International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Forecasting Semiconductor Raw Material Availability for Timely Production and Delivery Optimization

Kevin Darmawan[1], **Nina Tania Lestari**[2]*, **Winda Astuti**[3]
[1]Automotive and Robotics Program, Computer Engineering Department, BINUS ASO School of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480.
[2]Product Design Program, Industrial Engineering Department, BINUS ASO School of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480.
[3]Automotive and Robotics Program, Computer Engineering Department, BINUS ASO School of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480.
Email: [1]kevin.darmawan004@binus.ac.id, [2]nina.lestari@binus.edu, [3]wastuti@binus.edu

Corresponding Author*: nina.lestari@binus.edu

ABSTRACT:
The semiconductor industry relies on a consistent and efficient supply of raw materials, such as wafers, to meet fluctuating production demands. However, inaccurate demand forecasting often leads to inefficiencies, including overstocking, which ties up capital and increases storage costs, or understocking, which disrupts production schedules and results in lost orders. Ensuring an optimal supply of materials is critical to maintaining production efficiency, reducing waste, and improving overall profitability. To address this challenge, this study develops a forecasting model using time series analysis to predict future demand for critical raw materials. The models evaluated include ARIMA (Autoregressive Integrated Moving Average), ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables), SARIMA (Seasonal Autoregressive Integrated Moving Average), and SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables). The results indicate that the SARIMAX model outperforms the other methods by incorporating seasonal trends and external factors affecting demand, leading to more accurate predictions. This research aligns with the Sustainable Development Goals (SDGs), particularly SDG 9 (Industry, Innovation, and Infrastructure) and SDG 12 (Responsible Consumption and Production). By enhancing forecasting accuracy, semiconductor manufacturers can optimize resource allocation, minimize waste, and improve supply chain efficiency, contributing to more sustainable industrial operations. Additionally, better demand prediction ensures a stable supply of essential components, supporting technological advancements and infrastructure development. Implementing robust forecasting models enables semiconductor companies to enhance production continuity, reduce material waste, and promote sustainable manufacturing practices.
KEYWORDS: *Forecasting, ARIMA, ARIMAX, SARIMAX, SARIMA, Demand Prediction, SDG 9, SDG 12.*

INTRODUCTION

The semiconductor industry has been growing rapidly, with increasing demand for various types of wafers and raw materials [1] [2] [3][4]. However, some companies face challenges in predicting future demand accurately which leads to either overstocking, or understocking, which leads to lost orders and reduced profitability [5]. The unpredictable nature of demand fluctuations makes accurate forecasting a necessity for operating in the industry [6][7][8][9]. Although, some companies still struggle with achieving reliable accuracy, leading to inefficiencies in supply chain management.

Time series forecasting methods such as ARIMA, SARIMA, ARIMAX, and SARIMAX have been widely applied to predict demand in various industries. SARIMA is particularly known for handling seasonality and trends, while SARIMAX enhances the model by utilizing exogenous variables that may affect the forecast [10]. All methods have been successfully used in predicting various variables across multiple fields like agriculture and energy management [10][11]. Many studies have compared these models, showing mixed results depending on the dataset and the specific variables used.

