**Journal of Natural Sciences and Mathematics Research**
*J. Nat. Scien. & Math. Res.* Vol. 11 No. 2 (2025) 153-163
ISSN: 2460-4453 (Online); 2614-6487 (Printed)
Available online at http://journal.walisongo.ac.id/index.php/jnsmr

# Forecasting the unemployment rate in West Java Province using VARX and SVR methods

**Yuniar Kusuma Wardani[1*], and Sugiman[1]**
[1] Department of Mathematics, The Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Indonesia

* Corresponding author's e-mail: yuniarkw21@gmail.com

**ABSTRACT**
This study discusses the forecasting of the Open Unemployment Rate (OUR) in West Java Province using two time series approaches: Vector Autoregressive with Exogenous variables (VARX) and Support Vector Regression (SVR). The dataset consists of monthly observations from 2018 to 2023, including variables such as OUR, the Labor Force Participation Rate (LFPR), Gross Regional Domestic Product (GRDP), and the Human Development Index (HDI). Based on the optimal lag selection using the AIC, the VARX model produced the best lag configuration of (5,2), consisting of five lags for endogenous variables and two for exogenous variables. Meanwhile, the SVR model was developed through Grid Search to find the best parameter combination, resulting in a linear kernel with $C = 10$ and $\varepsilon = 0{,}1$. The evaluation results showed that the SVR model performed better than VARX, with MSE, RMSE, and MAPE values of 0.24, 0.49, and 6%, respectively, lower than those of the VARX model, which reached 0.68, 0.82, and 8.4%. SVR was selected as the best model and used to forecast the OUR until the end of 2025. The forecast results indicated a spike in OUR at the beginning of 2024 at 8.52%, followed by a declining trend that continues and stabilizes in the range of 7.96%-8.12% by the end of 2025. In conclusion, SVR outperforms VARX in predictive accuracy, while VARX remains useful for analyzing inter-variable relationships.
**Keywords:**
Forecasting; Open Unemployment Rate (OUR); VARX; SVR.

## Introduction

Unemployment is a crucial indicator in assessing the economic stability and social welfare of a region. A high unemployment rate indicates an imbalance between the number of job seekers and the availability of employment opportunities, as well as the inability of the economic system to absorb the existing labor force. In Indonesia, the national Open Unemployment Rate (OUR) in August 2023 was recorded at 5.32% (BPS, 2023). Meanwhile, West Java Province reported a higher rate of 7.44%, making it one of the provinces with the highest unemployment rate in the country.

The high unemployment rate in West Java is caused by various factors, such as the imbalance between the growth of the working-age population and the creation of job opportunities, the lack of workforce skills that meet industrial needs, and the impact of the COVID-19 pandemic, which has worsened labor market conditions (Baheramsyah, 2021). The government has launched several initiatives, such as training programs through Vocational Training Centers and the development of vocational education to improve job skills. However, to ensure that the policies implemented are well-targeted, accurate data and reliable forecasting models are required.

Forecasting the unemployment rate is essential for labor force planning and public policy development. One commonly used analytical approach for this purpose is time series analysis. Time series analysis is a statistical method used to evaluate data collected periodically over time, with the aim of identifying patterns, trends, cycles, and seasonal fluctuations in the data (Chatfield & Xing, 2019). This method allows for the identification of patterns, trends, and seasonality in historical data to predict future values (Hyndman & Athanasopoulos, 2021). In time series analysis, there are two main approaches: univariate and multivariate (Beeram & Kuchibhotla, 2021). The univariate approach relies on a single variable, whereas the multivariate approach incorporates multiple interrelated variables, often resulting in improved predictive accurac.

One of the multivariate methods used is the Vector Autoregressive Exogenous (VARX) model. VARX is an extension of the VAR model that allows for the analysis of dynamic relationships between endogenous and exogenous variables. This method is widely used in economics and finance due to its ability to capture the influence of external variables on key endogenous variables (Warsono et al., 2019). However, this method requires the assumption of stationarity to be fulfilled and is sensitive to the selection of the optimal lag length. In addition to classical statistical approaches, machine learning-based approaches such as Support Vector Regression (SVR) can also be used to forecast time series data. SVR is a variant of the Support Vector Machine (SVM) developed to handle regression problems. By utilizing kernel functions, SVR is capable of capturing non-linear patterns in the data and is resistant to overfitting issues (Ciaburro, 2018). A drawback of this method is the need for optimal hyperparameter tuning to ensure the model performs effectively.

