Received: 05/12/2025 | Revised: 04/02/2026 | Accepted: 12/04/2026
DOI: 10.69816/jgd.v3i1.48733

ORIGINAL ARTICLE

TAMALANREA
JOURNAL OF GOVERNMENT AND DEVELOPMENT

K-Means clustering of crime and socioeconomic factors as a basis for public policy formulation in Indonesia

Karina Khansa Thufailah[1]

Affiliation
[1]Department of Public Administration, Faculty of Social and Political Sciences, UIN SGD Bandung, Jawa Barat, Indonesia 40614.

Correspondence
deliana.ahsanti@gmail.com

Funding Information
This research did not receive any specific grants from any funding agencies.

Abstract
Crime is a multidimensional social phenomenon that is closely associated with regional socioeconomic conditions. This study aims to classify the 34 provinces of Indonesia based on combined crime and socioeconomic indicators using the K-Means clustering method to support evidence-based public policy formulation. The analysis utilised secondary data obtained from the Indonesian Central Bureau of Statistics (BPS), comprising three crime indicators—narcotics-related crimes, theft, and physical violence—and three socioeconomic indicators: the Open Unemployment Rate (OUR), Gross Regional Domestic Product (GRDP), and the Human Development Index (HDI). Prior to clustering, all variables were standardised, and the optimal number of clusters was determined using the Elbow and Silhouette methods, which identified a two-cluster solution (K = 2). The results revealed substantial disparities in crime and socioeconomic characteristics across provinces. Cluster 1, consisting of DKI Jakarta, North Sumatra, West Java, and East Java, exhibited higher levels of crime, economic activity, and unemployment. In contrast, Cluster 0 comprised the remaining provinces and was characterised by lower crime rates and less intensive socioeconomic dynamics. The findings indicate that provinces with higher unemployment and greater economic concentration tend to experience higher crime rates, supporting strain theory and previous empirical studies. This study contributes to the literature by integrating multiple crime and socioeconomic indicators within an unsupervised machine learning framework to identify provincial typologies that can inform differentiated, place-based policies for crime prevention and regional development in Indonesia.

Keywords
Crime, Socioeconomic Factors, K-Means Clustering, Public Policy, Unemployment, GRDP, HDI.

1 | INTRODUCTION

Crime is an important social indicator that reflects the effectiveness of public policies in maintaining social welfare, public security, and societal stability (Tumalavičius et al., 2017). Persistent disparities in crime rates across regions in Indonesia indicate that crime is not solely an issue of individual behavior but is closely associated with broader socioeconomic conditions (Sugiharti et al., 2022, 2023; Widyastaman & Hartono, 2022). Provinces with limited economic opportunities, high levels of social inequality, and inadequate access to public services tend to experience greater vulnerability to various forms of crime. Consequently, understanding the relationship between crime and socioeconomic conditions is essential for developing evidence-based policies that promote equitable and sustainable development.

Studies have consistently demonstrated that socioeconomic factors such as poverty, unemployment, and limited economic opportunities contribute to increased crime rates by generating economic pressure, uncertainty, and social exclusion (Sugiharti et al., 2022, 2023). These conditions may encourage individuals to adopt deviant behaviors as coping mechanisms or survival strategies in response to structural constraints (Vadera et al., 2013). Therefore, crime should be viewed not only as a legal or moral issue but also as a manifestation of socioeconomic inequality and the limited effectiveness of public policies in ensuring inclusive welfare distribution.

The relationship between socioeconomic conditions and crime has long been explained through strain theory. Individuals may experience psychological strain when they encounter barriers to achieving socially accepted goals through legitimate means (Ross, 2017; Van Laar et al., 2019). When access to economic resources and opportunities is restricted, some individuals may resort to alternative pathways outside established social norms, including criminal activities. This theoretical perspective suggests that crime emerges not merely from individual choices but also from structural inequalities embedded within society.

Variations in crime rates, including narcotics-related offenses, theft, and physical violence, are influenced by multiple socioeconomic factors (Anozi & Novianda, 2023; Cao et al., 2026; Contreras & Hipp, 2020). Unemployment is widely recognized as a significant determinant of economically motivated crimes because it increases financial insecurity and reduces access to legitimate sources of income. Individuals facing prolonged unemployment may become more susceptible to engaging in theft or participating in illicit activities, including drug distribution and abuse, either as a source of income or as a coping mechanism for economic stress.

In addition, Gross Regional Domestic Product (GRDP) reflects the economic capacity and productivity of a region and is closely associated with crime patterns (Santos et al., 2024). Regions characterized by low GRDP often experience limited economic activity, fewer employment opportunities, and wider income disparities, creating conditions conducive to property crimes such as theft. Economic deprivation may also contribute to aggressive forms of crime, including physical violence, as a consequence of prolonged social stress and frustration (Coccia, 2018; Wong, 2021). Meanwhile, the Human Development Index (HDI), which encompasses indicators of education, health, and living standards, provides insight into the quality of human capital within a region (Dasic et al., 2020; Rao & Min, 2018). Provinces with lower HDI levels frequently encounter challenges related to educational attainment, healthcare access, and overall quality of life, increasing the risk of criminal behavior. Limited educational opportunities may contribute to impulsive and violent behavior, while inadequate access to health services and rehabilitation programs may exacerbate drug-related crimes.

Although previous studies have extensively examined the influence of individual socioeconomic variables on crime rates, most have focused on linear relationships or isolated determinants (Contreras & Hipp, 2020; Santos et al., 2024; Van Laar et al., 2019). Relatively few studies have adopted a multidimensional perspective that simultaneously considers multiple crime indicators and socioeconomic characteristics to identify regional patterns of vulnerability. Furthermore, existing research in Indonesia has predominantly emphasized explanatory analyses,

while the use of unsupervised machine learning approaches for crime mapping and policy formulation remains limited.

