Khazanah Journal of Economics Vol (1) No (1) 60-71
DOI: https://doi.org/10.xxxxx/Economics.v1i1.2
https://journal.azkia.id/index.php/Economics
e-ISSN: xxxx-xxxx

The Relationship between Socioeconomic Factors and Crime at the Provincial Level in Indonesia: A Linear Regression Approach

Herdin Kristianjani Zebua[1*], Muhamad Fikri Zaelani[1], Rifqi Lutdan Padl[1], Irsyad Nur Hidayatulloh[1], Mahesa Adlan Falah[1]
[1]Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia
[*]Corresponding Author Email: Kristianjani@gmail.com

Received: March 09, 2026; Revised: March 29, 2026, Accepted: April 5, 2026

ABSTRACT

Background: Crime is a complex social phenomenon that is closely associated with socio-economic conditions. Understanding the relationship between economic factors and crime is important for developing evidence-based policies, particularly in the context of regional disparities in Indonesia.

Purpose: This study aims to examine the relationship between socio-economic factors and crime rates at the provincial level in Indonesia and to assess the extent to which these factors can explain variations in crime.

Methods: This study uses secondary data at the provincial level in Indonesia. The analysis was conducted using descriptive statistics, correlation analysis, data visualization, and multiple linear regression. The variables analyzed include crime rate, population, unemployment rate, poverty level, and education level.

Results: The results indicate a significant positive relationship between crime rates, population, and unemployment rate, which is consistent with the economic perspective on crime. In contrast, poverty and education show weak or inconsistent relationships with crime, suggesting a more complex interaction. The multiple linear regression model yields very limited explanatory power, with an R-squared (R^2) value of 0.03, indicating that the selected socio-economic variables explain only a small share of the variation in crime rates.

Conclusions: The findings suggest that crime cannot be sufficiently explained by the selected socio-economic indicators alone. Other factors, such as social cohesion, inequality, urban social conditions, and the effectiveness of law enforcement, may play a substantial role in shaping crime patterns. Therefore, a more comprehensive analytical approach and richer data are needed to support more effective crime prevention policies.

Research Contribution: This study contributes to the literature by showing the limited explanatory capacity of macro-level socio-economic variables in explaining crime rates in Indonesia and by highlighting the importance of incorporating broader social and institutional factors in future research.

Keywords: crime rate, socioeconomic factors, linear regression, correlation analysis, unemployment rate, prediction, indonesia.

INTRODUCTION

Worldwide, crime rates are viewed as a crucial social indicator, reflecting the health, stability, and success of a nation's development. Crime is not simply a series of individual acts, but rather a complex manifestation of the dynamic interaction between social structures, economic pressures, and the community environment. In

criminology, this relationship has long been at the center of theoretical and empirical debate. Classical theories such as Strain Theory explain that the stress of not being able to achieve life goals through legal means—often related to economic constraints—can drive individuals to engage in criminal activity (Pratt & Cullen, 2005). Meanwhile, Social Disorganization Theory emphasizes the importance of community factors, such as social cohesion and informal supervision, in suppressing deviant behavior (Sampson et al., 1997).

Understanding the multifaceted relationship between economic factors, social conditions, and crime rates is crucial for policymakers and law enforcement officials. An accurate understanding can form the basis for formulating more effective, efficient, and evidence-based crime prevention strategies. In the Indonesian context, the urgency of this study is even greater given Indonesia's archipelagic nature, with thousands of islands and significant socioeconomic diversity across provinces. Several national-level studies have attempted to measure the influence of economic variables, such as poverty and unemployment, on crime rates. However, the results still show significant variation depending on the study location, regional characteristics, and methodology used (Risma & Dariah, 2024; Soraya et al., 2024; Yuzani et al., 2024). This situation suggests that the relationship between socioeconomic factors and crime cannot be understood simply and requires more critical, contextual, and detailed analysis.

Against this background, this article aims to fill this gap by presenting a comprehensive analysis using interprovincial data in Indonesia. This study not only measures and visualizes the correlation between crime rates and several socioeconomic factors, such as population, education level, poverty, and unemployment, but also constructs a linear regression model to test the predictive power of these variables on crime rates. Furthermore, this study critically evaluates the performance of the resulting model, highlighting its limitations and the possibility of important variables not yet measured in the available data. By transparently presenting the findings and their limitations, this article hopes to stimulate a more in-depth discussion on the determinants of crime in Indonesia and provide a stronger basis for developing more holistic crime prevention policy recommendations.