Several previous studies have demonstrated the effectiveness of the VARX and SVR methods in time series forecasting. For example, Muschilati and Irsalinda (2020) applied the VARX model to forecast tourist visits in five regencies and municipalities in Yogyakarta. The results showed that the VARX (1,0) model provided reasonably good accuracy, with Mean Absolute Percentage Error (MAPE) values ranging from 6.60% to 11.19% across the regions. The lowest MAPE occurred in Bantul Regency at 6.60%, suggesting that the model is suitable for supporting tourism planning. Meanwhile, Sinaga et al. (2025) used the VARX model to forecast the prices of dragon fruit agricultural products in North Sumatra. In that study, the VARX (1,1) model yielded fairly accurate results. The MAPE values were -22% for large dragon fruit and -27.6% for small dragon fruit. Although the negative values indicate overestimation, the model was still considered informative in describing price dynamics based on the exogenous factors used.

On the other hand, the SVR method has also demonstrated excellent performance. Sadya (2022) used SVR to forecast the unemployment rate in Indonesia and found that the SVR model with parameters $\gamma = 83$ and $C = 81$ achieved high accuracy compared to other methods. These results confirm that SVR can effectively capture unemployment patterns based on historical data. Furthermore, Ishlah et al. (2023) applied SVR for stock price forecasting and found that using a linear kernel with parameters $C = 1$ and $\varepsilon = 0.11$ produced a model with minimal prediction error. This proves that SVR is not only flexible but also capable of delivering precise results in the context of volatile financial data.

This study aims to implement and compare the performance of the VARX and SVR methods in forecasting the unemployment rate in West Java Province. The results of this research are expected to provide recommendations for a more accurate forecasting method, which can be used as a basis for formulating more responsive and data-driven employment policies.

## Methods

### *Data dan Tools*

It is important to note that several variables in this study, particularly GRDP and HDI, were originally available in annual or quarterly formats and were converted to monthly observations using linear interpolation. While this approach enables consistency in the temporal structure of the data, it has inherent limitations that may affect the accuracy of the VARX model. Linear interpolation assumes a constant rate of change between known data points, which can artificially induce smoothness in the series and potentially distort the true temporal dynamics of the variables. This smoothing effect may mask short-term fluctuations or seasonal patterns that could be relevant for unemployment forecasting. Additionally, interpolated values do not represent actual observed measurements, which may reduce the model's ability to capture genuine relationships between variables. However, this limitation is less critical for the SVR model, as it focuses primarily on pattern recognition rather than structural interpretation of inter-variable relationships. Future studies may benefit from using higher-frequency original data or alternative interpolation methods such as cubic splines to better preserve the underlying dynamics of the economic indicators.

This study uses secondary data consisting of monthly observations of the Open Unemployment Rate (OUR), Labor Force Participation Rate (LFPR), Gross Regional Domestic Product (GRDP), and Human Development Index (HDI) in West Java Province. The data were obtained from official publications of Statistics Indonesia for the period from January 2018 to December 2023. Since not all variables were available in monthly format, linear interpolation was performed to standardize the time intervals across variables. The software used in this study includes EViews and Python. EViews was utilized to build the VARX model due to its capabilities in handling time series data and providing comprehensive statistical testing features, such as stationarity tests, optimal lag selection, and model diagnostics. Meanwhile, the SVR method was implemented using Python with the scikit-learn library for modeling and hyperparameter tuning, as well as pandas and numpy for data processing.

*Analysis Procedures*

*Descriptive Statistical Analysis*

Descriptive statistical analysis in this study was conducted to describe the general characteristics of each variable used. This analysis includes measures of central tendency such as the mean, as well as measures of dispersion such as the standard deviation, along with the minimum and maximum values. Through this analysis, an initial overview of the distribution and trends of the data was obtained before proceeding to further analysis.

*VARX Model*

In the VARX analysis stage, the first step is to conduct a stationarity test using the Augmented Dickey-Fuller (ADF) method. The data are considered stationary if the ADF test statistic is more negative than the critical value at the 5% significance level or if the p-value is below 0.05. If the data does not meet this criterion, differencing is applied until the data becomes stationary. Once stationarity is achieved, the optimal lag is determined based on the lowest value of the Akaike Information Criterion (AIC) to identify the best lag length for the VARX model. Next, a Granger Causality test is conducted to examine the causal relationship between endogenous and exogenous variables, where a p-value less than 0.05 indicates a significant causal relationship. Parameter estimation is then carried out to determine the coefficient values of each variable associated with the lagged values of the endogenous and exogenous variables. Based on these results, the VARX model can be constructed using the following equation.

$$z_t = \mu + \sum_{i=1}^{p} \phi_i z_{(t-i)} + \sum_{j=0}^{q} \theta_j x_{t-j} + e_t$$

where $z_t$ is the vector of endogenous variables at time $t$, $x_t$ is the exogenous variable, $\mu$ is the constant, $\phi_i$ and $\theta_j$ are the model coefficients, $p$ and $q$ are the optimal lags, and $e_t$ is the error term (Sukono et al., 2023). The residuals of the VARX model are then evaluated through several diagnostic tests, including the autocorrelation test (p-value $>$ 0.05 indicates no autocorrelation), the heteroskedasticity test (p-value $>$ 0.05 indicates no heteroskedasticity), and the normality test (p-value $>$ 0.05 indicates that the residuals are normally distributed).