To address this gap, this study employs the K-Means clustering method to classify Indonesian provinces based on a combination of crime indicators and socioeconomic variables, including unemployment rates, GRDP, and HDI. Unlike previous studies that primarily investigate causal relationships between individual variables, this study focuses on identifying latent patterns and similarities among provinces through data-driven clustering. The novelty of this study lies in three key aspects. First, it integrates multiple crime dimensions—namely narcotics offenses, theft, and physical violence—with key socioeconomic indicators within a single analytical framework. Second, it applies an unsupervised machine learning approach to generate evidence-based provincial typologies that capture the heterogeneity of crime and socioeconomic conditions across Indonesia. Third, it translates the resulting clusters into actionable policy insights by proposing differentiated intervention strategies tailored to the specific characteristics of each provincial group.

Accordingly, this study aims to examine the relationship between socioeconomic indicators and variations in crime rates across Indonesian provinces, classify provinces into two clusters based on combined crime and socioeconomic indicators, and identify the distinguishing characteristics of each cluster. The findings are expected to provide a more comprehensive understanding of regional crime patterns and serve as a foundation for developing targeted and context-specific public policies to reduce crime and improve social welfare in Indonesia.

2 | LITERATURE REVIEW

2.1 | Crime and Socioeconomic Conditions

Crime refers to actions that violate established legal norms and social values, thereby disrupting social order and community well-being (Sloane & Choi, 2016). Beyond its legal dimension, crime represents a complex social phenomenon that is closely associated with various structural and socioeconomic problems. High crime rates often indicate deficiencies in social welfare, unequal access to economic opportunities, and ineffective public policies aimed at promoting social inclusion (Kidd, 2017).

In Indonesia, crime remains a persistent public concern due to its adverse consequences for public safety, economic development, and social stability (Purnomo et al., 2023). The occurrence of criminal activities is influenced not only by individual characteristics but also by broader contextual factors, including poverty, unemployment, income inequality, educational attainment, and regional economic performance. Consequently, understanding crime requires a multidimensional perspective that considers the interaction between socioeconomic conditions and criminal behaviour.

Previous studies have demonstrated that socioeconomic deprivation increases the likelihood of criminal activity by generating financial pressure, social exclusion, and limited access to legitimate economic opportunities (DeMarco et al., 2021; Koku, 2015; Reiman & Leighton, 2020). Unemployment, for example, reduces individuals' income-generating capacity and may increase incentives to engage in economically motivated crimes. Similarly, disparities in income distribution and inadequate access to education and healthcare can weaken social cohesion and increase vulnerability to deviant behaviour.

2.2 | Theoretical Foundation

Social Disorganisation Theory

The Social Disorganisation Theory explains that crime rates tend to be higher in communities characterised by socioeconomic disadvantage, residential instability, and weak social institutions (Kubrin, 2009). According to this theory, structural conditions such as poverty, population mobility, and social inequality reduce a community's capacity to exercise informal social control, thereby increasing the likelihood of criminal behaviour.

Social disorganisation occurs when community networks, shared values, and collective efficacy are weakened, resulting in diminished social cohesion and reduced ability to regulate individual behaviour (Kingston et al., 2009). In areas experiencing economic hardship, residents often have limited resources and lower levels of community participation, which can create an environment conducive to crime (Benjumea Mejia et al., 2024).

This theoretical perspective is particularly relevant to the Indonesian context, where substantial disparities in socioeconomic development exist across provinces. Differences in employment opportunities, income levels, educational attainment, and quality of public services may contribute to variations in crime rates among regions (Sugiharti et al., 2022; Trisnawati & Khoirunurrofik, 2019). Therefore, Social Disorganisation Theory provides a useful framework for understanding how regional socioeconomic conditions shape crime patterns and can inform the formulation of targeted public policies.

Cluster Theory

Cluster theory emphasises that geographical areas or economic entities tend to form groups based on similarities in structural characteristics, economic activities, and patterns of interaction (McCann, 2019). Clusters emerge through the concentration of economic resources, shared development characteristics, and common institutional environments.

Within the context of regional analysis, the cluster approach facilitates the identification of groups of regions with similar socioeconomic profiles and development challenges (Majumder et al., 2022; Scaramuzzino et al., 2019). Regions sharing comparable characteristics often exhibit similar behavioural patterns and policy needs. Consequently, cluster analysis can serve as an effective tool for supporting evidence-based policymaking by enabling governments to design interventions that are tailored to the specific characteristics of each group.

2.3 | Empirical Studies on Socioeconomic Factors and Crime

Numerous empirical studies have investigated the relationship between socioeconomic factors and crime in Indonesia. Most of these studies have employed econometric approaches to examine the influence of individual socioeconomic indicators on crime rates. Existing studies consistently demonstrate that socioeconomic conditions play a substantial role in shaping regional crime patterns. However, most prior research has focused primarily on estimating causal relationships among variables using regression-based approaches.

Sugiharti et al. (2022) analysed the relationship between economic indicators and crime rates in five Indonesian provinces with the highest crime rates—DKI Jakarta, North Sumatra, East Java, South Sulawesi, and West Java—during the period 2010–2019 using panel data analysis. Their findings revealed that unemployment rates, income inequality, and population density positively and significantly affect crime rates. These results suggest that increasing economic inequality and limited employment opportunities contribute to higher levels of criminal activity.

Similarly, Amelia & Ridwan (2024) employed a Random Effects Model (REM) to examine the effects of socioeconomic factors on crime rates across one municipality and nine districts in Bengkulu Province from 2018 to 2022. Their findings indicated that education levels, unemployment, and income inequality significantly influence crime rates, whereas poverty does not exhibit a significant effect. Higher levels of education were associated with lower crime rates, while increases in unemployment and income inequality corresponded with higher levels of criminal activity.

Using a Fixed Effects Model (FEM), Trisnawati & Khoirunurrofik (2019) investigated the relationship between socioeconomic factors and crime rates across 32 Indonesian provinces. The study found that educational attainment has a negative and significant effect on crime, whereas unemployment has a positive and significant effect. Economic growth and poverty, however, did not significantly influence crime rates. These findings indicate that improvements in educational outcomes and labour market conditions are essential components of crime prevention strategies.

