

# Optimization of Sales Turnover Forecasting with the SARI-MAX Model Approach to Predict Revenue Trends Case Study: Sun Thai Tea

Putri Awaliyah Rahmah<sup>1\*</sup>, Novriyenni<sup>2</sup>, Melda Pita Uli Sitompul<sup>3</sup>

<sup>1,2,3</sup>*Informatic Engineering, STMIK Kaputama*  
[putriawaliyar@gmail.com](mailto:putriawaliyar@gmail.com)<sup>1\*</sup>, [Novriyenni.sikumbang@gmail.com](mailto:Novriyenni.sikumbang@gmail.com)<sup>2</sup>, [meldasitompul19@gmail.com](mailto:meldasitompul19@gmail.com)<sup>3</sup>

## Abstract

This study aims to apply the SARIMAX model to forecast the daily revenue of Sun Thai Tea by incorporating external variables such as weather and promotions, and to compare its performance with the SARIMA model. The daily sales data used exhibit a weekly seasonal pattern and are influenced by external factors, making SARIMAX a suitable choice due to its ability to accommodate exogenous variables. The SARIMAX model employs optimal parameters obtained through grid search, with an order of (2,1,0) and a seasonal order of (0,1,0,7), resulting in a Mean Absolute Percentage Error (MAPE) of 25.06%, which is more accurate compared to the SARIMA model's MAPE of 68.54%. The findings indicate that SARIMAX, when configured with appropriate parameters, can provide more accurate daily revenue predictions and support more efficient operational decision-making.

**Keywords:** Turnover; Forecasting; SARIMAX; SARIMA; Exogenous Variables

## 1. Introduction

This Sun Thai Tea is an MSME in the field of beverage sales located on Proklamasi Simpang Harapan Street, Stabat District, Langkat Regency. In its operations, the business has not implemented adequate data analysis to identify sales trends or predict turnover, so production decisions still rely on previous daily revenue and are less responsive to changes in demand. The financial report for the April-November 2024 period recorded an average daily turnover of IDR 5,263,388.31. The lowest turnover was recorded on August 23, 2024 of IDR 1,412,000, and the highest on August 13, 2024 of IDR 11,901,500. These fluctuations create uncertainty in production planning and in some cases cause losses, with the highest value being IDR 1,241,250 on August 14, 2024. This condition shows the need for a more effective strategy to maintain financial stability and improve operational efficiency.

In modern business, sales forecasting has become a critical component for operational efficiency and strategic decision-making. Accurate forecasting helps with marketing planning, stock management, and production adjustments to market demand. According to [1] proper forecasting allows companies to anticipate demand, control costs, and improve logistics efficiency. In addition, according to [2] forecasting accuracy is highly influential in resource allocation and business strategies that are adaptive to market dynamics. One of the commonly used methods in time series forecasting is SARIMA (*Seasonal Autoregressive Integrated Moving Average*) which combines AR, MA, and differentiating components, as well as seasonal elements. However, this method does not consider the influence of external variables on sales. To overcome this, the SARIMAX (*Seasonal Autoregressive Integrated Moving Average with Exogenous Variables*) model was developed which adds exogenous variables to improve the accuracy of predictions.

The superiority of SARIMAX in capturing seasonal patterns and external variables is supported by research by [3] which shows that this model is better than SARIMA in predicting short-term rainfall in Payerageung, Tasikmalaya, with an MSE of 110.89, a MAE of 86.22, and a MAPE of 44.05% at a 95% confidence level. This advantage is also reinforced by a study [4] that tested SARIMAX in predicting monthly natural gas production and consumption in the United States. Evaluation using RMSE and MAPE shows that SARIMAX is superior in capturing seasonal patterns and the influence of external variables, so it is able to predict trends until 2025. If this seasonal pattern continues, it is estimated that natural gas production will increase by 16% and consumption by 24% by 2025. These findings provide important inputs for future sustainable energy planning and policy.

## 2. Literature Review

### 2.1. Forecasting

Forecasting is an analytical process used to predict future events based on historical data and information available today. According to [5] forecasting serves as a key tool in strategic decision-making to minimize uncertainty and maximize efficiency in various fields, including business, economics, and science. Forecasting theory is based on the principle that patterns and trends in historical data can be used to make reliable predictions about the future. [6] explained that the forecasting method seeks to capture the relationship in deterministic and stochastic data in order to estimate future values with various quantitative and qualitative approaches. According to [7], there are several basic principles in forecasting that must be considered to improve model accuracy, including:

1. Stationarity in Time Series Data Analysis

In the analysis of time series data, one of the main characteristics to consider is stationarity. Data that is stationary has a relatively stable variability over time, both in terms of mean and variance.