This research uses a quantitative methodology based on secondary data, with the entire analysis process conducted using the standardized and reliable Python ecosystem. The dataset used is a CSV file named Crime Rate vs. Socioeconomic Factors.csv and is cross-sectional, representing data from various provinces in Indonesia. The initial stage of the research began with data ingestion, which involved loading the dataset into a DataFrame structure using the Pandas library. Subsequently, column names were standardized from Indonesian to English to improve clarity, consistency, and ease of analysis. For example, the variable "Crime Rate" was changed to "Crime Rate," "Education Level" to "Education Level," and "Province" to "State." In the next stage, data integrity was checked to ensure there were no missing values that could potentially affect the validity of the statistical analysis results. The results indicated that the dataset used in this study was complete.

To obtain an initial overview of the data characteristics, this study applied descriptive statistical analysis to each numeric variable. This analysis was performed using built-in Pandas functions, such as .describe(), to generate statistical summaries containing the mean, standard deviation, minimum, quartiles, and maximum. This step is crucial for understanding the general pattern of data distribution and detecting possible outliers or anomalies that could impact the

results of further analysis, particularly correlation analysis and regression modeling. Next, the linear relationship between the numeric variables was measured using the Pearson correlation coefficient, with a value ranging from -1, indicating a perfect negative correlation, to +1, indicating a perfect positive correlation. This calculation was performed using the .corr() method in a Pandas DataFrame. The results were then visualized in heatmap form using the Seaborn library to facilitate quick and intuitive identification of the direction and strength of the relationship between variables.

In addition to statistical analysis, this study also conducted in-depth visual analysis to explore data patterns that may not be fully captured by summary statistics. Scatter plots were used to visualize the relationship between the Crime Rate and each predictor variable, allowing for clearer observation of the general trend, data distribution, and the possibility of outliers. In addition, a bar chart is used to compare the average crime rate between provinces, providing a geographic overview of the distribution of crime rates in Indonesia. This visual approach complements statistical analysis to strengthen the interpretation of the relationship patterns between variables.

To test the simultaneous influence of independent variables on crime rates, this study constructed a multiple linear regression model with the crime rate as the dependent variable. Generally, the model is formulated as the equation $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$, where Y represents the crime rate, β_0 is the intercept or model constant, β_1 through β_4 are regression coefficients indicating the partial effect of each independent variable, and ϵ is the error term or residual. In this model, the independent variables used include Religious Population, Education Level, Poverty Rate, and Unemployment Rate. The model training process was carried out using the LinearRegression class from the scikit-learn library. To test the model's generalization ability, the data was divided into 80% training data and 20% testing data using the train_test_split function, so that the evaluation was carried out on data that was not used during training.

Model performance is evaluated using two main metrics, namely Mean Squared Error (MSE) and R-squared (R^2). MSE is used to measure the average squared error between predicted and actual values, where a lower MSE value indicates better model performance. Meanwhile, R^2 is used to measure the proportion of total variation in the dependent variable that can be explained by the regression model, with a value range between 0 and 1. The higher the R^2 value, the greater the model's ability to explain data variation. Through this combination of stages, the study seeks to provide an empirical picture of the relationship between socioeconomic factors and crime rates, while also assessing the extent to which the linear regression model is able to adequately explain the phenomenon.

RELATED STUDIES

The study of the relationship between socioeconomic factors and crime has been a central topic in empirical criminology. In general, previous literature can be grouped into several main theoretical streams. Structural theories emphasize that crime is a consequence of macro-socioeconomic conditions, which create social pressures. In a comprehensive meta-analysis, Pratt and Cullen Pratt & Cullen (2005) confirmed that macro-level predictors have a significant and consistent relationship with crime rates. However, there are also perspectives that focus more on the community level. One of the most significant contributions comes from the work of Sampson, Raudenbush, and Earls (1997), who introduced the concept of

"collective efficacy." They argue that social cohesion at the neighborhood level and residents' willingness to act for the common good have a greater power to reduce crime than simply socio-demographic or economic characteristics. The following is a literature review of factors relevant to the dataset used in this analysis, in the context of research in Indonesia:

1. Unemployment Rate: Consistent with global literature, numerous studies in Indonesia have found a positive relationship between unemployment and crime. Research conducted in Jakarta and West Java, as well as at the national level, indicates that rising unemployment correlates with increased crime (Risma & Dariah, 2024; Yuzani et al., 2024). This is reinforced by analysis in the Special Region of Yogyakarta, which identified unemployment as a significant factor driving crime (Anggaresa, 2023).
2. Poverty Rate: The relationship between poverty and crime in Indonesia has yielded mixed results. Some studies identify poverty as a driver of crime, particularly for property crimes on the island of Java [7]. However, other studies have found that the effect of poverty may be insignificant or even exhibit a complex relationship, depending on the type of crime or the regional context analyzed (Wicaksono & Suharto, 2023; Yuzani et al., 2024).
3. Education Level: The relationship between education level and crime is often counterintuitive. Although education is expected to reduce crime, several studies in Indonesia have found weak or even positive correlations, suggesting that education level may be a proxy for other variables, such as urbanization or greater social complexity in a region (Telaumbanua et al., 2024).
4. Demographic/Religious Factors (Religious Population): In general, the criminology literature also examines the role of religious institutions and levels of religiosity as a means of social control that promotes moral values. However, the Religious Level variable in this dataset serves more as a demographic indicator of population size, rather than as a measure of per capita religiosity or frequency of participation in religious activities (Leite et al., 2023). Therefore, this variable in this analysis tends to reflect population density in a region.