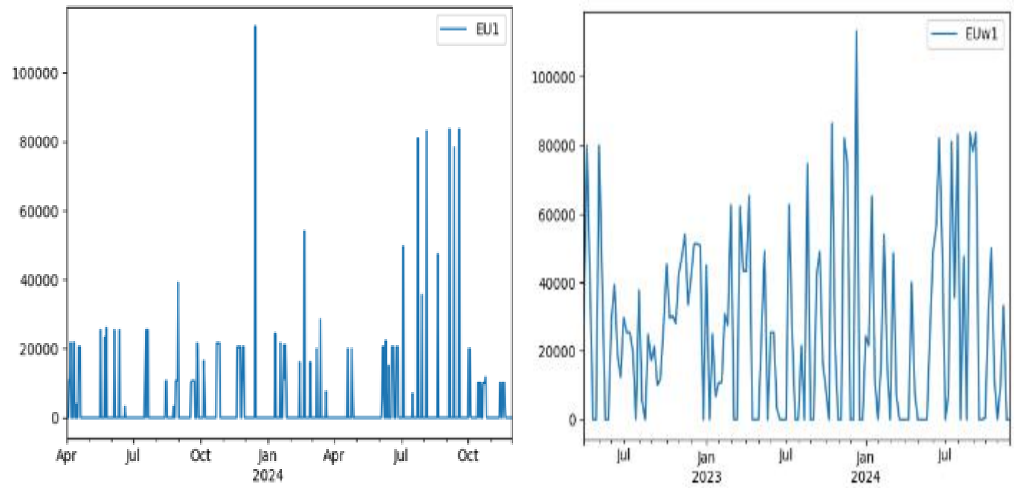
Several studies explore the comparison of SARIMAX and SARIMA in time series forecasting. For instance, a study by Jan Banaś and Utnik-Banaś (2021) applied ARIMA, SARIMA, and SARIMAX to forecast timber prices. The results show that the SARIMAX model with a MAPE of 1.46% outperforms SARIMA and ARIMA model with the accuracy of 1.90 and 2.05% respectively [12]. Another study by Palanisamy Manigandan et al. (2021) compares the SARIMAX and SARIMA model for forecasting natural gas production and consumption in the United States. The study shows that the SARIMAX

model with an MAPE of 15.93% for production and 24.36% for consumption outperforms the SARIMA model with an accuracy of 125.66% for production and 69.45% for consumption in the test set [13].
A study case where the SARIMA model sometimes outperforms the SARIMAX model is shown in another study by Junyu He et al. (2022). In the study, the SARIMAX model outperforms the SARIMA model during the model fitting stage, however the SARIMA model outperforms 6 out of 8 times during the model forecasting stage [14]. Another case about forecasting international passenger arrivals by Wai Hong Kan Tsui and Faruk Balli (2015) shows a case where the SARIMAX model sometimes outperform the SARIMA model, the SARIMA model sometimes outperforms the SARIMAX model, and both the SARIMA and SARIMAX model having the same accuracy [15]. The study case about forecasting the monthly average rainfall for Kogi State by Ibrahim A. and Musa A. O. (2023) shows that the SARIMAX model slightly shows a better accuracy of 81.92% over the SARIMA model with an accuracy of 81.86% [16]. Finally, the study case where an ARIMAX model performs better than the ARIMA model is shown in the study by Wiwik Anggraini et al. (2015) where including the variation of calendar effect in the form of dummy variables make the ARIMAX model outperform the ARIMA model [17].
In this work, the time series forecasting model applied to predict the demand for some certain wafer material over the next 6 months, by using the ARIMA, ARIMAX, SARIMAX, and SARIMA models. The models are developed to determine which method is better suited for the specific demands of the semiconductor industry. Based on the experimental results, the hypothesis is that the SARIMAX is more likely to be preferable in this case based on the studies above.

2. METHODOLOGY

2.1 Data

In this work, the data is provided by a certain anonymous company that will be referred to as Company X. The data received is about daily demands of certain wafer materials that came from Europe, which will be referred to as EU1. In the system, Company X needs a number to pre-order the wafer material 6 months before receiving it from Europe. However, the amount of demand required for the wafer materials is varying every day. Therefore, with the current data, time series models will be created to forecast into the future.
The graph shown in Figure 1 (a) shows the daily data of EU1 from April of 2022 to November of 2024. It consists of 975 individual data with a lot of random and irregular data with the presence of a lot of zeros.
The 0s in the data may cause the SARIMA and SARIMAX model to struggle and since the 0s are actual real data and not missing data, therefore getting rid of them will be data lost. To reduce the amount of 0s in the data, the data is converted into weekly data instead of daily data to lessen the amount of 0s by combining the data.
When converted into weekly data, the dataset is reduced to 140 and the 0s are reduced as shown in Figure 1 (b). Transforming the daily dataset into a weekly dataset reduces the random noise and simplifies the models. Seasonality that was previously hard to detect due to daily fluctuations, may also be more visible.