According to [8], time series data is said to be stationary if its statistical characteristics (mean, variance, and autocorrelation) remain constant all the time. Meanwhile, according to [8] a non-stationary time series is characterized by a trend or change in variance over time, which causes predictive models to become unstable.

2. Autocorrelation and Pattern Identification in Data

One of the important aspects of time series analysis is autocorrelation, which is the relationship between the value of data in a period and the previous period. Autocorrelation can be helpful in understanding how past data affects future values, thus allowing for more accurate forecasting. Using tools such as *the Autocorrelation Function (ACF)* and *the Partial Autocorrelation Function (PACF)*, patterns contained in the data can be identified, including the correlation between specific periods. This information is very useful in determining the parameters of a forecasting model such as ARIMA or SARIMA.

3. Identify Seasonal and Cyclical Patterns in Data

Many time series data show repeating patterns over a given period, known as *seasonality* and *cycle*. Seasonal patterns are often short-term and fixed, such as increased sales during the holidays or fluctuations in the number of customers in the monthly cycle. Meanwhile, cyclical patterns usually occur in the long term and are more influenced by economic factors or business trends. Forecasting models such as SARIMA (*Seasonal ARIMA*) are specifically designed to capture seasonal patterns in data, allowing for more accurate analysis of repetitive patterns occurring in a given period.

4. Evaluation of the Accuracy of the Forecasting Model

The success of a forecasting model is determined not only by its complexity but also by the accuracy of the forecast produced.

To assess the extent to which the model is able to predict actual data, several evaluation metrics are used, including, *The Mean Absolute Percentage Error (MAPE)*, which measures the percentage of error between the actual value and the predicted value, with the smaller the MAPE value indicates a more accurate model. As a measure of forecasting accuracy, MAPE has criteria that indicate that the smaller the MAPE value, the higher the level of forecasting accuracy [9].

**Table 1:** MAPE Accuracy Criteria

MAPE Values	Accuracy Criteria
<10%	Highly Accurate Forecasting Models
10% - 20%	Accurate Forecasting Models
20% - 50%	Forecasting Models Are Quite Accurate
>50%	Inaccurate Forecasting Models

Source : [9]

The entire document should be in Times New Roman. The font sizes to be used are specified in Table 1.

### 2.2. Forecasting Methods

According to [9], forecasting methods can be divided into two main approaches, namely:

1. Quantitative Forecasting

Quantitative forecasting uses mathematical and statistical techniques based on historical data to make predictions. This approach is often applied to data with clear patterns and having a sufficient number of observations to be analyzed numerically.

Quantitative forecasting methods consist of a variety of models, including:

- a) *Autoregressive Integrated Moving Average (ARIMA)*

This model is used to analyze time series data without seasonal patterns by combining autoregressive (AR), *differentiating (I)*, and *moving average (MA)* components to capture patterns in historical data.

- b) *Seasonal ARIMA (SARIMA)*

As an extension of ARIMA, this model considers seasonal patterns in the data to produce more accurate predictions in cases with periodic fluctuations.

- c) *Seasonal ARIMA with Exogenous Variables (SARIMAX)*

This model further develops SARIMA by incorporating external variables that can affect forecasting patterns, such as economic factors, advertising, or market trends.

- d) *Exponential Smoothing Method*

This technique prioritizes forecasting based on weighted exponential averages from historical data, making it more responsive to recent trend changes.

#### e) Time Series Regression Model

This approach uses statistical regression to predict future values based on the relationship between the target variable and other influential factors.

According to [8], quantitative methods are superior to data-driven forecasting because they produce more objective results and can be evaluated using predictive error metrics such as *Mean Absolute Percentage Error* (MAPE), *Root Mean Square Error* (RMSE), and *Mean Absolute Error* (MAE).

### 2. Qualitative Forecasting

Qualitative forecasting relies more on the intuition, experience, and opinions of experts in making predictions, especially when historical data is not available or its relevance is limited. This approach is used in situations where market changes are difficult to predict using historical data alone, such as when facing technological changes, new regulations, or rapidly evolving social trends.

### 2.3. Model SARIMA (*Seasonal Autoregressive Integrated Moving Average*)

*Seasonal Autoregressive Integrated Moving Average* (SARIMA) is a time series forecasting model that is an extension of ARIMA with mempertimbangkan pola seasonality in the data. According to [5] this model is particularly suitable for use in data that have periodic fluctuations, such as seasonal sales or weather patterns.