*SVR Model*

Next, for the SVR method, the procedure begins with data normalization using the Min-Max Scaling method to scale the data within the range of 0 to 1. The lag structure for the SVR model was determined based on the optimal lag configuration identified by the VARX model, namely (5,2), consisting of five lags for endogenous variables and two lags for exogenous variables. This approach was adopted to maintain consistency in the temporal structure of inputs across both models and to leverage the information content suggested by the AIC-based lag selection. However, it should be acknowledged that VARX-based lag selection is optimized for linear autoregressive structures and may not necessarily represent the optimal configuration for SVR, which is capable of capturing non-linear patterns. The lag structure determines the dimensionality of the feature space: specifically, the model uses a total of 16 input features constructed as follows: OUR(t-1), OUR(t-2), OUR(t-3), OUR(t-4), OUR(t-5), LFPR(t-1), LFPR(t-2), LFPR(t-3), LFPR(t-4), LFPR(t-5), GRDP(t), GRDP(t-1), GRDP(t-2), HDI(t), HDI(t-1), and HDI(t-2). Each row in the training dataset represents a time point, with these 16 lagged values serving as independent variables (features) and OUR(t) as the dependent variable (target). Alternative lag selection methods specifically tailored for non-linear models could potentially improve SVR performance and warrant exploration in future research.

After normalization, the data is split into training and testing sets with a 70:30 ratio, without shuffling, in order to preserve the temporal order (Vrigazova, 2021). The SVR model is then trained using various parameter combinations through Grid Search and by selecting the appropriate kernel (linear, polynomial, or radial basis function). Once the optimal parameters are found, the SVR prediction function can be formulated using the following equation.

$$f(x) = \langle w.x \rangle + b$$

where $f(x)$ is the prediction function of SVR, $x$ is the input vector, $w$ is the weight vector with dimension $l$, and $b$ is the bias (Rahmawati, 2023).

*Model Accuracy*

Forecasting model accuracy is assessed to measure the precision of predictions by calculating the error between actual values and forecasted results. Three commonly used metrics are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). MSE measures the average of the squared differences between actual and predicted values, where a smaller value indicates better forecasting performance. RMSE is the square root of MSE and represents the prediction error directly, with smaller values indicating more accurate predictions. Meanwhile, MAPE measures the average percentage error relative to the actual values; thus, the smaller the MAPE value, the more accurate the model is considered to be. The calculation of each of these errors can be performed using the following formulas.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100\%$$

where $n$ is the number of data points, $Y_i$ is the actual value at time $i$, and $\widehat{Y}_i$ is the predicted value at time $i$.

*Forecasting*

The evaluation results indicating that one of the models has the lowest MSE, RMSE, and MAPE values designate it as the best-performing model for forecasting the unemployment rate. This optimal model is then used to forecast the unemployment rate for the subsequent period, namely the years 2024 and 2025. This forecasting process provides a quantitative overview of the potential future unemployment trends based on historical patterns and the dynamics of the influencing economic variables.

## Results and Discussions

*Descriptive Statistics*

The following table presents a summary of the descriptive statistical analysis results of the data used in this study.

**Table 1.** Descriptive Statistics

| Variable | Maximum Value | Minimum Value | Mean | Standard Deviation |
|---|---|---|---|---|
| Open Unemployment Rate (OUR) ($Y_1$) | 10.46 | 7.44 | 8.43 | 0.80 |
| Labor Force Participation Rate (LFPR) ($Y_2$) | 66.92 | 62.16 | 65.28 | 1.15 |
| Gross Regional Domestic Product (GRDP) ($X_1$) | 55794 | 40273 | 46847.54 | 4519.84 |
| Human Development Index (HDI) ($X_2$) | 74.37 | 71.30 | 72.70 | 0.85 |

Based on the descriptive statistical analysis presented in Table 1, the Open Unemployment Rate (OUR) during the observation period showed relatively small variation, ranging from 7.44% to 10.46%, with an average of 8.43% and a standard deviation of 0.80%. The Labor Force Participation Rate (LFPR) was relatively stable, ranging from 62.16% to 66.92%, with a mean of 65.28% and a standard deviation of 1.15%. The Gross Regional Domestic Product (GRDP) showed considerable economic variation, with values ranging from 40,273 to 55,794 billion rupiah, an average of 46,847.54 billion, and a standard deviation of 4,519.84 billion. Meanwhile, the Human Development Index (HDI) was fairly consistent, ranging from 71.30 to 74.37, with an average of 72.70 and a standard deviation of 0.85.

*Stationarity Test of the Data*

Prior to model estimation, data stationarity must be ensured. The table 2 presents the results of the data stationarity test using the Augmented Dickey-Fuller (ADF) test.