(a) (b)

Fig. 1. Daily and Weekly Data (a) Daily Data (b) Weekly Data

2.2 Time Series Models

Time Series Models are often used to do forecasting in the future and some common statistical models that are used for forecasting are the SARIMA (Seasonal Autoregressive Integrated Moving Average) and SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) models which are extensions of the ARIMA (Autoregressive Integrated Moving Average) model [18].

2.3 ARIMA

The ARIMA model is one of the most used models when it comes to time series forecasting with statistical models [11]. The ARIMA model can be expressed as Eq. (1).

$$\phi_p(G)(1-G)^d X_t = \gamma_q(G)e_t \tag{1}$$

From Eq. (1), $\phi_p(G)$ is the AR equal to as shown as in Eq. (2), $\gamma_q(G)e_t$ is the MA equal to as shown as in Eq. (3), and the integration is represented in Eq. (4).

$$1-\phi_1 G-\phi_2 G^2-\phi_3 G^3-\cdots-\phi_p G^p \tag{2}$$

$$1-\gamma_1 G-\gamma_2 G^2-\gamma_3 G^3-\cdots-\gamma_p G^p \tag{3}$$

$$(1-G)^d \tag{4}$$

2.4 SARIMA

SARIMA is an extension of the ARIMA model, which adds a seasonal parameter to find seasonal patterns [13]. It is represented as Eq. (5) in which p (AR) is the number of autoregressive terms that looks for the relationship between an observation and its previous values, d (I) is the number of differentiation used to make series stationary by subtracting previous values from the current ones, and q (MA) is the number of moving average terms, in which the relation between an observation and a residual error from a moving average model is applied to lagged values.

$$\phi_p(G)\phi_P(G^s)(1-G^s)^d(1-G^s)^D X_t = \gamma_q(G)\omega_Q(G^s)e_t \tag{5}$$

The P, D, and Q are the seasonal autoregression ($\phi_P(G^s)$) as shown as in Eq. (6), the seasonal differentiation shown as Eq. (7), and the seasonal moving average ($\omega_Q(G^s)$) equivalent to Eq. (8), but with seasonality of s, which is the number of periods in each seasonal cycle.

$$1+\phi_1 G^s+\phi_2 G^{2s}+\phi_3 G^{3s}+\cdots+\phi_P G^{Ps} \tag{6}$$

$$(1-G^s)^D \tag{7}$$

$$1+\theta_1 G^s+\theta_2 G^{2s}+\theta_3 G^{3s}+\cdots+\theta_Q G^{Qs} \tag{8}$$

2.5 ARIMAX and SARIMAX

ARIMAX and SARIMAX, is an extension of the ARIMA and SARIMA model that includes exogenous variables (X) from Eq. (9). ARIMAX is represented from Eq. (10), while SARIMAX is represented from Eq. (11) [11][17].

$$\beta_k X_{k,t} \tag{9}$$

$$\phi_p(G)(1-G)^d X_t = \beta_k X_{k,t} + \gamma_q(G)e_t, \quad (10)$$

$$\phi_p(G)\phi_P(G^s)(1-G^s)^d(1-G^s)^D X_t = \beta_k X_{k,t} + \gamma_q(G)\omega_Q(G^s)e_t. \quad (11)$$

Exogenous variables are additional variables outside of the target variable that influence the predictions. With secondary data, ARIMAX and SARIMAX can do more complex forecasting problems. The X variable in the research uses 4 main exogenous variables which are the stock data of Sat Nusapersada Tbk PT, Galva Technologies Tbk PT, Telkom Indonesia, and the exchange rate of IDR to USD.

2.6 Autocorrelation and Partial Autocorrelation Functions

The autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) are statistical tools commonly used to find the parameters for the Moving Average (MA) and Auto Regression (AR) value from the ARIMA model. The ACF determines the correlation between the observed value and the time lagged and shows how strongly the current value X_t is related to past values X_{t-k}, where k is the value of lag. The PACF measures the correlation between X_t and X_{t-k}, while removing the influence of intermediate lags [19].

