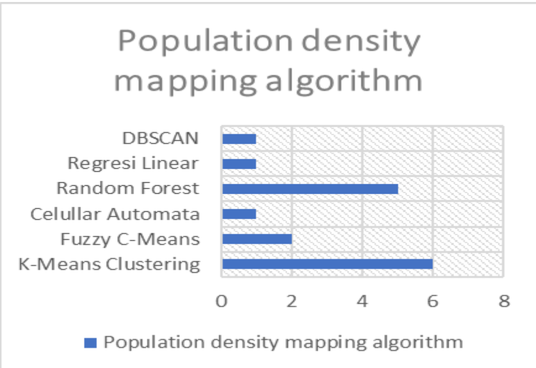
TABELV. RESEACH METHODS

No	Author	Methods																		
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Jingang Li et.al	v																		
2	Kopczewska et.al															v				
3	Philia Christi Latue et.al		v																	
4	Mangapul Siahaan				v	v														
6	Preddy Marpaung et.al				v															
7	Frisma Handayanna, Sunarti				v															
8	Basrid et.al					v			v											
9	Preddy .M et.al				v															
10	Rizki Fadillah et.al						v	v												
11	Gilang Yudistira Hilman et.al				v				v											
12	Nana Suarna et.al				v															
13	Abdul Karim et.al						v													
14	Martin Lnenicka et.							v												
15	Hasi Bagan et.al									v										
16	Tianjun Wu et.al										v									
17	Kevan E, et.al											v								
18	Forrest R. Stevens et.al										v		v							
19	Lionel Gueguen et.al													v						
20	Litao Wang et.al														v					

21	Nirav N. Patel et.al															v					
22	Edith Darin et.al											v									
23	Yaping Liu et.al											v									
24	John R. Ellis et.al																v				
25	Yimeng Song et.al																	v			
26	Marita Prasetyani et.al																		v		
27	Jinchao Song									v				v							
28	Yao Yao et.al							v			v										
29	Miao Wang et.al																				v
	Jumlah	1	1	1	6	1	2	2	1	2	5	1	2	1	1	1	1	1	1	1	1

Based on the data explained by the table above, A is PDI, B is Kernel Density, C is Crisp-Dm, D is K-means Clustering, E is linear regression, F is Fuzzy C-Means, G is GIS, H is SIG, I is GWR, J is Random Forest, K is SPD, L is Remote sensing, M is VHR, N is Remote Sensing, O is DBSCAN, Q adalah auto gluon, R is web scraping, S is cellular automata. In this study, based on this research, K-Means Clustering is the most used algorithm.

Fig 1. Comparison of the Use of Algorithms in Population Density Mapping



3.3 Research Implications

Based on the results of the study that has been conducted, there are several important implications related to the application of data mining in population density mapping:

1. Efficiency in Infrastructure Planning
- With more accurate density mapping, the government can more effectively plan infrastructure development such as roads, health facilities, and schools in areas experiencing rapid population growth.
2. Spatial Optimization and Urbanization Policy
- Studies conducted in Ambon City show an increase in settlement density in the last decade. This requires the need for a better urban planning management strategy in order to avoid congestion and land shortages[4].
3. Mapping Crime and Disease Spread
- Analysis conducted in Semarang City revealed a relationship between population density and crime rates. By utilizing the K-Means Clustering method, areas with high crime rates can be identified to optimize security policies[11].
- Meanwhile, in Riau Province, the Fuzzy C-Means Clustering method is used to map the spread of HIV/AIDS disease, which is certainly very helpful in the distribution of health resources more effectively(Fadillah et al., 2024).

4. Conclusion

In this study, a systematic review of various data mining techniques used in population density mapping has been carried out. Through the analysis of the literature that has been reviewed, it was found that the data mining method plays an important role in improving the accuracy and efficiency of population analysis, as well as providing deeper insights into population distribution patterns.

1. The main results of this study show that:
- K-Means Clustering is the most widely applied method for grouping areas based on population density, thus contributing to decision-making related to spatial planning and infrastructure distribution.
- Kernel Density Estimation (KDE) has proven to be effective in analyzing trends in population density changes over time, which can be leveraged to design more sustainable urbanization policies.
- Random Forest and Linear Regression demonstrate high predictive ability in projecting population growth, which can be used to support long-term development planning.
- Cellular Automata offers a simulation-based approach that is able to represent population density dynamics based on environmental factors and spatial planning policies.

