BAREKENG: Journal of Mathematics and Its Applications
December 2024 Volume 18 Issue 4 Page 2247–2262
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss4pp2247-2262

PREDICTION OF UNIT VALUE INDEX OF EXPORTS OF SITC 897 JEWELRY AND PRECIOUS GOODS GROUP IN INDONESIA

Grace Lucyana Koesnadi[1], Bagas Shata Pratama[2], Dzuria Hilma Qurotu Ain[3], Elly Pusporani[4*], M. Fariz Fadillah Mardianto[5]

[1,2,3,4,5]*Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga*
Jl. Dr. Ir. H. Soekarno, Surabaya, Jawa Timur, 60115, Indonesia

*Corresponding author's e-mail: *elly.pusporani@fst.unair.ac.id*

Article History:

Received: 23rd January 2024
Revised: 16th May 2024
Accepted: 12th July 2024
Published: 14th October 2024

Keywords:

ARIMA;
Export;
Fourier Series Estimator;
SITC;
SVR.

ABSTRACT

Export is an international trade activity that plays an important role in the economic progress in Indonesia. One of Indonesia's leading commodities that dominate the export market is jewelry. In export activities, the export unit value index is an important component that serves to describe the development of export commodity prices. This unit value index always changes every time and fluctuates. This research conducts a comparative analysis of the performance of parametric method, non-parametric method, and machine learning, specifically, ARIMA, Fourier series estimator, and Support Vector Regression (SVR). This study aims to evaluate the effectiveness of various methods in improving prediction accuracy for the unit value index of the SITC code 897 in Indonesia. The research data used is secondary data including monthly export unit value index data with SITC code 897 in Indonesia obtained from the Central Bureau of Statistics. The data divided into 90% training data and 10% testing data. The methods used in this analysis are ARIMA, Fourier series estimator, and SVR. The best model obtained from each method is ARIMA (1,1,1) with MAPE of 10.92%, Fourier series estimator with MAPE of 8.47%, and an SVR RBF kernel function with MAPE of 3.73%. The results of this study obtained the best method for predicting the unit value index of SITC code 897 is SVR with an RMSE value of 8.288 and very good prediction accuracy.

How to cite this article:

G.L. Koesnadi, B.S. Pratama, D.H.Q. Ain, E. Pusporani, and M.F.F. Mardianto, "PREDICTION OF UNIT VALUE INDEX OF EXPORTS OF SITC 897 JEWELRY AND PRECIOUS GOODS GROUP IN INDONESIA," *BAREKENG: J. Math. & App.*, vol. 18, iss. 4, pp. 2247-2262, December, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

International trade is very important in the progress of development because it acts as a driving force in the growth of the nation's economy [1]. In a macroeconomic perspective, export activities are the most important variable in determining a country's economic system. Usually, the higher the export figure, the more open the country's economy will be in the flow of the world economy [2]. As one of the developing countries, Indonesia has opened itself to take part in export activities to strengthen its economic position in the global market [3]. The purpose of export development is to support the improvement of global competitiveness of Indonesian products. Therefore, in the future, export activities will play an important role to spearhead the domestic economy [3]. In export activities, the export unit value index or export price index is an important component. The unit value index is a value that serves to describe the development of prices of export commodity groups [4]. This index not only reflects the price development of export commodity groups [4], but also plays a significant role in assessing Indonesia’s economic performance. The Export Unit Value Index is specifically used to calculate Gross Domestic Product (GDP) at fixed prices and calculate the terms of trade. These applications underscore its importance in monitoring and analyzing Indonesia’s export dynamics, particularly as the country seeks to enhance its economic resilience and competitiveness on the global stage. The index referred to here is compiled according to the Standard International Trade Classification (SITC) code [5].

One of the mainstay non-oil and gas commodities that increase the value of Indonesia's exports while also being a leading commodity with a large market share in several export destination countries is the jewelry industry [6]. The Ministry of Industry of the Republic of Indonesia stated that the value of Indonesia's jewelry exports in the first semester of 2022 has increased compared to the same period in 2021 [7]. During January-June 2021, Indonesia's jewelry industry exports reached USD 1.23 billion. The figure almost doubled to USD 2.37 in January-June 2022. The large contribution of the jewelry industry sector to Indonesia's export value was also seen during the Covid-19 pandemic in 2020. While other sectors experienced a decline, the jewelry industry sector actually increased by 76%. Jewelry commodities are categorized under the SITC code 897, which includes jewelry, gold and silver, and other articles made from precious or semi-precious materials. This classification is part of the Standard International Trade Classification (SITC) system, which is used to categorize traded goods for statistical purposes. According to data from the official website of the Central Bureau of Statistics, the SITC 897 export unit value index for the jewelry group, gold and silver, and other items of precious or semi-precious materials, not detailed in November 2022 increased by 38.80 percent compared to October 2022, from 112.37 to 151.16 [5]. When compared to November 2021, the SITC 897 export Unit Value index in November 2022 increased by 7.19 percent.

The commodity of the jewelry industry sector has a large contribution to Indonesia's economic growth, but the condition of the Unit Value Index cannot be ascertained and can change at any time, therefore, a prediction is needed to analyze the value of the Unit Value Index for the next few periods. Therefore, researchers are interested in predicting the SITC 897 Export Unit Value Index using the ARIMA parametric method, the nonparametric regression method based on the Fourier series estimator and the machine learning method with the Support Vector Regression (SVR) algorithm. The ARIMA model is used as the simplest parametric method in time series data analysis compared to other methods with assumptions that must be met. Next, the Fourier series estimator is used as one of the nonparametric regression methods in analyzing time series data whose patterns are unknown or fluctuating and seasonal. The estimation of the unknown Fourier coefficients can be done by determining the optimal λ (smoothing parameter) value and the minimum GCV value. In addition, researchers also conducted another comparative analysis with a machine learning approach, namely using the SVR algorithm chosen because it is a nonparametric method without requiring assumptions so that it can be used to overcome the limitations of regression analysis with time series data. The SVR method is a predictive model that can be used to predict nonlinear time series data and is part of a machine learning method known as Support Vector Machine (SVM).

Previous research conducted by [8] titled "Prediction of non-oil and gas export prices in Indonesia based on the Fourier Series Estimator Method and Support Vector Regression" which shows that the SVR model with the Radial Basis Function (RBF) kernel shows better accuracy than the Fourier series estimator model with a Mean Absolute Percentage Error (MAPE) value of 9.29% [9]. This study aims to conduct a comparative analysis of the performance of parametric methods, nonparametric methods, and machine learning in predicting the unit value index of SITC code 897 in Indonesia. The novelty of this research is to add a parametric comparison method with the ARIMA method using updated export data with urgency in accordance with current economic conditions. This research also supports the realization of Sustainable

Development Goals (SDGs) eighth goal, namely Decent Work and Economic Growth. The results of this study are expected to be a reference for investors and business people in the field of jewelry and precious goods trade. In addition, it can also be an appropriate reference for decision makers to strive for Indonesia's economic progress through export activities.

