Using A Time Series Forecasting Model, Predict Maritime Crime

Syachrul Arief[1], Soufi Jayanti Ningsih[2]

[1]Geospatial Information Authority of Indonesia, Jl. Raya Jakarta Bogor Km 46 Cibinong Bogor, Indonesia.

[2]Republic Indonesia of Defence University, Jl. Raya Jakarta Bogor Km 46 Cibinong Bogor, Indonesia.

Corresponding Author: syachroel@gmail.com

Abstract: - Indonesia is very disadvantaged by the many illegal activities on the sea route. One of the most vulnerable areas in Indonesia is the Riau Islands. This study aims to analyze marine crimes in the Riau Islands Province. For this reason, we ask a research question: How to predict the number of cases of marine crimes in the future along with the magnitude of the risk posed and the necessary handling? To answer, we designed mixed-method research. We conducted a literature study, collected data from agencies that handle sea transportation, customs, patrols, and law enforcement in the Riau Islands maritime area, and analyzed the data obtained. These results are used to fill out AHP forms and interviews to determine the magnitude of the risk posed by each type of marine crime. Finally, we forecast and analyze the results to formulate conclusions and recommendations. In this study, we use the ARIMA model to predict the number of marine crimes incidents in the future and descriptive analysis to determine the resulting risk classification. The findings of this study indicate that in the next year, it is predicted that there will be 91 events with the highest incidence in April, namely 11 events with a relative risk of 6.3. These findings imply that the Riau Islands Region is an area that is vulnerable to severe marine crimes. Therefore, it is recommended to take measures that anticipate the occurrence of violations early, such as the installation of warning sensor equipment and other adequate sensing technologies.

Key Words*: — Analytic Hierarchy Process, Marine Crime, Riau Islands, Time Series Forecasting.*

I. Introduction

Marine crimes, also known as illegal activities carried out by sea, cover all aspects or fields related to the sea that violate applicable laws and criminal law. Marine crimes consist of various types, including shipping, fisheries, piracy, narcotics, etc. This marine crime can pose multiple threats to national defense and security, such as food security, the environment, energy security, economic security, and human security, so the problem of marine crimes is one of the main problems for the government to pay more attention to maritime security, including the obligation to manage natural resources in the sea. [1] .

Manuscript revised December 09, 2022; accepted December 10, 2022. Date of publication December 11, 2022.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

Indonesia loses hundreds of trillions of rupiah every year for this marine crime [2]. One of the vulnerable areas is the Riau Archipelago. The Riau Archipelago is one of Indonesia's provinces with abundant natural resources. It is located in a strategic sea and air transportation traffic lane between the Malacca Strait, South China Sea, and the Karimata Strait. This area is also directly next to three countries: Singapore, Vietnam, and Malaysia, where the economic opportunities increase. This condition also opens up opportunities for crime to arise in the area. Unfortunately, research on adverse risks and threats in the region is still limited.

The government and related parties have tried to overcome marine crimes by implementing various technologies, environmental assessments, and studies of related Indonesia's strategic security. However, these efforts need to be enriched in advance, given the high number of cases of marine crimes in Indonesian waters. For example, in 2018, Polairud found 1054 criminal cases at sea. Then, during 2019, they handled 442 criminal acts at sea and found 617 instances throughout 2020. Therefore, in this study, the researcher proposes one way to expect and control the number of cases of marine crimes. Since such cases occurred in Indonesian waters, mainly in the Riau

Islands region, we focus the research on the zonal subject that occurred by analyzing time series forecasting.

Time series forecasting techniques are famous for predicting several problems and expecting practical actions in the future. Some of them are in building energy consumption [3], analyzing and predicting oil and piracy [4], showing the Covid-19 pandemic [5], predicting sea traffic [6], predicting the specific demand. Moreover, such an analytical approach can project and indicate Military aircraft spare parts [7] and build models to forecast defense spending [8]. However, in previous studies, we see rarely studies discussing predicting the number of marine crime incidents in the future. Thus, this research is expected to assist the related parties in making policies and decisions to deal with marine crimes in Indonesian waters, especially in the Riau Islands region.