2.4 | K-Means Clustering in Crime and Regional Analysis

K-Means clustering is an unsupervised machine learning method that partitions observations into homogeneous groups based on similarities among multiple variables. The method aims to minimise within-cluster variation while maximising differences between clusters, thereby facilitating the identification of underlying patterns in complex datasets.

Hapsari & Widodo (2017) applied the K-Means algorithm to classify crime vulnerability across 34 provinces in Indonesia using police-reported crime data. Their findings demonstrated that Indonesian provinces could be grouped into distinct clusters based on similarities in crime characteristics. The study highlighted the effectiveness of K-Means clustering in identifying spatial patterns of crime and supporting regional crime mapping. Nevertheless, the analysis did not incorporate socioeconomic variables that may contribute to variations in crime rates.

Uzcategui-Salazar & Lillo (2022) utilised K-Means clustering to group regions according to a range of socioeconomic indicators. Their findings showed that the method effectively identifies patterns of similarity among regions and provides useful information for policy planning and resource allocation. However, the study did not include crime indicators as part of the clustering process.

These studies demonstrate the utility of K-Means clustering for regional analysis. However, previous applications have generally focused either on crime indicators or on socioeconomic variables separately, limiting their ability to capture the multidimensional nature of crime.

2.5 | Research Gap and Novelty

Despite extensive research on the relationship between socioeconomic factors and crime in Indonesia, several gaps remain in the existing literature. First, most previous studies have relied on regression-based approaches to examine the influence of individual socioeconomic variables on crime rates, with limited attention to identifying latent patterns among regions. Second, prior studies employing K-Means clustering have typically focused exclusively on either crime indicators or socioeconomic characteristics, without integrating both dimensions within a unified analytical framework. As a result, the heterogeneity of provincial conditions and the interaction between crime and socioeconomic factors remain insufficiently explored. Third, existing studies often concentrate on specific provinces or regions, limiting the generalisability of their findings at the national level.

This study addresses these limitations by applying the K-Means clustering method to classify all provinces in Indonesia based on a combination of crime indicators and socioeconomic variables. Specifically, the analysis integrates narcotics-related crimes, theft, and physical violence with key socioeconomic indicators, including unemployment rates, Gross Regional Domestic Product (GRDP), and the Human Development Index (HDI).

The novelty of this study lies in three main contributions. First, it adopts a multidimensional approach by simultaneously incorporating crime and socioeconomic indicators into the clustering process. Second, it applies an unsupervised machine learning technique to identify hidden patterns and provincial typologies that may not be captured through conventional regression models. Third, it generates evidence-based cluster classifications that can support the formulation of more targeted, equitable, and context-specific public policies for crime prevention and regional development in Indonesia.

3 | METHODS

This study employed a quantitative descriptive research design using a cluster analysis approach based on the K-Means algorithm. The quantitative approach facilitates the systematic and objective collection, processing, and analysis of numerical data to explain and describe patterns related to specific phenomena (Taherdoost, 2022). In the context of this study, the K-Means clustering method was utilised to classify Indonesian provinces according to similarities in crime rates and socioeconomic conditions. This approach enables the identification of homogeneous

groups of provinces that share comparable characteristics, thereby providing a foundation for evidence-based public policy formulation.

3.1 | Data Mining

Data mining refers to the systematic process of extracting meaningful information, patterns, and hidden relationships from large datasets. Rather than merely summarising existing data, data mining aims to generate new knowledge that may not be readily observable through conventional analytical techniques. Within the framework of Knowledge Discovery in Databases (KDD), data mining encompasses a series of interrelated stages, including data collection, data preprocessing, pattern identification, and knowledge interpretation (Shu & Ye, 2023).

The application of data mining techniques in regional socioeconomic studies offers substantial advantages in identifying complex relationships among variables and revealing underlying patterns that can support policy development. In this study, data mining was employed to identify provincial groupings based on similarities in crime and socioeconomic indicators.

3.2 | Clustering

Clustering is an unsupervised learning technique that partitions a set of observations into several groups based on the degree of similarity among their attributes. Observations with similar characteristics are assigned to the same cluster, whereas observations with distinct characteristics are placed in different clusters. Unlike supervised classification methods, clustering does not require predefined labels or categories, allowing natural structures within the data to emerge (Ali et al., 2019).

In regional studies, clustering analysis is particularly useful for identifying groups of regions with similar socioeconomic and crime characteristics. The resulting clusters can facilitate comparative analysis and provide insights into regional disparities that may require differentiated policy interventions.

3.3 | K-Means Algorithm

K-Means is one of the most widely used partition-based clustering algorithms within the unsupervised learning paradigm. The algorithm aims to divide observations into a predetermined number of clusters by minimising the within-cluster sum of squares while maximising the separation between clusters (Chaudhry et al., 2023).

The K-Means algorithm operates iteratively by assigning each observation to the nearest cluster centroid based on a distance measure, typically Euclidean distance. After all observations are assigned, the centroid of each cluster is recalculated according to the mean values of the variables within the cluster. This process continues until the cluster assignments stabilise and no further significant changes occur.

In this study, K-Means clustering was selected due to its computational efficiency, simplicity of implementation, and ability to identify homogeneous provincial groups based on multidimensional socioeconomic and crime indicators.

3.4 | Data Sources

This study utilised secondary data obtained from the Indonesian Central Bureau of Statistics (Badan Pusat Statistik/BPS). The population consisted of all 34 provinces in Indonesia. A total sampling technique was employed, whereby all provinces with complete data for the selected variables were included in the analysis.

The dataset comprised three crime indicators, namely narcotics-related crimes, theft, and physical violence, along with three socioeconomic indicators: the Open Unemployment Rate (OUR), Gross Regional Domestic Product (GRDP), and the Human Development Index (HDI). These variables were selected based on theoretical considerations and empirical evidence indicating their relevance to regional crime patterns.