2. RESEARCH METHODS

2.1 ARIMA

The ARIMA (Autoregressive Integrated Moving Average) method is a periodic series analysis method known as Box-Jenkins. This method comes from the combination of Autoregressive (AR) and Moving Average (MA) models. According to Box-Jenkins, the ARIMA method consists of four stages: identification of the time series method, estimation of parameters for alternative methods, testing the method, and possible prediction values **[10]**. The assumption of stationarity is an assumption that must be met in modeling time series. Non-stationary series can be transformed into stationary series by Box-Cox transformation and differencing **[11]**. Non-stationarity in time series data can include non-constant mean or variance. The AR (Autoregressive) model of order p or written AR(p) states that the observation at time t is linearly related to the previous time observation $t-1, t-2, \dots, t-p$. The equation form of the AR model can be written in **Equation (1) [12]**.

$$Y_t = \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \cdots + \phi_1 Y_{t-p} + a_t \tag{1}$$

with Y_t is the time series data at period t, Y_{t-i} is the time series data in the $(t-i)$-th period, a_t is the error in period t, and ϕ_i is the i-th order AR model parameter for index $i = 1,2,\dots,p$ with assumption a_t mutually independent between Y_{t-j} and Y_{t-k} for $j \neq k$.

The MA (Moving Average) model is used to explain an event where an observation at time t is expressed as a linear combination of a number of residuals. The series is also called a q-level moving average or written MA(q) with the form of the equation that can be written in **Equation (2) [12]**.

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{2}$$

with Y_t is the time series data at period t, a_t is the error in period t, a_{t-i} is the error in the $(t-i)$-th period, and θ_i is the i-th order MA model parameter for index $i = 1,2,\dots,q$. In practice, the MA model only occurs at lag 1 and lag 2.

While the ARMA model is a combination of AR and MA models which can be written with the notation ARMA (p, q). The equation form of the ARMA model at order p and q can be written in **Equation (3) [12]**.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{3}$$

with Y_t is the time series data at period t, Y_{t-i} is the time series data in the $(t-i)$-th period, a_t is the error in period t, a_{t-j} is the error in the $(t-j)$-th period, ϕ_i is the i-th order AR model parameter for index $i = 1,2,\dots,p$ and θ_j is the j-th order MA model parameter for index $j = 1,2,\dots,q$.

ARIMA (p, d, q) model with p as the operator order of AR, d as the differencing order, and q as the operator order of MA. Time series data that has been stationary after differencing d times is by calculating the difference between the observation and the previous observation **[13]**. The general form of the ARIMA model equation according to Wei (2006) can be written in **Equation (4) [12]**.

$$\phi_i(B)(1-B)^d \mathrm{Y}_t = \theta_j(B) a_t \tag{4}$$

with Y_t is the time series data at period t, B as the backshift operator, $\phi_i(B) = \left(1 - \phi_i B^i\right)$ for index $i = 1,2,\dots,p$ are the parameters of the i-th order AR model, and $\theta_j(B) = \left(1 - \theta_j B^j\right)$ are the parameters of the j-th order MA model for index $j = 1,2,\dots,q$.

2.2 Nonparametric Regression

Nonparametric regression is a method to determine the pattern of the relationship between predictor variables and response variables that are not known in the form of functions so that they can be linear or nonlinear. Nonparametric regression is very concerned about flexibility and only assumes a smooth function form [14]. Through the nonparametric regression approach, there are no assumptions that must be met in modeling as in parametric regression. The nonparametric regression model can be expressed in Equation (5) [14].

$$y_i = \eta(x_i) + \varepsilon_i \,;\, \varepsilon_i \sim N(0, \sigma^2) \tag{5}$$

with $\eta(x_i)$ is a nonparametric function whose unknown shape adjusts the shape of the nonparametric estimator, x_i is the i-th predictor variable, and ε_i is the residual of the i-th observation which is assumed to be IIDN (Identically Independent Distribution Normal). There are several techniques to estimate nonparametric regression, including kernel estimator, spline estimator, wavelet estimator, orthogonal series estimator, histogram estimator, and Fourier series estimator [15].

2.3 Fourier Series Estimator

The Fourier series is a trigonometric polynomial function that has a very high degree of flexibility. The Fourier series estimator is generally used when the data used after the investigation is unknown and there tends to be a seasonal pattern. The Fourier series estimator is based on two parameters, namely the oscillation parameter as a representation of the bandwidth and the Fourier parameter. The Fourier series is well used to describe curves that show sine and cosine waves [12]. For example, with observed data (t_i, y_i) that follows a nonparametric time series model, it can be written in Equation (6) [16].

$$y_i = m(t_i) + \varepsilon_i \tag{6}$$

with $i = 1,2, \dots, n$ which indicates the number of observations and $\varepsilon_i \sim IIDN(0, \sigma^2)$. Regression function $m(t_i)$ in Equation (6) is unknown and will be estimated with the Fourier series estimator approach which can be written in Equation (7) [16].

$$m(t_i) = \beta_0 + \sum_{j=1}^{\lambda} (\alpha_j \cos 2\pi j t_i + \beta_j \sin 2\pi j t_i) \tag{7}$$

with λ is the oscillation parameter, α_j and β_j is the j-th regression coefficient parameter. In the Fourier series estimator approach, it is important to determine the optimal value of λ to obtain the best model. If Equation (7) is substituted for each of the following y_i in Equation (6) then Equation (8) is obtained [16].

$$y_i = \beta_0 + \sum_{j=1}^{\lambda} (\alpha_j \cos 2\pi j t_i + \beta_j \sin 2\pi j t_i) + \varepsilon_i;\ \varepsilon_i \sim IIDN(0, \sigma^2) \tag{8}$$

In matrix equation form, the nonparametric regression model with Fourier series estimator can be expressed in Equation (9) [16].

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}; \varepsilon_i \sim N(0, \sigma^2) \tag{9}$$

Based on the Ordinary Least Square (OLS) method, the estimator for the parameter vector $\boldsymbol{\beta}$ in equation (10) is obtained [16].

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{10}$$

by substituting the estimator for the parameter vector $\boldsymbol{\beta}$ in Equation (10) into Equation (9), the regression curve estimator in Equation (9) is obtained as follows.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{11}$$

so that the estimation form of the nonparametric regression equation based on the Fourier series estimator in Equation (12) is obtained [16].

