

ORIGINAL

Space-Time Autoregressive Integrated Moving Average (STARIMA) Modeling for Predicting Criminal Cases of Motor Vehicle Theft in Surabaya, Indonesia

Modelado de media móvil integrada autorregresiva espacio-temporal (STARIMA) para predecir casos penales de robo de vehículos motorizados en Surabaya, Indonesia

Arip Ramadan¹ , Dwi Rantini^{2,3}  , Yohanes Manasye Triangga², Ratih Ardiati Ningrum^{2,3} , Fazidah Othman⁴ 

¹Information System Study Program, School of Industrial and System Engineering, Telkom University Surabaya Campus. Surabaya, 60231, Indonesia.

²Data Science Technology Study Program, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga. Surabaya, 60115, Indonesia.

³Research Group of Data-Driven Decision Support System, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga. Surabaya, 60115, Indonesia.

⁴Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya. Malaysia.

Cite as: Ramadan A, Rantini D, Manasye Triangga Y, Ningrum RA, Othman F. Space-Time Autoregressive Integrated Moving Average (STARIMA) Modeling for Predicting Criminal Cases of Motor Vehicle Theft in Surabaya, Indonesia. Data and Metadata. 2024; 3:.621. <https://doi.org/10.56294/dm2024.621>

Submitted: 01-02-2024

Revised: 12-7-2024

Accepted: 27-12-2024

Published: 28-12-2024

Editor: Adrián Alejandro Vitón Castillo 

Corresponding Author: Dwi Rantini 

ABSTRACT

Introduction: motor vehicle theft poses significant challenges in urban areas, particularly in large metropolitan cities like Surabaya, Indonesia's second-largest city. Surabaya's strategic economic role makes it a hotspot for criminal activities, including motor vehicle theft, driven by various socio-economic factors.

Method: this study utilizes the Space-Time Autoregressive Integrated Moving Average (STARIMA) model to predict motor vehicle theft cases across five sub-regions in Surabaya, covering the period from January 2019 to December 2023. The STARIMA model, which incorporates both temporal and spatial dependencies, offers a more robust framework for crime prediction compared to traditional models like ARIMA. The results show that STARIMA effectively captures the spatio-temporal dynamics of crime, providing valuable insights for law enforcement to develop targeted strategies that enhance public safety.

Results: the model's performance was evaluated using the Root Mean Square Error (RMSE), indicating its suitability for accurate and actionable crime forecasting in Surabaya. Based on the RMSE value, the best model obtained is STARIMA (1,1,2) with a Uniform Location weighting matrix.

Conclusions: this STARIMA (1,1,2) model, it is used to predict motor vehicle theft incidents in West Surabaya, Central Surabaya, South Surabaya, East Surabaya, and North Surabaya. The forecast value carried out is for a period of five months into the future. Case predictions for the next five months show fluctuations in each region of Surabaya, with the highest regions in succession being North Surabaya, East Surabaya, and South Surabaya.

Keywords: Crime; STARIMA; Space-Time Analysis; Spatial; Temporal.

RESUMEN

Introducción: el robo de vehículos de motor plantea importantes problemas en las zonas urbanas, en particular en las grandes ciudades metropolitanas como Surabaya, la segunda ciudad más grande de Indonesia. El papel económico estratégico de Surabaya la convierte en un foco de actividades delictivas, incluido el robo de vehículos de motor, impulsadas por diversos factores socioeconómicos.

Método: este estudio utiliza el modelo de media móvil integrada autorregresiva espacio-temporal (STARIMA) para predecir los casos de robo de vehículos de motor en cinco subregiones de Surabaya, que abarcan el período de enero de 2019 a diciembre de 2023. El modelo STARIMA, que incorpora dependencias tanto temporales como espaciales, ofrece un marco más sólido para la predicción del delito en comparación con los modelos tradicionales como ARIMA. Los resultados muestran que STARIMA captura de manera eficaz la dinámica espacio-temporal del delito, lo que proporciona información valiosa para que las fuerzas del orden desarrollen estrategias específicas que mejoren la seguridad pública.

Resultados: el rendimiento del modelo se evaluó utilizando el error cuadrático medio (RMSE), lo que indica su idoneidad para realizar pronósticos precisos y prácticos de delitos en Surabaya. Según el valor RMSE, el mejor modelo obtenido es STARIMA (1,1,2) con una matriz de ponderación de ubicación uniforme.

Conclusiones: este modelo STARIMA (1,1,2) se utiliza para predecir los incidentes de robo de vehículos de motor en Surabaya occidental, Surabaya central, Surabaya meridional, Surabaya oriental y Surabaya norte. El valor de pronóstico realizado corresponde a un período de cinco meses en el futuro. Las predicciones de casos para los próximos cinco meses muestran fluctuaciones en cada región de Surabaya, siendo las regiones con mayor incidencia en sucesión Surabaya Norte, Surabaya Este y Surabaya Sur.

Palabras clave: Crimen; STARIMA; Análisis Espacio-Temporal; Espacial; Temporal.