Based on this literature review, there is a need to conduct an analysis that integrates the macroeconomic factors available in the dataset. This study aims to fill this gap by using provincial-level data and explicitly addressing the limitations of models in capturing complex factors at the community level, such as “collective efficacy” (Sampson et al., 1997), which are not captured in this dataset.

RESULTS AND DISCUSSION

This section presents the results of all stages of the analysis, from identifying relationships between variables to critical evaluation of the predictive model built.

Correlation Analysis and Visualization

Correlation analysis provides an initial overview of the linear relationship between variables. A correlation heatmap visualizes the correlation coefficient between each pair of variables, with warmer colors (red) indicating a strong positive correlation and cooler colors (blue) indicating a negative correlation.

1. Crime Rate vs. Population (Religious Population): There is a strong positive correlation (r=0.67). This indicates that provinces with higher populations tend to have higher crime rates [1]. It should be noted that the Religious

Population variable here serves as a demographic indicator of population size.

2. Crime Rate vs. Unemployment Rate: There is a moderate positive correlation (r=0.36), indicating that higher unemployment rates lead to higher crime rates, although the relationship is not as strong as with population [3], [5].
3. Crime Rate vs. Poverty Rate: A very weak negative correlation (r=−0.14). This relationship is barely significant and challenges common assumptions.
4. Crime Rate vs. Education Level: The weak positive correlation (r=0.29) is an anomalous result, implying that the relationship between the two is indirect or that other variables are involved.

The scatterplots below visualize the direct relationship between Crime Rate and each socioeconomic factor.