2.7 Augmented Dickey Fuller Test

For the ACF and PACF functions to be determined to have a correlation with the corresponding data, the Augmented Dickey Fuller (ADF) test is implemented. The ADF test is a statistical procedure used to determine if the data is stationary or non-stationary. If the probability value (p-value) of the results show that it is below 0.05, then the null hypothesis is rejected and so the data is stationary. However, if the p-value from the data shows that it is above 0.05, then the null hypothesis is unable to be rejected, and the data is non-stationary [11][20].

2.8 Parameters

From the ACF and PACF functions shown in Figure 2, the parameters that are used for the ARIMA and ARIMAX models are ARIMA (2,0,2) and ARIMAX (2,0,2)x based on the value on lag 2. The ACF and PACF plot does not show that there is seasonality, however when finding the best parameters through the method of grid search, SARIMA (2,0,2) $(2,0,2)^{11}$ shows a similar result and SARIMAX (2,0,2) $(2,0,2)^{11}$x gives a more consistent accuracy. ARIMA (1,1,2) and a couple more parameters also show to be a potentially viable variable, however the p-value on differentiation zero shows 0.015 which means that the data is stationary. However, for comparison, ARIMA (1,1,2), ARIMAX (1,1,2)x, SARIMA (1,1,2) $(2,0,2)^{11}$, and SARIMAX (1,1,2) $(2,0,2)^{11}$x from Figure 3 is compared with due to some similarities on the prediction values.