**Table 2.** Stationarity Test of the Data

| Variable | Level (*p-value*) | Diff-1 (*p-value*) | Diff-2 (*p-value*) | Description |
|---|---|---|---|---|
| OUR ($Y_1$) | 0.0380 | - | - | Stationary at Level |
| LFPR ($Y_2$) | 0.4362 | 0.1496 | 0.0105 | Stationary at Diff-2 |
| GRDP ($X_1$) | 0.9949 | 0.5250 | 0.0000 | Stationary at Diff-2 |
| HDI ($X_2$) | 0.9963 | 0.4510 | 0.0001 | Stationary at Diff-2 |

Based on the ADF test, the OUR variable is already stationary with p-value of 0.0380. In contrast, the LFPR variable becomes stationary only after the second differencing (p-value 0.0105). GRDP also shows initial non-stationarity and becomes stationary after the second differencing (p-value 0.0000). A similar result is observed for the HDI variable, which becomes stationary only after the second differencing, with p-value of 0.0001. Thus, all variables have met the stationarity assumption after the necessary differencing process.

*Optimum Lag Selection*

The optimal lag selection was carried out using the Akaike Information Criterion (AIC), where the smallest value indicates the best lag. The following table presents the optimal lag along with the corresponding AIC values.

**Table 3.** Optimum Lag along with AIC Values

| Optimal Lag | AIC |
|---|---|
| 0 | -4.276892 |
| 1 | -7.078185 |
| 2 | -7.890350 |
| 3 | -8.582111 |
| 4 | -8.562557 |
| 5 | -8.655584 |

Based on the table 3, the optimal lag is lag 5 as it has the smallest AIC value of -8.655584.

*Granger Causality Test*

To examine the causal relationships among variables, a Granger Causality test was conducted. The test indicates that most pairs of variables in the system do not exhibit causal relationships, as shown by p-values greater than 0.05. For example, between OUR and LFPR, the p-values are 0.6462 and 0.1150, respectively, indicating no causal relationship. Similar results are observed between GRDP and OUR (p-values 0.5402 and 0.3955), HDI and OUR (p-values 0.7569 and 0.4783), as well as GRDP and LFPR (p-values 0.6474 and 0.0783). However, there is a one-way causal relationship from LFPR to HDI (p-value $0.0228 < 0.05$) and from GRDP to HDI (p-value $0.0039 < 0.05$).

*Optimal Lag Selection*

The optimal lag determination for the VARX model is carried out by testing various lag combinations for both endogenous and exogenous variables, using the AIC value as the selection criterion. The lag combination that produces the lowest AIC value is considered the most suitable, as it effectively captures the dynamic relationships among variables without causing overfitting or losing important information. Based on the test results, the VARX (5,2) model yielded the lowest AIC value of -8.762824.

*Parameter Estimation of the VARX (5,2) Model*

The estimation of the VARX (5,2) model aims to determine the influence of the lagged values of endogenous and exogenous variables on the current value of the endogenous variables. Based on the estimation results, two equations are obtained for OUR and LFPR as follows.

$$\begin{aligned}
OUR_t = {} & -0.002104 + 0.664529OUR_{t-1} - 0.27932LFPR_{t-1} + 0.312512OUR_{t-2} \\
& + 0.250667LFPR_{t-2} - 0.483298OUR_{t-3} + 0.095649LFPR_{t-3} \\
& + 0.286328OUR_{t-4} - 0.048719LFPR_{t-4} - 0.437865OUR_{t-5} \\
& - 0.011688LFPR_{t-5} + 0.000350GRDP_t - 1.857166HDI_t + 0.000222GRDP_{t-1} \\
& - 0.506934HDI_{t-1} - 0.000320GRDP_{t-2} + 1.143798HDI_{t-2}
\end{aligned}$$

$$\begin{aligned}
LFPR_t = {} & -0.000103 + 0.019636OUR_{t-1} + 1.004360LFPR_{t-1} + 0.360089OUR_{t-2} \\
& + 0.383567LFPR_{t-2} - 0.212139OUR_{t-3} - 0.769551LFPR_{t-3} \\
& - 0.345655OUR_{t-4} + 0.074614LFPR_{t-4} + 0.434327OUR_{t-5} \\
& + 0.020244LFPR_{t-5} - 0.000158GRDP_t + 1.541896HDI_t - 0.0000689GRDP_{t-1} \\
& + 0.817217HDI_{t-1} - 0.0000483GRDP_{t-2} - 0.099815HDI_{t-2}
\end{aligned}$$

*Diagnostic Testing of the VARX (5,2) Model*

Diagnostic testing is conducted to ensure that the model meets classical assumptions such as the absence of autocorrelation, normality of residuals, and stationarity. Based on the Q-Q plot, the residuals of OUR and LFPR mostly follow the diagonal line, indicating that both are approximately normally distributed (Figure 1). The results of the Jarque-Bera test also support this, with p-values of 0.3317 for OUR and 0.6602 for LFPR, both greater than 0.05, indicating that the residuals of both variables are normally distributed.