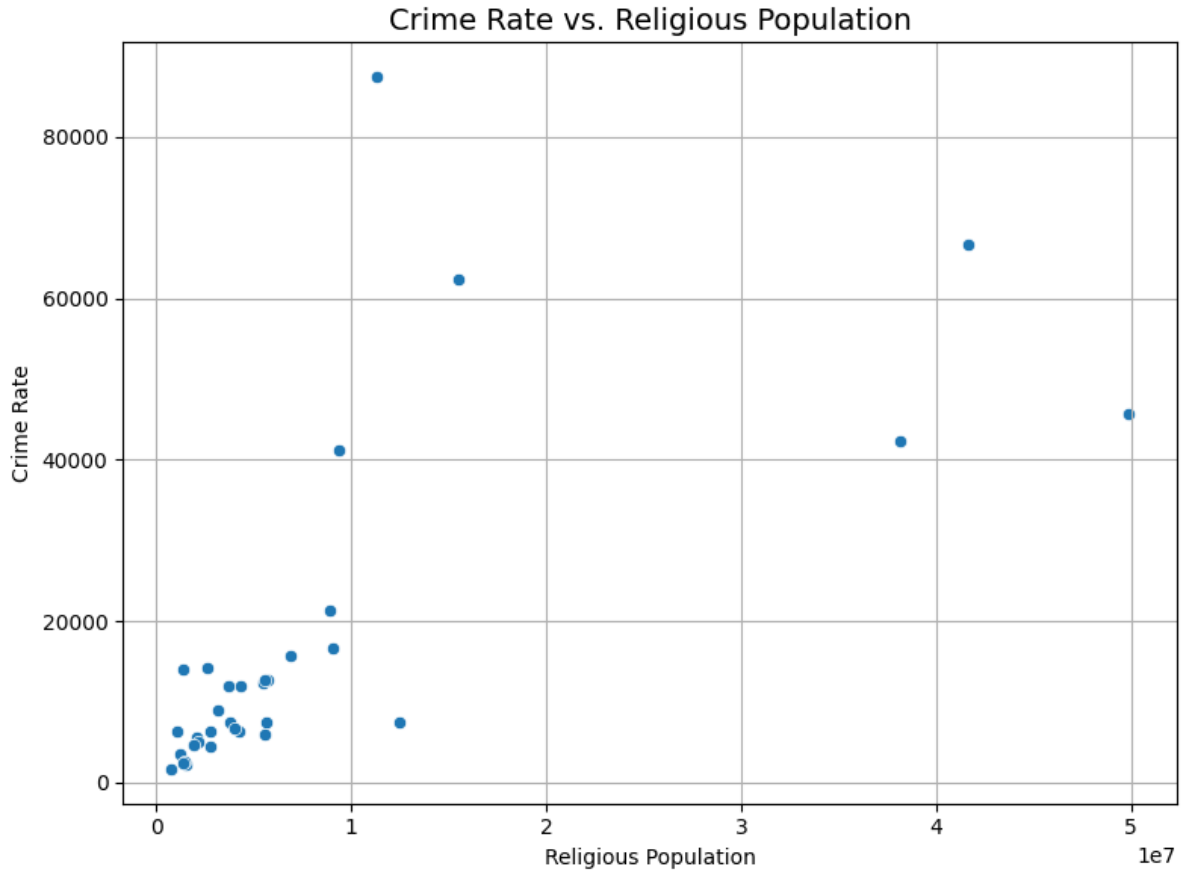

Figure 1. Scatterplot of Crime Rate vs. Religious Population

Figure 1 show a clear positive trend. As population increases, crime rates also tend to increase. The data points in the upper right corner indicate provinces with very large populations but also have very high crime rates, serving as interesting outliers.

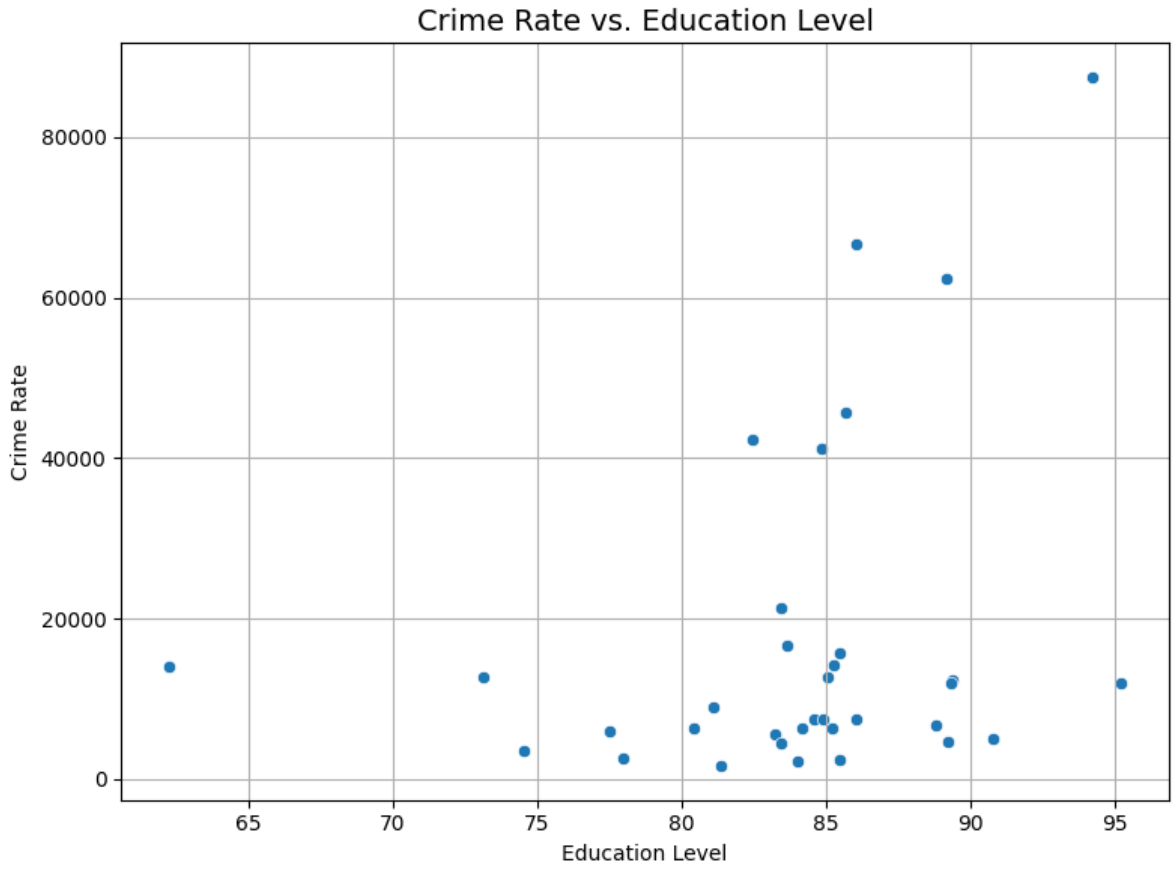

Figure 2. Scatterplot of Crime Rate and Education Level

Figure 2 show distribution of data points in this plot appears very random, although there is a slight positive trend. This visually supports the finding of a weak correlation and confirms the hypothesis that education level is not a major predictor of crime rates.