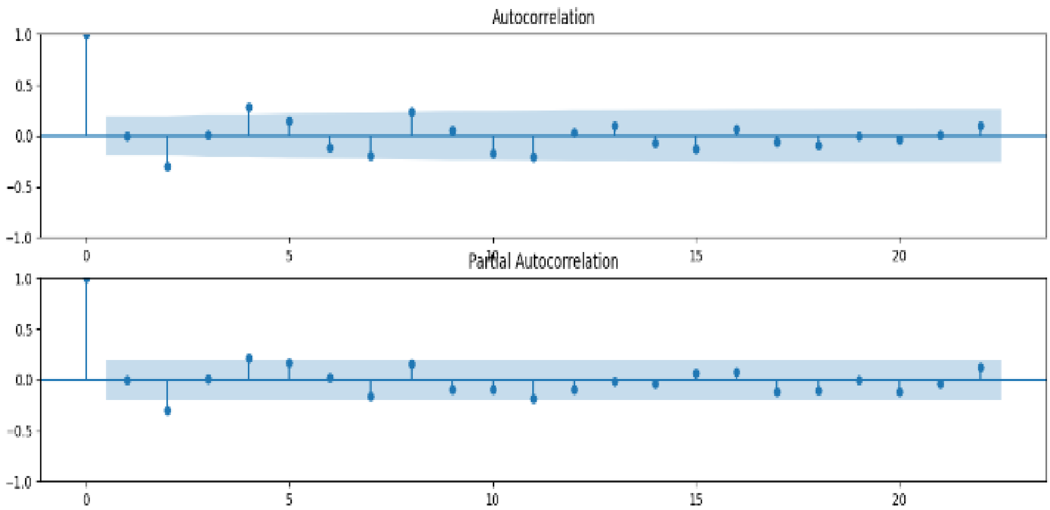


Fig. 2. ACF and PACF Plot, i=0

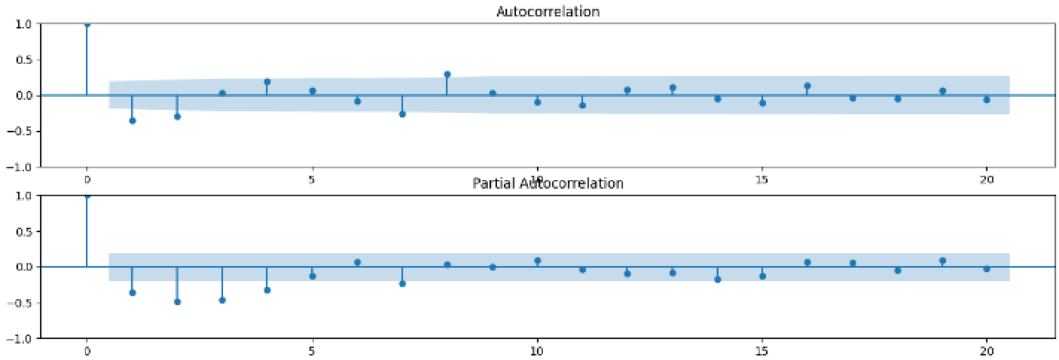


Fig. 3. ACF and PACF Plot, i=1

2.9 Forecasting

After getting the parameters and fitting the models, the output of the forecast is then able to be compared to the testing data and the forecasting data from Company X after a few adjustments. First, the models will forecast 26 steps ahead of time to predict 6 months ahead. Then the data that is forecasted is to be converted into monthly data. After getting the monthly data, it is then possible to get the accuracy of each individual model compared to the real test data and the forecasted data from Company X.

3. RESULTS

3.1 Forecast Results

The parameters from the proposed systems are then used to do forecasting into the testing data for comparison. There are a couple of different parameters used and three different months of output, September, October, and November, in which all of them are the number of demands forecasted in the sixth month.

Figure 4, 5, 6, 7, 8, and 9 shows the testing data, Company X's forecasting data, SARIMAX (2,0,2) (2,0,2)11x shown as SX1, SARIMA (2,0,2) (2,0,2)11 shown as S1, ARIMAX (2,0,2)x shown as AX1, ARIMA (2,0,2) shown as A1, SARIMAX (1,1,2) (2,0,2)11x shown as SX2, SARIMA (1,1,2) (2,0,2)11 shown as S2, ARIMAX (1,1,2)x shown as AX2, and ARIMA (1,1,2) shown as A2.