SARIMA is denoted as

$$\text{SARIMA}(p, d, q) \times (P, D, Q, s) \quad (1)$$

with:

- $p$  : Order of *autoregressive* (AR) components.
- $d$  : Degree of differentiation to create stationary data.
- $q$  : Order of the moving *average* (MA) component.
- $P, D, Q$  : Seasonal components of AR, *differentencing*, and MA.
- $s$  : Seasonal periods in the data (for example, 12 for monthly data, 7 for weekly daily data).

According to research [6] the SARIMA model works effectively in overcoming seasonal patterns by performing seasonal differentiation to eliminate periodic trends before carrying out the forecasting process. This allows the model to capture long-term trends and seasonal patterns simultaneously.

### 2.4. Model SARIMAX (*SARIMA with Exogenous Variables*)

*Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX) is a development of SARIMA that allows the existence of exogenous variables (X) in the model. According to [8] this approach is most effective when there are external factors that affect the target variable, such as weather, promotion, or economic policy.

SARIMAX is notated as

$$\text{SARIMAX}(p, d, q) \times (P, D, Q, s) + X_t \quad (2)$$

with:

- $p$  : Order of *autoregressive* (AR) components.
- $d$  : Degree of differentiation to create stationary data.
- $q$  : Order of the moving *average* (MA) component.
- $P, D, Q$  : Seasonal components of AR, *differentencing*, and MA.
- $s$  : Seasonal periods in the data (for example, 12 for monthly data, 7 for weekly daily data).
- $X_t$  : Exogenous variables that affect the prediction of dependent variables

According to [10] SARIMAX works better than SARIMA in situations with: seasonal patterns are not only influenced by internal factors but also by external factors such as economic or weather changes.

The SARIMAX model has the main advantage in considering external variables that affect the data being predicted [5]. By incorporating external factors, SARIMAX is able to significantly improve forecasting accuracy compared to SARIMA [11]. The model is also more flexible in a variety of sectors, including retail, finance, and energy, with external factors often playing a role dalam tren data historis [7]. Namun, SARIMAX has a higher level of complexity than SARIMA because it requires the selection of appropriate exogenous variables and additional analysis to understand their impact on the target variable [6]. In addition, these models require a greater amount of data to ensure a valid relationship between exogenous variables and target variables, so in some cases, the use of SARIMAX may not always provide a significant increase in accuracy if the selected exogenous variable does not have a strong correlation with the target variable [5].

### 2.5. Variabel

A variable in a study is a characteristic, attribute, or measure that can change and affect the results of the research. According to [11], variables are grouped into two main types, namely independent variables and dependent variables, which have causal relationships in statistical and experimental analysis. Dependent variables are variables that are predicted or described in a study or statistical model. In the context of time series forecasting, dependent variables are the main variables that want to be predicted based on historical data. According to [11], dependent variables are influenced by internal and external factors that can be analyzed using various statistical and *machine learning methods*. In this study, the dependent variable is Sun Thai Tea's sales turnover, which is measured in the form of daily revenue. These variables are influenced by internal factors such as historical sales trends as well as external factors that can affect consumer behavior.

An exogenous variable is a variable that affects a dependent variable but is not affected by the dependent variable itself in a statistical model. In the SARIMAX model, exogenous variables are used to improve the accuracy of the prediction by including relevant external factors. According to [7], exogenous variables can be economic, social, or environmental data that affect time series patterns.

In this study, the exogenous variables used included daily weather and the promotion of Sun Thai Tea. Daily weather was chosen because the drinks sold are cold drinks that tend to experience an increase in demand in hot weather. According to research [5], weather variables such as temperature can contribute to improving the accuracy of forecasting models in the food and beverage sector. In addition, the marketing promotion factor also has a great influence on sales, as discount strategies and advertising campaigns can increase the number of customers in a given period [8].

## 2.6. Income

Income has various meanings according to various views. Some definitions of income from various sources are as follows. According to accounting science, Russel Swanburg (2000) in [12] defines income as revenue from the sale of products and services. The Indonesian Institute of Accountants explains that income is income obtained from a company's activities, which is often referred to by other terms such as Sales. Meanwhile, Sukirno (2000) in [12] stated that income is a very crucial element in the world of trade, because in running a business, it is important to know the amount of income obtained during the business.

## 3. Research Methods

Detailed This study uses a quantitative approach with a *time series analysis method*. The main focus of the research is to optimize the forecasting of Sun Thai Tea's daily turnover through the SARIMAX (*Seasonal Autoregressive Integrated Moving Average with Exogenous Variables*) model, which integrates exogenous variables, namely daily weather conditions and promotional activities. The SARIMA model is used as a comparator to evaluate the extent to which the addition of exogenous variables can improve the accuracy of predictions.