Crime indicators represent the prevalence of criminal activities within each province, while socioeconomic indicators capture dimensions of economic performance, labour market conditions, and human development. The

integration of these variables enables a comprehensive analysis of the relationship between crime and socioeconomic conditions across Indonesian provinces.

3.5 | Data Analysis Stages

The analytical process followed several sequential stages. First, all variables were standardised using z-score transformation to eliminate differences in measurement scales and prevent variables with larger numerical values from exerting disproportionate influence during the clustering process.

Second, the optimal number of clusters was determined using cluster validation techniques, including the Elbow Method and the Silhouette Method. The evaluation results indicated that a two-cluster solution (K = 2) provided the most appropriate balance between cluster compactness and separation. Third, the K-Means algorithm was applied to classify provinces into two groups based on similarities in crime and socioeconomic indicators. The clustering process was conducted iteratively until the cluster centroids converged and stable group assignments were obtained.

Finally, the resulting clusters were interpreted by examining the centroid values and characteristic profiles of each group. Provinces assigned to Cluster 1 were characterised by relatively higher crime rates and were generally associated with densely populated regions and major economic centres. In contrast, provinces included in Cluster 0 exhibited lower crime rates and were typically characterised by lower population density and less intensive socioeconomic activity compared with provinces in Cluster 1.

3.6 | Dataset

The dataset used in this study consisted of provincial-level observations that integrated crime and socioeconomic indicators for all 34 provinces in Indonesia. Each observation represented a province and included six variables derived from average values over the selected observation period.

The three crime variables comprised narcotics-related crimes, theft, and physical violence, while the socioeconomic variables included the Open Unemployment Rate (OUR), Gross Regional Domestic Product (GRDP), and the Human Development Index (HDI). These variables collectively provided a multidimensional representation of provincial characteristics and served as the basis for the K-Means clustering analysis. Table 1 presents the structure of the dataset used in this study, including the variables, measurement units, and data sources.

Table 1. Variables and data sources used in the K-means clustering analysis.

Province	Narcotics	Theft	Physical Violence	OUR (%)	GRDP	HDI
DKI Jakarta	5,252	22582	7855	6.53	3443026	83.55
North Sumatra	5,308	22377	10381	5.89	1050995	75.13
West Java	2,411	14443	6581	7.44	2625226	74.24
South Sulawesi	2,583	9678	3767	4.33	652648	74.60
East Java	5,036	10562	6792	4.88	2953547	74.65
South Sumatra	1,991	6614	3767	4.11	629169	73.18
Lampung	1,446	5284	2222	4.23	448851	72.48
Riau	1,596	5320	2041	4.23	1026472	74.95
Central Java	1,995	3213	1504	5.13	1695622	73.39
DI Yogyakarta	631	1163	765	3.69	180696	81.09
Aceh	1,476	2878	1651	6.03	227018	74.70
North Sulawesi	201	1772	10381	6.10	171969	75.04
East Nusa Tenggara	45	2526	5242	3.14	128523	68.40
West Sumatra	1,274	2600	2490	5.94	312769	75.64
Banten	775	1869	844	7.52	814122	75.77
Central Sulawesi	547	2465	2025	2.95	347139	71.66

Jambi	739	1854	1275	4.53	293780	73.73
Papua	216	5557	2657	2.67	81731.7	63.01
Bali	829	1299	833	2.69	274358	78.01
Bengkulu	416	1552	1042	3.42	96583.1	74.30
Riau Islands	366	950	917	6.80	331645	79.08
West Kalimantan	826	1687	600	5.05	274469	70.47
South Kalimantan	1535	1406	773	4.31	269192	74.66
West Nusa Tenggara	708	2095	1019	2.80	166395	72.37
East Kalimantan	645	1382	1024	5.31	843571	78.20
West Papua	114	2420	1565	5.38	61576.4	67.47
Southeast Sulawesi	382	740	415	3.15	176180	72.94
Central Kalimantan	645	964	415	4.10	208783	73.73
Gorontalo	109	355	1453	3.06	51366.7	71.25
Maluku	151	973	1931	6.31	58487.8	72.75
West Sulawesi	265	401	447	2.27	58572.1	69.80
Bangka	413	509	277	4.56	102527	74.09
North Maluku	137	243	1030	4.31	85143.1	70.98
North Kalimantan	294	343	336	4.01	147279	72.88

Source: Indonesian Central Bureau of Statistics (Badan Pusat Statistik/BPS) (2025).

4 | RESULTS AND DISCUSSION

4.1 | Cluster Distribution of Indonesian Provinces

The K-Means clustering analysis classified the 34 provinces in Indonesia into two distinct groups based on similarities in crime and socioeconomic indicators (Table 2). The optimal number of clusters ($K = 2$) was determined using the Elbow and Silhouette methods, which indicated that a two-cluster solution provided the best balance between cluster compactness and cluster separation. This classification reveals the existence of substantial heterogeneity in provincial characteristics and highlights the need for differentiated policy approaches to address regional disparities.

Table 2. Distribution of Indonesian provinces by K-Means cluster.

Province	Narcotics	Theft	Physical Violence	OUR (%)	GRDP	HDI	Cluster
DKI Jakarta	5,252	22582	7855	6.53	3443026	83.55	1
North Sumatra	5,308	22377	10381	5.89	1050995	75.13	1
West Java	2,411	14443	6581	7.44	2625226	74.24	1
South Sulawesi	2,583	9678	3767	4.33	652648	74.60	0
East Java	5,036	10562	6792	4.88	2953547	74.65	1
South Sumatra	1,991	6614	3767	4.11	629169	73.18	0
Lampung	1,446	5284	2222	4.23	448851	72.48	0
Riau	1,596	5320	2041	4.23	1026472	74.95	0
Central Java	1,995	3213	1504	5.13	1695622	73.39	0
DI Yogyakarta	631	1163	765	3.69	180696	81.09	0
Aceh	1,476	2878	1651	6.03	227018	74.70	0
North Sulawesi	201	1772	10381	6.10	171969	75.04	0
East Nusa Tenggara	45	2526	5242	3.14	128523	68.40	0
West Sumatra	1,274	2600	2490	5.94	312769	75.64	0
Banten	775	1869	844	7.52	814122	75.77	0
Central Sulawesi	547	2465	2025	2.95	347139	71.66	0
Jambi	739	1854	1275	4.53	293780	73.73	0
Papua	216	5557	2657	2.67	81731.7	63.01	0
Bali	829	1299	833	2.69	274358	78.01	0
Bengkulu	416	1552	1042	3.42	96583.1	74.30	0