*Data Preprocessing*

In the preprocessing stage, normalization and data splitting were performed. Normalization was carried out using the Min-Max method to scale the variables into the range [0,1], in order to avoid the influence of scale differences that could affect the performance of the SVR. After normalization, the data was divided into two parts: 70% for training and 30% for testing, allowing the model's accuracy to be evaluated on previously unseen data. The following is an example of Min-Max Normalization.

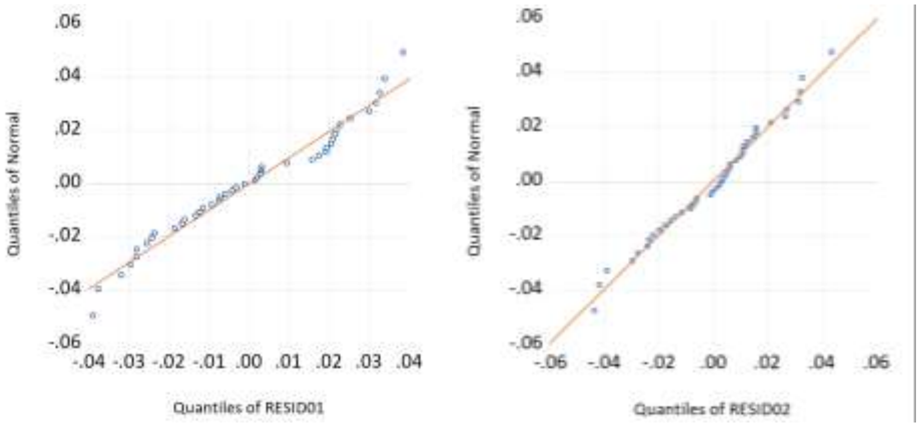


**Figure 1.** Q-Q Plot of Normal Residuals for OUR and LFPR

Min-Max Normalization for the OUR variable:

$$Y_{1(scaled)} = \frac{8 - 7.44}{10.46 - 7.44} = 0.1854$$

Min-Max Normalization for the LFPR variable:

$$Y_{2(scaled)} = \frac{65.03 - 62.16}{66.92 - 62.16} = 0.6029$$

Min-Max Normalization for the GRDP variable:

$$X_{1(scaled)} = \frac{40273 - 40273}{55794 - 40273} = 0$$

Min-Max Normalization for the HDI variable:

$$X_{2(scaled)} = \frac{71.3 - 71.3}{74.37 - 71.3} = 0$$

*Parameter Estimation of the SVR Model*

The SVR model was built with a lag structure of (5,2), consisting of five lags for the endogenous variables (OUR and LFPR) and two lags for the exogenous variables (GRDP and HDI), referring to the optimal lag from the VARX model to maintain input consistency. The training process used Grid Search to determine the best parameter combination across three types of kernels: linear, polynomial, and Radial Basis Function (RBF). The parameters tuned included the value of $C$ with test values [0.01; 0.1; 1; 10; 100], $\varepsilon$ with test values [0.01; 0.1; 1; 10; 100], and degree (for the polynomial kernel) with test values of 2 and 3.

The best results were obtained using the linear kernel with $C = 10$ and $\varepsilon = 0.1$. The model utilized a total of 16 weight features $w$ derived from the combination of lagged variables, and produced an SVR model with a bias of -0.032279. The final equation includes the weights of each lagged variable contributing to the prediction of OUR, with the largest contributions coming from the lagged values of OUR and GRDP. The following is the final equation showing the weights of each lagged variable in predicting OUR.

$$\begin{aligned} f(x) = {} & 0.812252OUR_{t-1} - 0.088997LFPR_{t-1} + 0.286469OUR_{t-2} + 0.067304LFPR_{t-2} \\ & - 0.062221OUR_{t-3} + 0.123744LFPR_{t-3} - 0.169679OUR_{t-4} \\ & + 0.077602LFPR_{t-4} - 0.068264OUR_{t-5} - 0.042254LFPR_{t-5} + 0.049792GRDP_t \\ & + 0.059831HDI_t + 0.017045GRDP_{t-1} + 0.041141HDI_{t-1} - 0.014666GRDP_{t-2} \\ & + 0.011831HDI_{t-2} - 0.032279 \end{aligned}$$

*Model Accuracy*

The actual data and forecasted results for the Open Unemployment Rate (OUR) are presented for the period from April 2022 to December 2023. To evaluate the performance of the models in forecasting the data, accuracy measurements were conducted using three indicators: MSE, RMSE, and MAPE. These three indicators are used to compare the error levels between the predictions generated by the SVR and VARX models and the actual data, in order to determine which model provides the most accurate forecasts.

