Trend Analysis and Prediction of Violence Against Women and Children Cases in Jakarta Based on the Victim's Education Level Using ARIMA and SARIMA Method

Zaqi Kurniawan[1]*, Rizka Tiaharyadini[2],Arief Wibowo[3] ,Rusdah[4]
Department of Technology and Information [1], [2], [3], [4]
Universitas Budi Luhur
Jakarta, Indonesia
zaqi.kurniawan@budiluhur.ac.id[1], rizka.tiaharyadini @budiluhur.ac.id [2],arief.wibowo@budiluhur.ac.id [3] rusdah@budiluhur.ac.id [4]

Abstract— **Violence against women and children remains a critical social issue in Jakarta, Indonesia, where densely populated urban areas often correlate with increased risks of domestic abuse. The urgency of addressing this problem lies in its direct impact on public health, education, and community well-being. This study uses time series prediction models to examine and anticipate trends in the number of reported incidents of violence against women and children in Jakarta. Using publicly accessible data from Jakarta Open Data and the National Commission for the Protection of Women and Children, we applied the ARIMA and SARIMA Models. Key variables included in the dataset are the data period, education level, and total number of victims Using three performance indicators—MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and RMSE (Root Mean Square Error)—to assess model accuracy the ARIMA model performed better than the SARIMA model. SARIMA recorded an RMSE of 80.26, an MAE of 66.21, and an undefined MAPE because of zero values in the real data, while ARIMA specifically obtained an RMSE of 32.22, an MAE of 32.09, and a MAPE of 5.19%. These results suggest that the non-seasonal ARIMA model is more suitable for this dataset. The study contributes to policy planning and early intervention strategies by offering a data-driven approach to predicting trends in violence within urban contexts**.

Keywords— ***Arima, Child Violence, Education, Sarima, Gender-Based and Child-Directed Violence***

I. INTRODUCTION

Violence against women and children is a significant societal that impacts communities worldwide [1]. According to research and reports over the years, this issue has escalated significantly in Indonesia, particulary in urban areas such as Jakarta. In addition to inflicting immediate harm on victims, these violent crimes also have broader social implications, perpetuating cycles of social unrest, poverty and inequality [2]. Jakarta, the capital and most populous city of the country, faces significant challenges in addressing gender-based violence due to its complex cultural dynamics, socioeconomic inequalities, and large population [3]. According to research, stigma, fear, and systemic shortcomings in addressing the issue are the primary reasons why violence often goes unreported [4]. In Jakarta, domestic violence is a prevalent form of abuse that disproportionately affects women and children and reflecting deeper societal issue that require immediate attention [5]. Such violence jeopardizes the foundation of Indonesian society foundation by undermining public health and the welfare of future generations. The importance of addressing this issue cannot be overstated. Protection is essential to ensure the rights and well-being of women and children, who often represent the most vulnerable segments of society [6]. Research shows that violence has a wide range of impacts, including economic instability for families and communities, disruption of schooling, and psychological trauma [7]. Furthermore, the abuse of the rights of women and children reinforces inequality and hinders Indonesia's progress toward social justice and sustainable development [8]. Civil society, law enforcement, and legislators must give this issue careful consideration. Jakarta can lead the nation in preventing violence by raising awareness and establishing a strong legislative framework. According to [9], Addressing the underlying causes and societal factors is essential for developing remedies that are specific to urban issues. Creating effective strategies necessitates a comprehensive understanding of the causes of violence against women and children. Research indicates that education level is a significant factor contributing to vulnerability to violence. Research has repeatedly shown that limited educational attainment is strongly associated with a higher risk of experiencing sexual violence and other abusive behaviors, both within households and in broader social contexts [10].

In urban areas such as Jakarta, where socioeconomic

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

disparities exacerbate vulnerabilities, this issue is particularly evident [11]. Studies empharize that education is a protective factor that gives people the skills and imformation they need to identify and avoid abusive circumstances.

Research by [12], studies emphasize that education is a protective factor that equips people who possess the abilities and knowledge required to recognize and steer clear of harmful circumstances. Similar results were obtained by [13], research has shown that low levels of education increase the risk of early marriage, which exposes young women to prolonged cycles of abuse. Furthermore, in Jakarta's high-density neighborhoods, the combination of low parental education and financial hardship creates an environment conducive to violence against children. [14]. Community programs that raise awareness and empower vulnerable groups to break the cycle of abuse must be implemented alongside efforts to enhance access to education for women and children. For effective intervention and prevention of violence against women and children, it is essential to analyze patterns and make predictions. Through these assessments, policymakers and stakeholders can identify trends, root causes, and areas of increased vulnerability. This proactive approach aids in resource allocation and the development of targeted policies that address specific risk factors. Research indicates that predictive modeling can be utilized to identify high-risk situations and facilitate prompt actions, potentially saving lives [15]. Similar to [16], advanced analytics, including image and video analysis, can be utilized to predict violent incidents, particularly in cyberspace, where a significant number of cases involving women and children originate. Data-driven insights are especially crucial in cities like Jakarta, where diverse socioeconomic and cultural factors necessitate complex strategies. According to research by [17], the ability to predict potential surges in violence enables the development of community-specific education and awareness initiatives. According to [18], consistent monitoring of violence trends ensures that policies remain relevant and adaptable to evolving societal contexts. By utilizing historical data, authorities can enhance their strategies and concentrate their efforts where they are most needed.

There are two widely used models for evaluating and forecasting time series data seasonal and non-seasonal ARIMA techniques. Their use is especially appropriate for examining trends and patterns in incidents of violence against women and children. Each model has unique benefits; while ARIMA accounts for seasonal changes, it captures broad patterns, allowing for comprehensive long-term data forecasting. Because of their effectiveness in handling complex datasets, these models have been widely utilized in research. An examination of monthly crime data, including incidents of violence against women and children, demonstrated the utility of SARIMA and its ability to accurately capture seasonal patterns [19]. In a similar vein, the ability of ARIMA to dynamically forecast in response to external factors, such as policy changes and global outbreaks, was demonstrated. It was used to anticipate changes in violent crime rates [20]. The integration of ARIMA and SARIMA into violence prediction models offers several key benefits. These models excel at forecasting future patterns based on historical data, which significantly reduces uncertainty [21].

This study is to investigate trends in the prevalence of violence against women and children in Jakarta, with a particular focus on how academic achievement influences susceptibility. This study analyzes historical data to identify trends and correlations that demonstrate the impact of education on an individual's vulnerability to violence. Comprehending these patterns is essential for developing interventions and effective policies that address the underlying causes of violence in cities such as Jakarta. The ARIMA and SARIMA methods are also used in this study to increase its scope. SARIMA accounts for seasonality and captures recurring patterns influenced by factors such as cultural or economic cycles, while ARIMA provides a robust framework for modeling and forecasting extensive time-series data. By utilizing these complementary methodologies, the research can deliver accurate, long-term predictions of future violent incidents while considering periodic fluctuations.