Riau Islands	366	950	917	6.80	331645	79.08	0
West Kalimantan	826	1687	600	5.05	274469	70.47	0
South Kalimantan	1535	1406	773	4.31	269192	74.66	0
West Nusa Tenggara	708	2095	1019	2.80	166395	72.37	0
East Kalimantan	645	1382	1024	5.31	843571	78.20	0
West Papua	114	2420	1565	5.38	61576.4	67.47	0
Southeast Sulawesi	382	740	415	3.15	176180	72.94	0
Central Kalimantan	645	964	415	4.10	208783	73.73	0
Gorontalo	109	355	1453	3.06	51366.7	71.25	0
Maluku	151	973	1931	6.31	58487.8	72.75	0
West Sulawesi	265	401	447	2.27	58572.1	69.80	0
Bangka	413	509	277	4.56	102527	74.09	0
North Maluku	137	243	1030	4.31	85143.1	70.98	0
North Kalimantan	294	343	336	4.01	147279	72.88	0

Source: Indonesian Central Bureau of Statistics (Badan Pusat Statistik/BPS) (2025). *Note: The identified clusters represent groups of provinces with similar crime patterns and socioeconomic characteristics and provide a basis for differentiated public policy interventions.*

The clustering results show that Cluster 1 consists of four provinces, namely DKI Jakarta, North Sumatra, West Java, and East Java. These provinces are characterised by high levels of economic activity, large populations, rapid urbanisation, and intensive population mobility. In contrast, the remaining 30 provinces are grouped into Cluster 0, which generally comprises regions with lower levels of urbanisation, smaller populations, and less concentrated economic activity.

The distribution of provinces across the two clusters indicates that crime patterns in Indonesia are not randomly distributed but tend to concentrate in economically strategic regions. This finding suggests that provinces functioning as national economic centres face more complex social challenges than other regions (Sparrow et al., 2020; Yeh & Chen, 2020). Therefore, understanding the characteristics of each cluster is essential for developing targeted interventions that reflect the specific needs and vulnerabilities of different provincial groups.

Fig. 1 illustrates the geographical distribution of the two clusters across Indonesia. The map shows that provinces included in Cluster 1 are concentrated in highly urbanised areas, whereas provinces in Cluster 0 are more geographically dispersed. This spatial pattern further emphasises the relationship between urbanisation, economic concentration, and crime vulnerability.

Fig. 1. Spatial Distribution of Crime and Socioeconomic Clusters Across Indonesian Provinces. *Note: The figure illustrates the provincial distribution of the two clusters identified through K-Means clustering based on combined crime and socioeconomic indicators.*

4.2 | Crime Characteristics Across Clusters

The comparison of crime indicators between clusters reveals significant differences in the incidence of narcotics-related crimes, theft, and physical violence. Table 3 presents the average values of crime indicators for each cluster, demonstrating that Cluster 1 consistently records substantially higher crime rates than Cluster 0 across all crime categories.

Table 3. Average crime indicators by cluster across Indonesian provinces.

Crime Rate		
Indicators	**Cluster 0**	**Cluster 1**
Narcotics	778.33	4,501.75
Theft	2,337.07	17,491
Physical Violence	1,823.6	7,902.25

Note: Values represent the mean number of narcotics-related crimes, theft cases, and physical violence incidents for each cluster.

The average number of narcotics-related offences in Cluster 1 reaches 4,501.75 cases, which is considerably higher than the average of 778.33 cases observed in Cluster 0. Similarly, the average number of theft cases in Cluster 1 is 17,491 cases, while Cluster 0 records an average of 2,337.07 cases. Physical violence also exhibits a pronounced disparity, with Cluster 1 reporting an average of 7,902.25 cases compared with 1,823.60 cases in Cluster 0.

These findings indicate that provinces characterised by high levels of economic activity and urbanisation tend to experience greater exposure to crime. Higher population density and increased social interaction create more opportunities for criminal behaviour and may weaken informal social control mechanisms. Moreover, large urban centres often attract migrants seeking economic opportunities, leading to increased competition for resources and employment, which can contribute to social tensions and criminal activity. The substantial differences in crime rates between the two clusters demonstrate that crime is a multidimensional phenomenon influenced by broader structural conditions rather than solely by individual factors. Consequently, crime prevention strategies should address the socioeconomic contexts in which criminal activities occur.

4.3 | Socioeconomic Characteristics Across Clusters

Table 4 summarises the average socioeconomic indicators for each cluster, including the Open Unemployment Rate (OUR), Gross Regional Domestic Product (GRDP), and the Human Development Index (HDI). The results indicate clear disparities between the two clusters, reflecting differences in economic performance, labour market conditions, and human development.

Table 4. Average socioeconomic indicators by cluster across Indonesian provinces.

Socio-Economic Rate		
Indicators	**Cluster 0**	**Cluster 1**
OUR	4.4	6.18
GRDP	340,555	2,581,999
HDI	73.35	76.89

Note: OUR refers to the Open Unemployment Rate (%), GRDP refers to Gross Regional Domestic Product, and HDI refers to the Human Development Index.

Provinces in Cluster 1 generally exhibit higher GRDP values than those in Cluster 0, indicating that these provinces serve as major economic centres within Indonesia. In addition, Cluster 1 records relatively higher HDI values, reflecting better access to education, healthcare, and public services. Despite these favourable conditions, Cluster 1 also demonstrates higher unemployment rates than Cluster 0.