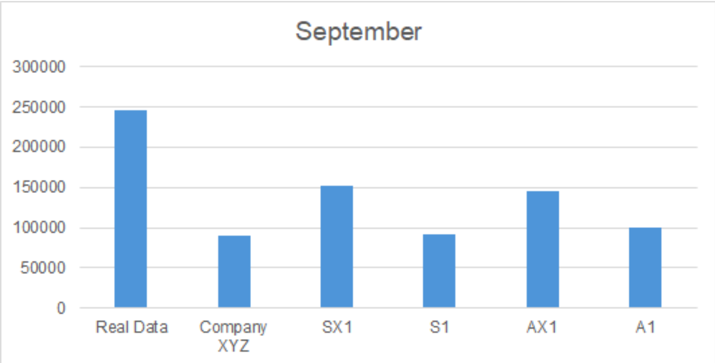


Fig. 4. September Forecast 1

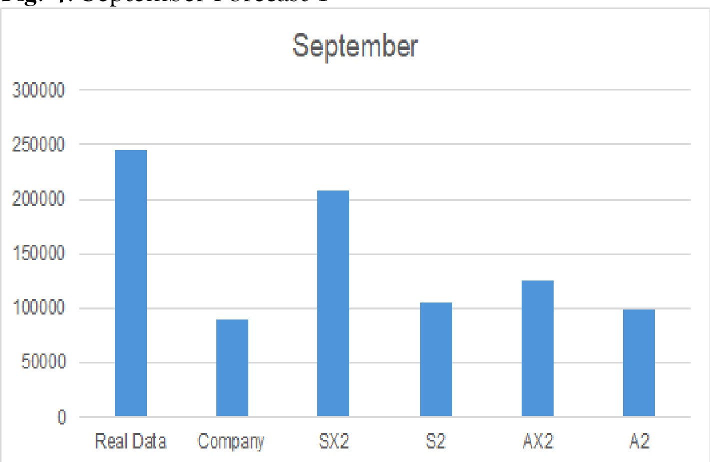


Fig. 5. September Forecast 2

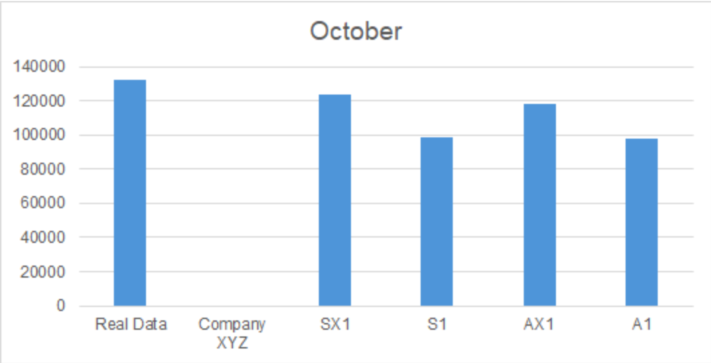


Fig. 6. October Forecast 1

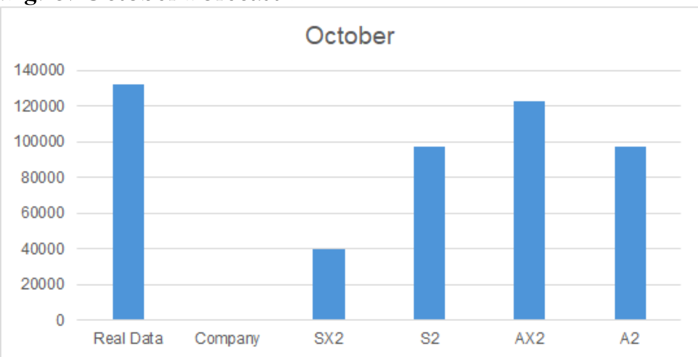


Fig. 7. October Forecast 2

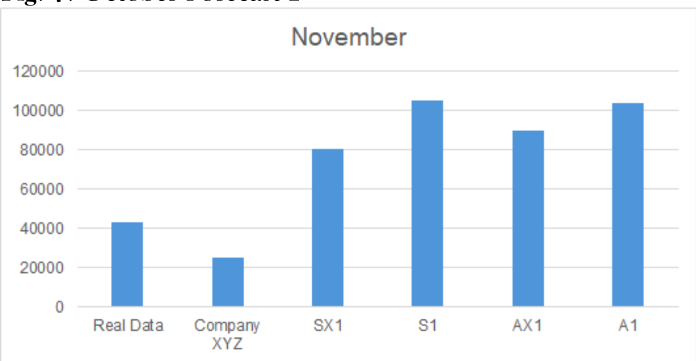


Fig. 8. November Forecast 1

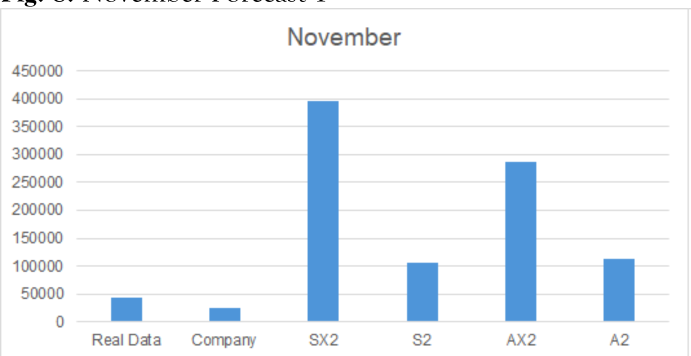


Fig. 9. November Forecast 2

The forecasting compared to the error of Company X's is defined by using the formula

$$\text{Improvement Percentage} = \frac{\text{Ex} - \text{En}}{\text{Ex}} * 100, \qquad (11)$$

where Ex is the error from Company X's prediction, and En is the error from the other predictions. With the formula, the improvement percentage is calculated from the total error of September, October, and November.

The forecasting result shown in Table 1, SARIMAX (2,0,2) (2,0,2)11x results in a 54.43% accuracy improvement when compared to the prediction from Company X even when the company's prediction has the least amount of error in November when compared to the others. Similarly, ARIMAX (2,0,2) x improves the prediction by 47.54% compared to the company prediction. SARIMA (1,1,2) (2,0,2)11 is the next highest improving model with a 22.23% accuracy improvement, followed by ARIMA (2,0,2) with a 20.92% accuracy improvement, SARIMA (2,0,2) with an 18.43% accuracy improvement, ARIMA (1,1,2) with an 18.04% improvement. There are also worse models than the Company prediction which are ARIMAX (1,1,2)x and SARIMAX (1,1,2) (2,0,2)11x with a decrease in performance of 21.72% and 57.87% respectively.