II. Materials and methods

The data used in this study is data on the number of cases of marine crimes that occurred in the Riau Islands in the 2018-2021 period, detailed monthly (January 2018-November 2022). Furthermore, the data was analyzed and obtained what types of marine crimes occurred in the Riau Islands. Then, the data acquisition results are used to determine the magnitude of the risk to national defense and security by using the AHP technique and interviewing experts.

Finally, the number of marine crimes and the extent of the threat is predicted using time series forecasting analysis. The purpose of this prediction is to see how large the number of incidents of marine crimes in the Riau Islands will be in the future, along with the magnitude of the risks that are possibly caused to defense and security. In conducting this analysis and forecasting, the tool used is the statistical software XLSTAT. To explore the analytical techniques used in this paper, we need to briefly describe some of the time series techniques and the techniques associated with their testing.

2.1 Time Series

A time series is a sequential collection of data points, usually measured in the parent of time. Time series is mathematically defined as a vector set t), t=0,1,2,⋯ where t represents the elapsed time, and the variable x(t) is a random variable. Measurements made during an event in a time series are arranged in better chronological order [2], [3]. Time series are generally influenced by four main components that can be distinct from the observed data. These components are Trend, Cyclical, Seasonal, and Irregular [4], [5]. Irregular components are also known as residuals. These components are described as follows.

2.1.1Trend

A trend is a general movement that a time series exhibits throughout the observation without considering seasons and deviations. A trend can also be defined as long-term development in a time series (i.e., upwards, downwards, or upwards, downwards, or stagnates) [6].

2.1.2 Seasonal

Seasonality is defined as a repeating pattern in a time series [6].

2.1.3 Cyclical

A cycle is up and down in a time series without a fixed frequency. In contrast to the seasons, the amplitude and duration of the process vary from time to time [6].

2.1.4 Irregular

Some residual values remain after the trends, and cyclic oscillations are calculated and eliminated. These values are sometimes high enough to cover trends and seasons. In this case, the term outlier refers to these residuals, and robust statistics are usually applied to address them. [7].

2.2 Stationarity

One of the essential characteristics of a time series is stationary. Therefore, most statistical forecasting methods assume that the time series is stationary through transformations. Statistical properties (such as mean, variance, auto-correlation) of a stationary time series do not change over time. Therefore, the stationary time series is easier to model and predict. However, in practice, time series usually exhibit a combination of trend and seasonal patterns and are therefore not stationary [8]. For this purpose, time series is transformed, adjusted seasonally, made trend-stationary by eliminating the trend, or made difference-stationary with the possibility of repeated differences (i.e., calculating the difference between successive observations).

2.3 Forecasting a Time Series

Forecasting is the main reason for doing time series analysis, which is the basic idea of trying and using observations from past data to predict future possibilities. The forecasting model is a functional representation that describes the time series [9]. Today, forecasting is an established and essential pillar in many disciplines that require a means to “forecast” the future by examining past observations. Time series forecasting aims to predict how the time series will develop over time. Forecasting accuracy is highly dependent on the forecasting method used. Therefore, selecting the most suitable forecasting method is a crucial part. One of the methods used in forecasting is the ARIMA model.

The ARIMA model is a variant of the ARMA model that relaxes the stationary time series requirements through differentiation. In 1938, H. Wold laid the foundation for using the ARMA model for time series [10]. The ARMA model combines the autoregressive AR(p) model and the moving-average MA(q) model. The latter is represented by 'I' (for integrated) in ARIMAodel. As indicated by the term autoregressive, the observed variable is represented as a linear combination of its past values. The ARIMA method is suitable for short-term forecasting, the results are easy to interpret, and the model is formed more quickly. In contrast to the autoregressive model, the moving-average MA(q) model uses the past forecast error to represent the observed values as a linear combination. The parameters that determine the ARIMA model are order p from the AR model, order q from the MA model, and degree d differencing.

The following are the steps in forecasting using the ARIMA model [11], [12].

a. Check for data stationary. If the data is not stationary, then the data must be stationary by differencing.
b. Parameter estimation and identify the best model.
c. Diagnostic checking.
d. Forecasting.