INTRODUCTION

Surabaya, Indonesia's second-largest city, is a major metropolitan hub with extensive infrastructure, making it a crucial economic center. However, this also makes it a prime target for various crimes, particularly motor vehicle theft. The city's large population and economic significance attract many people to live and work there, but the imbalance between job availability and population growth has led to an increase in criminal activities as some residents resort to crime to meet their needs.⁽¹⁾ In 2022, Surabaya experienced a 28,51 % increase in criminal cases, with motor vehicle theft being one of the most prevalent, resulting in significant financial losses and heightened public insecurity.⁽²⁾

Motor vehicle theft is a pervasive issue in urban areas, posing significant challenges for law enforcement agencies tasked with ensuring public safety. Traditional time series models like ARIMA (Autoregressive Integrated Moving Average) have been widely used for crime forecasting; however, they primarily focus on temporal dependencies and often neglect the spatial relationships between different regions.⁽³⁾ Urban crime exhibits complex spatial and temporal patterns, often concentrated in specific areas due to factors such as economic conditions, population density, and social interactions.^(4,5)

To overcome these limitations, the Space-Time Autoregressive Integrated Moving Average (STARIMA) model was developed, which extends the ARIMA model by incorporating spatial dependencies.⁽⁶⁾ The STARIMA model is particularly well-suited for predicting crimes that exhibit spatial clustering, such as motor vehicle theft, where incidents in one area may influence the likelihood of similar crimes occurring in neighboring regions.^(3,7,8,9) Recent studies have demonstrated the STARIMA model's superior performance over traditional models in capturing the spatio-temporal dynamics of urban crime.⁽⁷⁾

This study aims to apply the STARIMA model to predict motor vehicle theft cases across various regions in Surabaya for the upcoming months. By integrating spatial weight matrices that account for the influence of neighboring areas, the STARIMA model provides a more accurate and comprehensive approach to crime forecasting, making it a valuable tool for law enforcement agencies in cities like Surabaya. The predictions generated by the model are expected to offer critical insights for developing targeted crime prevention strategies, thereby enhancing public safety and optimizing resource allocation.

METHOD

Data Set

In this study, primary data were obtained from the Surabaya Police Station, covering motor vehicle theft cases across five sub-regions—Central Surabaya, East Surabaya, West Surabaya, North Surabaya, and South Surabaya—from January 2019 to December 2023. The data used in this study are numerical data. Then, to obtain research results, several statistical tests are needed to draw conclusions. Thus, this research is the quantitative research.

Space-Time Autoregressive Integrated Moving Average (STARIMA) Model

Before analysis and modeling are carried out, time series data must meet the stationarity test. The stationarity test that must be met is stationary in the variance and stationarity in the mean. In this research, the Box-Cox plot is used for stationarity of variance. Data needs to be transformed if a value of $\lambda \neq 1$ is obtained.⁽³⁾ Meanwhile, the test used to determine stationarity in the mean is Augmented Dickey-

Fuller (ADF) test. The assumption test is met if the p-value of each test is less than significance level α .⁽⁷⁾

The STARIMA model was employed for forecasting space-time series data, which is particularly suitable for scenarios where both temporal and spatial dependencies exist.⁽³⁾ STARIMA performs forecasting by incorporating these dependencies, where spatial factors play a significant role in the propagation of motor vehicle theft cases across different regions in Surabaya. To analyze the STARIMA model, there are three orders that need to be determined first: p , which is the temporal autoregressive order identified based on the Space-Time Autocorrelation Function (STACF) plot; q , which is the moving average order identified based on the Space-Time Partial Autocorrelation Function (STPACF) plot; and d , which is the differencing order.⁽⁷⁾ Next, it is determined whether the parameters in the model are significant or not. In the space-time model, the best parameter estimate to use is maximum likelihood estimation (MLE).⁽⁴⁾ To determine the best model, the root mean square error (RMSE) was used. The model with the lowest RMSE value is the best model.⁽⁵⁾

As STARIMA models focus on both temporal and spatial forecasting, a spatial weight matrix is utilized to incorporate spatial dependencies. Two types of spatial weight matrices were considered in this research: the uniform matrix and the inverse matrix. If using the uniform matrix, the weight for each neighboring location is the same, while if using the inverse matrix, it uses an inverse comparison with the actual distance so that the weight of each location will not be the same.⁽¹⁰⁾ The STARIMA model is expressed as follows:

$$\nabla Z_i(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} W_l \nabla Z_i(t-k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W_l a_i(t-k) + a_i(t)$$

In this equation, $\nabla Z_i(t)$ represents the differenced time series for location i at time t , capturing the changes in motor vehicle theft rates over time. The parameters p and q denote the orders of the autoregressive and moving average components, respectively. The indices k and l represent the time lag and spatial lag, respectively, with λ_k and m_k indicating the spatial orders for the autoregressive and moving average terms. The parameters ϕ_{kl} and θ_{kl} are the coefficients that quantify the influence of past values and errors, respectively, both in time and across space. The spatial weight matrix W_l defines the influence of neighboring regions on the crime rate at location i , with $a_i(t)$ representing the random error term at time t .