The selection of the SARIMAX model is based on its ability to capture seasonal patterns and historical trends, as well as considering the influence of external variables on turnover fluctuations. The integration of exogenous variables is expected to be able to produce a more accurate forecasting model than the SARIMA model which only uses historical data. The stages of research carried out include:

### 3.1. Data Collection

Daily turnover data is the main information that acts as a dependent variable and will be predicted using the SARIMAX model. Data was obtained from Sun Thai Tea's daily transaction records during the period April to November 2024 (231 observation days). Daily turnover reflects the amount of business revenue on a daily basis, which is influenced by various internal and external factors.

Meanwhile, promotion data and daily weather were used as exogenous variables in this study. Promotional data was obtained from Sun Thai Tea's marketing records. Promotions include discounts, *bundling* packages, and other marketing activities carried out on certain days. Effective promotions are expected to increase the number of purchases.

Weather data is obtained from the *AccuWeather* weather monitoring site and is categorized into three types, namely:

- 1) Rain: Daily temperature with a maximum temperature below 28°C or 25°C a low minimum temperature, indicating rainfall.
- 2) Cloudy: Daily temperatures range between and drinking temperatures below 28°C – 31°C, with cloudy skies but no rain.
- 3) Heat: Daily temperature above 31°C, sunny and dry weather.

### 3.2. Preprocessing Data

At this stage, several processes are carried out as follows:

- a) Turnover data is checked to identify *missing values* and *outliers* that can affect model performance.
- b) The stationarity test was performed using the *Augmented Dickey-Fuller* (ADF) Test to determine whether the data need differentiating process to be stationary.

The *Augmented Dickey-Fuller* (ADF) test is performed using the following mathematical formula.

$$\Delta Y_t = \alpha + \beta_t + \gamma Y_{t-1} + \sum_{i=1}^p \delta_i \Delta Y_{t-i} + \epsilon_t \quad (3)$$

with:

$Y_t$  : the variable that is tested for stationarity (in this case, daily turnover)

$\Delta Y_t = Y_t - Y_{t-1}$  : *differentiating* first to eliminate the trend

$\alpha$  : constant in the model

$\beta_t$  : deterministic trend coefficient (if any)

$\gamma$  : the coefficient of the root unit that is being tested for existence

$p$  : the amount of lag used to overcome autocorrelation

$\delta_i$  : coefficient of the lag of the first difference  $\Delta Y_{t-i}$

$\epsilon_t$  : error term (*white noise*)

- If the data is not stationary, *differentencing is done* to eliminate trends that may interfere with the model's accuracy.

### 3.3. SARIMAX Model Selection

- a) The selection of SARIMAX parameters is carried out using the *Autocorrelation Function* (ACF) and *Partial Autocorrelation Function* (PACF) to determine the optimal value and  $(p, d, q)(P, D, Q, s)$
- b) Exogenous parameters are tested using regression to see the extent to which temperature and promotion affect turnover.
- c) The model is optimized by using *Grid Search* to determine the best combination of parameters.

### 3.4. SARIMAX Modeling

a) The SARIMAX model is built with exogenous variables in mind temperature and promotion. The SARIMAX model is notated as follows.

$$Y_t = \Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Y_t + X_t\beta + \theta_q(B^s)\theta_q(B)\varepsilon_t \quad (4)$$

where:

- $\phi_1$  : Autoregressive (AR) component
- $\theta_1$  : Moving Average (MA) component
- $\Phi_1$  : Seasonal Autoregressive (SAR) component
- $\Theta_1$  : Seasonal Moving Average (SMA) component
- $X_t\beta$  : Exogenous Variables (Weather and Promotion)

b) The coefficients of exogenous variables are calculated to determine the contribution of temperature and promotion to turnover forecasting.

c) The SARIMA model is also built as a comparator to evaluate the influence of exogenous variables on prediction accuracy. The SARIMA model is notated as follows.

$$Y_t = \Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Y_t + \theta_q(B^s)\theta_q(B)\varepsilon_t \quad (5)$$

where:

- $\phi_1$  : Autoregressive (AR) component
- $\theta_1$  : Moving Average (MA) component
- $\Phi_1$  : Seasonal Autoregressive (SAR) component
- $\Theta_1$  : Seasonal Moving Average (SMA) component

### 3.5. Model Evaluation

- a) The model evaluation was carried out with *Mean Absolute Percentage Error (MAPE)* as the main metric to assess the accuracy of the prediction with the following criteria.
- b) The MAPE value of SARIMAX is compared to the MAPE of SARIMA to see the increased accuracy obtained from exogenous variables.

To describe the process carried out in building a forecasting model, a *flowchart is used* as a visual representation of the modeling stages that are carried out. This flowchart covers the entire process from data collection to evaluation of prediction results. Each stage is arranged sequentially to ensure methodological clarity in the application of the SARIMAX model as well as comparative validation of the SARIMA model.