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{\lambda} (\hat{\alpha}_j \cos 2\pi j t_i + \hat{\beta}_j \sin 2\pi j t_i) \tag{12}$$

The optimal oscillation parameter in the Fourier series estimator is obtained using the Generalized Cross Validation (GCV) method which can be seen from the minimum GCV value. The GCV method is generally defined as follows [16].

$$GCV(\lambda) = \frac{MSE(\lambda)}{(n^{-1}\text{trace}[\mathbf{I} - \boldsymbol{H}(\lambda)])^2} \tag{13}$$

with

$$MSE(\lambda) = n^{-1}\sum_{i=1}^{n}\left(y_i - \widehat{m}(t_i)\right)^2 \tag{14}$$

and

$$\boldsymbol{H}(\lambda) = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \tag{15}$$

2.4 Support Vector Regression (SVR)

The SVR method is the result of SVM modification in overcoming regression problems, where the output of SVR is represented as a real and continuous number **[17]**. The SVR method can be used to determine the best hyperplane as a regression function and minimize the probability of error by maximizing the boundary. The concept of the SVR method is to have training data $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$ where $x_i \in \mathbb{R}^d$ is the i-th input vector with $i = 1,2,\dots, n, d$ is the dimension and y_i is the goal or target value. The general equation of the SVR model can be written in **Equation (16) [18]**.

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\varphi(\boldsymbol{x}) + b \tag{16}$$

where $\boldsymbol{w}$ is a dimensionless weight vector $n, \varphi(\boldsymbol{x})$ is a function that maps x in a space with n dimension, and b is a constant. To obtain a good generalization of the regression function $f(x)$, the next step is to minimize the smallest $\boldsymbol{w}$ criterion. Based on this, the solution to the hyperplane optimization problem is as follows **[18]**.

$$min_w \frac{1}{2} \,||\,\boldsymbol{w}||^2 \text{ with} \begin{cases} y_i - \boldsymbol{w}^T\varphi(\boldsymbol{x_i}) - b \ \leq \varepsilon \text{ for } i = 1,..,n \\ \boldsymbol{w}^T\varphi(\boldsymbol{x_i}) + b - y_i \ \leq \varepsilon \text{ for } i = 1,..,n \end{cases} \tag{17}$$

The factor $||\boldsymbol{w}||^2$ called regulation. Minimizing $||\boldsymbol{w}||^2$ will make a function as flat as possible, so as to control the function capacity. In the regression function f, it is assumed that all points within the range $f(\boldsymbol{x}) \pm \varepsilon$ are feasible, and points outside the range are infeasible, so the slack variables ξ and ξ^* are added to overcome the non-conforming limitations of the optimization problem. Furthermore, **Equation (17)** can be transformed in the following form.

$$min_w \frac{1}{2} \,||\boldsymbol{w}||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i{}^*) \tag{18}$$

The optimization problem has the following constraint function **[18]**.

$$\begin{aligned} y_i - \boldsymbol{w}^T\varphi(\boldsymbol{x_i}) - b &\leq \varepsilon + \xi_i\,, \qquad i = 1,..,n \\ \boldsymbol{w}^T\varphi(\boldsymbol{x_i}) + b - y_i &\leq \varepsilon + \xi_i{}^*, \qquad i = 1,..,n \end{aligned} \tag{19}$$

by solving the optimization problem, thus obtained.

$$w = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\varphi(\boldsymbol{x}_i) \tag{20}$$

The data contained in the vector x from the input space can be mapped to a higher dimensional feature space with the function φ which is approximated by a kernel function so that the SVR function can be defined as follows **[18]**.

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) + b \tag{21}$$

To solve the high-dimensional nonlinear problem, what needs to be done is $K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$ with the selected kernel function in SVR presented in **Table 1 [8]**.

Table 1. Kernel Function

Kernel Type	Formula
Linear	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i{}^T - \boldsymbol{x}_j)$
Polynomial	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i . \boldsymbol{x}_j + 1)^p$
Radial Basis Function (RBF)	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma\, (\boldsymbol{x}_i - \boldsymbol{x}_j)^2\right)$
Sigmoid	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(g\boldsymbol{x}_i{}^T\boldsymbol{x} + 1)$

The kernel function used in this study is the RBF kernel which is defined as follows.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma\, (\boldsymbol{x}_i - \boldsymbol{x}_j)^2\right) \tag{22}$$

with

$$\gamma = \frac{1}{2\sigma^2} \tag{23}$$

Therefore, the use of SVR with RBF kernel type requires the determination of 3 parameters namely cost (C), gamma (γ), and epsilon (ε). In finding the optimal parameter values, one of the methods that can be used is the grid search method [8][18].

2.5 Best Model Selection

Measuring the accuracy of prediction results is done to determine the accuracy between actual data and predicted data. In measuring the accuracy of prediction results, there are several measurement measures used, namely RMSE and MAPE [19].

1. **Root Mean Square Error (RMSE)**

RMSE is the size of the error rate obtained from the prediction results. The smaller the RMSE value or closer to 0, the better the accuracy. The calculation of RMSE is to subtract the predicted value from the actual value, square it, add up the total results, and divide it by the amount of prediction data [20]. Furthermore, RMSE can be expressed mathematically as follows [20].

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{n}} \tag{24}$$

where n is the amount of data, y_t is the actual data in period t, and $\hat{y}_t$ is the predicted data in the-t th period.

2. **Mean Absolute Percentage Error (MAPE)**

The MAPE value is a measure of the average absolute error of the prediction results compared to the actual data in percentage. MAPE value measurement is done by dividing the absolute error of each period by the actual data value of that period in percentage form, then averaged for each forecast period. The MAPE value can be expressed in Equation (25) [21].

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n}\, x\ 100\% \tag{25}$$

where n is the amount of data, y_t is the actual data in period t, and $\hat{y}_t$ is the predicted data in the t-th period.

In this study, the MAPE value is used as a measure of comparing the prediction results between several time series models used. The MAPE value criteria are shown in Table 2 [22].