Table 1 Root Mean Squared Error of the Forecasts

Month	Company X	SX1	S1	AX1	A1	SX2	S2	AX2	A2
September	155520	93014	153683	99532	146261	37451	140725	119568	146312
October	132300	8684	33778	14283	34694	92433	35117	9800	34889
November	17820	37595	61847	46532	60742	352622	61852	242649	69305
Total Error	305640	139293	249307	160347	241697	482506	237694	372017	250506
Improvement Percentage	0%	54.42%	18.43%	47.54%	20.92%	-57.87%	22.23%	-21.72%	18.04%

4. CONCLUSION

In conclusion, this work focuses on forecasting the demand for wafer materials on a specific company, where the number six months ahead is needed to know for a precise pre-order number. Through comparing the error of the forecasting prediction of multiple time series forecasting models with the company data and prediction, the implementation of the time series forecasting models resulted in some forecasts being very close and better than the company's predictions. From several tests from the models, SARIMAX (2,0,2) (2,0,2)11x showed the best results for the predictions, however there are also results that show a lot of errors, for instance, SARIMAX (1,1,2) (2,0,2)11x which misses the prediction on November by 352622 demands. The results imply that the exogenous variable when added with seasonality has the better overall prediction when the data is complex and non-linear, with limited amounts of data. Overall, this work contributes towards the forecasting of data, where in the long term it may have a positive impact when compared to forecasting data manually. For future studies, more data and data types may be implemented, as well as adding a proper user interface for a more reliable way to read the forecast.

Acknowledgement

Appreciation towards the faculty of Automotive and Robotics and Product Design Engineering at BINUS ASO School of Engineering, Bina Nusantara University for the support given throughout this work. The author expresses his deepest gratitude Ir. Winda Astuti, S.T., M.Sc., Ph.D., as the academic supervisor, who assisted and provided guidance throughout the entire work and ensuring the work being a success, and Nina Tania Lestari, S.T., M.T., as the academic supervisor, who also assisted with data and guided the author throughout the work.

REFERENCES

[1] Ochonogor, Kenechukwu Nwuka, Gbolahan Solomon Osho, Cyril O Anoka, and Oluwagbemiga Gbenga Ojumu. "The COVID-19 Pandemic and Supply Chain Disruption: An Analysis of the Semiconductor Industry's Resilience." Vol. 6, 2023. www.ijtsre.orgwww.ijtsre.org.

[2] Padwal, Prasad. "Semiconductor Chip Market Size, Share Industry & Industry Growth 2030," June 2023.

[3] Choi, Jaehee. "Growth of the Chinese Semiconductor Industry," July 2023. https://doi.org/10.58445/rars.252.

[4] Samarasinghe, Tharanga. "Review of the Global Semiconductor Industry: Tharanga Samarasinghe Review of the Global Semiconductor Industry," 2024.

[5] Alnahhal, Mohammed, Batin Latif Aylak, Muataz Al Hazza, and Ahmad Sakhrieh. "Economic Order Quantity: A State-of-the-Art in the Era of Uncertain Supply Chains." Sustainability (Switzerland). Multidisciplinary Digital Publishing Institute (MDPI), July 1, 2024. https://doi.org/10.3390/su16145965.

[6] Gerlach, Richard, Garth Tarr, Eric J Beh, Mahdi Abolghasemi, and Eric Beh. "Demand Forecasting in Supply Chain: The Impact of Demand Volatility in the Presence of Promotion," 2019. https://doi.org/10.48550/arXiv.1909.13084.

[7] Taylor, Sean J, and Benjamin Letham. "Forecasting at Scale," September 27, 2017. https://doi.org/10.7287/peerj.preprints.3190v2.

[8] Syntetos, Aris A., and John E. Boylan. "The Accuracy of Intermittent Demand Estimates." International Journal of Forecasting 21, no. 2 (April 2005): 303–314. https://doi.org/10.1016/j.ijforecast.2004.10.001.

[9] Ho ' ˜, S L, and M Xie. "THE USE OF ARIMA MODELS FOR RELIABILITY FORECASTING AND ANALYSIS." Computers Ind. Engng. Vol. 35, 1998.