To increase the study's scope, the study uses the Auto-Regressive Integrated Moving ARIMA and SARIMA methodologies. SARIMA accounts for seasonality and captures recurrent patterns influenced by factors such as cultural or economic cycles, while ARIMA provides a robust framework for modeling and forecasting extensive time-series data. This research aims to deliver precise, long-term predictions of potential violent incidents while considering periodic fluctuations, thanks to these complementary methodologies. The simultaneous application of ARIMA and SARIMA ensures accurate modeling of both historical and projected patterns, offering stakeholders valuable insights. Policymakers, social workers, and community leaders can utilize these findings to develop targeted initiatives, allocate resources effectively, and implement timely interventions to mitigate violence against women and children in Jakarta. By integrating trend analysis with comprehensive predictive modeling, this research project seeks to bridge the knowledge gap between the past and the present, ultimately contributing to a more secure and equitable society..

II. Related Work

Several studies have explored the application of time series forecasting methods to analyze and predict social issues, particularly those related to violence and public safety. The ge SARIMA and ARIMA models have been widely utilized due to their effectiveness in capturing temporal patterns in socio-criminal data. Research by [22], The SARIMA model was employed to forecast crime trends in the Philippines, demonstrating its ability to effectively manage monthly seasonality in crime reports, particularly those related to violence against women and children. Similarly, [23] compared hybrid neural networks with ARIMA in predicting child-line calls in Kenya, highlighting the practical utility of ARIMA for short-term forecasting of social issues. In Zambia, research by [24], discovered that parameters including marital status and educational attainment were important predictors of sexual and gender-based assault cases when ARIMA models were used to

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

anticipate them. Likewise, [25] [26] applied the ARIMA model to study the impact of the COVID-19 lockdown on criminal behavior in Dhaka and noted that reduced mobility and educational disruptions had complex implications for crime trends.

Despite these efforts, existing studies rarely focus on Indonesia, particularly Jakarta, as a geographic context, and none explicitly analyze the role of victims' education levels as a central variable in forecasting cases of violence against women and children. This presents a critical gap in the literature, especially considering the urban complexity and population density of Jakarta, which may influence both incidence rates and the effectiveness of predictive models. Therefore, the current study addresses this gap by applying ARIMA and SARIMA models to publicly available case reports in Jakarta, aiming to assess whether victims' education levels correlate with and improve the accuracy of time series forecasts of cases. This localized and education-sensitive approach offers a novel contribution to predictive violence modeling in Southeast Asia.

III. RESEARCH METHOD

This research employs the methodology utilized within the ARIMA and SARIMA model frameworks The framework consists of five core phases: initial data preparation, selecting an appropriate model, estimating relevant parameters, conducting diagnostic assessments, and evaluating the predictive performance. These research stages are depicted in Figure 1.

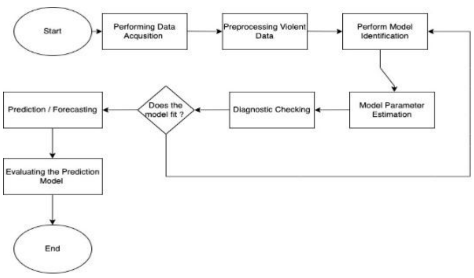


Fig. 1. Phases of The Research Process

A. Gathering of Datadata

The Open Data Jakarta website and the National Commission for the Protection of Women and Children were the two primary sources of data used in this research (https://satudata.jakarta.go.id/data/korban-pppa) The dataset encompasses the period from January to December 2024, covering a full 12 months. These sources offer thorough and trustworthy information on recorded cases of violence against women and children in Jakarta, ensuring the authenticity and applicability of the data for trend analysis and predictive modeling. The features or variables included in the dataset are data_periode, education_level, and total_victims. A total of 216 records were collected. Figure 2 contains the specifics of the victims' violence dataset.

	periode_data	pendidikan	jumlah
0	202406	NaN	72
1	202406	Perguruan Tinggi	21
2	202406	SD / Sederajat	16
3	202406	SMA / Sederajat	26
4	202406	SMP / Sederajat	4

Fig. 2. Victims Violent Dataset

B. Preprocessing

Preprocessing the victim violence data is the next stage following data collection. The purpose of this process is to prepare raw data for use as input in modeling, ultimately resulting in a higher-quality end model [27]. To guarantee data quality and prepare for analysis, the dataset underwent several crucial pre-processing stages. To preserve the integrity of the dataset, missing values were identified and addressed using appropriate imputation techniques. Subsequently, the dataset was normalized by employing one-hot encoding to convert categorical variables, such as types of violence, into numerical values [28]. Finally, the data was normalized to ensure consistent scaling of numeric features, thereby facilitating compatibility with time-series modeling methods such as ARIMA and SARIMA.

C. Identification of The Model

This research outlines the model selection process through a series of structured steps:

1) Plotting the is allowed them to determine whether it was stationar, specifically looking at whether the variance or mean showed stationarity. The time series data was subsequently divided down into a number of smaller parts in order to assess how each part affected the data series as a whole. Two models are typically used: Both multiplicative and additive decomposition.

2) It is possible to utilize the Augmented Dickey-Fuller Test to utilized alongside data visualization to assess whether the test statistic falls below the critical value in a stationary dataset [29]. To assess whether the time series data maintains consistent statistical properties over time, we examine the p-value. A result below 0.05 suggests stability in the data pattern, whereas a higher value implies that the data exhibits changing behavior over time. When such inconsistency is present, differencing can be used to normalize fluctuations in the average or variability. The process of calculating the shift or variation in observation values is referred to as differencing [30]. The acquired difference is rechecked to determine if it remains consistent. When data is stationary, it indicates that there is neither growth nor decline. As a result, regardless of time or oscillation variance, data variations are either continuously steady or center around a constant average value. [31]. Equation (1) describes the differencing equation

$$X'_t = X_t - X_{t-1} \quad (1)$$

Where :
X_t': First differencing
X_t : X value at order t
X_{t-1} : X value at order *t-1*

(3) Plots illustrating overall and lag-specific correlations serve as vital instruments in time series analysis for determining the appropriate model once the data has been made stationary. This phase involves selecting the model order by analyzing the ACF and PACF plots, particularly in the context of seasonally patterned data.

D. Evaluating Model Parameters

This phase involves determining the values of Moving Average, Autoregressive, as well as seasonal and non-seasonal parameters, followed by assessing their statistical significance. A model is considered to have failed the test if any of its parameters are not significant. To develop a model with relevant parameters, any unnecessary parameters will be removed.