Table 2. MAPE Value Criteria

MAPE Value	Criteria
$\mathbf{MAPE} \le \mathbf{10\%}$	Very good accuracy
$\mathbf{10\%} < \mathbf{MAPE} \le \mathbf{20\%}$	Good accuracy
$\mathbf{20\%} < \mathbf{MAPE} \le \mathbf{50\%}$	Sufficient accuracy
$\mathbf{MAPE} > \mathbf{50\%}$	Poor accuracy

2.6 Data Description

The data used in this study are data on the SITC 897 code export Unit Value index in Indonesia for the gold and jewelry group in the form of monthly data from the Indonesian Central Bureau of Statistics obtained through the official website bps.go.id from January 2018 to August 2023. Empirical studies show that the division of research data 80-90% training data and 10-20% testing data obtained the best prediction results [9]. The division of 90% training data and 10% testing data in this study is data for the period January 2018 to January 2023 as much as 61 data as training data and data for the period February 2023 to August 2023 as much as 7 data as testing data. In the analysis process, training data is used for modeling, while testing data is used to compare prediction results with actual data with the aim of knowing the goodness of the prediction model obtained.

2.7 Step Analysis

The complete procedure of the analysis method in this research is detailed as follows.

1. Determine the characteristics and descriptive statistics of export Unit Value index data of SITC code 897 in Indonesia.
2. Divide the research data into training and testing data.
3. Perform data modeling and prediction with a parametric approach based on the ARIMA model with the following steps:
   a. Conduct a data stationarity test.
   b. Performing Box-Cox transformation.
   c. Perform the differencing process.
   d. Identify several possible ARIMA models based on ACF and PACF plots.
   e. Conduct parameter significance tests for several ARIMA models in (c).
   f. Perform model diagnostic tests on ARIMA models that fulfill step (d), including residual normality test, white noise test, and homoscedasticity test.
   g. Determine the best ARIMA model that satisfies step (f) with the smallest MSE value and create the ARIMA model equation.
   h. Predict the testing data based on the best ARIMA model.
4. Perform data modeling and prediction with a nonparametric approach based on the Fourier series estimator with the following steps:
   a. Determine the GCV and MSE formulas based on the Fourier series estimator results.
   b. Determine the optimal smoothing parameter (λ) based on the minimum GCV value.
   c. Determine the estimator model on the training data with a non-parametric regression approach using the Fourier series estimator based on the optimal λ value.
   d. Calculating the goodness of estimator model criteria using R squared, RMSE, and MAPE.
   e. Predict the testing data based on the best Fourier series estimator model obtained in (c).
5. Predict data using a machine learning approach with the Support Vector Regression (SVR) algorithm with the following steps:
   a. Performed linearity relationship detection test.
   b. Determine the significant lag on the PACF plot as the predictor variable.
   c. Perform initial modeling using training data.
   d. Perform the SVR parameter tuning process with the grid search method.
   e. Determine the type of kernel function and loss function used for prediction.
   f. Perform SVR modeling on training data using optimal parameters with the grid search method.
   g. Predict the testing data using the best SVR model obtained in (f).
6. Calculate the RMSE and MAPE values of the testing data obtained based on the best model from each of the three methods in steps (3) – (5).
7. Compare the prediction results of the ARIMA model, Fourier series estimator, and SVR based on the smallest RMSE and MAPE values.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

The data used in this study are monthly data on the SITC 897 export Unit Value index from January 2018 to August 2023. The following is a summary of monthly data on the SITC 897 export Unit Value index in **Table 3**.

Table 3. Descriptive Statistics

Data	Total Data	Minimum	Mean	Maximum
Training	61	80.73	135.61	393.27
Testing	7	118.02	138.63	149.9
All	68	80.73	135.92	393.27

Moreover, the monthly time series data of the SITC 897 export Unit Value index from January 2018 to August 2023 are graphically showed in **Figure 1**.

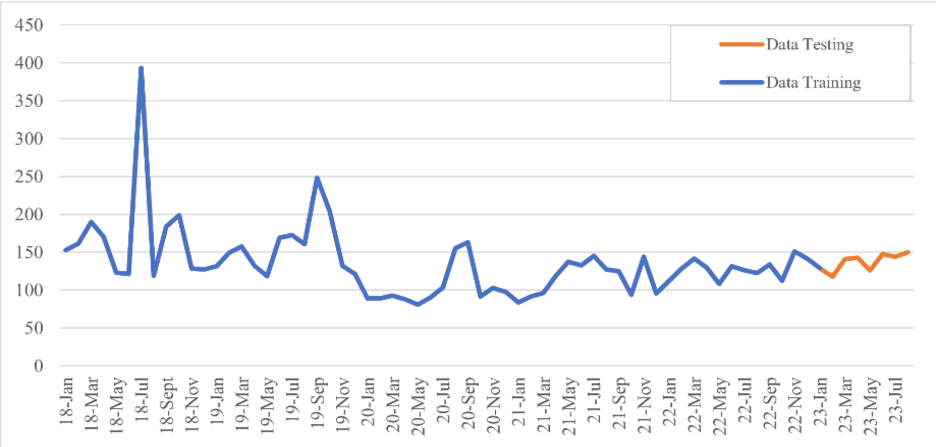
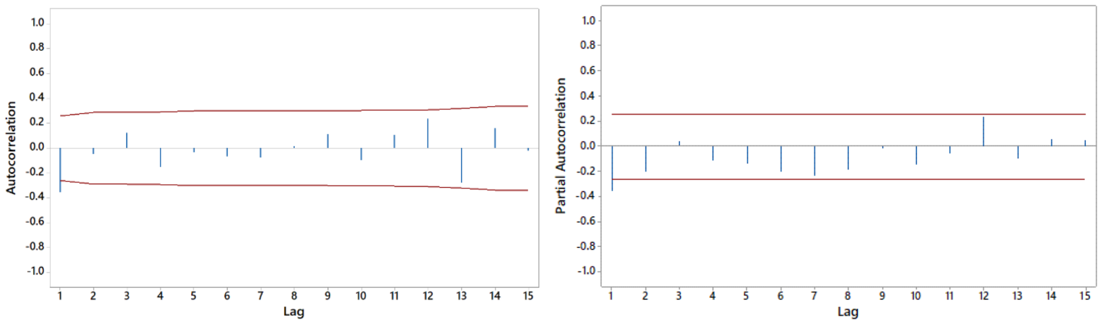


Figure 1. Time Series of Training and Testing Data

Based on **Figure 1**, it can be seen that in general, the Unit Value index of SITC 897 exports tends to fluctuate every month with the highest value in July 2018 with an index of 393.27. Furthermore, the Unit Value index of SITC 897 exports fluctuates up and down and tends to increase until January 2023.

3.2 Modeling Results with ARIMA

In selecting the best ARIMA model, there are several conditions and assumptions that must be met, including data stationarity test, parameter significance test, white noise test on model residuals which means there is no correlation between residuals, normally distributed residuals, and constant variance. In the stationarity test, it was found that the research data using training data was not stationary, both in mean and variance. To overcome data non-stationarity, Box-Cox transformation is performed once, namely Y_t^{-1} and differencing the data once to achieve stationarity. Furthermore, the ARIMA model on data that has been stationary will be identified using its ACF and PACF plots presented in **Figure 2**.