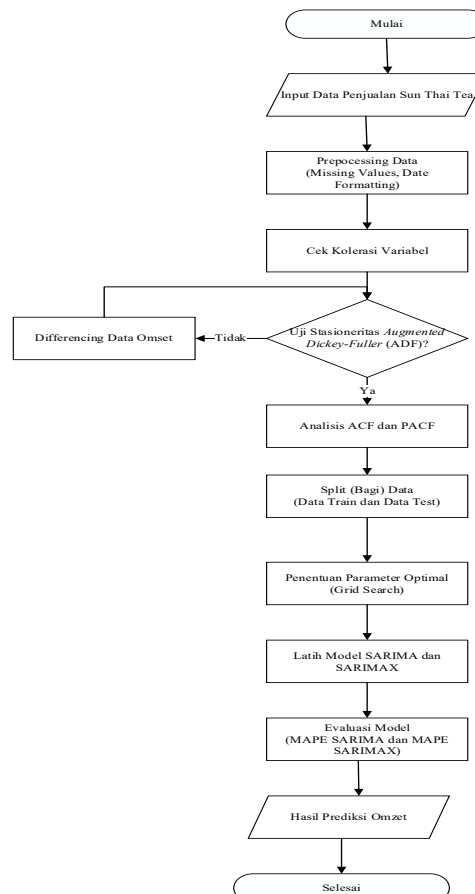


Fig. 1 : Revenue Forecasting Modeling Flowchart

## 4. Result and Discussion

In The results of this study were obtained through the stages of implementation of time series forecasting models, namely SARIMA (*Seasonal Autoregressive Integrated Moving Average*) and SARIMAX (SARIMA with eXogenous variables), in predicting daily sales turnover in Sun Thai Tea MSMEs. The selection of these two models is based on the characteristics of the data that have seasonal patterns and are likely to be influenced by external factors such as weather and promotional activities. The SARIMA model is used as a comparison model without involving exogenous variables, while the SARIMAX model is developed by incorporating external variables.

### 4.1. Collecting Data

The data used in this study is data on the daily sales turnover of Sun Thai Tea MSMEs during the period April to November 2024, with a total of 231 observation days. This data is obtained from daily transaction records that are systematically recorded by the business management. Each entry reflects the total daily revenue.

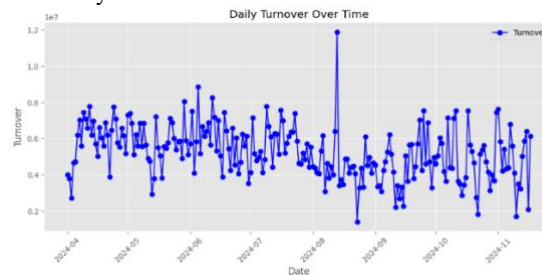


Fig.2 : Revenue Chart Period April-November 2024

The daily turnover graph indicates consistent and recurring fluctuations, particularly in the weekly cycle, reflecting seasonal patterns in sales behavior. The graph serves as an initial representation of the characteristics of the data before the modeling process is carried out. The entire modeling process was carried out using the Python programming language that supports numerical and statistical computing, by utilizing the statsmodels library for statistical model building, pandas for data manipulation, matplotlib for graph visualization, and numpy for numerical calculations. The process starts from the exploration and pre-processing stages of data, followed by stationarity tests using *Augmented Dickey-Fuller (ADF)*, *analysis of Autocorrelation Function (ACF)* and *Partial Autocorrelation Function (PACF)*, search for optimal parameters using *grid search*, to evaluation of model performance based on *Mean Absolute Percentage Error (MAPE)* values.

This systematic and structured approach in the implementation of this model aims to obtain predictive models that are not only accurate but also capable of capturing daily sales dynamics influenced by internal and external factors. The final results of this modeling provide a solid basis for comparing the effectiveness of the SARIMA and SARIMAX models in the context of MSME turnover forecasting based on historical data and environmental factors.

### 4.2. Stationarity Test with (*Augmented Dickey-Fuller*) ADF

The initial stage in time series modeling is to test whether the data is stationary. The test was carried out using *the Augmented Dickey-Fuller (ADF)* method with the aim of identifying the presence of unit roots in time series data. The p-value of the ADF test results is used as the basis for decision-making. If the value is  $p\text{-value} > 0.05$ , then the data is considered non-stationary, and needs to be transformed, such as *differentiating*.