E. Examining the Diagnosis

In this study, diagnostic checking establishes whether the model is appropriate and practical for forecasting. The features of the suitable conjectural model should resemble those of the original data. This assessment is conducted through a utilizing model diagnostics, a standard distribution test and a white noise diagnostic test. According to the optimal model, the residuals produced should exhibit characteristics of white noise or conform to a normal distribution.

F. Forecasting

Two models—ARIMA and SARIMA—are used in this study's prediction stage. A method for time series analysis called the ARIMA model makes use of autocorrelation and the fluctuation of time series residuals. The structure of the ARIMA model comprises three components: Autoregressive (AR), Moving Average (MA), and Integrated (I) models. The Integrated component indicates the order of differencing required to transform non-stationary data into a stationary series. Equation (2) presents the standard structure of the ARIMA model.

$$\Phi_p(B)\nabla^d Y_t = \xi + \Theta_q(B)\varepsilon_t \quad (2)$$

Where :

Φ_p : Autoregressive parameters
B : Backward sliding operator
d : Differencing Parameter
Y_t : Observation value at time t
ξ : Constant parameters
Θ_q : Moving average parameters
ε_t : Residual value (*error*

The SARIMA model, on the other hand, estimates future variables and identifies patterns in historical data by utilizing time-series data. The SARIMA framework is defined by the notation (p, d, q)(P, D, Q)[s], where 's' denotes the seasonal cycle length, and all parameters—p, d, q, P, D, and Q—are expressed as whole numbers. Equation (3) illustrates how this equation reflects the SARIMA model in its general version.

$$\Phi_p(B^s)\Phi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (3)$$

Where :

$\Phi_p(B)$: Non-seasonal autoregressive level
$\Phi_p(B_s)$: Seasonal autoregressive rate
$(1-B)_d$: Non-seasonal differencing level
$(1-B_s)_D$: Seasonal differencing level
$\Theta_q(B)$: *Non-seasonal moving average*
$\Theta_q(B_s)$: *Seasonal moving average*
Y_t : Actual data t-th
ε_t : t-period error

The ARIMA and SARIMA techniques were selected for this study partly because they have been extensively utilized and researched across various domains, including comparisons of the ARIMA and SARIMA algorithms for machine learning-based predictions. As per the study's findings, the sea level rise prediction models ARIMA and SARIMA exhibit exceptional performance, attaining a noteworthy forecast accuracy with a lower confidence level of 92.78%[32]. A later phase of the study involved utilizing ARIMA and SARIMA approaches to estimate the number of domestic passengers departing from Tanjung Perak Port. According to the findings, the SARIMA method was found to be the better strategy for this forecasting, while the analysis employing the ARIMA method produced a lower accuracy rating of 16.15%. [33]. Another study explored the performance of ARIMA and SARIMA models in forecasting crude oil prices through comparative analysis. The evaluation results show that both ARIMA and SARIMA have RMSE values of 1.905 Over the upcoming seven-day period, ARIMA forecasts a price of 86.230003, slightly outperforming SARIMA's estimate of 86.260002. The findings of this study are intended to assist policymakers in making informed decisions regarding the utilization of crude oil [34].

A notable feature that sets ARIMA and SARIMA apart from many other prediction methods is that ARIMA can produce forecasts independently, without incorporating outside influencing factors [35]. ARIMA aims to establish a robust statistical correlation between a variable's historical values and its expected future values, enabling the model to be utilized for forecasting. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was developed as an extension of the ARIMA framework. SARIMA is designed to analyze seasonal or recurrent data patterns at specified intervals, such as quarterly, semi-annually, and annually. By utilizing historical data, this predictive model seeks to establish relationships between expected variables to generate forecasts. Building upon earlier studies and leveraging the advantages of these techniques, this research utilizes ARIMA and SARIMA to estimate cases of violence involving women and children in Jakarta, with educational level considered as a contributing element.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

G. Analysis of The Findings

Analyze the forecasting operations' outcomes. The number of victims of violence against women and children in Jakarta will be predicted by a graph that displays the forecasting findings system effectiveness is evaluated using performance indicators like Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) to determine if enhancements are required to better align with user needs.. The MSE calculation is detailed in Equation (4).

$$MSE = \sum_{t=1}^{n} \frac{(A_t - F_t)}{n} \quad (4)$$

The following formula, meanwhile, can be used to get the Root Mean Square Error (RMSE): (5) .

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(A_t - F_t)^2}{n}} \quad (5)$$

Where:
A_t : Original data value
F_t : Forecasting data value
n : Amount of data

Mean Squared Error quantifies prediction accuracy by averaging the squares of the deviations between estimated outcomes and observed values. Typically, the MSE is used to assess the accuracy of forecasting models. However, the error rate RMSE is frequently used as a benchmark metric to gauge the reliability of prediction outcomes. Predictions are considered more accurate when the values of MSE or RMSE are smaller or approach zero. Perfect predictions occur when the results equal zero.

IV. Results And Discussion

In this chapter, the findings of the analysis are presented along with their implications for violence against women and children in Jakarta. The results are organized to provide a comprehensive understanding of the patterns observed in the dataset, their relationship to educational attainment, and the effectiveness of the ARIMA and SARIMA models in predicting outcomes. Each section highlights the efficiency of the forecasting techniques, significant trends, and seasonal fluctuations, integrating quantitative data with contextual observations.

A. Data Acquisition

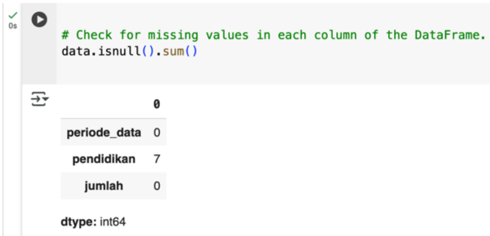
The research dataset contains 72 records collected between January and December 2024, covering a 12-month period. It includes three main columns: which indicates the monthly time frame of the data; which classifies the victims' educational attainment (e.g., primary, secondary, or higher education) to investigate its connection to vulnerability to violence; and shows the number of victims reported for each education level in the corresponding month.

This information, which has been obtained from the National Commission for the Protection of Women and Children and Open Data Jakarta, provides a strong basis for trend analysis and predictive modeling.

B. Preprocessing

1) Missing Value Handling

The dataset on women and children who have been victims of violence in Jakarta must undergo preprocessing after data collection. Missing values are identified and removed as part of this process. Figure 3 below illustrates the results of the missing value analysis.