(a) (b)

Figure 2. ACF and PACF Plots of Stationary Data
(a) ACF Plot, (b) PACF Plot

Figure 2 show the ACF and PACF plots that used to identify the ARIMA model, which are both plots showing significant correlation at lag 1, suggesting the presence of AR and MA components. Based on that results, three potential models are identified including ARI (1,1), IMA (1,1) and ARIMA (1,1,1). The selection of these specific models is grounded in the observed significant lag and the theoretical understanding that ARI models focus on autoregressive terms, IMA models on moving average terms, and ARIMA models on the combined influence of autoregressive and moving average components. The selection of the best ARIMA model based on parameter significance test, model diagnostic test, and RMSE criteria is summarized in Table 4.

Table 4. ARIMA Parameter Estimation of SITC 897 Monthly Export Unit Value Index

ARIMA Model		Parameter Significance	RMSE	Residuals
ARI (1,1)	Deterministic	AR 1 = 0.005	0.00184	White Noise
		Constant = 0.916		
	Probabilistic	AR 1 = 0.005	0.00182	White Noise
IMA (1,1)	Deterministic	MA 1 = 0.000	0.00179	No White Noise
		Constant = 0.864		
	Probabilistic	MA 1 = 0.000	0.00179	No White Noise
ARIMA (1,1,1)	Deterministic	AR 1 = 0.000	0.00176	White Noise
		MA 1 = 0.000		
		Constant = 0.366		
	Probabilistic	AR 1 = 0.006	0.00176	White Noise
		MA 1 = 0.000		

***Data source**: Results of Minitab*

Based on Table 4, the best model is the probabilistic ARIMA (1,1,1) model because all parameters are significant ($p - value < 0.05$), fulfill the White Noise assumption, and have the smallest RMSE value compared to other models. After that, the residual normality test and the homoscedasticity test were conducted, and it was found that the model had met both assumptions. Therefore, the ARIMA (1,1,1) model is the best parametric model to predict the SITC 897 Monthly Export Unit Value Index presented in Equation (26).

$$
\begin{aligned}
\phi_p(B)(1-B)^d Y_t &= \theta_q(B)a_t(1-\phi_1 B)\dot{Y}_t = (1-\theta_q B)a_t \\
\phi_1(B)(1-B)^1 \dot{Y}_t &= \theta_1(B)a_t(1-\phi_1 B)\dot{Y}_t = (1-\theta_1 B)a_t \\
\dot{Y}_t - \phi_1 B\dot{Y}_t &= a_t - \theta_1(B)a_t \\
\dot{Y}_t &= 0.432\dot{Y}_{t-1} + a_t - 0.8965a_{t-1}
\end{aligned} \tag{26}
$$

Based on Equation (26), the prediction results of the SITC 897 Monthly Export Unit Value Index for the next seven periods are presented in Table 5 and Figure 3.

Table 5. Prediction of SITC 897 Monthly Export Unit Value Index with ARIMA Model (1,1,1)

Period	Actual Data	Prediction	Lower Limit of 95% Confidence Interval	Upper Limit of 95% Confidence Interval
Feb – 23	118.02	126.8436	88.3408	224.8353
Mar – 23	141.33	125.8608	84.4409	247.0356
Apr – 23	142.69	125.4411	83.0896	255.8461
May – 23	126.38	125.2608	82.4063	260.9808
June – 23	147.89	125.1830	81.9504	264.9638
July – 23	144.17	125.1495	81.5847	268.5501
Aug – 23	149.90	125.1351	81.2579	272.0126

***Data source**: Results of Minitab*

Figure 3 shows a time series graph along with the results of data prediction for the period February 2023 to August 2023. Furthermore, the MAPE (Mean Absolute Percentage Error) value will be calculated and the results of the MAPE calculation are 10.9%, so it can be categorized that the accuracy of the prediction results is good.

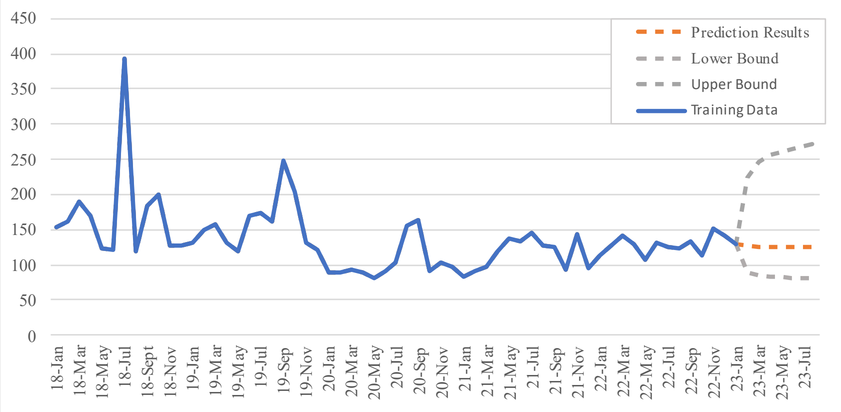


Figure 3. Time Series Graph of Prediction Results with ARIMA Model (1,1,1)

3.3 Modeling Results with Fourier Series Estimator

The nonparametric regression model with Fourier series estimator has one oscillation or smoothing parameter (λ). The selection of the optimal smoothing parameter lambda (λ) is obtained by the GCV method so that the resulting estimator is also optimal. Plots and GCV values obtained using R software for in sample or training data are presented in **Figure 4** and **Table 6**.

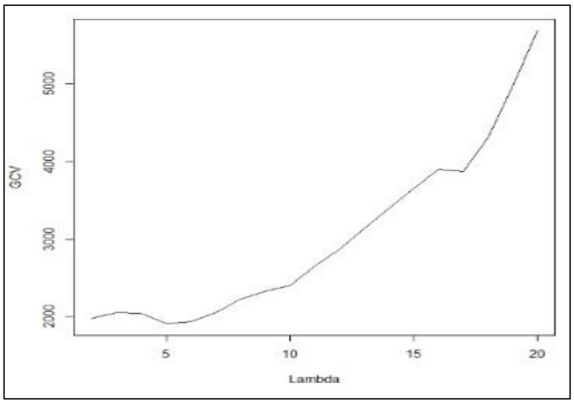


Figure 4. Plot of GCV Against *Lambda* Parameter (λ)

Figure 4 shows that the GCV values reach the lowest value when lambda is at point 5. The GCV value is shown in detail in the **Table 6**.