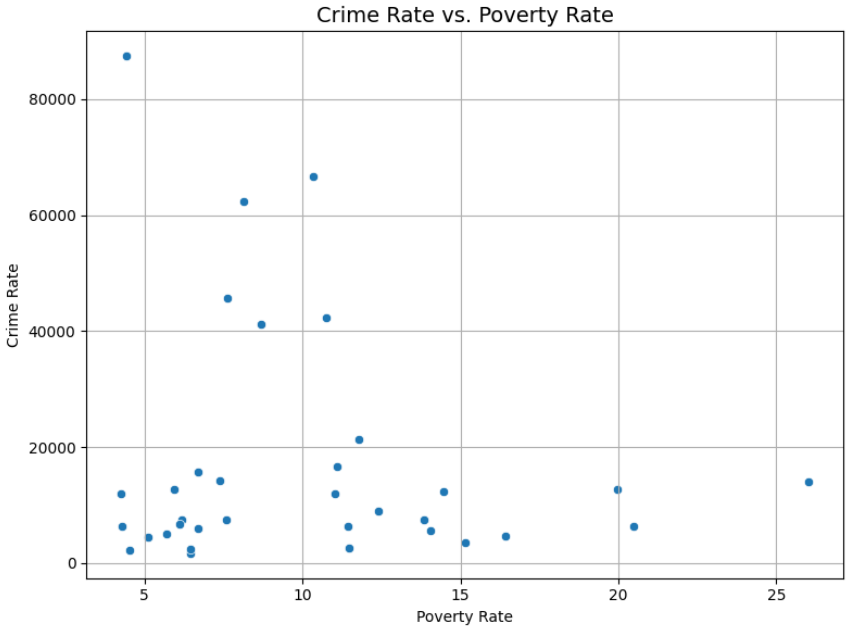

Figure 3. Scatterplot of Crime Rate and Poverty Rate

Figure 3 show a highly variable distribution of data with no clear pattern, visually proving that the linear relationship between poverty and crime rates in this dataset is very weak or non-existent.

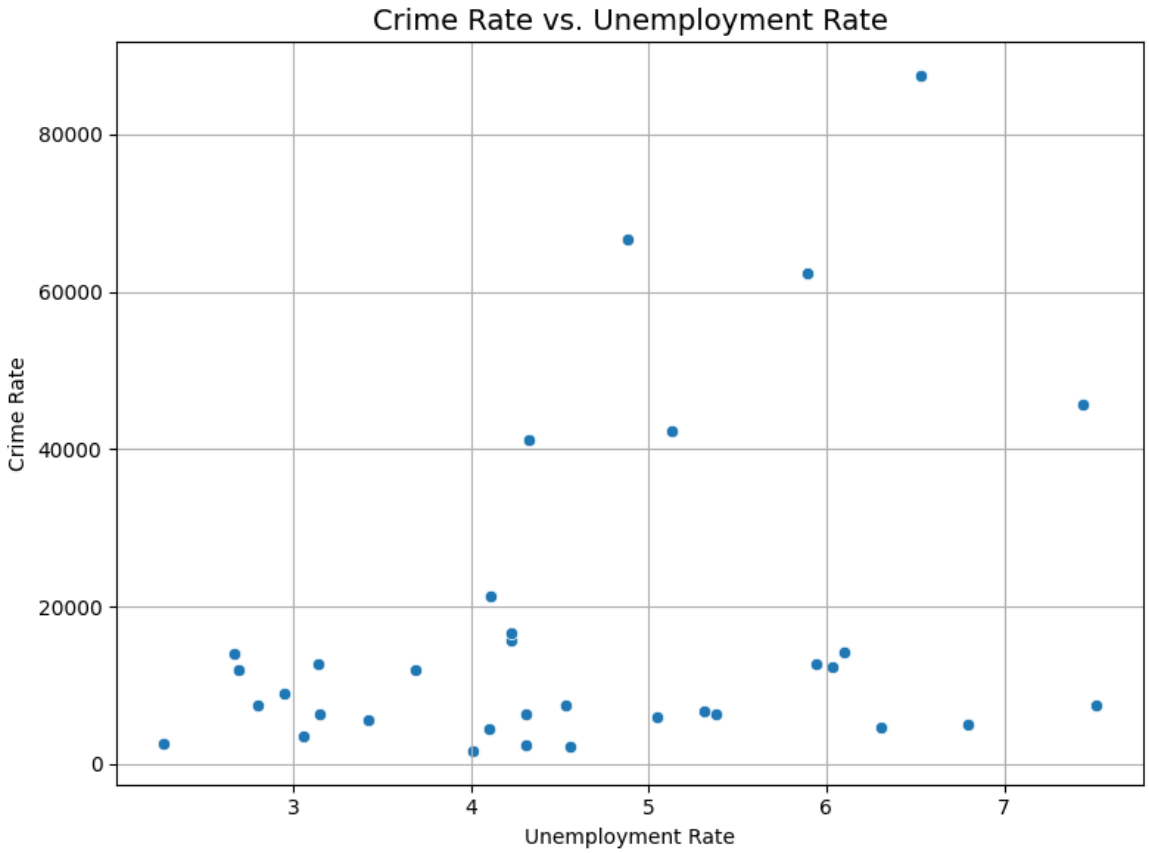

Figure 4. Scatterplot of Crime Rate and Unemployment Rate

Figure 4 shows a clearer positive trend than the poverty and education plots. As unemployment increases, there is a tendency for crime rates to increase, which is consistent with the economic distress theory.