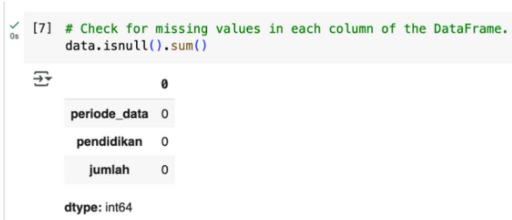
```
# Check for missing values in each column of the DataFrame.
data.isnull().sum()
```

	0
periode_data	0
pendidikan	7
jumlah	0

dtype: int64

Fig. 3. Findings from the dataset's missing value check

The education_level column contains seven rows with missing values, as indicated by a missing value check. To address this issue, the rows with missing values will be removed from the dataset. This approach eliminates potential biases or inaccuracies that could arise from using incomplete data, ensuring that only complete and accurate information is utilized in the study. Although this results in a proportional decrease in the dataset size, it helps maintain the overall quality and reliability of the data for subsequent modeling and analysis. As a result, an effective way to handle the missing data is by eliminating the rows in which they appear; The outcome of this process are illustrated in Figure 4.



```
[7] # Check for missing values in each column of the DataFrame.
data.isnull().sum()
```

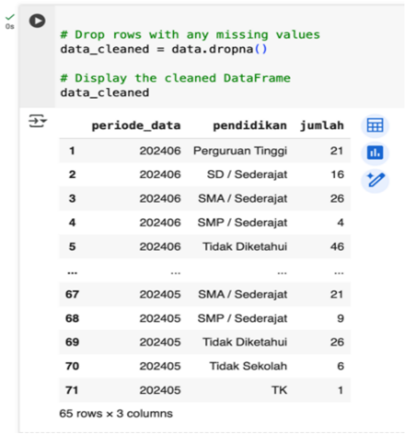
	0
periode_data	0
pendidikan	0
jumlah	0

dtype: int64

Fig. 4. Results of handling missing value dataset

In order to address the issue of missing values in 65 records, data in empty rows or columns (null values) is deleted. This process reduces the overall amount of data. Figure 5 illustrates the specifics of the data frame after the missing values have been addressed.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025



```
# Drop rows with any missing values
data_cleaned = data.dropna()

# Display the cleaned DataFrame
data_cleaned
```

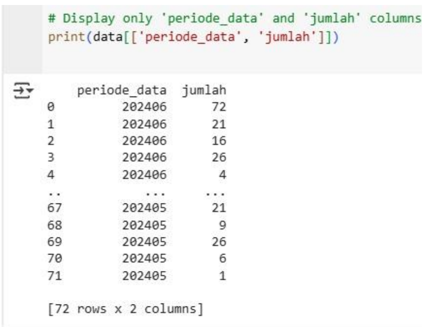
	periode_data	pendidikan	jumlah
1	202406	Perguruan Tinggi	21
2	202406	SD / Sederajat	16
3	202406	SMA / Sederajat	26
4	202406	SMP / Sederajat	4
5	202406	Tidak Diketahui	46
...	...	...	...
67	202405	SMA / Sederajat	21
68	202405	SMP / Sederajat	9
69	202405	Tidak Diketahui	26
70	202405	Tidak Sekolah	6
71	202405	TK	1

65 rows × 3 columns

Fig. 5. Results of treating a dataset with missing values

2) Data Integration

During the data integration phase, the time-series prediction model was constructed using the `data_period` and `total_victims` columns. The time dimension, represented by `data_period`, is crucial for establishing a timeline and identifying trends over time. The number of victims in each period serves as the target variable, provided by `total_victims`, and is essential for predicting future occurrences. The dataset is streamlined for time-series modeling by focusing on these two key columns, ensuring that the ARIMA and SARIMA models can effectively detect trends and seasonal variations in the data.



```
# Display only 'periode_data' and 'jumlah' columns
print(data[['periode_data', 'jumlah']])

    periode_data  jumlah
0         202406      72
1         202406      21
2         202406      16
3         202406      26
4         202406       4
..           ...     ...
67        202405      21
68        202405       9
69        202405      26
70        202405       6
71        202405       1

[72 rows x 2 columns]
```

Fig. 6. Outcomes of the dataset following the data integration phase

C. *Model Identification*

The model identification stage will involve several procedures, including creating time-series charts, utilizing the ADF procedure to test for stationarity test, and producing Autocorrelation Function and Partial Autocorrelation Function plots. The objectives of this analysis are to identify the underlying patterns in the data, evaluate stationarity, and determine the appropriate sequence for differencing and model parameters.

1) Create a time series plot

The data_period, month, and total_victims columns were utilized to create a time-series figure that illustrates patterns in violence cases against women and children over time. The data_period column presents the data in chronological order, the month column assists in identifying recurring trends, and the total_victims column indicates the total number of incidents recorded during each period. The visualization suggests potential seasonal trends, highlighting variations in the number of victims over several months, with notable peaks at certain times. This insight is crucial for understanding periodic fluctuations and will enhance the accuracy of predictive models such as ARIMA and SARIMA.

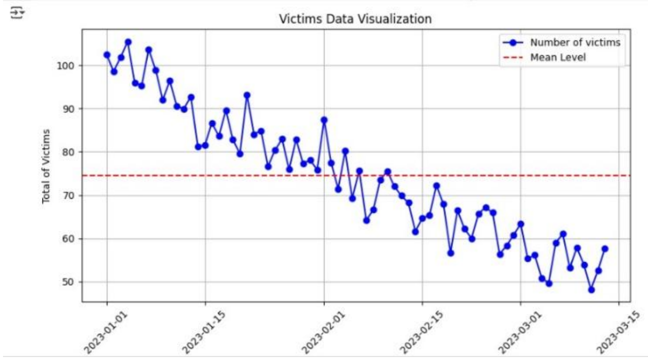


Fig. 7. Time series plot using the number of victims dataset

2) Time series stationarity check using ADFf

This study employed the Augmented Dickey-Fuller test to determine whether the time-series data is stationary, which is a crucial prerequisite for constructing reliable ARIMA and SARIMA models. To assess the stationarity of the korban (victim_count) time-series data, the Augmented Dickey-Fuller (ADF) test was applied. At the 1% (-3.526), 5% (-2.903), and 10% (-2.589) significance levels, the test result of -9.351 is significantly below the critical values. Furthermore, the p-value (8.29e-16) is considerably lower than the conventional limit of 0.05, offering robust support for rejecting the presence of a unit root. According to these findings, the time-series data is already stationary, indicating that further differencing is unnecessary to stabilize the trend. Given that stationarity is an essential condition for ARIMA and SARIMA modeling, these results confirm that the dataset is suitable for predictive modeling without any modifications. The results of the Augmented Dickey-Fuller (ADF) test can be found in Table 1, with its visualization presented in Figure 8.