Table 6. GCV Value Against *Lambda* Parameter (λ)

λ	GCV	λ	GCV
1	1998.485	11	2658.633
2	1979.351	12	2868.944
3	2051.237	13	3139.086
4	2047.216	14	3399.770
5	1914.30	15	3659.206
6	1936.572	16	3901.067
7	2053.851	17	3874.508
8	2229.251	18	4307.577
9	2331.148	19	4977.093
10	2403.862	20	5687.453

***Data source**: Results of OSS-R*

Based on the results in **Table 6**, the minimum GCV value is 1914.3 with the optimum smoothing parameter (λ) value of 5. Then the estimator model is obtained with a nonparametric regression approach based on the Fourier series estimator in **Equation (27)**.

$$\hat{Y}_t = \beta_0 + \alpha_1 \cos 2\pi t_r + \beta_1 \sin 2\pi t_r + \alpha_2 \cos 4\pi t_r + \beta_2 \sin 4\pi t_r + \alpha_3 \cos 6\pi t_r + \beta_3 \sin 6\pi t_r + \alpha_4 \cos 8\pi t_r + \beta_4 \sin 8\pi t_r + \alpha_5 \cos 10\pi t_r + \beta_5 \sin 10\pi t_r \quad (27)$$

Based on the model in **Equation (27)**, the estimation of the Unit Value index model of SITC 897 exports in Indonesia is as follows.

$$\hat{Y}_t = 135.610 + 22.276 \cos 2\pi t_r + 21.087 \sin 2\pi t_r - 6.028 \cos 4\pi t_r + 7.747 \sin 4\pi t_r + 9.652 \cos 6\pi t_r + 5.313 \sin 6\pi t_r - 3.788 \cos 8\pi t_r + 15.008 \sin 8\pi t_r - 18.993 \cos 10\pi t_r - 6.507 \sin 10\pi t_r \quad (28)$$

The model in **Equation (28)** has a good training model criteria value measured using MAPE and RMSE values. The MAPE value is 15.8652% which can be categorized as good prediction accuracy and the RMSE value is 35.8628. The movement of the SITC 897 export Unit Value index in Indonesia can be predicted using the Fourier series estimator model. The graph of estimated and actual data on training data modeling with the Fourier series estimator model is presented in **Figure 5**.

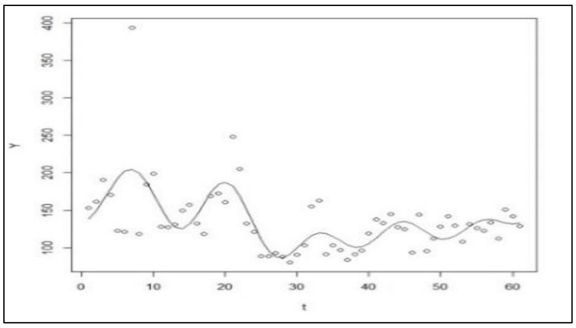


Figure 5. Plot of Estimation Model with Fourier Series Estimator

Figure 5 shows the plot of the estimation results with the optimal smoothing parameter with the t index indicating the data period and the y index indicating the value of the export value index. The dot plot shows the actual data for the training data, while the line plot is the prediction graph. The plot produces an estimate that is not too rough (under-smooth) and not too smooth (over-smooth) or it can be said that the estimation curve is the optimal estimation result. That is indicated by the estimated curve being close to the actual data points reflecting the model's ability to accurately capture the underlying trend. However, the curve does not fit the data point so tightly to every fluctuation in the data, thereby avoiding the risk of overfitting. This balance between following the trend and not overfitting the noise demonstrates an appropriate trade-off between bias and variance, ensuring that the estimates are neither too coarse nor too smooth. Furthermore, the model is used to predict the testing data, the prediction results of the SITC 897 export Unit Value index for the period February 2023 to August 2023 using the Fourier series estimator are presented in **Table 7**.

Table 7. Prediction Results with Fourier Series Estimator

Period	Actual Data	Prediction
February 2023	118.02	132.9449
March 2023	141.33	133.0010
April 2023	142.69	133.0564
May 2023	126.38	133.1111
June 2023	147.89	133.1650
July 2023	144.17	133.2182
August 2023	149.90	133.2707

***Data source**: Results of OSS-R*

Based on **Table 7**, the calculation of the MAPE value is 8.466%, so it can be categorized that the prediction results are very good. So, the model is suitable as a reference for predicting the SITC 897 Monthly Export Unit Value Index in the next 7 periods.

3.4 Modeling Results with Support Vector Regression

Before modeling with Support Vector Regression, it is necessary to test the detection of linearity relationship using White test and Terasvirta test. The White test is created to identify nonlinearity in neural network models. Similarly, the Terasvirta test, which also detects nonlinearity, is developed from neural network models and is part of a group of tests based on the Lagrange Multiplier approach with Taylor expansion [23]. Based on the test results on the SITC code 897 of unit value index data, the results show that the data contains a nonlinear pattern and the PACF plot shows a significant lag at lag 1. Next, the best kernel function is selected by looking at the smallest MAPE and RMSE values of the training data and it is found that the best model for nonlinear models is the radial kernel function, with a Root Means Square Error (RMSE) of 42.904 and a Mean Average Percentage Error (MAPE) of 14.883% so that the model is accurate for long-term predictions [18]. The plot results of actual and predicted data are presented in Figure 6.

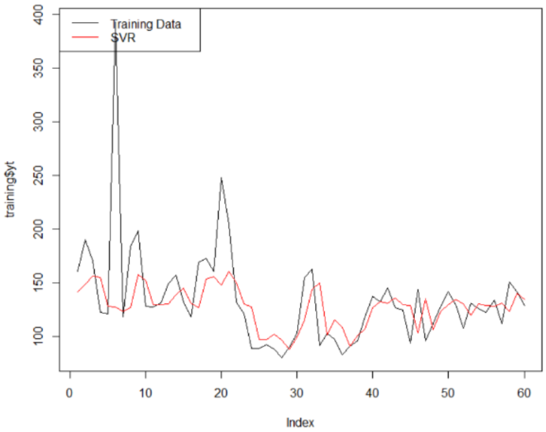
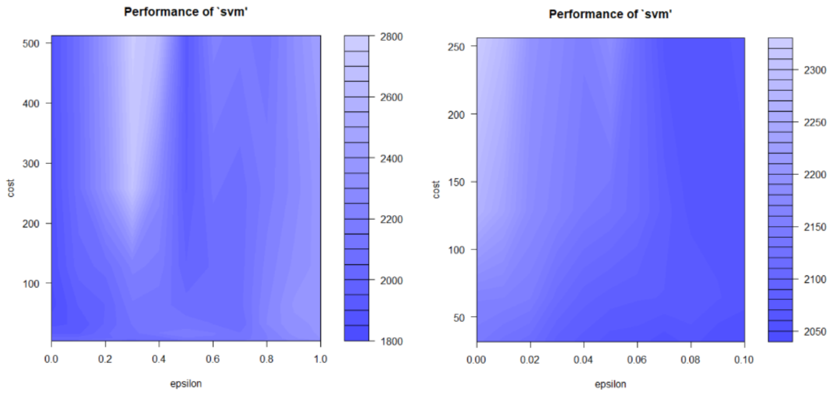