Based on the table 4, it can be seen that the VARX model's forecast results closely matched the actual data in the early part of the period, for example in April 2022, with a difference of only 0.03. However, from the end of 2022 through 2023, the performance of the VARX model began to decline, as indicated by increasing deviations from the actual data and consistently lower forecast values. In contrast, the SVR model produced more stable and consistent forecast values, remaining around 8 throughout the forecasting period. Although SVR tended to slightly overestimate the actual values, it maintained a stable prediction pattern. Statistically, SVR demonstrated better evaluation performance compared to VARX. This is evident from the model accuracy metrics, where the SVR model achieved an MSE of 0.24 lower than the VARX model's 0.68. Similarly, the RMSE for SVR was 0.49, while VARX recorded 0.82. Additionally, the MAPE for SVR was smaller at 6%, compared to VARX at 8.4%. These three indicators show that the SVR model has lower error rates and better predictive performance, leading to the conclusion that SVR outperforms VARX in forecasting the OUR values during the analyzed period. To better illustrate the comparison between the forecasted values and actual data, the following graph visualizes the prediction results of the VARX and SVR models against the actual OUR data over the period.

**Table 4.** Actual Data and Forecasted OUR Results (VARX and SVR)

| No. | Time | Actual Data | Forecast (VARX) | Forecast (SVR) |
|---|---|---|---|---|
| 1. | March 2022 | 8.22 | 8.22 | 8.43 |
| 2. | April 2022 | 8.16 | 8.19 | 8.38 |
| 3. | May 2022 | 8.17 | 8.26 | 8.42 |
| 4. | June 2022 | 8.22 | 8.38 | 8.49 |
| 5. | July 2022 | 8.27 | 8.52 | 8.58 |
| 6. | August 2022 | 8.31 | 8.59 | 8.66 |
| 7. | September 2022 | 8.31 | 8.57 | 8.71 |
| 8. | October 2022 | 8.27 | 8.4 | 8.7 |
| 9. | November 2022 | 8.2 | 8.13 | 8.66 |
| 10. | December 2022 | 8.11 | 7.78 | 8.58 |
| 11. | January 2023 | 8 | 7.41 | 8.5 |
| 12. | February 2023 | 7.89 | 7.05 | 8.41 |
| 13. | March 2023 | 7.78 | 6.75 | 8.33 |
| 14. | April 2023 | 7.67 | 6.57 | 8.26 |
| 15. | May 2023 | 7.58 | 6.5 | 8.19 |
| 16. | June 2023 | 7.5 | 6.5 | 8.14 |
| 17. | July 2023 | 7.45 | 6.54 | 8.1 |
| 18. | August 2023 | 7.44 | 6.59 | 8.08 |
| 19. | September 2023 | 7.47 | 6.59 | 8.1 |
| 20. | October 2023 | 7.55 | 6.49 | 8.14 |
| 21. | November 2023 | 7.69 | 6.27 | 8.23 |
| 22. | December 2023 | 7.88 | 5.98 | 8.35 |
| **MSE** | | | 0.68 | 0.24 |
| **RMSE** | | | 0.82 | 0.49 |
| **MAPE** | | | 8.4% | 6% |

The evaluation results show that the SVR model outperforms in terms of accuracy and efficiency in predicting the Open Unemployment Rate (OUR). Subsequently, the SVR model with a linear kernel ($C = 10, \varepsilon = 0.1$) was used to forecast the OUR for the period from January 2024 to December 2025. By maintaining the same lag structure (5 for endogenous variables and 2 for exogenous variables), the forecasting was carried out iteratively using historical data up to December 2023. The following presents the OUR forecast results from the SVR model for the specified period (Figure 2).