Diagnostic tests were performed to ensure the validity of the STARIMA model. The Ljung-Box test was used to assess whether the residuals exhibited white noise, satisfied when the p-value exceeded the significance level α .⁽¹¹⁾ The Shapiro-Wilk test was conducted to verify the normality of the residuals, also satisfied when the p-value was greater than α .⁽¹²⁾ The model's performance was evaluated by comparing the Root Mean Square Error (RMSE) of the predictions against actual out-of-sample data, determining whether the inclusion of spatial factors improved the accuracy of predicting motor vehicle theft cases across Surabaya's regions.⁽¹³⁾

RESULTS AND DISCUSSION

Exploratory Data Analysis

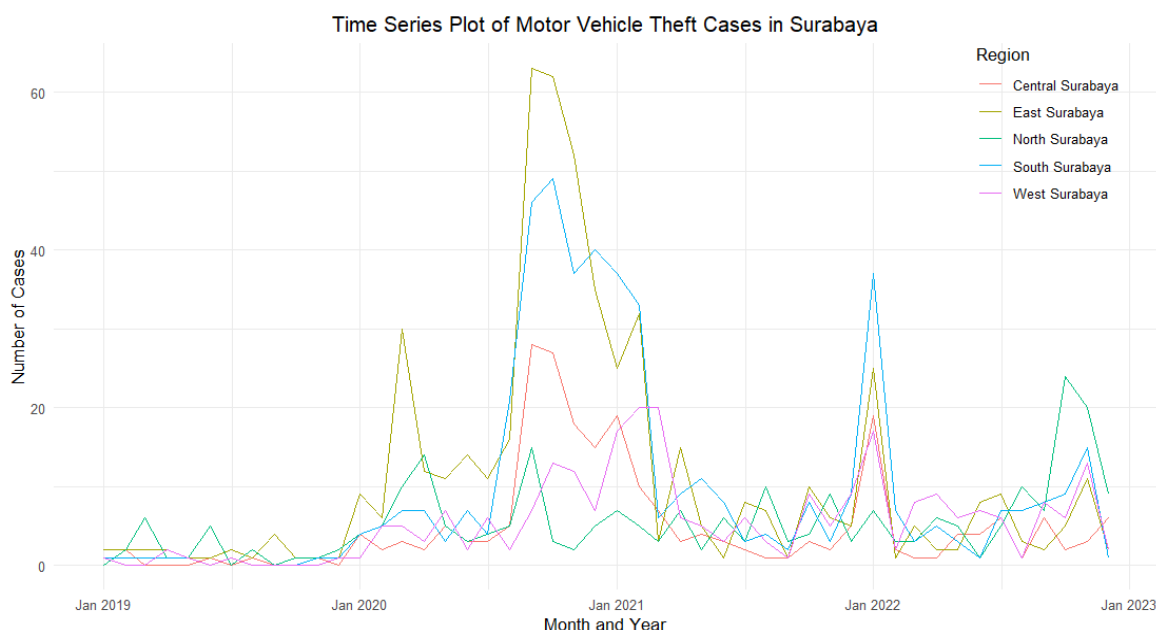


Figure 1. Time series plot of monthly motor vehicle theft cases across five regions in Surabaya (2019-2023)

As illustrated in figure 1, the temporal patterns of theft incidents exhibit notable variations, with distinct spikes that highlight periods of increased criminal activity. East Surabaya emerges as the region with the most significant fluctuations, particularly around early 2021, where theft cases surged dramatically. This spike coincides with the relaxation of COVID-19 restrictions, likely leading to increased mobility and opportunities for criminal activities.⁽¹⁴⁾ As a major hub for commerce and trade, East Surabaya's economic activity may have contributed to higher theft incidents, further emphasizing the correlation between economic factors and crime rates. Similarly, South Surabaya also experienced a notable rise in cases, underscoring the vulnerability of densely populated and economically active regions to such crimes. In contrast, West, Central, and North Surabaya displayed more stable trends with fewer fluctuations, suggesting either fewer opportunities for theft or more effective crime prevention measures. Additionally, the periodic spikes in theft cases, especially during holiday seasons, further emphasize the need for region-specific crime prevention strategies and continuous monitoring throughout the year.⁽¹⁵⁾

Spatial Weighting Matrix

The weights used in this study are uniform location weights and inverse distance location weights. These two weights are used to obtain significant Moran Index results and the best STARIMA model with the appropriate location weights. The uniform location weight matrix is W_A , which shows the uniform value among the regions. Meanwhile, the inverse distance location weight uses the actual distance between locations which is then normalized. This weight is obtained from the inverse of the actual distance, with the latitude and longitude position of each region used as the center or centroid. The inverse distance location weight matrix is W_B , which gives smaller coefficient values for longer distances and vice versa. These two matrices are used to assess the spatial relationship between areas in Surabaya in the context of this study.

$$W_A = \begin{pmatrix} 0 & 0.333 & 0.333 & 0 & 0.333 \\ 0.250 & 0 & 0.250 & 0.250 & 0.250 \\ 0.333 & 0.333 & 0 & 0.333 & 0.333 \\ 0 & 0.333 & 0.333 & 0 & 0.333 \\ 0.333 & 0.333 & 0 & 0.333 & 0 \end{pmatrix}$$

$$W_B = \begin{pmatrix} 0 & 0.379 & 0.238 & 0.170 & 0.214 \\ 0.455 & 0 & 0.176 & 0.169 & 0.200 \\ 0.323 & 0.199 & 0 & 0.247 & 0.231 \\ 0.287 & 0.238 & 0.307 & 0 & 0.167 \\ 0.330 & 0.256 & 0.262 & 0.152 & 0 \end{pmatrix}$$

Spatial Autocorrelation

Table 1. Moran's I result test			
I	E (I)	Var (I)	P-Value
-0,1027	-0,2500	0,0065	0,0342

Based on table 1, it can be seen that there is spatial autocorrelation in motor vehicle theft cases in each region of Surabaya with a p-value (0,0342) < α (0,05), which means that the level of motor vehicle theft cases in an area is influenced by the level of motor vehicle theft cases in other areas.