- Fuzzy C-Means allows for more flexible analysis in regional grouping, which can be applied in mapping social aspects such as disease spread and crime rates.
- DBSCAN is very effective in identifying high-density population clusters without having to determine the number of clusters in advance. This makes it an ideal method for detecting irregular spatial distributions as well as overcoming noise in population mapping. This method is particularly useful in urban planning and business location analysis, as it is able to distinguish between densely populated areas and less significant outlier areas.

2. The implications of the results of this study include:
   - The application of data mining techniques in population density mapping can improve the efficiency of infrastructure planning, especially in the allocation of resources and public services in areas with rapid population growth.
   - The data mining method can be used as a support tool in evidence-based policy-making, especially in the health, security, and urban spatial management sectors.
   - The combination of various data mining methods can result in more comprehensive and adaptive analytical models, allowing the identification of priority areas in development and mitigation of social risks.

As a recommendation for future research, the integration of data mining techniques with Geographic Information Systems (GIS) needs to be further explored to improve the accuracy of population mapping. In addition, the development of machine learning-based models, especially deep learning techniques, can be a more sophisticated approach in predictive analysis of population density trends.

Overall, this study confirms that the use of data mining in population density mapping can be an effective approach to support data-based, sustainable, and responsive regional planning to demographic dynamics.

References

[1] D. Guo and X. Zhu, “Origin-destination flow data smoothing and mapping,” *IEEE Trans Vis Comput Graph*, vol. 20, no. 12, 2014, doi: 10.1109/TVCG.2014.2346271.

[2] J. Li, J. Li, Y. Yuan, and G. Li, “Spatiotemporal distribution characteristics and mechanism analysis of urban population density: A case of Xi’an, Shaanxi, China,” *Cities*, vol. 86, pp. 62–70, Mar. 2019, doi: 10.1016/j.cities.2018.12.008.

[3] K. Kopczewska, M. Kubara, and M. Kopyt, “Population density as the attractor of business to the place,” *Sci Rep*, vol. 14, no. 1, p. 22234, Dec. 2024, doi: 10.1038/s41598-024-73341-8.

[4] P. C. Latue, S. E. Manakane, and H. Rakuasa, “Analisis Perkembangan Kepadatan Permukiman di Kota Ambon Tahun 2013 dan 2023 Menggunakan Metode Kernel Density,” *Blend Sains Jurnal Teknik*, vol. 2, no. 1, pp. 26–34, Jun. 2023, doi: 10.56211/blendsains.v2i1.272.

[5] M. Siahaan, “Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 3, pp. 316–324, Dec. 2022, doi: 10.32736/sisfokom.v11i3.1453.

[6] P. Kepadatan, P. Berdasarkan, J. Penduduk, K. Medan, P. Marpaung, and R. F. Siahaan, “Penerapan Algoritma K-Means Clustering Untuk,” 2021.

[7] F. Handayanna and S. Sunarti, “Penerapan Algoritma K-Means Untuk Mengelompokkan Kepadatan Penduduk Di Provinsi DKI Jakarta,” *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 50–55, Mar. 2024, doi: 10.52158/jacost.v5i1.477.

[8] A. A. H. C. St. M. R. Basri, “Sistem Informasi Geografis Pemetaan dan Prediksi Pertumbuhan Penduduk Menggunakan Regresi Linear,” *Bulletin of Information Technology (BIT)*, vol. 4, no. 2, pp. 171–177, 2023, doi: 10.47065/ bit.v3i1. 633.

[9] Preddy, P. Marpaung, I. Pebrian, and W. Putri, “Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 6, no. 2, pp. 64–70, 2023.

[10] R. Fadillah, Sarjon Defit, and Sumijan, “Implementasi Data Mining untuk Pemetaan Persebaran Infeksi Human Imunodeficiency Virus di Provinsi Riau,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 1, pp. 151–158, May 2024, doi: 10.37859/coscitech.v5i1.6712.

[11] G. Yudistira Hilman, B. Sasmito, and A. Putra Wijaya, “PEMETAAN DAERAH RAWAN KRIMINALITAS DI WILAYAH HUKUM POLTABES SEMARANG TAHUN 2013 DENGAN MENGGUNAKAN METODE CLUSTERING,” 2015.

[12] N. Suarna and W. Prihartono, “ANALISIS JUMLAH PENDUDUK MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN KABUPATEN/KOTA DI INDONESIA,” 2024. [Online]. Available: https://www.bps.go.id/.

[13] A. Karim *et al.*, “Pemetaan untuk Strategi Dakwah di Kota Semarang Menggunakan Pendekatan Data Mining (Mapping for Da’wah Strategy in Semarang City Using Data Mining Approach),” *Jurnal Dakwah Risalah*, vol. 32, no. 1, p. 40, Jun. 2021, doi: 10.24014/jdr.v32i1.12549.