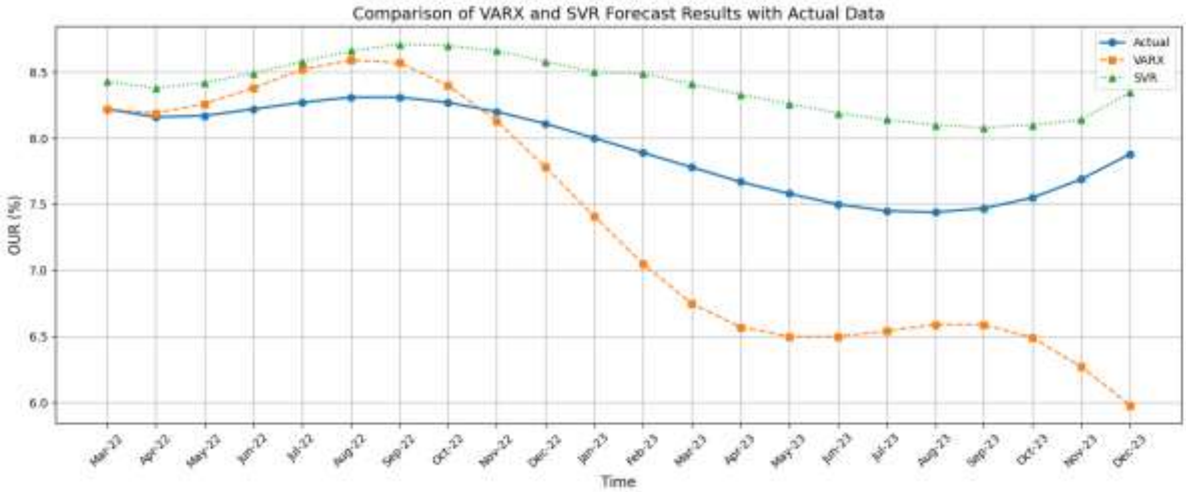


**Figure 2.** Comparison of VARX and SVR Forecast Results. The blue line represents actual OUR data, the orange line shows VARX predictions, and the green line displays SVR predictions. Shaded areas indicate the prediction error bands (±RMSE) for each model

*Forecasting Results*

**Table 5.** OUR Forecast Results

| No. | Time | Forecast Result |
|---|---|---|
| 1. | January 2024 | 8.52 |
| 2. | February 2024 | 8 |
| 3. | March 2024 | 7.83 |
| 4. | April 2024 | 7.94 |
| 5. | May 2024 | 8.12 |
| 6. | June 2024 | 8.12 |
| 7. | July 2024 | 8.12 |
| 8. | August 2024 | 8.10 |
| 9. | September 2024 | 8.08 |
| 10 | October 2024 | 8.05 |
| 11. | November 2024 | 8.03 |
| 12. | December 2024 | 8 |
| 13. | January 2025 | 7.99 |
| 14. | February 2025 | 7.98 |
| 15. | March 2025 | 7.98 |
| 16. | April 2025 | 7.99 |
| 17. | May 2025 | 7.99 |
| 18. | June 2025 | 8 |
| 19 | July 2025 | 8.01 |
| 20. | August 2025 | 8.01 |
| 21. | September 2025 | 8.01 |
| 22. | October 2025 | 8 |
| 23. | November 2025 | 7.98 |
| 24. | December 2025 | 7.96 |

Based on the SVR forecast results using a linear kernel, Open Unemployment Rate (OUR) shows a gradual decline (Table 5). In 2024, OUR, which was initially 8.52% in January, decreases to 8% by the end of the year. Entering 2025, OUR is projected to continue displaying a stable trend, ranging between 7.96% and 8.01%. Although there are slight fluctuations in certain months, overall, OUT demonstrates a declining and stabilizing pattern. To better illustrate this trend, the following chart presents a comparison between the forecasted OUR and the actual OUR data (BPS).

Based on the Figure 3, the actual OUR data (from BPS) is represented by the blue line, showing a stable downward trend from around 6.95% in January 2024 to approximately 6.74% in January 2025. On the other hand, the forecasted OUR, indicated by the orange line, consistently remains above the actual OUR values throughout the period. Throughout 2024, the forecasted values fluctuate around 8% to 8.12%, with a slight decrease to 8% in December. Entering 2025, the forecasted OUR gradually declines from 7.99% in January to 7.96% in December. In contrast, the BPS data shows a consistent decline from 6.96% to around 6.73% over the same period. This means that the gap between the forecasted results and BPS data remains in the range of 1.2 to 1.3 points throughout 2025. Overall, the actual BPS data indicates a more stable and lower trend, while the forecasted trend shows a gradual decline from early 2024 to the end of 2025, with some minor fluctuations.