TABLE I. STATIONARITY ASSESSMENT VIA ADF METHOD

Metric	Values
Test Statistics	-9.351093
p-value	8.291883e-16
Lags Used	0
Number of Observations Used	71
Critical Value (1%)	-3.526005
Critical Value (5%)	-2.903200
Critical Value (10%)	-2.588995

3) Correlation Analysis and Lag-based Correlation Analysis

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

After ensuring that since the data exhibits stationarity, plots of correlation and partial correlation over lags are generated. Seasonal patterns in the data are identified by determining the appropriate order of the model.

a) Autocorrelation Function

The order of the Moving Average (MA) model is determined using the Autocorrelation Function (ACF) plot. The purpose of this study is to assess whether the data is stationary in terms of its mean. The autocorrelation plot reveals a strong serial dependence, with autocorrelation values exceeding 0.5 for the first seven lags and progressively declining thereafter. Several lags fall outside the confidence interval, indicating a non-random pattern likely attributable to trends or seasonality. The results of the autocorrelation analysis are illustrated in Figure 9 below.

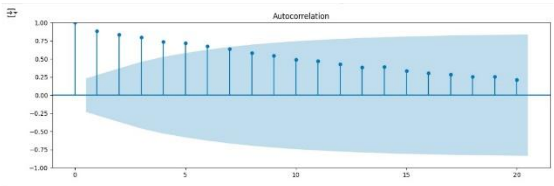


Fig. 8. Results of ACF Plots

b) Partial Autocorrelation Function

The Partial Autocorrelation Function (PACF) plot is utilized to identify the appropriate Autoregressive (AR) model. In Figure 10, the PACF reveals significant associations at lags 1 and 2 (-0.9 and 0.75, respectively), while all other lags fall within the confidence interval. This suggests that the time series primarily exhibits short-term dependence up to lag 2, after accounting for the effects of earlier lags. Given the steep decline observed after lag 2, these data may be best modeled by an Autoregressive model of order 2 (AR(2)). Higher-order AR terms may not be necessary, as the absence of substantial correlations beyond lag 2 indicates a lack of long-term dependencies.

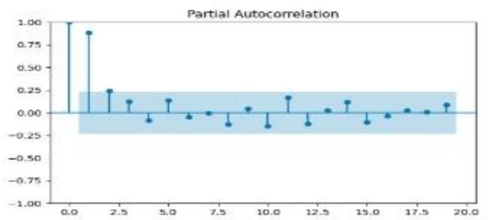
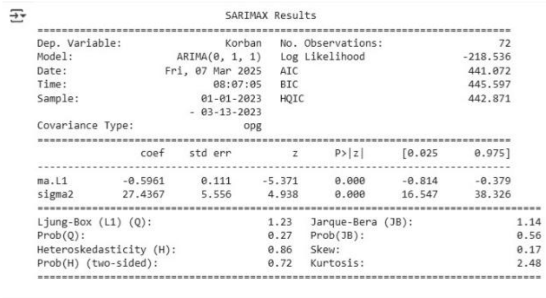


Fig. 9. Results of PACF Plots

D. Model Parameter Estimation

Model components encompassing periodic effects, regular patterns, error smoothing (MA), and lagged dependencies (AR) are considered all estimated at this stage, and their respective significance is evaluated. A model is deemed to fail the test if its parameters are not statistically significant. To assess and compare statistical models, researchers often rely on the Akaike and Bayesian Information Criteria (AIC and BIC) selecting ARIMA models and determining model parameters for both ARIMA and SARIMA. Below are the parameter estimations for the SARIMA model.

Figure 11 below presents the results of the ARIMA model significance test. An ARIMA (0,1,1) model was applied to the dataset, which consists of 72 observations. The results indicate a p-value of 0.000 and a moving average (MA) coefficient of -0.5961 at lag 1, demonstrating statistical significance. The variance of the residuals is represented by the sigma² value, which is 27.4367. Model fit is assessed using the model selection criteria: AIC (445.597), BIC (445.597), and HQIC (442.871). Diagnostic tests reveal that with a Jarque-Bera p-value of 0.72, the residuals exhibit characteristics consistent with a normal distribution. These findings suggest that while the model effectively captures short-term interdependence, it lacks an autoregressive component and is well-fitted.



```
                               SARIMAX Results
==============================================================================
Dep. Variable:                 Korban   No. Observations:                   72
Model:                 ARIMA(0, 1, 1)   Log Likelihood                -218.536
Date:                Fri, 07 Mar 2025   AIC                            441.072
Time:                        08:07:05   BIC                            445.597
Sample:                    01-01-2023   HQIC                           442.871
                         - 03-13-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5961      0.111     -5.371      0.000      -0.814      -0.379
sigma2        27.4367      5.556      4.938      0.000      16.547      38.326
===================================================================================
Ljung-Box (L1) (Q):                   1.23   Jarque-Bera (JB):                 1.14
Prob(Q):                              0.27   Prob(JB):                         0.56
Heteroskedasticity (H):               0.86   Skew:                             0.17
Prob(H) (two-sided):                  0.72   Kurtosis:                         2.48
===================================================================================
```

Fig. 10. Estimated Parameters of the ARIMA Model

The significant test results for the Sarima model are displayed in Figure 20 below. A Seasonal ARIMA (SARIMA) framework, specifically the SARIMAX (2,1,1)x(2,1,1,12), was applied to a dataset comprising 72 observations. The autoregressive (AR) and seasonal autoregressive (AR.S) terms were found to be insignificant (p-values < 0.05); however, the moving average (MA) and seasonal moving average (MA.S) terms were significant, indicating that the MA components are more effective at capturing short-term dependencies. Model diagnostics reveal that the residuals are normally distributed (Jarque-Bera p-value = 0.33) and show no meaningful evidence of correlation across lags, as indicated by a Ljung-Box p-value of 0.59. Overall, the model fits the data well, although there may be opportunities for improvement by reassessing the AR factors.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

```
self._init_dates(dates, freq)
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                            Korban   No. Observations:                   72
Model:             SARIMAX(2, 1, 1)x(2, 1, 1, 12)   Log Likelihood                -182.323
Date:                           Sat, 08 Mar 2025   AIC                            378.646
Time:                                   04:05:32   BIC                            393.188
Sample:                               01-01-2023   HQIC                           384.322
                                    - 03-13-2023
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1786      0.191     -0.937      0.349      -0.552       0.195
ar.L2         -0.2409      0.180     -1.337      0.181      -0.594       0.112
ma.L1         -0.7672      0.141     -5.454      0.000      -1.043      -0.492
ar.S.L12      -0.2989      0.869     -0.344      0.731      -2.003       1.405
ar.S.L24      -0.0995      0.511     -0.195      0.846      -1.102       0.903
ma.S.L12      -0.4262      0.964     -0.442      0.659      -2.316       1.464
sigma2        24.5356      5.302      4.628      0.000      14.144      34.927
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                 2.20
Prob(Q):                              0.59   Prob(JB):                         0.33
Heteroskedasticity (H):               0.48   Skew:                             0.47
Prob(H) (two-sided):                  0.10   Kurtosis:                         2.88
===================================================================================
```