Figure 6. Comparison of Actual and Predicted of Training Data of SVR Method

After forming the initial model with the best model using the radial kernel function, the SVR parameter tuning process will be carried out using the grid search method with two tuning processes using loose grid and finer grid. The results of tuning using loose grid and finer grid and obtained the optimal grid search method is the finer grid method by finding the optimal parameters in the neighborhood of the loose grid $C = 32, \varepsilon = 0.1$. Then, the results of the tuning plot with both grid search methods are presented in Figure 7.



(a) **(b)**

Figure 7. Results of Tuning Plot of Grid Search Method
(a) Loose Grid, (b) Finer Grid

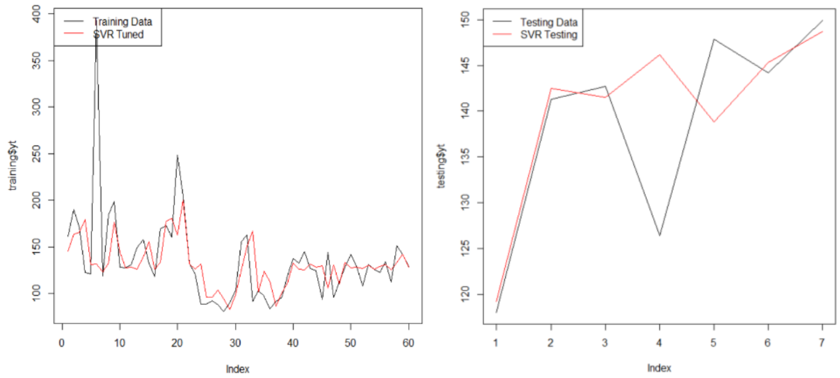
The plot results are seen in the dark area and show that in that area both parameters show optimal results. After performing the tuning process with grid search and obtaining the optimal parameter combination, the modeling of the monthly export Unit Value index which became the training and testing data was carried out again using SVR with the radial kernel function presented in Table 8.

Table 8. RMSE and MAPE of Prediction with SVR

Data Type	RMSE	MAPE
Training Data	41.441	13.8%
Testing Data	8.288	3.73%

***Data source**: Results of OSS-R*

The prediction results of applying the SVR model to the training data and testing data are visualized on the plot which shows that the prediction results do not differ much from the actual data. Furthermore, the MAPE accuracy rate is less than 10% which is included in the highly accurate category. The plot results of predictions with actual data for training and testing can be presented in Figure 8.



(a) (b)

Figure 8. Comparison Plot of Predicted and Actual Data of SVR Method (a) Comparison Plot of Training Data, (b) Comparison Plot of Testing Data

3.5 Comparison of ARIMA, Fourier Series, and SVR Prediction Results

Based on the analysis that has been done with the three methods, namely ARIMA, Fourier series estimator, and SVR, a comparison of the prediction results of the best model for each method is presented in **Figure 9** and a comparison of the MAPE and RMSE values in **Table 9**.

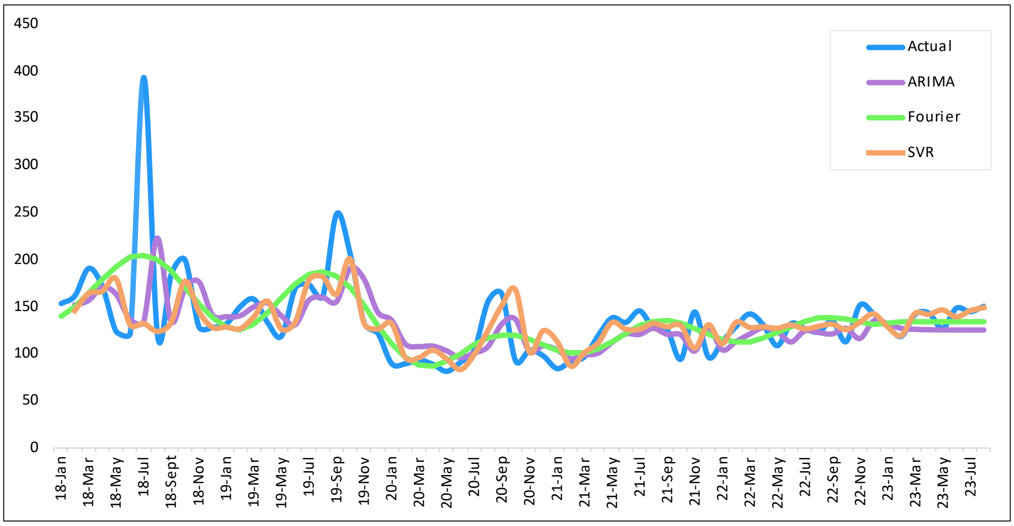


Figure 9. Comparison Plot of Testing Data Prediction of The Three Research Methods

Based on **Figure 9**, the Fourier series and SVR methods are in the category of excellent prediction accuracy as shown in **Table 9**, but the plot of prediction results using SVR is able to predict fluctuation patterns in actual data more accurately, such as the occurrence of Unit Value index fluctuations in February 2023, March 2023, July 2023, and August 2023. Meanwhile, the prediction results with the ARIMA and Fourier Series methods tend to produce linear forecasts and do not follow the pattern in the actual data.

Table 9. Comparison of RMSE and MAPE of The Three Methods

Methods	RMSE	MAPE	Accuracy
ARIMA	17.3478	10.92%	Good
Fourier Series Estimator	12.2097	8.47%	Very good
SVR	**8.2880**	**3.73%**	**Very Good**

***Data source**: Results of OSS-R*

Based on **Table 9**, it can be seen that the prediction results with the Support Vector Regression (SVR) approach are more accurate than other methods. Therefore, the SVR method is obtained as the best method in predicting the unit value index of SITC code 897 exports in Indonesia with the smallest RMSE and MAPE values.