Stationary Test

Data is considered stationary if it has a constant mean and variance over time, so that the situation in the past remains relevant to the situation in the present and future. To achieve stationarity, several transformations and tests are performed. One of the first steps is to perform logarithmic transformation, especially to handle zero or non-positive values in the data of motor vehicle theft cases in Surabaya, as recommended by Atkinson et al.⁽¹⁶⁾ This transformation helps in stabilizing the variance and facilitates further analysis.

$$Z_t^* = \log(Z_t + 4)$$

After logarithmic transformation, a Box-Cox Test was conducted to test for stationarity of variance. The test results show that some regions require additional transformation to achieve stationarity of variance and mean.

Table 2. Lambda result of Box-Cox test and transformation

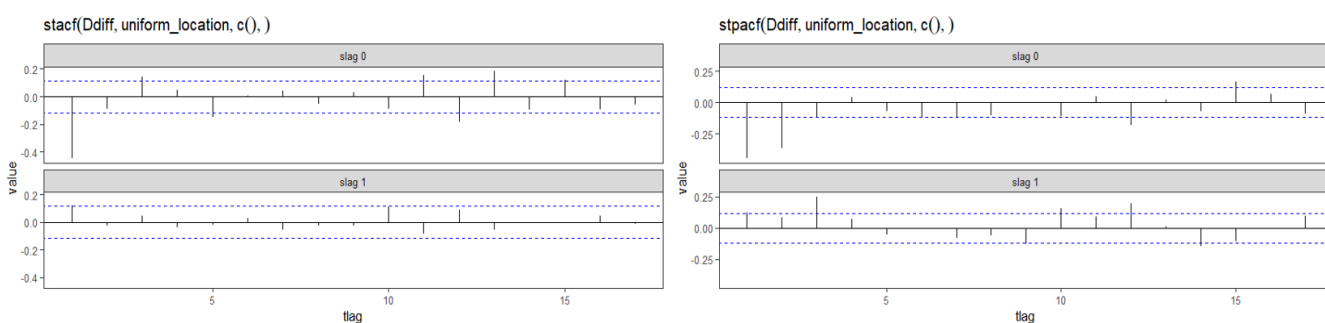
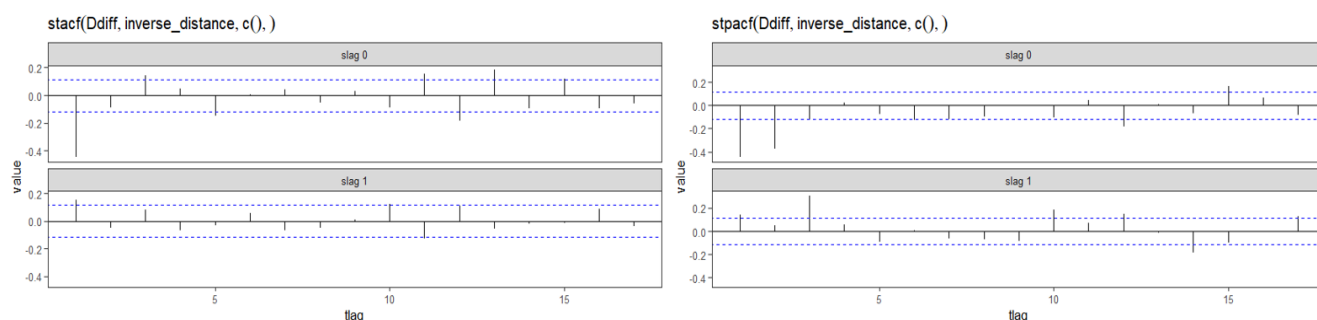
Region	Transformation Box-Cox 1 (Z_t^{***})		Transformation Box-Cox 2 (Z_t^{***})		
	λ	Transformation	λ	Transformation	λ
West Surabaya	1	Z_t^*	1	Z_t^*	1
Central Surabaya	1	Z_t^*	1	Z_t^*	1
South Surabaya	0,5	$\sqrt{Z_t^*}$	0,5	$\sqrt{Z_t^*}$	1
East Surabaya	-3	Z_t^{*-3}	1	Z_t^*	1
North Surabaya	0,5	$\sqrt{Z_t^*}$	1	Z_t^*	1

Table 3. Augmented Dickey-Fuller test results

Region	Before Differencing		After Differencing	
	τ	p-value	τ	p-value
West Surabaya	-2,1225	0,5253	-4,1502	0,0100
Central Surabaya	-2,1783	0,5027	-4,2672	0,0100
South Surabaya	-2,2615	0,4691	-3,8460	0,0225
East Surabaya	-2,0700	0,5465	-4,3023	0,0100
North Surabaya	-2,6310	0,3197	-5,3352	0,0100

Based on the results presented in table 2 and table 3, it is evident that three regions in Surabaya—Central Surabaya, South Surabaya, and North Surabaya—had lambda (λ) values that were not equal to 1 during the initial Box-Cox transformation test. As a result, these regions required further transformations to stabilize the variance. Specifically, since the initial transformation did not achieve stationarity in all regions, a second Box-Cox transformation was applied, resulting in the data being transformed until it reached the appropriate stationary form, denoted as Z_t^{***} . This second transformation ensured that all regions, including those with initial λ values different from 1, achieved the required stationarity in variance. Following this, the Augmented Dickey-Fuller (ADF) test was applied to assess stationarity in mean. Initially, the data were non-stationary in mean, as indicated by p-values greater than 0,05. However, after differencing, the p-values for all regions dropped below 0,05, confirming that the data had become stationary in mean. This combined process of multiple transformations and differencing was essential to ensure that the motor vehicle theft data from all regions in Surabaya were suitable for reliable time series analysis.