[10] Perez-Guerra, Uri H., Rassiel Macedo, Yan P. Manrique, Eloy A. Condori, Henry I. Gonzáles, Eliseo Fernández, Natalio Luque, Manuel G. Pérez-Durand, and Manuel García-Herreros. "Seasonal Autoregressive Integrated Moving Average (SARIMA) Time-Series Model for Milk Production Forecasting in Pasture-Based Dairy Cows in the Andean Highlands." PLoS ONE 18, no. 11 November (November 1, 2023). https://doi.org/10.1371/journal.pone.0288849.

[11]Alharbi, Fahad Radhi, and Denes Csala. “A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach.” Inventions 7, no. 4 (December 1, 2022). https://doi.org/10.3390/inventions7040094.
[12]Banaś, Jan, and Katarzyna Utnik-Banaś. “Evaluating a Seasonal Autoregressive Moving Average Model with an Exogenous Variable for Short-Term Timber Price Forecasting.” Forest Policy and Economics 131 (October 1, 2021). https://doi.org/10.1016/j.forpol.2021.102564.
[13]Manigandan, Palanisamy, MD Shabbir Alam, Majed Alharthi, Uzma Khan, Kuppusamy Alagirisamy, Duraisamy Pachiyappan, and Abdul Rehman. “Forecasting Natural Gas Production and Consumption in United States-Evidence from Sarima and Sarimax Models.” Energies 14, no. 19 (October 1, 2021). https://doi.org/10.3390/en14196021.
[14]He, Junyu, Xianyu Wei, Wenwu Yin, Yong Wang, Quan Qian, Hailong Sun, Yuanyong Xu, Ricardo J. Soares Magalhaes, Yuming Guo, and Wenyi Zhang. “Forecasting Scrub Typhus Cases in Eight High-Risk Counties in China: Evaluation of Time-Series Model Performance.” Frontiers in Environmental Science 9 (January 12, 2022). https://doi.org/10.3389/fenvs.2021.783864.
[15]Tsui, Wai Hong Kan, and Faruk Balli. “International Arrivals Forecasting for Australian Airports and the Impact of Tourism Marketing Expenditure.” Tourism Economics 23, no. 2 (2017): 403–28. https://doi.org/10.5367/te.2015.0507.
[16]Ibrahim, A., and A. O. Musa. “ON THE PERFORMANCE OF SARIMA AND SARIMAX MODEL IN FORECASTING MONTHLY AVERAGE RAINFALL IN KOGI STATE, NIGERIA.” FUDMA JOURNAL OF SCIENCES 7, no. 6 (December 19, 2023): 24–31. https://doi.org/10.33003/fjs-2023-0706-2095.
[17]Anggraeni, Wiwik, Retno Aulia Vinarti, and Yuni Dwi Kurniawati. “Performance Comparisons between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study.” In Procedia Computer Science, 72:630–637. Elsevier, 2015. https://doi.org/10.1016/j.procs.2015.12.172.
[18]Zhang, G Peter. “Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model.” Neurocomputing. Vol. 50, 2003. www.elsevier.com/locate/neucom.
[19]Dong, Yichen, Siyi Li, and Xueqin Gong. “Time Series Analysis: An Application of ARIMA Model in Stock Price Forecasting” 2017. https://doi.org/10.2991/iemss-17.2017.140
[20]Manoj, Kumar. “AN APPLICATION OF TIME SERIES ARIMA FORECASTING MODEL FOR PREDICTING SUGARCANE PRODUCTION IN INDIA,” April 2014. 9(1):81 – 94

Author Contributorship
Kevin Darmawan contributed to Research Design and Development, and Analysis, Manuscript Preparation; **Nina Tania Lestari** contributed to Assisting in Research Design, Data Collection; **Winda Astuti** contributed to Assisting in Research Design, Collaborative Efforts.

Data Availability
The Authors: Kevin Darmawan, Nina Tania Lestari, Winda Astusti, we demonstrate our commitment to data openness and transparency. To facilitate further research, we have made the data used in their study publicly available.
The data used by the author can be opened via the link below:
- https://drive.google.com/drive/folders/1_N92wEyq2Ta2o7M_78VaLYZgl-QhcL14?usp=drive_link