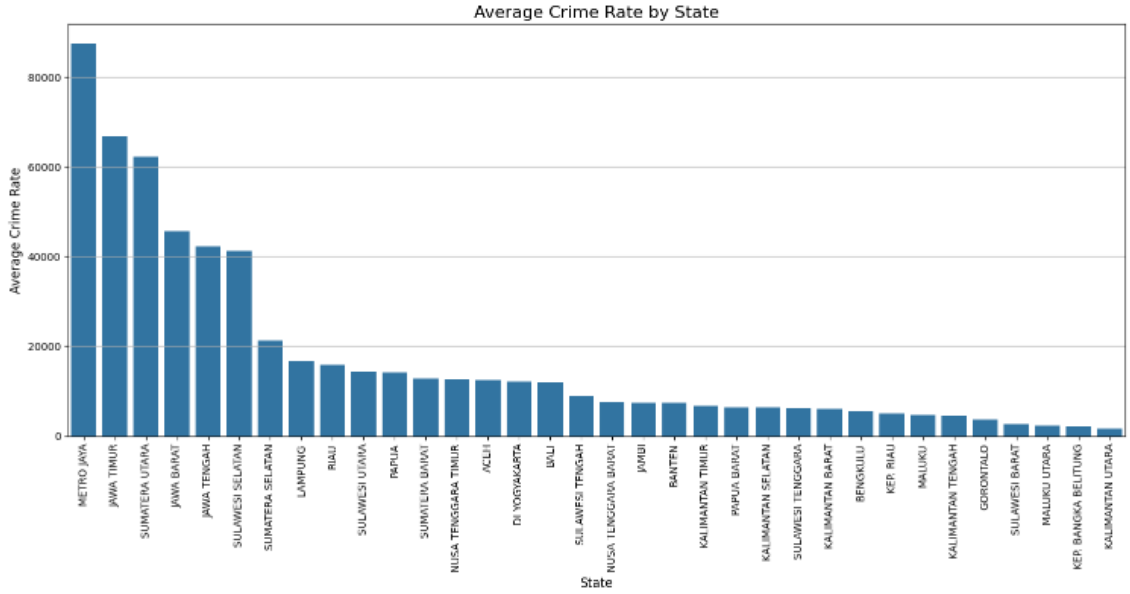

Figure 5. Bar Chart of Average Crime Rate per Province

Figure 5 show bar chart clearly shows significant geographic variation in crime rates. The provinces with the highest average crime rates are NORTH SUMATRA, DKI JAKARTA, EAST JAVA, and WEST JAVA, while the lowest are NORTH KALIMANTAN, GORONTALO, and WEST PAPUA.

Linear Regression Model Results

A multiple linear regression model was constructed to measure the independent contribution of each variable to the crime rate. The results of the model coefficients are presented in Table 1.

Table 1. Linear Regression Model Coefficients

Factor	Coefficient	Interpretation with Ceteris Paribus Assumptions
Religious Population	0.001076	Every one-unit increase in population is predicted to increase the crime rate by approximately 0.001 cases, holding other factors constant.
Education Level	604.21	Every 1% increase in education level predicts an increase in the crime rate of approximately 604 cases, holding other factors constant.
Poverty Rate	150.05	Every 1% increase in poverty level predicts an increase in the crime rate of approximately 150 cases, holding other factors constant.
Unemployment Rate	1537.44	Every 1% increase in unemployment rate predicts a significant increase in the crime rate of approximately 1,537 cases, holding other factors constant.

The much larger coefficient of Unemployment Rate (1537.44) compared to other variables clearly identifies unemployment as the strongest predictor in this model.

Critical Evaluation

Although the model was successfully trained, its evaluation metrics provide the most critical insight into its validity:

1. R-squared (R2) Score: 0.03
2. Mean Squared Error (MSE): 148.3 Million

The very low R2 value (0.03) is the most crucial finding. This means that the linear regression model built is only able to explain about 3% of the total variation in the crime rate between provinces. Practically speaking, this model has very weak predictive power, and the high MSE value confirms that the model's predictions are far from the actual values.