STARIMA Model Identification

**Figure 2.** STACF & STPACF plot with uniform location weight matrix**Figure 3.** STACF & STPACF plot with inverse distance weight matrix

The STARIMA model has two orders, namely time order and spatial order. Generally, the spatial order is limited to one because each region in Surabaya is neighboring in one city, so that shifts between locations remain in the same region, as well as to facilitate interpretation.⁽¹⁷⁾ The STARIMA model was identified using two types of weight matrices, namely inverse distance weights and uniform location weights. Model identification is done by analyzing the STACF and STPACF plots of the differenced data, looking at the first 10 lags.

Based on the STACF and STPACF plots using inverse distance and uniform location weighting, we observed an exponential decay in STACF values over the first 10 lags, indicating reduced autocorrelation as lag increases, consistent with STARMA model characteristics. The STPACF plots show a consistent decline at spatial lags 0 and 1, followed by cutoffs at time lags, suggesting appropriate model order. Therefore, the recommended STARIMA models for this data are STARIMA (1,1,1), STARIMA (2,1,1), STARIMA (1,1,2), and STARIMA (2,1,2), with AR and MA orders up to 2 based on significant decreases in STACF and STPACF values. STACF and STPACF plot with uniform location weight matrix and with inverse distance weight matrix are given in figure 2 and figure 3, respectively.

Parameter Estimation

Utilizing the previously constructed spatial weighting matrices, parameter estimation for each STARIMA model combination was conducted using the Maximum Likelihood Estimation (MLE) method. To determine the significance of the estimated parameters, a t-test was performed. The results of the parameter estimation and t-tests for each model combination are presented in table 4 through table 9.

Table 4. Estimation parameter model STARI(1,1)

Weight Matrix	Parameter	Estimation	t_{test}	t_{table}	Description
Uniform Location	ϕ_{10}	-0,4519	-7,9883	1,9690	Significant
	ϕ_{11}	0,1170	1,1724	1,9690	Not Significant
Inverse Distance	ϕ_{10}	-0,4450	-7,6828	1,9690	Significant
	ϕ_{11}	0,1637	1,7079	1,9690	Not Significant

Table 5. Estimation parameter model STIMA(1,1)

Weight Matrix	Parameter	Estimation	t_{test}	t_{table}	Description
Uniform Location	θ_{10}	-0,5900	-9,0811	1,9690	Significant
	θ_{11}	0,2595	2,3630	1,9690	Significant
Inverse Distance	θ_{10}	-0,5950	-9,1493	1,9690	Significant
	θ_{11}	0,2615	2,5285	1,9690	Significant

Table 6. Estimation parameter model STARIMA(1,1,1)

Weight Matrix	Parameter	Estimation	t_{test}	t_{table}	Description
Uniform Location	ϕ_{10}	-0,1695	-1,5793	1,9690	Not Significant
	ϕ_{11}	-0,0007	-0,0032	1,9690	Not Significant
	θ_{10}	-0,4128	-3,1933	1,9690	Not Significant
	θ_{11}	0,2002	0,8310	1,9690	Not Significant
Inverse Distance	θ_{10}	-0,1543	-1,2417	1,9690	Not Significant
	θ_{11}	0,1272	0,5202	1,9690	Not Significant
	θ_{10}	-0,4188	-2,8537	1,9690	Not Significant
	θ_{11}	0,1132	0,4077	1,9690	Not Significant

Table 7. Estimation parameter model STARIMA(2,1,1)

Weight Matrix	Parameter	Estimation	t_{test}	t_{table}	Description
Uniform Location	ϕ_{10}	-0,5196	-3,3658	1,9690	Significant
	ϕ_{11}	-0,1706	-0,5662	1,9690	Not Significant
	ϕ_{20}	-0,2332	-2,4343	1,9690	Significant
	ϕ_{21}	-0,1080	-0,6556	1,9690	Not Significant

Inverse Distance	θ_{10}	-0,1092	-0,6412	1,9690	Not Significant
	θ_{11}	0,4630	1,3953	1,9690	Not Significant
	ϕ_{10}	-0,5360	-2,7526	1,9690	Significant
	ϕ_{11}	-0,0938	-0,2832	1,9690	Not Significant
	ϕ_{20}	-0,2468	-2,0984	1,9690	Significant
	ϕ_{21}	-0,1079	-0,6460	1,9690	Not Significant
	θ_{10}	-0,0830	-0,3837	1,9690	Not Significant
	θ_{11}	0,3968	1,0683	1,9690	Not Significant

Table 8. Estimation parameter model STARIMA(1,1,2)

Weight Matrix	Parameter	Estimation	t_{test}	t_{table}	Description
Uniform Location	ϕ_{10}	-0,4388	-1,9890	1,9690	Significant
	ϕ_{11}	-0,3447	-1,0286	1,9690	Not Significant
	θ_{10}	-0,1694	-0,7136	1,9690	Not Significant
	θ_{11}	0,6713	1,8059	1,9690	Not Significant
	θ_{20}	-0,1513	-0,8874	1,9690	Not Significant
	θ_{21}	-0,1890	-0,8290	1,9690	Not Significant
Inverse Distance	ϕ_{10}	-0,5242	-2,4897	1,9690	Significant
	ϕ_{11}	-0,3011	-0,9721	1,9690	Not Significant
	θ_{10}	-0,0588	-0,2538	1,9690	Not Significant
	θ_{11}	0,6886	1,9281	1,9690	Not Significant
	θ_{20}	-0,2129	-1,2924	1,9690	Not Significant
	θ_{21}	-0,1497	-0,7282	1,9690	Not Significant