III. RESULTS AND DISCUSSION

In this study, to determine the ARIMA model (p,d,q), the value of d is first determined by performing a stationarity test on the data. This process is carried out in several stages as follows.

3.1 Testing the data stationarity

3.1.1. Data on the number of incidents (NOI) of marine crime

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Jumlah Kejadian	47	0	47	0,000	21,000	6,404	5,531

Dickey-Fuller test (ADF(stationary) / k: 3 / JML KEJADIAN):

Tau (Observed value)	-2,783
Tau (Critical value)	-3,484
p-value (one-tailed)	0,199
alpha	0,05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.

Fig.1. Unit Root Test Results of NOI.

As a note, "Jumlah Kejadian" is the number of incidents of marine crime. The results of the unit root test carried out obtained a probability value (p-value) of 0.199, which is greater than the significant level of α=0.05, which means that there is not enough evidence to reject the null hypothesis (H0 is accepted), which means that the data has a unit root. Thus, this indicates that the data is not stationary.

3.2 Data on the relative risk index (RRI)

The unit root test results obtained a probability value (p-value) of 0.281, greater than the significant level α=0.05, which means there is not enough evidence to reject the null hypothesis, which means that the data has a unit root. Thus, this indicates that the data is not stationary.

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
RISKRELATIF	47	0	47	0,000	9,369	2,448	2,548

Dickey-Fuller test (ADF(stationary) / k: 3 / RISKRELATIF):

Tau (Observed value)	-2,570
Tau (Critical value)	-3,484
p-value (one-tailed)	0,281
alpha	0,05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.

Fig.2. Unit Root Test Results of RRI

As a note, "Risk Relatif" is the relative risk index of incidents (RRI). It means that the data has a unit root and is not stationary. The unit root test obtained a probability value (p-value) of 0.199, greater than the significant level of α=0.05.

3.3 Differencing Process

If the data obtained is not stationary, then differencing is performed. On the other hand, if the data obtained is stationary, proceed to the next stage.

From the results of stage (a), it is known that the data on the number of maritime crime cases obtained and the information on the possible risks caused are not stationary. So, the data must be stationary first by differencing by one lag (d=1). The following are the results of the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) correlograms from data that has been differentiated by one lag (d=1).

3.3.1. Data on the number of incidents (NOI) of marine crime

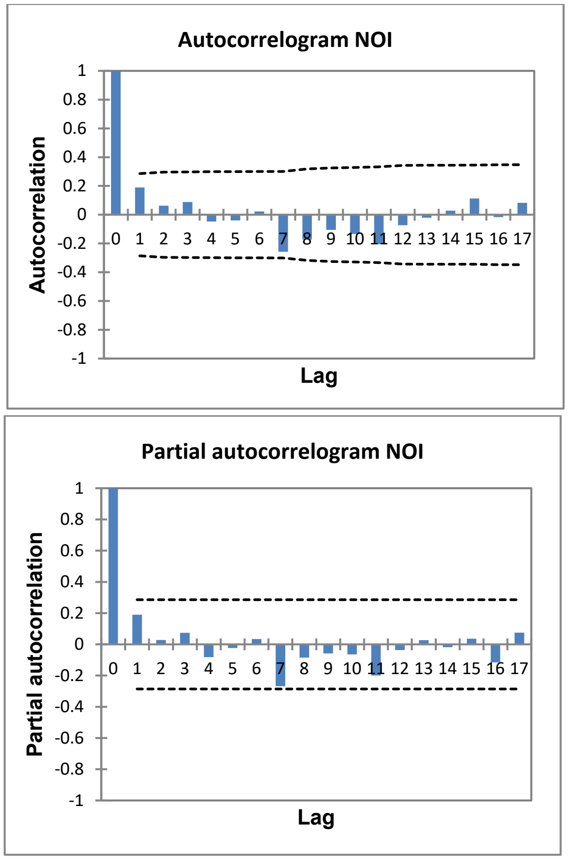
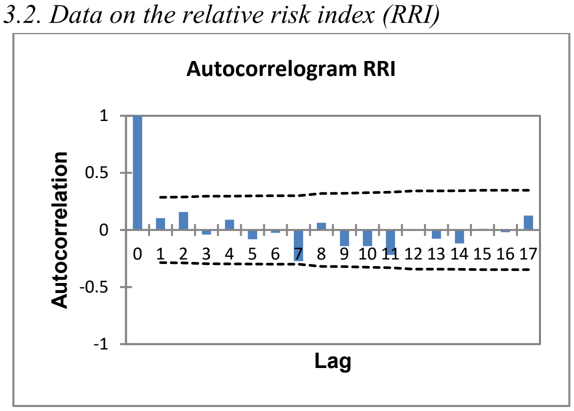