Why does this model fail to explain most of the variation? This is due to several fundamental limitations:

1. Underspecified Model: This model likely ignores key variables that are actually the main drivers of crime, a phenomenon known as omitted variable

bias. Variables such as police budget, number of law enforcement personnel, urbanization rate, and the economic inequality index (Gini ratio) are not included in the dataset, yet play a significant role in various studies [6]

2. Non-Linear Relationships: The relationships between the variables may be non-linear or involve complex interactions, which cannot be captured by a simple linear regression model.
3. Community-Level Contextual Factors: This model uses only aggregate data at the provincial level. Modern criminological theory emphasizes the importance of community- or neighborhood-level factors, such as "collective efficacy" (mutual trust and willingness to act together for the common good) [2]. These unmeasured social variables are often more influential in reducing crime than macroeconomic data alone.

Discussion

The research results indicate that the relationship between socioeconomic factors and crime rates at the provincial level in Indonesia is not uniform. Correlation findings indicate that the population variable has a fairly strong positive relationship with the crime rate (r = 0.67), while the unemployment rate shows a moderate positive relationship (r = 0.36). Conversely, the poverty rate shows a very weak negative correlation (r = -0.14), while the education level shows a weak positive correlation (r = 0.29). This pattern confirms that crime cannot be explained solely through economic dimensions, but rather results from a more complex interaction between demographic, economic, and social factors.

The relatively strong positive relationship between population and crime rates indicates that regions with larger populations tend to experience higher crime rates (Sugiharti et al., 2023). This finding can be understood from the perspective of the concentration of social and economic activities in denser areas, which in turn increases the opportunities for social interaction, conflict, and lawlessness. In the context of large provinces such as DKI Jakarta, West Java, East Java, and North Sumatra, high crime rates can also be attributed to greater levels of urbanization, population mobility, and social heterogeneity compared to regions with smaller populations (Silver, 2024). Thus, population not only serves as a demographic indicator but also reflects the complexity of the social environment that can influence crime rates.

Meanwhile, the positive relationship between unemployment and crime rates supports the theoretical argument that economic pressure can drive increased criminal behavior (Aryal, 2024). When access to formal employment is limited, some individuals may face greater economic pressure, thus increasing the risk of engaging in illegal activities. This finding aligns with the economics of crime perspective, which views crime as a response to limited legitimate economic opportunities. However, the strength of this relationship is moderate, so unemployment cannot be positioned as the sole primary determinant. This suggests that economic pressure is indeed relevant, but its influence still interacts with other factors such as regional characteristics, social control, and the effectiveness of local institutions (Shui et al., 2025).

Contrary to common expectations, the poverty variable in this study does not show a strong linear relationship with crime rates (Bourne et al., 2025a). The very weak negative correlation indicates that poverty, in this dataset, is not a direct explanation of inter-regional variations in crime. This finding is important because it confirms that the simple assumption that "the poorer an area, the higher its crime rate" does not always hold empirically (Kefayat & Thill, 2025; Zandiatashbar &

Laurito, 2023). Most likely, the effect of poverty on crime is indirect and mediated by other factors, such as income inequality, limited access to public services, urbanization, or weak institutional capacity. Therefore, poverty needs to be understood as one element within a broader social structure, not as a single variable that automatically increases crime rates.

The relationship between education level and crime also shows a non-simple pattern. The weak positive correlation may be considered counterintuitive, as education is often theoretically associated with increased economic opportunities and a reduced tendency for deviant behavior (Bourne et al., 2025b). This result suggests that the effect of education on crime is likely indirect. Higher education in a province does not automatically reduce crime if other factors, such as population density, economic opportunities, or urban dynamics, predominately influence social conditions (Yasin et al., 2024). Therefore, the education variable in this study seems more appropriately understood as part of the social structure interacting with other variables, rather than as a single, deterministic predictor.

The results of multiple linear regression further emphasize this complexity. Although the model shows that the unemployment rate has the largest coefficient compared to other variables, the most significant finding lies in the very low coefficient of determination, $R^2 = 0.03$. This means the model is only able to explain approximately 3% of the variation in crime rates between regions. The very large Mean Squared Error (MSE) value also indicates that the model's predictions have a high deviation from the actual values (Hodson, 2022). Methodologically, this indicates that the socioeconomic variables used in the model are insufficient to adequately capture crime dynamics. In other words, the regression model has very limited explanatory and predictive power.

This model's weaknesses can be explained by several factors. First, the model likely suffers from underspecification because it does not include other important, theoretically relevant variables, such as urbanization rate, economic inequality, law enforcement capacity, security budgets, and residential density. Second, the relationship between socioeconomic factors and crime may not be linear, and therefore cannot be fully explained by simple multiple linear regression. Third, the use of aggregate data at the provincial level makes the model less able to capture contextual factors at the community or neighborhood level, even though modern criminological theory emphasizes the importance of social cohesion, informal supervision, and collective efficacy in shaping crime rates. Thus, the model's poor performance does not merely indicate analytical failure, but rather reveals that crime is a multidimensional phenomenon that requires a more comprehensive analytical approach (Wikström & Kroneberg, 2022).