Table 9. Estimation parameter model STARIMA(2,1,2)

Weight Matrix	Parameter	Estimation	t_{test}	t_{table}	Description
Uniform Location	ϕ_{10}	-0,8132	-3,9583	1,9690	Significant
	ϕ_{11}	-0,0010	-0,0027	1,9690	Not Significant
	ϕ_{20}	-0,2211	-1,7443	1,9690	Not Significant
	ϕ_{21}	-0,0743	-0,2942	1,9690	Not Significant
	θ_{10}	0,2199	0,9575	1,9690	Not Significant
	θ_{11}	0,2866	0,7231	1,9690	Not Significant
	θ_{20}	-0,2957	-1,6441	1,9690	Not Significant
Inverse Distance	θ_{21}	0,1852	0,6128	1,9690	Not Significant
	ϕ_{10}	-0,7040	-2,6420	1,9690	Significant
	ϕ_{11}	0,1473	0,3753	1,9690	Not Significant
	ϕ_{20}	-0,2458	-1,5282	1,9690	Not Significant
	ϕ_{21}	-0,1747	-0,5448	1,9690	Not Significant
	θ_{10}	0,0886	0,2957	1,9690	Not Significant
	θ_{11}	0,1614	0,3511	1,9690	Not Significant
	θ_{20}	-0,1955	-0,8416	1,9690	Not Significant
	θ_{21}	0,3408	0,8116	1,9690	Not Significant

Using the t-table to determine parameter significance, it was found that the STIMA (1,1) model with the specified weighting matrix yielded significant results. Although some parameters in other models were also significant, lower-order models demonstrated better stability. In this study, non-significant parameters were retained in the model as the primary focus was on model adequacy, assessed through RMSE values to ensure accurate forecasting. Therefore, the significance testing of parameters can be disregarded.⁽¹⁸⁾

Diagnostic Check

Assumption of Residual White Noise

Table 10. Test of residual white noise assumptions			
Model	Weight	p-value	Decision
STARI (1,1)	Uniform Location	0,3940	White Noise
	Inverse Distance	0,4187	White Noise
STIMA (1,1)	Uniform Location	0,5656	White Noise
	Inverse Distance	0,5727	White Noise
STARIMA (1,1,1)	Uniform Location	0,6001	White Noise
	Inverse Distance	0,5979	White Noise
STARIMA (2,1,1)	Uniform Location	0,5922	White Noise
	Inverse Distance	0,7007	White Noise
STARIMA (1,1,2)	Uniform Location	0,4282	White Noise
	Inverse Distance	0,5382	White Noise
STARIMA (2,1,2)	Uniform Location	0,4282	White Noise
	Inverse Distance	0,5382	White Noise

Table 10 demonstrates that all model combinations, using both uniform location and inverse distance weighting matrices, have p-value > α (0,05). This indicates that there is no autocorrelation between residuals, meaning the residuals satisfy the white noise assumption.

Multivariate Normal Residual Assumption

Table 11. Multivariate normal residual assumption test			
Model	Weight	p-value	Decision
STARI (1,1)	Uniform Location	0,5051	Normal Multivariate
	Inverse Distance	0,5024	Normal Multivariate
STIMA (1,1)	Uniform Location	0,5242	Normal Multivariate
	Inverse Distance	0,5089	Normal Multivariate
STARIMA (1,1,1)	Uniform Location	0,5270	Normal Multivariate
	Inverse Distance	0,5104	Normal Multivariate
STARIMA (2,1,1)	Uniform Location	0,5209	Normal Multivariate
	Inverse Distance	0,5081	Normal Multivariate
STARIMA (1,1,2)	Uniform Location	0,5207	Normal Multivariate
	Inverse Distance	0,5047	Normal Multivariate
STARIMA (2,1,2)	Uniform Location	0,5207	Normal Multivariate
	Inverse Distance	0,5047	Normal Multivariate

Table 11 shows that all model combinations, using both uniform location and inverse distance weighting matrices, have p-values > α (0,05). This indicates that the residuals meet the assumption of normal distribution.

The Best Model Selection

Table 12. RMSE values of all models		
Model	Weight	RMSE
STARI (1,1)	Uniform Location	73,4610
	Inverse Distance	16,8320
STIMA (1,1)	Uniform Location	19,7030
	Inverse Distance	40,4990
STARIMA (1,1,1)	Uniform Location	20,5680
	Inverse Distance	18,1570

STARIMA (2,1,1)	Uniform Location	18,4330
	Inverse Distance	21,1980
STARIMA (1,1,2)	Uniform Location	11,7510
	Inverse Distance	203,0640
STARIMA (2,1,2)	Uniform Location	6 708,2560
	Inverse Distance	4,58873E+26

Forecasting Out-Sample

Before carrying out forecasting, information is provided regarding the transformations applied to this research data. Information is given in table 13.