Fig.3. Correlogram of ACF and PACF dif 1 Data of NOI

Based on Figure 3, for each lag, the ACF and PACF coefficient values are quite low and close to zero, or in other words, the ACF and PACF coefficient values are relatively small. Thus, the pattern of ACF coefficient values with a differencing level of 1 (d=1) indicates the data is stationary.

3.3.2. Data on the relative risk index (RRI)



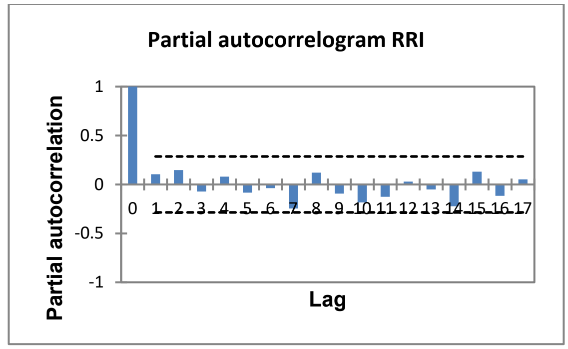


Fig.4. Correlogram ACF dan PACF dif 1 Data of RRI

Based on the data in Figure 4, for each lag, the ACF and PACF coefficient values are quite low and close to zero, or in other words, the ACF and PACF coefficient values are relatively small. Thus, the pattern of ACF coefficient values with a differencing level of 1 (d=1) indicates the data is stationary.

3.4 Identify the best ARIMA model.

After checking the data stationarity, the ARIMA model was identified for the data on the number of cases of marine crimes by looking at the ACF and PACF correlograms. Following the general form of ARIMA (p, d, q), the estimated model is,

1. ARIMA (1,1,1) model with AR(1), differencing level of one (d=1), and MA(1),
2. ARIMA (1,1,2) model with AR(1), differencing level of one (d=1), and MA(2),
3. ARIMA (2,1,1) model with AR(2), differencing level of one (d=1), and MA(1), and
4. ARIMA (2,1,2) model with AR(2), differencing level of one (d=1), and MA(2).

We can use the MAPE, MSE, AIC, or AICC values to determine the best model from the several models estimated above. The AICC (Akaike's Information Criterion Bias Corrected) value was used to determine the best model in this study. The best model is the model that has the smallest AICC value.

3.4.1. Data on the number of incidents (NOI) of marine crime

The following are the results of testing the best model from the data on the number of cases of maritime crimes for the estimated model.

Table.1. The Best ARIMA Model of NOI

p	q	AICC
1	1	251,891
1	2	255,010
2	1	255,414
2	2	259,059

The smallest AICC value is in the ARIMA (1,1,1) model based on the obtained AICC value. So the best model for NOI data from the four models tested is the ARIMA model (1,1,1).

3.4.2. Data on the relative risk index (RRI)

The following are testing results of the best model from the data on the possible risks (relative risk data) for the estimated model.