Fig. 11. Estimated Parameters of the SARIMA Model

E. *Examining The Diagnosis*

Diagnostic verification follows the estimation of parameters for the ARIMA and SARIMA models to determine their appropriateness and effectiveness in forecasting. The results of the diagnostic testing on the ARIMA model, which includes an examination of density plots and residuals, are presented in Fig. 13. The histogram of the residuals reveals a right-skewed distribution, indicating that, rather than being perfectly normally distributed, the residuals are primarily centered around zero, with a few significant outliers. Since all lags fall within the confidence intervals, The residual ACF plot reveals no significant correlation patterns over time. This observation is reinforced by the results of the Ljung-Box test, which produces high p-values of 0.998861 and 0.999984, indicating a lack of serial dependence, suggesting that the residuals are uncorrelated and that the model effectively captures the time series patterns. The Ljung-Box statistical check was conducted to evaluate whether residuals show correlation across lags, and the findings suggest the model's adequacy in capturing time-dependent structure is well-specified, as the residuals behave like white noise.

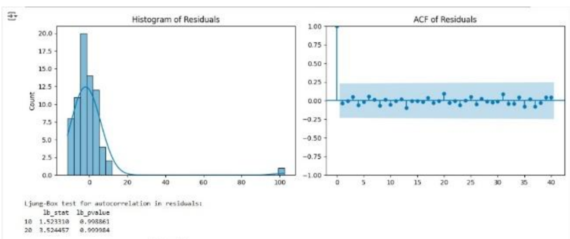


Fig. 12. Diagnostic Evaluation Outcomes for the ARIMA Model

Furthermore, diagnostic testing has been conducted on the SARIMA model. Potential outliers may be present, as the histogram of the residuals exhibits a right-skewed distribution, with most residuals concentrated near zero and a few extreme values. Since the majority of lag values fall remains inside the margin of estimation, the ACF plot of the residuals shows no significant autocorrelation.

According to the results of the Ljung-Box test, the lb_stat values at lags 10 and 20 are 2.283699 and 17.713411, respectively, with p-values of 0.993667 and 0.606281. There is no significant autocorrelation in the residuals, since both p-values exceed the threshold, the null hypothesis remains valid. This indicates that our SARIMA model is suitable for forecasting, as it accurately represents the time series structure and the residuals behave like white noise .

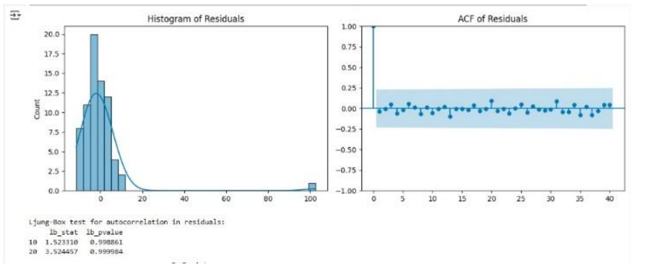


Fig. 13. SARIMA Model Diagnostics Checking Results

F. *Forecasting*

Following diagnostic testing, the forecast for each model is produced by contrasting the expected and actual values from the time series data. Forecasts become more accurate as a result of this process. Prediction results are obtained by applying ARIMA and SARIMA models to a reserved portion of the dataset designated for validation. To evaluate the correspondence between the actual and expected data, predictions are formulated. Figures 15 and 16, illustrating the forecasts generated by the ARIMA and SARIMA models respectively, demonstrate that the fluctuations in the predicted values closely resemble those observed in the recorded dataset.

The study on ARIMA-based predictions for child sexual abuse cases is illustrated in the graph, analyzing predicted values relative to empirical records, predicted cases, and the confidence interval. Initially, the training data and actual cases closely align in early 2025, with cases starting at nearly zero and progressively increasing. However, the actual data exhibits a more pronounced upward trend and begins to diverge from the forecast around March 2025. The ARIMA model predicts approximately 60 instances by June 2025, which underestimates the true trend, as actual cases exceed 80. As the confidence interval widens, it reflects increasing uncertainty, indicating that future projections may lose accuracy over time.

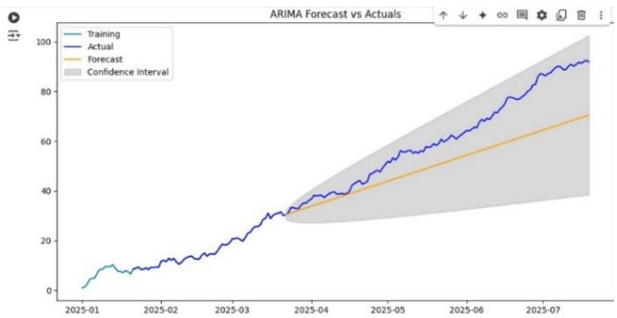


Fig. 14. ARIMA Model Training Data Forecasting Results

The chart provides a comparative analysis of training data,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

actual data, and forecasted values to evaluate the predictive capability of the SARIMA (2,1,1) (1,1,1,12) model in estimating cases of child sexual violence. The actual data represents the real occurrences of such cases, while the training data comprises historical events used to develop the SARIMA model. Initially, from January to March 2025, the training data and actual instances align closely, showing a steady increase from 0 to approximately 30 cases. Beginning in April 2025, when the model commences its predictions, the forecasted values closely approximate the actual data, suggesting that the SARIMA model with parameters (2,1,1)(1,1,1,12) successfully models the underlying trend along with the recurring seasonal patterns in the data.

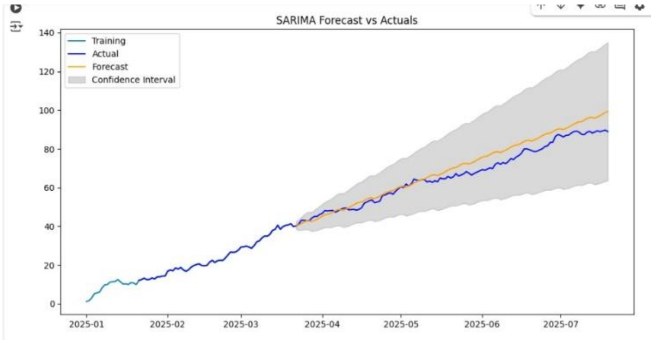


Fig. 15. SARIMA Model Training Data Forecasting Results

The actual instances show slight variations but usually fall within the predicted confidence interval when examined more closely. The prognosis somewhat overestimated at about 100 instances, but actual cases increase significantly from May to July 2025, reaching about 90 by that time. This implies that while the SARIMA model includes the general increasing trend, it significantly exaggerates the anticipated rise. In comparison to normal ARIMA, the model's order parameters (2,1,1,12) better account for seasonal impacts and short-term shocks, guaranteeing greater adaptability to periodic oscillations.