Table 13. Tranformation form and inverse transformation		
Region	Transformation Form	Inverse Transformation
West Surabaya	Z_t^*	$Z_t = Z_t^* - 4$
Center Surabaya	Z_t^*	$Z_t = Z_t^* - 4$
South Surabaya	$\sqrt{Z_t^*}$	$Z_t = (Z_t^{***})^2 - 4$
East Surabaya	Z_t^{*-3}	$Z_t = (Z_t^{**})^{-1/3} - 4$
North Surabaya	$\sqrt{Z_t^*}$	$Z_t = (Z_t^{**})^2 - 4$

The STARIMA (1,1,2) model, applied with uniform location weighting, was used to forecast motor vehicle theft cases across five regions in Surabaya, using out-of-sample data from March 2023 to December 2023. Since the data utilized for modeling were transformed to achieve stationarity, the forecasted results required inverse transformations to return the data to their original scale, as outlined in the table above.

The inverse transformations vary across the regions, depending on the specific transformation applied during the modeling phase. For instance, South Surabaya, which initially underwent a square root transformation, required squaring after adjustment to restore the original data scale, while Central Surabaya, which had a straightforward transformation, simply added a constant value post-modeling.

The forecasting results, shown in figure 4, demonstrate that the STARIMA (1,1,2) model effectively captures the temporal and spatial patterns of historical motor vehicle theft data across the regions. The model performed well in areas like East and Central Surabaya, where theft patterns are more pronounced and variable, whereas regions with more stable or less variable data, such as West and North Surabaya, showed slightly less precision. These findings underscore the model’s robustness in predicting crime trends and highlight its potential utility for assisting law enforcement agencies in Surabaya with resource allocation and strategic planning.

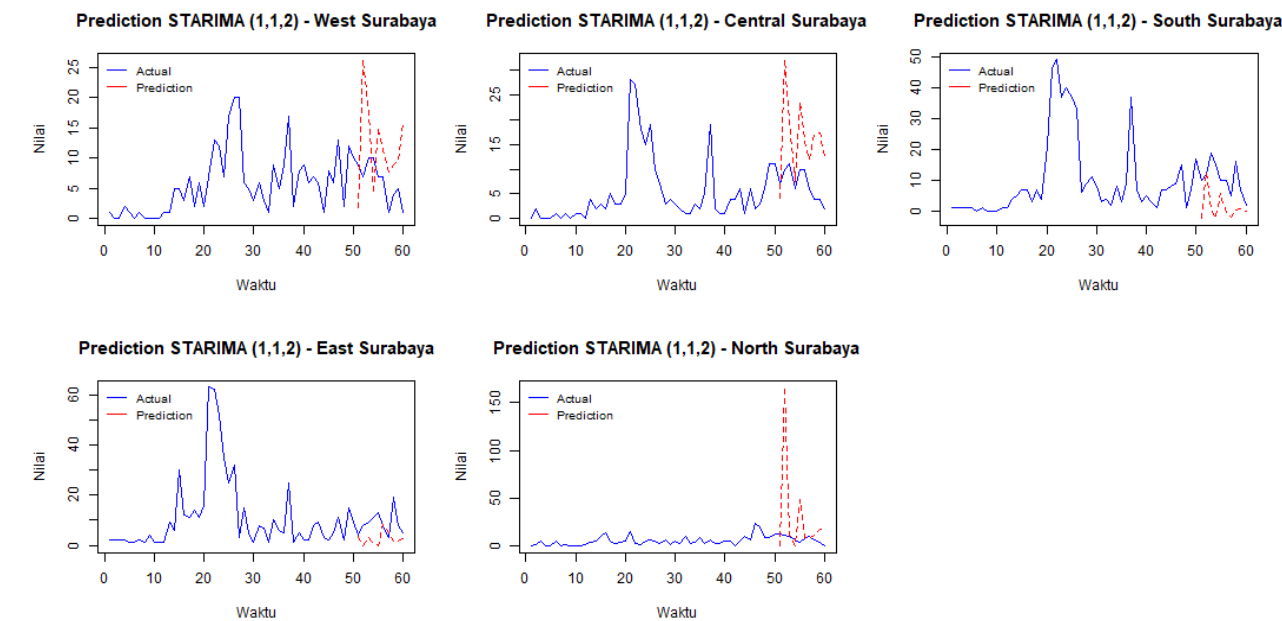


Figure 4. Time series plot prediction

Figure 4 shows the forecasting results using the STARIMA (1,1,2) model with uniform location weighting. The model successfully captures the temporal patterns of historical data across the five regions of Surabaya: West, Central, South, East, and North. The plots indicate that the model is fairly accurate in predicting motor vehicle theft cases in Surabaya. The best or closest predictions to the actual data are observed in South Surabaya and East Surabaya, where the model closely follows the actual data trends. However, some data points still do not fully match or deviate slightly from the actual data, possibly due to the model estimating higher or lower values compared to the actual trends. This may be due to sudden spikes or drops in the historical patterns that the model did not fully anticipate. Overall, the model provides a good prediction of the temporal patterns in each region of Surabaya, effectively capturing the major trends in motor vehicle theft cases across the city.