Here is the *Dickey-Fuller Augmented Test (ADF)* program code

```
# Perform ADF test on Turnover
adf_test = adfuller(df["Turnover"])
adf_result = {
    "ADF Statistic": adf_test[0],
    "p-value": adf_test[1],
    "Critical Values": adf_test[4],
}
```

From the code above, here is the output of the *Augmented Dickey-Fuller (ADF)* test

```
{'ADF Statistic': np.float64(-2.049760521280087),
 'p-value': np.float64(0.2651896988750524),
 'Critical Values': {'1%': np.float64(-3.460849270544952),
 '5%': np.float64(-2.87495318813585),
 '10%': np.float64(-2.5739190539191745)}}
```

Fig. 3: Output of *Dickey-Fuller Augmented Test Results (ADF)* before *Differentiating*

The results of the initial test showed that the p-value was large and the statistical ADF value  $0.2652 > 0.05$  was higher than the critical value, so the turnover data was not stationary. Therefore, a differentiating transformation is carried out once.

After the transformation, the *Augmented Dickey-Fuller* (ADF) test is re-performed to ensure the data has achieved stationarity:

```
# ADF Test after differencing
adf_test_diff =
adfuller(df["Turnover_diff"].dropna(
))
adf_result_diff = {
"ADF Statistic": adf_test_diff[0],
"p-value": adf_test_diff[1],
"Critical Values": adf_test_diff[4],
```

The results of the *Augmented Dickey-Fuller* (ADF) test after *differencing* are as follows.

```
{'ADF Statistic': np.float64(-7.642737203152065),
'p-value': np.float64(1.8771242341126767e-11),
'Critical Values': {'1%': np.float64(-3.460849270544952),
'5%': np.float64(-2.87495318813585),
'10%': np.float64(-2.5739190539191745)}}
```

Fig. 4 : Output of *Dickey-Fuller Augmented Test* Results (ADF) after *Differencing*

The results show that the p-value is , so it can be concluded that the data is stationary and can be continued to the modeling stage.  $0.000000000018771 < 0.05$

### 4.3. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) Analysis

After the data is declared stationary, the next step is to perform an *analysis of the Autocorrelation Function* (ACF) and *Partial Autocorrelation Function* (PACF). The purpose of this analysis was to identify the dominant lag structure, which is the basis for determining the *autoregressive* (AR) and *moving average* (MA) parameters of the model.

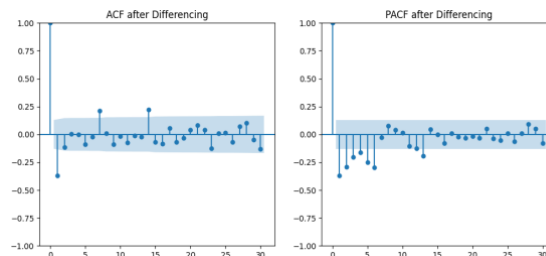


Fig. 5: ACF and PACF charts after *Differencing*

After *differencing* once ( $d = 1$ ), stationary characteristics were checked through *Autocorrelation Function* (ACF) and *Partial Autocorrelation Function* (PACF) graphs. The results of the analysis of the ACF and PACF charts after *differencing* are as follows:

1. ACF (*Autocorrelation Function*)
  - a) The autocorrelation value immediately drops sharply after the 0th lag.
  - b) There is no pattern of *gradual decay* as before *differencing*.
  - c) Most lags are within the limits of the confidence interval  $\rightarrow$  indicate an insignificant autocorrelation.
  - d) Indicates that the data has been free of trends and is approaching a stationary nature.
2. PACF (*Partial Autocorrelation Function*)
  - a) Significant PACF values are seen in some of the initial lags, especially in the 1st lag to around the 5th lag.
  - b) The pattern of PACF fluctuations after the initial lag does not form a systematic pattern and is mostly within confidence limits.
  - c) Indicates that the data does not have a strong direct autocorrelation at high lag.

After the first differentiation ( $d = 1$ ), the data showed stationary characteristics. The ACF and PACF graphs show that the differencing process has succeeded in eliminating trends in the data.

### 4.4. Determining Optimal Parameters with *Grid Search*

To obtain optimal model parameters, a grid search process was carried out by testing various combinations of values. The combination evaluation was carried out based on  $(p, d, q)$ ,  $(P, D, Q, s)$ , the *Akaike Information Criterion* (AIC) and *Mean Absolute Percentage Error* (MAPE) values.