Table.2. The Best ARIMA Model of RRI

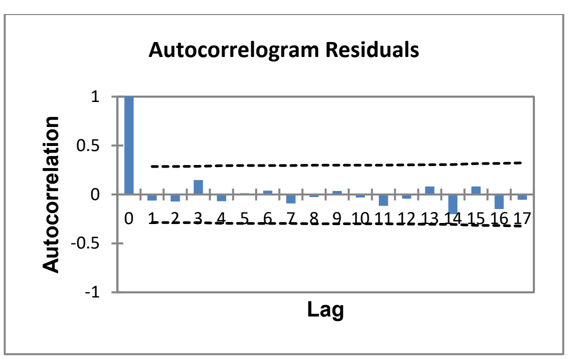
p	q	AICC
1	1	195,813
1	2	199,847
2	1	200,411
2	2	204,263

The smallest AICC value is in the ARIMA (1,1,1) model based on the obtained AICC value so that the best model for the possible risk data (RRI) from the four models tested is the ARIMA model (1,1,1).

Table.3. Predicted Results of the Number of Marine Crimes and Possible Risks

Month	NOI	RRI
Dec-21	7	4,97
Jan-22	5	1,90
Feb-22	6	2,66
Mar-22	11	5,03
Apr-22	11	6,30
May-22	6	2,81
Jun-22	8	5,41
Jul-22	7	3,31
Aug-22	7	4,50
Sep-22	5	2,50
Oct-22	8	3,53
Nov-22	2	1,71
Dec-22	8	5,79
Total	91	50

3.5. Perform diagnostic checks



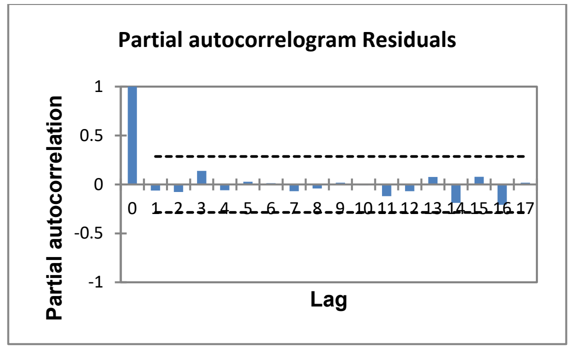


Fig.5. Correlogram of ACF and PACF of Residual NOC

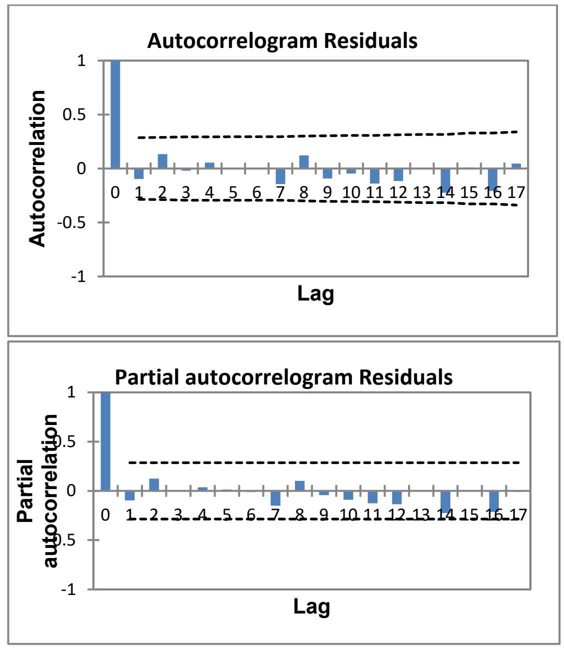


Fig.6. Correlogram of ACF and PACF of Residual RRI

Diagnostic checks are carried out to know the fit of the model. If the residual is white noise, the selected model can be feasible to use so that the model can explain the data well. Based on Figure 5 and Figure 6 (Residual NOI and Residual RRI) showed that the residuals are all in the Bartlett line, or the residuals are white noise. Because the residuals are already white noise, it is concluded that the ARIMA (1,1,1) model is suitable for use in forecasting.

3.6 Conduct forecasting

After obtaining the best model to predict the number of cases of marine crime and the magnitude of the risk that may be caused (relative risk) in the future, forecasting is carried out. Forecasting determines the number of marine crime cases and the magnitude of the risk that may be caused (relative risk) in the next 13 months.

Predicting the number of cases of marine crimes and the magnitude of the risk that may be caused (relative risk) to national defense and security for the Riau Islands region in the next 13 months is presented in Table.3.