The results of this study show that socioeconomic factors remain related to crime rates, particularly through population and unemployment variables. However, this relationship is limited and not strong enough to fully explain variations in crime. This finding emphasizes the importance of caution in formulating public policy. Crime prevention efforts cannot focus solely on macroeconomic indicators but need to be combined with interventions that consider the social context, regional governance, institutional quality, and community empowerment. Therefore, effective crime prevention strategies must be multidimensional, data-driven, and sensitive to the local characteristics of each region.

CONCLUSION

This study shows a relationship between several socioeconomic factors and crime rates at the provincial level in Indonesia, but the strength of this relationship varies. Population and unemployment rates have a clearer positive relationship with crime rates, while poverty and education levels show a weaker and inconsistent relationship. These findings suggest that crime cannot be understood simply through macroeconomic indicators alone, as more complex social dynamics underlie variations in crime rates across regions.

On the other hand, the results of multiple linear regression modeling show very limited predictive power, as reflected in the R^2 value of 0.03 and the high MSE value. This indicates that a significant portion of the variation in crime rates cannot be explained by the population, education, poverty, and unemployment variables used in this study. Therefore, crime is likely influenced by other factors not captured in the dataset, such as income inequality, urbanization rate, the effectiveness of law enforcement, the quality of institutions, and social conditions at the community level.

This study concludes that a more holistic approach is needed to understand and address crime in Indonesia. Future studies should consider using more diverse variables, more detailed data coverage, and the potential application of non-linear analytical models or spatial approaches to better explain crime dynamics. From a policy perspective, crime prevention efforts should not solely focus on improving economic conditions, but should also focus on strengthening institutions, enhancing the quality of the social environment, and developing community-based interventions tailored to the characteristics of each region.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to all parties who have contributed to the completion of this study.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors were involved in the development and design of the study. The processes of material preparation, data collection, data analysis, and manuscript drafting were carried out jointly by all authors. Each author reviewed and approved the final version of the manuscript.

FUNDING

The authors received no financial support for the research, authorship, and/or publication of this article.

DATA AVAILABILITY

Data sharing is not applicable to this article because no new data were created or analyzed in this study.

BIBLIOGRAPHY

Anggaresa, F. P. (2023). Unraveling Crime Dynamics in Indonesia: Exploring the Impact of Population Density, Poverty, Average Years of Schooling (RLS), and

Open Unemployment Rate (TPT) in Java in 2020. *WELFARE: Jurnal Ilmu Kesejahteraan Sosial*, *12*(1).

Aryal, G. R. (2024). Economic Inequality and Crime: A Socio-criminological Perspective. *Journal of Development and Social Engineering*, *10*(1), 76–85.

Bourne, P. A., Solan, I., & Thorpe, F. (2025a). Macroeconomic Deprivation and Violent Crime: An OLS and ARIMAX Analysis of Homicide, Poverty, and Unemployment in Jamaica, 1989-2023. *International Journal of Business Management Insight & Transformations*, *9*(1), 1–22.

Bourne, P. A., Solan, I., & Thorpe, F. (2025b). Macroeconomic Deprivation and Violent Crime: An OLS and ARIMAX Analysis of Homicide, Poverty, and Unemployment in Jamaica, 1989-2023. *International Journal of Business Management Insight & Transformations*, *9*(1), 1–22.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1–10.

Kefayat, E., & Thill, J.-C. (2025). Urban street network configuration and property crime: an empirical multivariate case study. *ISPRS International Journal of Geo-Information*, *14*(5), 200.

Shui, X., Zhang, M., Wang, Y., & Smart, P. (2025). Do climate change regulatory pressures increase corporate environmental sustainability performance? The moderating roles of foreign market exposure and industry carbon intensity. *British Journal of Management*, *36*(1), 223–239.

Silver, C. (2024). Rapid urbanization: the challenges and opportunities for planning in Indonesian cities. *The Indonesian Economy and the Surrounding Regions in the 21st Century: Essays in Honor of Iwan Jaya Azis*, 35–48.

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.

Wikström, P.-O. H., & Kroneberg, C. (2022). Analytic criminology: Mechanisms and methods in the explanation of crime and its causes. *Annual Review of Criminology*, *5*(1), 179–203.

Yasin, M. A., Hasan, M., Ali, F., & Ahmad, S. (2024). Understanding crime through socioeconomic lenses: Education and economic development. *Jahan-e-Tahqeeq*, *7*(3), 494–506.

Zandiatashbar, A., & Laurito, A. (2023). An empirical analysis of the link between built environment and safety in Chicago's transit station areas. *Journal of the American Planning Association*, *89*(2), 225–239.