[14] M. Lnenicka, J. Hovad, J. Komarkova, and M. Pasler, “A proposal of web data mining application for mapping crime areas in the Czech Republic,” in *ICSOFT-EA 2015 - 10th International Conference on Software Engineering and Applications, Proceedings; Part of 10th International Joint Conference on Software Technologies, ICSOFT 2015*, SciTePress, 2015, pp. 450–455. doi: 10.5220/0005558104500455.

[15] H. Bagan and Y. Yamagata, “Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data,” *GIsci Remote Sens*, vol. 52, no. 6, pp. 765–780, Nov. 2015, doi: 10.1080/15481603.2015.1072400.

[16] T. Wu *et al.*, “Disaggregating County-Level Census Data for Population Mapping Using Residential Geo-Objects with Multisource Geo-Spatial Data,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 13, pp. 1189–1205, 2020, doi: 10.1109/JSTARS.2020.2974896.

[17] K. Edinborough *et al.*, “A Neolithic population model based on new radiocarbon dates from mining, funerary and population scaled activity in the Saint-Gond Marshes region of North East France,” *Quaternary International*, vol. 586, pp. 121–132, Jun. 2021, doi: 10.1016/j.quaint.2021.03.001.

[18] F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem, “Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data,” *PLoS One*, vol. 10, no. 2, Feb. 2015, doi: 10.1371/journal.pone.0107042.

[19] L. Gueguen *et al.*, “Mapping Human Settlements and Population at Country Scale from VHR Images,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 10, no. 2, pp. 524–538, Feb. 2017, doi: 10.1109/JSTARS.2016.2616120.

[20] L. Wang *et al.*, “Mapping population density in China between 1990 and 2010 using remote sensing,” *Remote Sens Environ*, vol. 210, pp. 269–281, Jun. 2018, doi: 10.1016/j.rse.2018.03.007.

[21] N. N. Patela *et al.*, “Multitemporal settlement and population mapping from landsatusing google earth engine,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 35, no. PB, pp. 199–208, 2015, doi: 10.1016/j.jag.2014.09.005.

[22] E. Darin *et al.*, “Mapping refugee populations at high resolution by unlocking humanitarian administrative data,” *Journal of International Humanitarian Action*, vol. 9, no. 1, p. 14, Oct. 2024, doi: 10.1186/s41018-024-00157-6.

[23] Y. Liu, J. Wang, K. Yang, and A. Ochir, “Mapping livestock density distribution in the Selenge River Basin of Mongolia using random forest,” *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-61959-7.

[24] J. R. Ellis and N. B. Petrovskaya, “A computational study of density-dependent individual movement and the formation of population clusters in two-dimensional

spatial domains," *J Theor Biol*, vol. 505, Nov. 2020, doi: 10.1016/j.jtbi.2020.110421.

[25] Y. Song *et al.*, "Dynamic population mapping with AutoGluon," *Urban Informatics*, vol. 1, no. 1, Nov. 2022, doi: 10.1007/s44212-022-00017-x.

[26] M. Prasetyani, R. Rizal Isnanto, and C. Edi Widodo, "Article title 4 Exploring Visitor Density Trends in Rest Areas Through Google Maps Data and Data Mining Keywords." [Online]. Available: https://www.google.com/maps/search/rest+area/

[27] J. Song, X. Tong, L. Wang, C. Zhao, and A. V. Prishchepov, "Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach," *Landsc Urban Plan*, vol. 190, Oct. 2019, doi: 10.1016/j.landurbplan.2019.05.011.

[28] Y. Yao *et al.*, "Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data," *International Journal of Geographical Information Science*, vol. 31, no. 6, pp. 1220–1244, Jun. 2017, doi: 10.1080/13658816.2017.1290252.

[29] M. Wang, M. Yang, X. D. Yang, J. Chen, and B. Yang, "Data mining of national geographical census for decision-making in urban planning: A geo-simulation of urban size in Beijing, china," *Sensors and Materials*, vol. 35, no. 3, pp. 965–974, 2023, doi: 10.18494/SAM4241.

[30] R. Fadillah, Sarjon Defit, and Sumijan, "Implementasi Data Mining untuk Pemetaan Persebaran Infeksi Human Imunodeficiency Virus di Provinsi Riau," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 1, pp. 151–158, May 2024, doi: 10.37859/coscitech.v5i1.6712.