G. *Prediction Model Evaluation*

Following the prediction, the final step is to assess the outcomes of the forecasting process. Root Mean Square Error and Mean Squared Error calculations are used to test the model. The Root Mean Square Error (RMSE) serves as a metric to quantify the discrepancy between predicted outcomes and actual observations. When the RMSE score is minimal, it suggests strong agreement between the projected figures and the real-world measurements. Mean Square Error is a forecasting metric used to assess how accurate the forecasting results are. The accuracy of the forecasting findings increases with a smaller MSE value.

The prediction model's performance across several error measures is shown by the evaluation results. The model's predictions often deviate from the actual values by about 32.23 units, according to the Root Mean Square Error (RMSE) of 32.23; bigger errors are penalized more severely. The average absolute difference between the predicted and actual values is around 32.09 units, as indicated by the Mean Absolute Error (MAE) of 32.09. Mean Absolute Percentage (MAPE) of 5.19%, on the other hand, suggests that the model's predictions perform quite well, with an average error of 5.19% about the actual values. Table 2, which follows, displays the specifics of the model evaluation results.

TABLE II. EVALUATION OF THE PREDICTION MODELS OF SARIMA MODELS, AND THE ARIMA

Method	ARIMA	SARIMA
RMSE	32.22	80.26
MAE	32.09	66.21
MAPE	5.19	-

V. CONCLUSIONS

According to modeling conducted to forecast the incidence of violence against women and children in Jakarta. The models that performed the best were ARIMA (0,1,1) and SARIMA (0,1,1) (2,2,1) 12. The evaluation's findings indicate that the ARIMA model predicts violence against women and children in Jakarta more accurately than the SARIMA model. The ARIMA model demonstrates greater prediction accuracy and consistency, as evidenced by its significantly lower RMSE (32.22) and MAE (32.09) compared to SARIMA's RMSE of 80.26 and MAE of 66.21. Furthermore, the ARIMA model's MAPE of 5.19% reflects a comparatively low percentage error, while the SARIMA model's MAPE could not be calculated, possibly due to the presence of zero or negative actual values in the data. According to these results, a simpler ARIMA model performs better on this forecasting task than its complex seasonal equivalent. Hybrid modeling approaches, including merging ARIMA with machine learning techniques or adding external variables, like socioeconomic indicators or law enforcement data, are suggested for further research, as they may increase prediction power

Furthermore, a more thorough seasonality analysis might help in improving SARIMA combinations' effectiveness. Effectiveness can also be enhanced by incorporating advanced techniques such as time series decomposition or dynamic regression models, which allow for a more nuanced insight into the hidden structures within the dataset. By exploring these methods, researchers may uncover additional insights that can lead to more accurate and reliable forecasts.

REFERENCES

[1] M. Myall, S. Morgan, and S. Scott, “Editorial: Domestic violence and abuse: increasing global and intersectional understanding,” *Frontiers in Health Services*, vol. 4, Sep. 2024, doi: 10.3389/frhs.2024.1465688.

[2] A. Bluschke, N. Faedda, J. Friedrich, and E. J. Dommett, “Editorial: Women in psychiatry 2023: ADHD,” *Front Psychiatry*, vol. 15, Aug. 2024, doi: 10.3389/fpsyt.2024.1447958.

[3] Y. F. Wismayanti, P. O'Leary, C. Tilbury, and Y. Tjoe, “Child sexual abuse in Indonesia: A systematic review of literature, law and policy,”

Child Abuse Negl, vol. 95, p. 104034, Sep. 2019, doi: 10.1016/j.chiabu.2019.104034.

[4] A. L. Wirtz *et al.*, "Development of a screening tool to identify female survivors of gender-based violence in a humanitarian setting: qualitative evidence from research among refugees in Ethiopia," *Confl Health*, vol. 7, no. 1, p. 13, Dec. 2013, doi: 10.1186/1752-1505-7-13.

[5] C. Stoicescu, B. Medley, E. Wu, N. El-Bassel, P. Tanjung, and L. Gilbert, "Synergistic effects of exposure to multiple types of violence on non-fatal drug overdose among women who inject drugs in Indonesia," *International Journal of Drug Policy*, vol. 129, p. 104486, Jul. 2024, doi: 10.1016/j.drugpo.2024.104486.

[6] N. Lepcha and S. Paul, "Exploring Violence Against Children Under Sustainable Development Goals," 2021, pp. 286–296. doi: 10.1007/978-3-319-95687-9_72.

[7] Ore-ofe Loveth Oluwajobi, Chidinma Favour Udechukwu, and Toluwanimi Oreoluwa Arogundade, "Understanding the impact of domestic violence on children's mental health and exploring effective intervention strategies," *World Journal of Advanced Research and Reviews*, vol. 23, no. 3, pp. 1405–1418, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2812.

[8] Riswanda, J. McIntyre-Mills, and Y. Corcoran-Nantes, "Prostitution and Human Rights in Indonesia: A Critical Systemic Review of Policy Discourses and Scenarios," *Syst Pract Action Res*, vol. 30, no. 3, pp. 213–237, Jun. 2017, doi: 10.1007/s11213-016-9393-4.

[9] G. Dhamija, P. Roychowdhury, and B. Shankar, "Does urbanization empower women? Evidence from India," *J Popul Econ*, vol. 38, no. 1, p. 27, Mar. 2025, doi: 10.1007/s00148-025-01085-4.

[10] A. Al Mutair *et al.*, "Domestic violence and childhood trauma among married women using machine learning approach: a cross-sectional study," *BMC Public Health*, vol. 25, no. 1, p. 1340, Apr. 2025, doi: 10.1186/s12889-025-22537-2.

[11] D. Rukmana and D. Ramadhani, "Income Inequality and Socioeconomic Segregation in Jakarta," 2021, pp. 135–152. doi: 10.1007/978-3-030-64569-4_7.

[12] O. O. Okedare, M. M. Salawu, and O. I. Fawole, "Intimate partner violence and quality of life of young women in urban slum and non-slum communities, Ibadan, Nigeria," *BMC Public Health*, vol. 25, no. 1, p. 1199, Mar. 2025, doi: 10.1186/s12889-025-22385-0.