4. CONCLUSIONS

Based on the results of the analysis that has been carried out on the prediction of the Unit Value Index of the jewelry and precious goods group with SITC code 897 using the ARIMA, Fourier Series and Support Vector Regression methods, it is found that the SVR method with the radial kernel function produces the smallest MAPE testing value compared to the other two methods, which is 3.73% with a very good prediction accuracy category and an RMSE of 8.288. This method also produces a prediction pattern that almost resembles actual data which tends to fluctuate up and down, while the other two methods produce data patterns that tend to be linear. Therefore, SVR can be used as the best method to predict the SITC code export unit value index for the period February 2023 to August 2023. The prediction results that have been carried out are expected to be adapted by investors and business people in the field of jewelry and precious goods trade as a reference for decision making in conducting transactions in the future period. In addition, it can also be an appropriate reference for decision makers to strive for Indonesia's economic progress through international trade activities.

ACKNOWLEDGMENT

The researcher would also like to thank Elly Pusporani, S.Si., M.Stat. and Dr. M. Fariz Fadillah Mardianto, S.Si., M.Si. as Multivariate Time Series Analysis (ARWM) Lecturers in the Bachelor Program in the Statistics Program Study at FST Universitas Airlangga for supporting and providing encouragement during the research so that this research can be completed properly.

REFERENCES

[1] F. Amelia, "Analisis Strategi Pemasaran Dalam Meningkatkan volume Penjualan Dalam Perspektif Ekonomi Islam (studi kasus pada Toko AGS kuningan)," *Chang. Think Jounal*, vol. 1, no. 1, pp. 50–59, 2018.

[2] A. Wardhana, "Analisis Faktor-Faktor yang Mempengaruhi Ekspor Nonmigas Indonesia ke Singapura Tahun 1990-2010," *J. Manaj. Dan Akunt.*, vol. 12, no. 2, pp. 99–102, 2011.

[3] K. A. S. Pramana and L. G. Meydianawathi, "Variabel-Variabel yang Mempengaruhi Ekspor Nonmigas Indonesia ke Amerika Serikat," *J. Ekon. Kuantitatif Terap.*, vol. 6, no. 2, pp. 98–105, 2013.

[4] L. Hugida and S. Sofian, "Analisis faktor-faktor yang mempengaruhi volatilitas harga saham (Studi pada perusahaan yang terdaftar dalam indeks LQ45 periode 2006–2009)," Universitas Diponegoro, Semarang, 2011.

[5] B. P. Statistik, "Indeks Unit Value Ekspor Menurut Kode SITC Bulan November 2022," Jakarta, 2022.

[6] M. A. D. Prabayanthi and I. A. N. Saskara, "ANALISIS DAYA SAING DAN FAKTOR-FAKTOR YANG MEMPENGARUHI EKSPOR PERHIASAN INDONESIA," *E-Jurnal Ekon. Pembang.*, vol. 10, no. 7, pp. 3075–3103, 2021.

[7] E. S. A. S. Agustin and S. Edhie, *Kajian Tengah Tahun INDEF 2022: Reformulasi Kemandirian Ekonomi di Tengah Dinamika Global*. INDEF, 2022.

[8] C. F. F. Purwoko, S. Sediono, T. Saifudin, and M. F. F. Mardianto, "Prediksi Harga Ekspor Non Migas di Indonesia Berdasarkan Metode Estimator Deret Fourier dan Support Vector Regression," *Inferensi*, vol. 6, no. 1, pp. 45–55, 2023.

[9] C. F. F. Purwoko, Sediono, T. Saifudin, and M. F. F. Mardianto, "Prediksi Harga Ekspor Non Migas di Indonesia Berdasarkan Metode Estimator Deret Fourier dan Support Vector Regression," *Inferensi*, vol. 6, no. 1, pp. 45–55, 2023.

[10] F. Danitasari, "Perbandingan Prediksi Sifat Hujan Bulanan antara Analisis Komponen Utama Model Arima dan Metode Probabilitas di Stasiun Meteorologi Pongtiku Tana Toraja," *J. Meteorol. Klimatologi dan Geofis.*, vol. 2, no. 2, pp. 207–215, 2015.

[11] A. Lusiani and E. Habinuddin, "Pemodelan Autoregressive Integrated Moving Average (ARIMA) Curah Hujan di Kota Bandung," *J. Sigma-Mu*, vol. 3, no. 2, pp. 9–25, 2011.

[12] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, Second Edi. New York: Pearson Adisson Wesley, 2006.

[13] Y. Susanto, "Pemodelan Curah Hujan dengan Pendekatan Model ARIMA, Feed Forward Neural Network dan Hybrid (ARIMA-NN) di Banyuwangi," *J. Sains Dan Seni ITS*, vol. 5, no. 2, pp. 145–150, 2016.

[14] Budiantara and I. Nyoman, "Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter," 2011.

[15] A. A. Anandari, *Analisis Regresi Deret Fourier: Aplikasi Data Curah Hujan*. Jawa Barat: CV Jejak (Jejak Publisher), 2023.

[16] N. Chamidah and B. Lestari, *Analisis Regresi Nonparametrik dengan Perangkat Lunak R*. Surabaya: Airlangga University Press., 2022.

[17] A. J. Smola and B. Scholkopf, "A Tutorial on Support Vector Regression," *Stat. Comput.*, vol. 12, pp. 199–222, 2004.

[18] R. P. Furi, J. Jondri, and D. Saepudin, "Prediksi Financial Time Series Menggunakan Independent Component Analysis Dan Support Vector Regression. Studi Kasus: Ihsg Dan Jii," *eProceedings Eng.*, vol. 2, no. 2, 2015.

[19] X. Guo, Y. Gao, Y. Li, D. Zheng, and D. Shan, "Short-Term Household Load Forecasting Based on Long-and Short-Term

Time Series Network," *Energy Reports*, vol. 7, pp. 58–64, 2021.

[20] M. Merfin and R. S. Oetama, "Prediksi Harga Saham Perusahaan Perbankan Menggunakan Regresi Linear Studi Kasus Bank BCA Tahun 2015-2017," *Ultim. J. Tek. Inform.*, vol. 11, no. 1, pp. 11–15, 2019, doi: 10.31937/ti.v11i1.1239.

[21] J. J. M. Moreno, A. P. Pol, A. S. Abad, and B. C. Blasco, "Using the R-MAPE Index as Resistant Measure of Forecast Accuracy," *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013.

[22] P. C. Chang, Y. W. Wang, and C. H. Liu, "The Development of a Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting," *Expert Syst. with Appl.*, vol. 32, no. 1, pp. 86–96, 2007.

[23] R. D. Khoirunnisa, W. Wibowo, and A. Suharsono., "Nonlinearity Test on Time Series Data Case Study: The Number of Foreign Tourist," 2016.