The coexistence of high GRDP, high HDI, and high unemployment rates suggests that economic growth and improvements in human development do not necessarily translate into equitable employment opportunities (Hung et al., 2026). Rapid economic expansion in metropolitan areas may attract large numbers of migrants, increasing labour market competition and creating disparities in income distribution (Heider et al., 2020; Rodríguez-Pose & Storper, 2020). Consequently, some individuals may experience economic insecurity despite residing in economically advanced regions. These findings highlight the complexity of the relationship between socioeconomic development and crime. While higher levels of economic development are generally associated with improved living standards, they may also generate social pressures and inequalities that increase vulnerability to criminal behaviour.

4.4 | Relationship between Crime and Socioeconomic Factors

The clustering results suggest that unemployment is one of the most important socioeconomic factors associated with crime patterns across Indonesian provinces. Provinces characterised by higher unemployment rates tend to exhibit higher incidences of theft and other economically motivated crimes. This finding supports previous studies that identify unemployment as a significant predictor of criminal activity.

From a theoretical perspective, the results are consistent with strain theory, which argues that limited access to legitimate economic opportunities can generate frustration and social pressure, thereby increasing the likelihood of criminal behaviour (Kingston et al., 2009). Individuals who experience economic hardship and face barriers to achieving socially accepted goals may resort to alternative means, including illegal activities, to satisfy their needs (Nassar & Elsayed, 2018). The relationship between GRDP and crime appears more complex (Purnomo et al., 2023; Widyastaman & Hartono, 2022). Although provinces in Cluster 1 demonstrate stronger economic performance, they also record significantly higher crime rates. This finding suggests that economic growth alone is insufficient to reduce crime. The benefits of economic growth may not be evenly distributed, resulting in persistent inequalities and social exclusion within economically advanced regions.

Similarly, the positive association between HDI and crime indicates that improvements in education, health, and living standards do not automatically guarantee lower crime rates (Anser et al., 2020; de Almeida et al., 2021; Zaman et al., 2019). In densely populated urban areas, the effects of social complexity, population mobility, and inequality may outweigh the protective effects of human development. Consequently, crime should be understood as a multidimensional phenomenon shaped by the interaction between socioeconomic conditions and urbanisation processes. The findings also support Social Disorganisation Theory, which emphasises that rapid urbanisation, residential mobility, and weakened social cohesion can reduce informal social control and increase opportunities for criminal behaviour (Kubrin, 2009). Therefore, effective crime prevention policies should consider both socioeconomic development and community-level social dynamics.

4.5 Policy Implications

The identification of two distinct provincial clusters highlights the importance of adopting differentiated and place-based approaches to crime prevention. The substantial variation in crime and socioeconomic characteristics across provinces suggests that uniform national policies may be insufficient to address the diverse challenges faced by different regions.

For provinces included in Cluster 1, policy interventions should prioritise the reduction of socioeconomic pressures associated with rapid urbanisation and high population density. In addition to strengthening law enforcement and public security measures, governments should implement programmes aimed at expanding employment opportunities, improving workforce skills, and reducing income inequality. Comprehensive strategies for narcotics prevention, rehabilitation, and community-based crime prevention are also essential in these regions. For provinces in Cluster 0, policy priorities should focus on maintaining social cohesion and preventing future increases in crime rates. Investments in education, healthcare, local economic development, and employment generation can strengthen social resilience and reduce long-term vulnerability to criminal activity. Enhancing

institutional capacity and improving access to public services are also critical for sustaining positive development outcomes.

The findings suggest that effective crime prevention requires an integrated approach that addresses both the immediate manifestations of crime and the underlying socioeconomic factors that contribute to criminal behaviour. Tailoring policies to the specific characteristics of each cluster can improve the efficiency and effectiveness of public interventions.

4.6 Limitations and Future Research

Several limitations should be considered when interpreting the findings of this study. First, the analysis relies exclusively on secondary data obtained from the Indonesian Central Bureau of Statistics (BPS). Although these data provide comprehensive coverage of all provinces, they may not fully capture the complexity of crime and its underlying determinants.

Second, this study incorporates a limited number of socioeconomic variables, namely unemployment, GRDP, and HDI. Other factors that may influence crime patterns, such as income inequality, population density, law enforcement capacity, social capital, and migration dynamics, were not included due to data limitations. Consequently, the resulting clusters may not reflect the full range of factors associated with regional crime variations. Third, the use of a two-cluster solution, although supported by the Elbow and Silhouette methods, may oversimplify the diversity of provincial characteristics across Indonesia. Future studies could explore alternative clustering solutions or compare different clustering algorithms, such as hierarchical clustering, DBSCAN, or Gaussian mixture models, to examine the robustness of the findings.

This study does not explicitly consider spatial relationships among provinces. Given that crime patterns may be influenced by geographical proximity and interregional interactions, future research should incorporate spatial analytical techniques to investigate the geographic dimensions of crime and socioeconomic disparities in Indonesia.

5 | CONCLUSION

This study successfully grouped 34 provinces in Indonesia into two clusters based on crime characteristics and socioeconomic conditions using the K-Means Clustering method. Cluster 1 consists of metropolitan provinces with high crime rates such as DKI Jakarta, West Java, and East Java, while Cluster 0 includes provinces with lower crime rates such as Banten, Jambi, and West Papua. The main findings show that high GRDP and HDI do not automatically correlate with low crime rates. In fact, provinces with high economic development tend to experience higher crime rates due to the complexity of urbanization, population density, and social pressures. The open unemployment rate (OUR) has been proven to be the most consistent factor influencing crime, especially economically motivated crimes, in line with strain theory and social disorganization.

This study makes an important contribution by simultaneously integrating socioeconomic and crime variables, resulting in a more comprehensive mapping than previous studies. The clustering results emphasize the importance of differentiated place-based policy approaches: Cluster 1 requires multidimensional interventions including job creation programs, drug rehabilitation, and strengthening Crime Prevention Through Environmental Design (CPTED), while Cluster 0 requires prevention strategies that focus on maintaining social stability and local economic development.