Based on the prediction results above, it is necessary to know the level of security or awareness of the possible risks arising from the prediction of the number of criminal acts in the Riau Islands region on national defense and security for the next 13 months. For this reason, the data normalization against the relative data risk is first performed in Table.3., and the results are presented in Table.4.

Table.4. Interpretation Results

Month	NOI	RRI	Relative Risk Aggregate Index (%)	Interpretation
Dec-21	7	4,97	71,02%	A2
Jan-22	5	1,9	4,14%	VS
Feb-22	6	2,66	20,70%	VS
Mar-22	11	5,03	72,33%	A2
Apr-22	11	6,3	100,00%	A3
May-22	6	2,81	23,97%	VS
Jun-22	8	5,41	80,61%	A2
Jul-22	7	3,31	34,86%	S
Aug-22	7	4,5	60,78%	A1
Sep-22	5	2,5	17,21%	VS
Oct-22	8	3,53	39,65%	S
Nov-22	2	1,71	0,00%	VS
Dec-22	8	5,79	88,89%	A2
Total	91	50	614,16%	

Based on Table 4, the interpretation of the data obtained from the results of the descriptive analysis of percentile data and acquired a qualitative scale division of risk, namely for the aggregative index of relative risk <30% categorized as very safe (VS), (30-50)% categorized as safe (S), (51-70)% categorized as alert 1 (A1), (71-90)% is categorized as alert 2 (A2), and for the relative risk aggressive index >90% is categorized as very alert (VA).

IV. Conclusion

The results of the data analysis indicate that the prediction of the number of cases of marine crimes in the Riau Islands in the future ranges from 2 to 11 instances of incidents per month. The prediction results also show that April is the month of highest risk relative to other months during the year ahead to predict national defense and security equal to 6.3 or the aggregate relative risk index equal to 100% (Very Alert). These findings indicate that the Riau Islands are vulnerable to severe marine crimes. Therefore, it is recommended to anticipate the occurrence of violations early, such as the installation of warning sensor equipment and adequate sensing technology.

References

[1]. A. Bauer, M. Züfle, N. Herbst, A. Zehe, A. Hotho, and S. Kounev, "Time Series Forecasting for Self-Aware Systems," Proceedings of the IEEE, vol. 108, no. 7. 2020.

[2]. A. Bauer, M. Züfle, N. Herbst, A. Zehe, A. Hotho, and S. Kounev, "Time Series Forecasting for Self-Aware Systems," Proceedings of the IEEE, vol. 108, no. 7. 2020.

[3]. D. C. Montgomery, C. L. Jennings, and M. Kulahci, Introduction to Time Series Analysis and Forecasting, vol. Second Edition. 2015.

[4]. H. Wold, "A Study in Analysis of Stationary Time Series," vol. 102, no. 2, pp. 295–298, 1939.

[5]. J. H. Cochrane, Time Series for Macroeconomics and Finance. University of Chicago, 1997.

[6]. K. W. Hipel and A. I. McLeod, Time Series Modelling of Water Resources and Environmental Systems. Amsterdam: Elsevier Science Pub Co, 1994.

[7]. M. S. Roem Sutrisno, I. Abbas, and B. Badaru, "Pertanggungjawaban Pidana Korporasi Dalam Tindak Pidana Di Bidang Perikanan," Journal of Lex Generalis (JLS), vol. 1, no. 7, 2020.

[8]. P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, vol. Third Editon. 2016.

[9]. R. A. Maronna, R. D. Martin, and V. J. Yohai, Robust Statistics: Theory and Methods. 2006.

[10]. R. Adhikari and R. K. Agrawal, An Introductory Study on Time Series Modeling and Forecasting. Germany: LAP Lambert Academic Publishing, 2013.

[11]. R. Adhikari and R. K. Agrawal, An Introductory Study on Time Series Modeling and Forecasting. Germany: LAP Lambert Academic Publishing, 2013.

[12]. Y. Ben Baccar, "Comparative Study on Time Series Forecasting Models," Université Paris-Saclay, 2019.