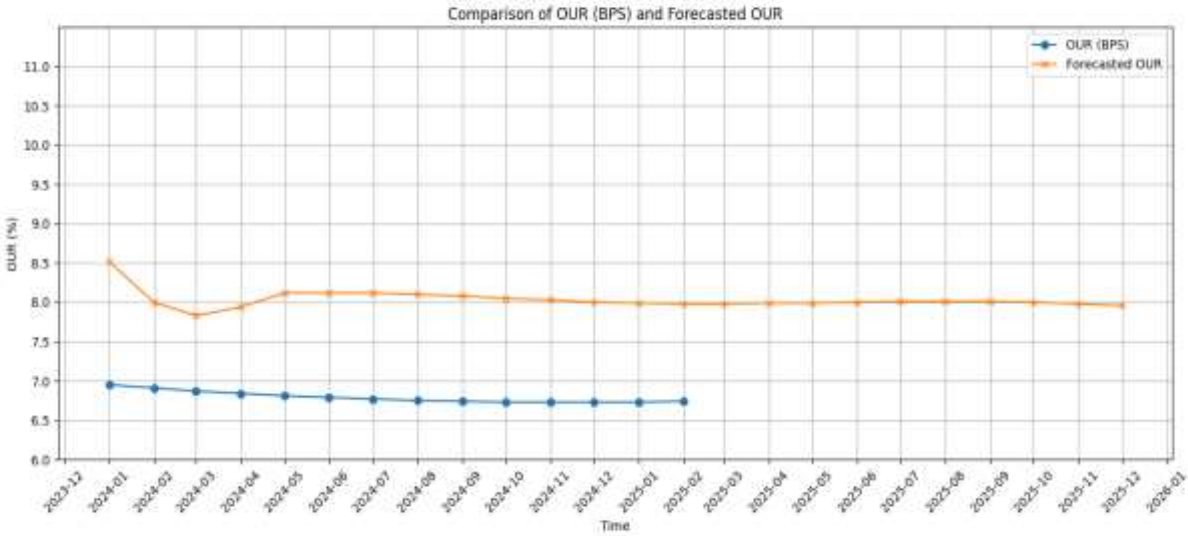


**Figure 3.** Comparison Chart of OUR (BPS) and Forecasted OUR

## Conclusion

Based on the forecasting analysis of the Open Unemployment Rate (OUR) using the VARX and SVR methods, it can be concluded that the SVR model with a linear kernel ($C = 10$, $\varepsilon = 0.1$) demonstrates better predictive performance compared to the VARX(5,2) model. This is evident from the accuracy evaluation results, where SVR achieved an MSE of 0.24, RMSE of 0.49, and MAPE of 6%, all lower than those of VARX, which recorded an MSE of 0.68, RMSE of 0.82, and MAPE of 8.4%. Therefore, SVR is considered more reliable for forecasting OUR. The SVR forecast for January 2024 to December 2025 indicates an initial decline followed by stabilization between 7.96% and 8.12%.

## Acknowledgments

The author expresses appreciation to the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang for their contributions and support during the implementation of this research. Gratitude is also extended to the reviewers for their valuable input and suggestions on this article.

## Conflicts of Interest

The author affirms that there are no conflicts of interest in this research.

## References

Badan Pusat Statistik (BPS). (2023, November 6). *Tingkat Pengangguran Terbuka (TPT) sebesar 5,32 persen dan Rata-rata upah buruh sebesar 3,18 juta rupiah per bulan*.

Baheramsyah. (2021, April 9). *BLK, Solusi Kurangi Pengangguran di Jawa Barat*. InfoPublik.

Beeram, S. R., & Kuchibhotla, S. (2021). *Time Series Analysis on Univariate and Multivariate Variables: A Comprehensive Survey*. https://doi.org/10.1007/978-981-15-5397-4_13

Chatfield, C., & Xing, H. (2019). *The Analysis of Time Series*. Chapman and Hall/CRC. https://doi.org/10.1201/9781351259446

Ciaburro, G. (2018). *Regression Analysis with R*. Packt Publishing.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

Ishlah, A. W., Sudarno, S., & Kartikasari, P. (2023). Implementasi Gridsearchcv Pada Support Vector Regression (Svr) Untuk Peramalan Harga Saham. *Jurnal Gaussian*, *12*(2), 276–286. https://doi.org/10.14710/j.gauss.12.2.276-286

Muschilati, E., & Irsalinda, N. (2020). Forecasting Tourist Visit Using the Vector Autoregressive Exogenous Method (VARX). *Jurnal Ilmiah Matematika*, *7*(2), 81. https://doi.org/10.26555/konvergensi.v7i2.19608

Rahmawati, E. (2023). *Penerapan Metode Support Vector Regression (Svr) Menggunakan Kernel Linear, Polinomial, Dan Radial Dengan Grid Search Optimization*.

Sadya, S. (2022). *Peramalan Tingkat Pengangguran di Indonesia Menggunakan Support Vector Regression dan Long Short-Term Memory*. Institut Teknologi Sepuluh Nopember.

Sinaga, S. A., Cipta, H., & Widyasari, R. (2025). Model Vector Autoregressive Exogenous (Varx) Dalam Hasil Usaha Tani Buah Naga. *Proximal: Jurnal Penelitian Matematika Dan Pendidikan Matematika*, *8*(1). https://doi.org/10.30605/proximal.v5i2.4332

Sukono, S., Ibrahim, R. A., Riaman, R., Hertini, E., Hidayat, Y., & Saputra, J. (2023). Forecasting the Effectiveness of COVID-19 Vaccination Using Vector Autoregressive with an Exogenous Variable: On the Cases of COVID-19 in Indonesia. *Discrete Dynamics in Nature and Society*, *2023*. https://doi.org/10.1155/2023/6285328

Vrigazova, B. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. *Business Systems Research*, *12*(1), 228–242. https://doi.org/10.2478/bsrj-2021-0015

Warsono, Russel, E., Wamiliana, Widiarti, & Usman, M. (2019). Vector autoregressive with exogenous variable model and its application in modeling and forecasting energy data: Case study of PTBA and HRUM energy. *International Journal of Energy Economics and Policy*, *9*(2), 390–398. https://doi.org/10.32479/ijeep.7223