Although providing a strong empirical foundation, this study has limitations related to its reliance on secondary data, the selection of K=2 which potentially simplifies regional heterogeneity, and the absence of spatial analysis. Future research should integrate spatial analysis to capture geographic autocorrelation, explore a greater number of clusters, and include contextual variables such as law enforcement effectiveness and social cohesion. Thus, this study contributes to the development of science while offering practical recommendations for the

formulation of more targeted, evidence-based, and contextual public policies in an effort to realize a safer and more prosperous Indonesia.

Acknowledgments

The authors would like to acknowledge Statistics Indonesia (Badan Pusat Statistik/BPS) for providing the secondary data used in this study.

Disclosure Statement

The authors declare that there are no conflicts of interest regarding the publication of this article.

Data Availability Statement

The data supporting the findings of this study are publicly available from Statistics Indonesia (Badan Pusat Statistik/BPS) through its official publications and databases.

References

Ali, M., Alqahtani, A., Jones, M. W., & Xie, X. (2019). Clustering and Classification for Time Series Data in Visual Analytics: A Survey. *IEEE Access*, *7*, 181314–181338. https://doi.org/10.1109/ACCESS.2019.2958551

Amelia, D., & Ridwan, P. (2024). Analysis of Education Level, Unemployment, Poverty and Income Inequality on Crime Rate in Bengkulu Province. *Jurnal Pendidikan Indonesia*, *5*(12), 693–706. https://doi.org/10.59141/japendi.v5i12.6319

Anozi, D. T., & Novianda, B. (2023). Socio-Economic and Property Crime Rate in Indonesia. *Economics Development Analysis Journal*, *12*(3), 305–318. https://doi.org/10.15294/edaj.v12i3.68829

Anser, M. K., Yousaf, Z., Nassani, A. A., Alotaibi, S. M., Kabbani, A., & Zaman, K. (2020). Dynamic linkages between poverty, inequality, crime, and social expenditures in a panel of 16 countries: two-step GMM estimates. *Journal of Economic Structures*, *9*(1), 43. https://doi.org/10.1186/s40008-020-00220-6

Benjumea Mejia, D. M., Chilton, J., & Rutherford, P. (2024). Collective urban green revitalisation: Crime control an sustainable behaviours in lower-income neighbourhoods. *World Development*, *177*, 106534. https://doi.org/10.1016/j.worlddev.2024.106534

Cao, X., Cui, Z., Ali, S., & Nazar, R. (2026). Health Crisis or Crime Surge? Assessing the Asymmetric Nexus Between Pandemic Uncertainty and Drug-Related Crimes. *Crime & Delinquency*, *72*(3), 791–828. https://doi.org/10.1177/00111287241248089

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, *15*(9), 1679. https://doi.org/10.3390/sym15091679

Coccia, M. (2018). Economic inequality can generate unhappiness that leads to violent crime in society. *International Journal of Happiness and Development*, *4*(1), 1. https://doi.org/10.1504/IJHD.2018.090488

Contreras, C., & Hipp, J. R. (2020). Drugs, Crime, Space, and Time: A Spatiotemporal Examination of Drug Activity and Crime Rates. *Justice Quarterly*, *37*(2), 187–209. https://doi.org/10.1080/07418825.2018.1515318

Dasic, B., Devic, Z., Denic, N., Zlatkovic, D., Ilic, I. D., Cao, Y., Jermsittiparsert, K., & Le, H. Van. (2020). Human development index in a context of human development: Review on the western Balkans countries. *Brain and Behavior*, *10*(9). https://doi.org/10.1002/brb3.1755

de Almeida, R. D. C., Ehrl, P., & Moreira, T. B. S. (2021). Social and Economic Convergence Across Brazilian States Between 1990 and 2010. *Social Indicators Research*, *157*(1), 225–246. https://doi.org/10.1007/s11205-021-02659-x

DeMarco, L. M., Dwyer, R. E., & Haynie, D. L. (2021). The accumulation of disadvantage: Criminal justice contact, credit, and debt in the transition to adulthood *. *Criminology*, *59*(3), 545–580. https://doi.org/10.1111/1745-9125.12286

Hapsari, D. P. T., & Widodo, E. (2017). Grouping of crime-prone areas in Indonesia using K-Means Clustering analysis. *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika Dan Nilai-Nilai Islami)*, *1*(1), 147–153.

Heider, B., Stroms, P., Koch, J., & Siedentop, S. (2020). Where do immigrants move in Germany? The role of international migration in regional disparities in population development. *Population, Space and Place*, *26*(8). https://doi.org/10.1002/psp.2363

Hung, H., Chiu, S., Yang, C., & Chiu, Y. (2026). Comparison of Economic and Human Development Efficiency in the Asia-Pacific Region: A Correlation Analysis in the Context of Sustainable Development Goals. *Sustainable Development*.