Forecasting the Next 5 Months

Month and Year	West Surabaya	Center Surabaya	South Surabaya	East Surabaya	North Surabaya
January 2024	3	3	1	4	2
February 2024	3	4	1	3	7
March 2024	4	2	3	3	4
April 2024	6	6	17	1	15
May 2024	4	2	1	10	2

Table 14 presents the 5-month forecast of motor vehicle theft cases across five regions in Surabaya using the STARIMA (1,1,2) model with uniform location weighting. The predictions indicate varying trends across the regions. West Surabaya is expected to see a significant rise in cases, peaking at 6 in April 2024, suggesting the need for increased security measures. South Surabaya shows the most substantial spike with 17 cases in April 2024, highlighting a serious security concern that demands special attention. Central and East Surabaya exhibit notable fluctuations, with Central Surabaya peaking at 6 cases in April 2024 and East Surabaya reaching 10 cases in May 2024, indicating the necessity for flexible security strategies. North Surabaya shows consistent increases, particularly in February and April 2024, with 7 and 15 cases respectively, requiring intensified preventive measures. Overall, these predictions provide critical insights into the regions that require heightened surveillance and preventive efforts, supporting the refinement of future security policies and strategies. This research has contributed to time-series data modeling by considering location effects. With the addition of location effects, the model can be made more accurate. The limitation of the research is that when adding spatial effects, only two weighting matrices are used, namely uniform and inverse. In further research, other spatial weighting matrices can be used.

CONCLUSIONS

The STARIMA (1,1,2) model with uniform location weights was selected as the best model based on the results of the model performance evaluation with an RMSE value of 11,751. The estimated Autoregressive (AR) and Moving Average (MA) parameters indicate that this model is able to capture good time autocorrelation patterns. The results of the diagnostic test indicate that this model has also met the assumptions of white noise and multivariate normal. Case predictions for the next five months show fluctuations in each region of Surabaya, with the highest regions in succession being North Surabaya, East Surabaya, and South Surabaya. These results can be a reference for law enforcement and crime prevention strategies in each region of Surabaya, especially areas with the highest predictions.

BIBLIOGRAPHIC REFERENCES

1. Satiawan PR, Tucunan KP, Azarine RY. The spatial configuration of crime in Surabaya. IOP Conf Ser Earth Environ Sci. 2019;340(1):012035. doi: <https://dx.doi.org/10.1088/1755-1315/340/1/012035>
2. Supangat S, Sholiq MM. The Utilization of Information System for Crime Rate Modelling in Surabaya Using K-means. Journal of Information Technology and Cyber Security. 2023;1(1):22-30.
3. Pfeifer PE, Deutsch SJ. A Three-Stage Iterative Procedure for Space-Time Modeling. Technometrics. 1980;22(1):35-47. doi: <http://www.jstor.org/stable/1268381>
4. Hakim S, Shachmurove Y. Spatial and Temporal Patterns of Commercial Burglaries. The American Journal of Economics and Sociology. 1996 Oct 1;55(4):443-56. doi: <https://doi.org/10.1111/j.1536-7150.1996.tb02643.x>

5. Ratcliffe J. The Hotspot Matrix: A Framework for the Spatio-Temporal Targeting of Crime Reduction. *Police Practice and Research*. 2004 Apr 1;5:5-23.
6. Berkani S, Guermah B, Zakroum M, Ghogho M. Spatio-temporal forecasting: A survey of data-driven models using exogenous data. *IEEE Access*. 2023;11:75191-214.
7. Giacomini R, Granger CWJ. Aggregation of space-time processes. *J Econom*. 2004;118(1-2):7-26.
8. Rantini D, Iriawan N, Irhamah. Fernandez-steel skew normal conditional autoregressive (FSSN CAR) model in stan for spatial data. *Symmetry*. 2021;13(4):545.
9. Rantini D, Iriawan N, Irhamah. Bayesian Mixture Generalized Extreme Value Regression with Double-Exponential CAR Frailty for Dengue Haemorrhagic Fever in Pamekasan, East Java, Indonesia. *J Phys Conf Ser*. 2021;1752(1):12022. doi: <http://dx.doi.org/10.1088/1742-6596/1752/1/012022>
10. Pfeifer PE, Deutsch SJ. Independence and sphericity tests for the residuals of space-time arma models. *Communications in Statistics-Simulation and Computation*. 1980;9(5):533-49.
11. Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika*. 1978;65(2):297-303.
12. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52(3-4):591-611.
13. Saha A, Singh KN, Ray M, Rathod S. A hybrid spatio-temporal modelling: an application to space-time rainfall forecasting. *Theor Appl Climatol*. 2020;142:1271-82.
14. Syamsuddin R, Fuady MIN, Prasetya MD, Umar K. The effect of the COVID-19 pandemic on the crime of theft. *Int J Criminol Sociol*. 2021;10:305-12.
15. Anggrayni A. The Effect of Economic Factors on Property Crime Rates. *Efficient: Indonesian Journal of Development Economics*. 2022 Jun 17;5(2). doi: <https://journal.unnes.ac.id/sju/efficient/article/view/51036>
16. Atkinson AC, Riani M, Corbellini A. The box-cox transformation: Review and extensions. 2021;
17. Wutsqa DU. Seasonal multivariat time series forecasting on tourism data by using var-gstar model. *Jurnal Ilmu Dasar*. 2010;11(1):101-9.
18. Kostenko A V, Hyndman RJ. Forecasting without significance tests. manuscript, Monash University, Australia. 2008;

FINANCING

The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum.

Data curation: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Yohanes Manasye Triangga.

Formal analysis: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Yohanes Manasye Triangga.

Research: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Fazidah Othman.

Methodology: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Yohanes Manasye Triangga.

Project management: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum.

Resources: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Fazidah Othman.

Software: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Yohanes Manasye Triangga.

Supervision: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum.

Validation: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Fazidah Othman.

Display: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Yohanes Manasye Triangga.

Drafting - original draft: Yohanes Manasye Triangga, Dwi Rantini.

Writing - proofreading and editing: Arip Ramadan, Dwi Rantini, Ratih Ardiati Ningrum, Fazidah Othman.