[13] Md. R. Kabir, S. Ghosh, and A. Shawly, "Causes of Early Marriage and Its Effect on Reproductive Health of Young Mothers in Bangladesh," *Am J Appl Sci*, vol. 16, no. 9, pp. 289–297, Sep. 2019, doi: 10.3844/ajassp.2019.289.297.

[14] D. Tunas and A. Peresthu, "The self-help housing in Indonesia: The only option for the poor?," *Habitat Int*, vol. 34, no. 3, pp. 315–322, Jul. 2010, doi: 10.1016/j.habitatint.2009.11.007.

[15] Mr. B. R. A. -, "Harnessing Technology and Data Analytics to Advance Prevention and Treatment in the Opioid Crisis," *International Journal For Multidisciplinary Research*, vol. 5, no. 5, Oct. 2023, doi: 10.36948/ijfmr.2023.v05i05.31737.

[16] Y. Cao *et al.*, "Recognize Human Activities from Partially Observed Videos," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2013, pp. 2658–2665. doi: 10.1109/CVPR.2013.343.

[17] J. W. Petty, "Research in Small-Firm Entrepreneurial Finance: A Note on Developing a Paradigm," *The Journal of Entrepreneurial Finance*, vol. 1, no. 1, pp. 88–90, Dec. 1991, doi: 10.57229/2373-1761.1114.

[18] A. A. Fagan *et al.*, "Scaling up Evidence-Based Interventions in US Public Systems to Prevent Behavioral Health Problems: Challenges and Opportunities," *Prevention Science*, vol. 20, no. 8, pp. 1147–1168, Nov. 2019, doi: 10.1007/s11121-019-01048-8.

[19] G. M. Campedelli, A. Aziani, and S. Favarin, "Exploring the Immediate Effects of COVID-19 Containment Policies on Crime: an Empirical Analysis of the Short-Term Aftermath in Los Angeles," *American Journal of Criminal Justice*, vol. 46, no. 5, pp. 704–727, Oct. 2021, 10.1007/s12103-020-09578-6.

[20] H. Seyidoglu, G. Farrell, A. Dixon, J. Pina-Sánchez, and N. Malleson, "Post-pandemic crime trends in England and Wales," *Crime Sci*, vol. 13, no. 1, p. 6, Mar. 2024, doi: 10.1186/s40163-024-00201-1.

[21] S. Yao *et al.*, "Predicting Land Use Changes under Shared Socioeconomic Pathway–Representative Concentration Pathway Scenarios to Support Sustainable Planning in High-Density Urban Areas: A Case Study of Hangzhou, Southeastern China," *Buildings*, vol. 14, no. 7, p. 2165, Jul. 2024, doi: 10.3390/buildings14072165.

[22] C. V. Redoblo, J. L. G. Redoblo, R. A. Salmingo, C. M. Padilla, and J. C. T. Arroyo, "Forecasting the influx of crime cases using seasonal autoregressive integrated moving average model," *International Journal of ADVANCED AND APPLIED SCIENCES*, vol. 10, no. 8, pp. 158–165, Aug. 2023, doi: 10.21833/ijaas.2023.08.018.

[23] S. Siamba, A. Otieno, and J. Koech, "Application of ARIMA, and hybrid ARIMA Models in predicting and forecasting tuberculosis incidences among children in Homa Bay and Turkana Counties, Kenya," *PLOS Digital Health*, vol. 2, no. 2, p. e0000084, Feb. 2023, doi: 10.1371/journal.pdig.0000084.

[24] M. K. Lubeya *et al.*, "Using the ARIMA Model to forecast sexual and gender-based violence cases reported to a tertiary hospital in Lusaka, Zambia," *PAMJ - One Health*, vol. 5, 2021, doi: 10.11604/pamj-oh.2021.5.4.27590.

[25] A. Anavatan and E. Y. Kayacan, "Investigation of femicide in Turkey: modeling time series of counts," *Qual Quant*, vol. 58, no. 3, pp. 2013–2028, Jun. 2024, doi: 10.1007/s11135-023-01619-6.

[26] S. Rashid, "Impact of COVID-19 on Selected Criminal Activities in Dhaka, Bangladesh," *Asian J Criminol*, vol. 16, no. 1, pp. 5–17, Mar. 2021, doi: 10.1007/s11417-020-09341-0.

[27] A. Luong *et al.*, "Comparison of Machine Learning Models to a Novel Score in the Identification of Patients at Low Risk for Diabetic Retinopathy," *Ophthalmology Science*, vol. 5, no. 1, p. 100592, Jan. 2025, doi: 10.1016/j.xops.2024.100592.

[28] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y.-I. Cho, "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, p. 2553, Aug. 2024, doi: 10.3390/math12162553.

[29] Z. Liang and M. T. Ismail, "Advanced CEEMD hybrid model for VIX forecasting: optimized decision trees and ARIMA integration," *Evol Intell*, vol. 18, no. 1, p. 12, Feb. 2025, doi: 10.1007/s12065-024-00984-x.

[30] Y. Shen *et al.*, "Near real-time corn and soybean mapping at field-scale by blending crop phenometrics with growth magnitude from multiple temporal and spatial satellite observations," *Remote Sens Environ*, vol. 318, p. 114605, Mar. 2025, doi: 10.1016/j.rse.2025.114605.

[31] R. Refinetti, "Non-stationary time series and the robustness of circadian rhythms," *J Theor Biol*, vol. 227, no. 4, pp. 571–581, Apr. 2004, doi: 10.1016/j.jtbi.2003.11.032.

[32] K. M. Wantzen, K.-O. Rothhaupt, M. Mörtl, M. Cantonati, L. G.-Tóth, and P. Fischer, "Ecological effects of water-level fluctuations in lakes: an urgent issue," *Hydrobiologia*, vol. 613, no. 1, pp. 1–4, Nov. 2008, doi: 10.1007/s10750-008-9466-1.

[33] S. Putri and A. Sofro, "Peramalan Jumlah Keberangkatan Penumpang Pelayaran Dalam Negeri di Pelabuhan Tanjung Perak Menggunakan Metode ARIMA dan SARIMA," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 10, no. 1, pp. 61–67, Apr. 2022, doi: 10.26740/mathunesa.v10n1.p61-67.

[34] V. P. Ariyanti and Tristyanti Yusnitasari, "Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 405–413, Mar. 2023, doi: 10.29207/resti.v7i2.4895.

[35] Elvina Catria, A. A. Putra, D. Permana, and D. Fitria, "Adding Exogenous Variable in Forming ARIMAX Model to Predict Export Load Goods in Tanjung Priok Port," *UNP Journal of Statistics and Data Science*, vol. 1, no. 1, pp. 31–38, Feb. 2023, doi: 10.24036/ujsds/vol1-iss1/10.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025