https://doi.org/10.1002/sd.71019
Kidd, S. (2017). Social exclusion and access to social protection schemes. *Journal of Development Effectiveness*, *9*(2), 212–244. https://doi.org/10.1080/19439342.2017.1305982
Kingston, B., Huizinga, D., & Elliott, D. S. (2009). A Test of Social Disorganization Theory in High-Risk Urban Neighborhoods. *Youth & Society*, *41*(1), 53–79. https://doi.org/10.1177/0044118X09338343
Koku, P. S. (2015). Financial exclusion of the poor: a literature review. *International Journal of Bank Marketing*, *33*(5), 654–668. https://doi.org/10.1108/IJBM-09-2014-0134
Kubrin, C. E. (2009). *Social Disorganization Theory: Then, Now, and in the Future* (pp. 225–236). https://doi.org/10.1007/978-1-4419-0245-0_12
Majumder, S., Kayal, P., Chowdhury, I. R., & Das, S. (2022). Regional disparities and development in India: evidence from Wroclow Taxonomy and K-means clustering. *GeoJournal*, *88*(3), 3249–3282. https://doi.org/10.1007/s10708-022-10805-2
McCann, P. (2019). Theories of agglomeration and regional economic growth: a historical review. In *Handbook of Regional Growth and Development Theories* (pp. 6–23). Edward Elgar Publishing. https://doi.org/10.4337/9781788970020.00007
Nassar, D. M., & Elsayed, H. G. (2018). From Informal Settlements to sustainable communities. *Alexandria Engineering Journal*, *57*(4), 2367–2376. https://doi.org/10.1016/j.aej.2017.09.004
Purnomo, S. D., Supriyo, D. A., Rusito, R., Anindito, T., Hariadi, W., & Jati, D. (2023). How Economic Indicator Drive Crime? Empirical Study in Developing Country, Indonesia. *International Journal of Economics and Financial Issues*, *13*(3), 94–99. https://doi.org/10.32479/ijefi.14309
Rao, N. D., & Min, J. (2018). Decent Living Standards: Material Prerequisites for Human Wellbeing. *Social Indicators Research*, *138*(1), 225–244. https://doi.org/10.1007/s11205-017-1650-0
Reiman, J., & Leighton, P. (2020). *The rich get richer and the poor get prison: Thinking critically about class and criminal justice*. Routledge.
Rodríguez-Pose, A., & Storper, M. (2020). Housing, urban growth and inequalities: The limits to deregulation and upzoning in reducing economic and spatial inequality. *Urban Studies*, *57*(2), 223–248. https://doi.org/10.1177/0042098019859458
Ross, C. E. (2017). *Social causes of psychological distress*. Routledge.
Santos, J. M. E., Moutinho, V. F., & Leitão, J. (2024). *Relationship between Economic Growth, Development and Both Violent and Property Crime: Evidence from 34 Countries*. https://doi.org/10.2139/ssrn.5006227
Scaramuzzino, C., Garegnani, G., & Zambelli, P. (2019). Integrated approach for the identification of spatial patterns related to renewable energy potential in European territories. *Renewable and Sustainable Energy Reviews*, *101*, 1–13. https://doi.org/10.1016/j.rser.2018.10.024
Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, *110*, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817
Sloane, D. C., & Choi, H. (2016). *Crime and Community Well-Being: The Role of Social Capital and Collective Efficacy in Increasing Safety* (pp. 87–99). https://doi.org/10.1007/978-3-319-29942-6_6
Sparrow, R., Dartanto, T., & Hartwig, R. (2020). Indonesia Under the New Normal: Challenges and the Way Ahead. *Bulletin of Indonesian Economic Studies*, *56*(3), 269–299. https://doi.org/10.1080/00074918.2020.1854079
Sugiharti, L., Esquivias, M. A., Shaari, M. S., Agustin, L., & Rohmawati, H. (2022). Criminality and Income Inequality in Indonesia. *Social Sciences*, *11*(3), 142. https://doi.org/10.3390/socsci11030142
Sugiharti, L., Purwono, R., Esquivias, M. A., & Rohmawati, H. (2023). The Nexus between Crime Rates, Poverty, and Income Inequality: A Case Study of Indonesia. *Economies*, *11*(2), 62. https://doi.org/10.3390/economies11020062
Taherdoost, H. (2022). What are Different Research Approaches? Comprehensive Review of Qualitative, Quantitative, and Mixed Method Research, Their Applications, Types, and Limitations. *Journal of Management Science & Engineering Research*, *5*(1), 53–63. https://doi.org/10.30564/jmser.v5i1.4538
Trisnawati, D., & Khoirunurrofik, K. (2019). Inter-Provincial Spatial Linkages of Crime Pattern in Indonesia: Looking at Education and Economic Inequality Effects on Crime. *Indonesian Journal of Geography*, *51*(2), 106. https://doi.org/10.22146/ijg.34026
Tumalavičius, V., Veikša, I., Načisčionis, J., Zahars, V., & Draskovic, V. (2017). Issues of the state and society security (Part I): Ensuring public security in the fight against crime. *Journal of Security and Sustainability Issues*, *6*(3). https://doi.org/10.9770/jssi.2017.6.3(7)
Uzcategui-Salazar, M., & Lillo, J. (2022). Assessment of social vulnerability to groundwater pollution using K-means cluster analysis. *Environmental Science and Pollution Research*, *30*(6), 14975–14992. https://doi.org/10.1007/s11356-022-22810-6
Vadera, A. K., Pratt, M. G., & Mishra, P. (2013). Constructive Deviance in Organizations. *Journal of Management*, *39*(5), 1221–1276. https://doi.org/10.1177/0149206313475816
Van Laar, C., Meeussen, L., Veldman, J., Van Grootel, S., Sterk, N., & Jacobs, C. (2019). Coping With Stigma in the Workplace:

Understanding the Role of Threat Regulation, Supportive Factors, and Potential Hidden Costs. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01879

Widyastaman, P. A., & Hartono, D. (2022). Economic Inequality Decomposition and Spatial Pattern of Crime in Indonesia. *Papers in Applied Geography*, *8*(3), 268–281. https://doi.org/10.1080/23754931.2021.1991842

Wong, P.-H. (2021). *Economic Deprivation: Approaches, Causes, and Consequences for Violent Conflicts* (pp. 187–197). https://doi.org/10.1007/978-3-319-95960-3_70

Yeh, A. G.-O., & Chen, Z. (2020). From cities to super mega city regions in China in a new wave of urbanisation and economic transition: Issues and challenges. *Urban Studies*, *57*(3), 636–654. https://doi.org/10.1177/0042098019879566

Zaman, K., Usman, B., Sheikh, S. M., Iswan, Khan, A., Kosnin, A. B. M., Rosman, A. S. Bin, Ismail, S., Ali, D. F., & Hishan, S. S. (2019). Managing crime through quality education: A model of justice. *Science & Justice*, *59*(6), 597–605. https://doi.org/10.1016/j.scijus.2019.08.004