The following code is used for parameter evaluation.

```
# Predict on test set
sarimax_pred =
sarimax_fit.predict(start=len(train)
, end=len(train) + len(test) - 1,
exog=test[["Weather_Category_Num",
"Promotion_Flag"]])
```

```
#Compute MAPE
sarimax_mape = mean_absolute_percentage_error(test["Turnover"],
sarimax_pred) * 100

# Update best SARIMAX model if better
if sarimax_mape < best_sarimax["MAPE"]:
    best_sarimax = {
        "AIC": sarimax_fit.aic,
        "MAPE": sarimax_mape,
        "order": param,
        "seasonal_order": param_seasonal,
    }
```

Here are the output of the best parameter search results with *grid search*.

```
Best SARIMAX Parameters: {'AIC': np.float64(5416.235184744857), 'MAPE': 25.06823194575481,
SARIMA MAPE : 68.5462566120604
   Model  Order Seasonal Order  MAPE
0 SARIMAX (2, 1, 0) (0, 1, 0, 7) 25.068232
1 SARIMA (2, 1, 0) (0, 1, 0, 7) 68.546257
```

**Fig.6** : Initialization of SARIMA and SARIMAX Modeling Parameters

The best results for the SARIMAX model show that:

1. SARIMAX with the best parameters
  - a) AIC (5416.23)  
Lower AIC values indicate that the SARIMAX model with this parameter is better in terms of fit for the data than other models with different parameters.
  - b) *Mean Absolute Percentage Error* (MAPE) (25.06%) value.  
The SARIMAX model produces a relatively smaller prediction error compared to the SARIMA model with a *Mean Absolute Percentage Error* (MAPE) (68.54%) value.
  - c) *Order* (2, 1, 0)  
This parameter shows the ARIMA model used in SARIMAX, i.e. with  $p = 2, d = 1, q = 0$
  - d) *Seasonal Order* (0, 1, 0, 7)  
This parameter indicates a seasonal component with  $p, d$ , and a seasonal period  $7P = 0, D = 1Q = 0$
2. SARIMA MAPE (68.54%)  
The SARIMA model has a higher *Mean Absolute Percentage Error* (MAPE) value, indicating that this model is less effective in predicting compared to SARIMAX.(68.54%),

SARIMAX with parameters for (2,1,0) orders and for (0,1,0,7) seasonal orders provides better performance compared to the SARIMA model. With a lower AIC value and a smaller *Mean Absolute Percentage Error* (MAPE) value, SARIMAX is suitable for use in daily sales turnover forecasting for Sun Thai Tea.

#### 4.5. Evaluation of Accuracy and Visualization of Prediction Results

Accuracy evaluation is carried out to measure the performance of the model in predicting turnover data. The prediction results are compared with actual data through the calculation of *Mean Absolute Percentage Error* (MAPE) and are displayed in the form of a table. The *Mean Absolute Percentage Error* (MAPE) values of the two models were compared to determine which model had a higher level of accuracy.

```
SARIMA Train MAPE: 28.06749068988782
SARIMAX Train MAPE: 20.78114075842919
SARIMA Test MAPE: 68.5462566120604
SARIMAX Test MAPE: 25.06823194575481
```

**Fig. 7** :Output *Mean Absolute Percentage Error* (MAPE) Value of Training Data and Test Data

**Table 2:** Table of Evaluation Results of SARIMA and SARIMAX Models

Day/Date	Turnover	SARIMAX Prediction	SARIMA Prediction	SARIMAX MAPE	SARIMA MAPE
0 2024-04-01	3997000	48615	0	20.781139%	28.067489%
1 2024-04-02	3808000	3369376	3224339	20.781139%	28.067489%
2 2024-04-03	2704000	3541317	3424931	20.781139%	28.067489%
3 2024-04-04	4670000	3341434	3414132	20.781139%	28.067489%
4 2024-04-05	4733000	3985962	3938098	20.781139%	28.067489%
...	...	...	...	...	...
226 2024-11-13	5015000	3708890	958071	25.068235%	68.546259%
227 2024-11-14	5861000	4218983	-1957399	25.068235%	68.546259%
228 2024-11-15	6411000	5117360	890730	25.068235%	68.546259%
229 2024-11-16	2076000	2907693	-1592016	25.068235%	68.546259%
230 2024-11-17	6142000	2984330	-3056433	25.068235%	68.546259%

The results of the evaluation showed that the SARIMAX model was superior in terms of accuracy compared to SARIMA, which was indicated by the lower Mean Absolute Percentage Error (MAPE) value. The SARIMAX model has a MAPE value of 20.78% in training data and 25.06% in test data, while the SARIMA model shows a MAPE value of 28.06% in training data and 68.54% in test data.

To see a visual comparison between the actual and predicted turnover of the two models, a graph was created using *matplotlib*.

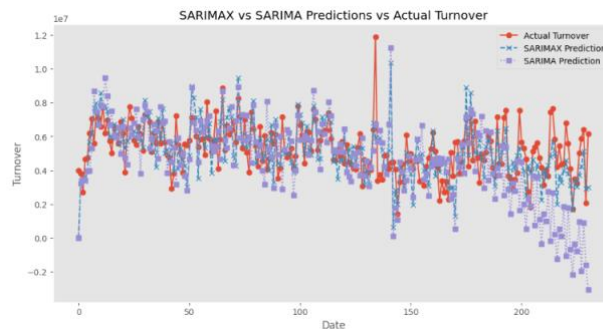


Fig.8 : Chart of Actual Comparison of Turnover with SARIMA and SARIMAX Predictions

The graph of the prediction results shows that the SARIMAX model is able to follow actual data patterns better than SARIMA, especially in periods of weather variation and promotion.

### 5. Conclusion

The SARIMAX model was successfully applied to forecast the daily turnover of Sun Thai Tea by taking into account the weekly seasonal component. The model is configured using the optimal parameters of *Grid Search results*, namely *orders (2,1,0)* and *seasonal orders(0,1,0,7)*, which are able to capture weekly sales patterns such as weekend turnover spikes.

The evaluation of the model's accuracy was carried out using the *Mean Absolute Percentage Error (MAPE)* metric, which is a measure of the average prediction error in the form of a percentage of the actual value. The SARIMAX model generates a MAPE value of 25,06%, which falls into the category of "Fairly Accurate Forecasting Model" according to the following classification:

MAPE Values	Accuracy Criteria
<10%	Highly Accurate Forecasting Models
10% - 20%	Accurate Forecasting Models
20% - 50%	Forecasting Models Are Quite Accurate
>50%	Inaccurate Forecasting Models

Source : [9]

These findings indicate that the SARIMAX model has high reliability as an analytical instrument in supporting the process of daily operational decision-making, production planning, and inventory management in a more systematic and measurable manner. This model is not only statistically superior, but also relevant in the business context, as it is able to anticipate actual sales patterns with a relatively small error rate. The advantage of SARIMAX lies in its ability to explicitly capture seasonal fluctuations as well as take exogenous variables into account, making it a more representative model of the complexity of sales patterns.

### References

- [1] R. Kelly, "UNDERSTANDING THE IMPORTANCE OF SALES FORECASTING IN IMPROVING ORGANIZATIONAL PERFORMANCE: A STUDY ON THE BAKERY INDUSTRY IN NIGERIA," *Interdiscip. Res. J. Manag. Soc. Sci.*, vol. 9, no. 1, pp. 91–104, 2022.
- [2] J. Boylan and A. Syntetos, "Advances in Intermittent-Demand Forecasting," *Foresight Int. J. Appl. Forecast.*, no. 64, pp. 4–8, 2022.
- [3] A. A. Maulana and H. Rosalina, "IMPLEMENTASI METODE SARIMAX UNTUK PREDIKSI CURAH HUJAN JANGKA PENDEK DI PAGERAGEUNG, TASIKMALAYA," *J. Sumber Daya Air*, vol. 20, no. 1, pp. 39–50, 2024.
- [4] P. Manigandan *et al.*, "Forecasting natural gas production and consumption in United States-evidence from SARIMA and SARIMAX models," *Energies*, vol. 14, no. 19, p. 6021, 2021.
- [5] F. Petropoulos *et al.*, "Forecasting: theory and practice," *Int. J. Forecast.*, vol. 38, no. 3, pp. 705–871, 2022.
- [6] R. Hyndman and G. Athanasopoulos, "Principles of Business Forecasting by Keith Ord & Robert Fildes Forecasting: Principles and Practice," 2012.
- [7] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*. John Wiley & sons, 2008.

- 
- [8] R. J. Hyndman, *Forecasting: principles and practice*. OTexts, 2018.
- [9] H. K. Juliarto, I. Purnamasari, and S. Prangga, "PERAMALAN PEREDARAN UANG KARTAL DI INDONESIA MENGGUNAKAN MODEL HYBRID SARIMAX-NEURAL NETWORK," *J. Gaussian*, vol. 12, no. 4, pp. 465–476, 2024.
- [10] R. J. Hyndman *et al.*, "Package 'forecast,'" *Online*] <https://cran.r-project.org/web/packages/forecast/forecast.pdf>, 2020.
- [11] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.
- [12] M. Tindangen, D. S. M. Engka, and P. C. Wauran, "PERAN PEREMPUAN DALAM MENINGKATKAN EKONOMI KELUARGA (STUDI KASUS: PEREMPUAN PEKERJA SAWAH DI DESA LEMOH BARAT KECAMATAN TOMBARIRI TIMUR KABUPATEN MINAHASA)," *J. Berk. Ilm. Efisiensi*, vol. 20, no. 03, 2020.