Data Science and Engineering (DSE) Record, Volume 6, issue 1.

STEEL SALES PREDICTION USING DEEP LEARNING AND TRADITIONAL FORECASTING TECHNIQUES

Patcharaporn Saguanchokvanich[1], Chompoonoot Kasemset [2] , and Trasapong Thaiupathump [3]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University
[3] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
patcharaporn_sag@cmu.ac.th

Abstract. Steel sales forecasting is a crucial element in the strategic planning of Chiang Mai Center Steel Co., Ltd. This study focuses on forecasting sales of product WR-44202050, the company's top-selling item, by comparing various forecasting models, including ARIMAX, LSTM, VARX, and Hybrid ARIMAX–MLP. Model performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The ARIMAX model achieved the highest accuracy, with MAE of 4.28, MSE of 18.28, RMSE of 4.28, and MAPE of only 0.01%. These results indicate that ARIMAX is the most suitable model for steel sales forecasting in this context, offering strong potential as a decision-support tool for production planning, inventory management, and strategic business operations.

Keywords: Forecasting, ARIMAX, VARX, ARIMAX-MLP, LSTM

1 Introduction

Steel sales forecasting is a critical component in the strategic planning of steel distribution companies, enabling informed decisions on production scheduling, inventory control, and procurement management. Chiang Mai Center Steel Co., Ltd., established in 2011, has grown from a regional distributor of construction steel products such as wire mesh, round steel bars, and structural steel into a major supplier serving eleven provinces in Northern Thailand. Over the past decade, the company has expanded its portfolio to include construction tools and materials, introduced value-added services such as cutting, bending, and drilling of steel, and upgraded its operations through the implementation of an Enterprise Resource Planning (ERP) system. In 2022, the company further strengthened its capacity by opening a warehouse in San Kamphaeng, Chiang Mai, tripling its storage space and expanding its fleet to enhance logistics efficiency. Today, its primary customers are government construction contractors involved in infrastructure projects such as roads, bridges, schools, and public buildings.

This research focuses on forecasting sales of WR-44202050, the company's best-selling round steel bar, which consistently accounts for a significant share of its revenue. The study examines the influence of macroeconomic indicators including steel import and export volumes, construction material price indices, and national GDP growth on sales trends. Four forecasting models are developed and compared: the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Long Short-Term Memory (LSTM) networks, Vector Autoregression with Exogenous Variables (VARX), and a Hybrid ARIMAX Multi Layer Perceptron (MLP) approach. Each model is trained and tested using historical sales data combined with relevant economic variables, and their performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

By systematically comparing statistical, deep learning, and hybrid approaches, this study aims to identify the most accurate and practically applicable forecasting method for steel sales in this business context. The findings are expected to provide a reliable decision-support tool for production planning, inventory management, and strategic operations, while also contributing to the broader understanding of time series forecasting in the steel industry.

2 Related Works

2.1 Incorporating Exogenous Variable

Venkatachary M. et al. [1] explains that steel demand forecasting necessitates the integration of external macroeconomic indicators such as gross domestic product (GDP) growth rates, employment levels, and import–export volumes, as these factors exert a direct influence on production processes, pricing strategies, and market demand.

Sonja S. and Shailesh T. [2] states that historical industrial production data and prevailing market trends play a critical role in identifying the cyclical nature of the steel industry, thereby serving as a foundation for formulating strategies and restructuring supply chains to align with market conditions.

The Office of Industrial Economics [3] emphasizes that global factors, including commodity price volatility and international trade policies, must be comprehensively considered to assess and mitigate the risks of potential market disruptions.

2.2 Traditional Forecasting

Christogonus Ugoh. [4] notes that the ARIMAX model, an extension of ARIMA, incorporates exogenous variables, thereby enabling the integration of economic indicators and policy changes into forecasting.

S. M. Ulyah. [5] highlights that VARX models are particularly well-suited for multivariate time series forecasting, as they capture the dynamic interplay between production levels, demand, and prices.

Venkatachary M. et al. [1] and Wiwik A. et al. [6] report that comparative studies consistently indicate the value of traditional forecasting models for their interpretability, adaptability, and scalability, making them applicable to both small-scale operations and large enterprises.

2.3 Deep Learning in Forecasting

Tian Guo et al. [7] and Thomas Shering et al. [8] demonstrate that deep learning methods, particularly Long Short-Term Memory (LSTM) networks, achieve strong performance in time-series forecasting due to their ability to capture long-term dependencies and handle multivariate, non-linear relationships. This capability is especially advantageous in volatile markets such as steel, where rapid fluctuations in demand and supply occur frequently.

Paulo S. G. de Mattos Neto et al. [9] and Wiwik Anggraeni et al. [6] report that hybrid approaches, such as combining ARIMAX with Multi-Layer Perceptrons (MLP), leverage the strengths of both seasonality handling from ARIMAX and non-linear modeling from MLP resulting in improved forecasting accuracy.

3 Data and Methodology

This chapter outlines the methodology for forecasting steel sales using traditional time-series and deep learning models. It covers data acquisition, preprocessing, feature engineering, model development, validation, and performance evaluation, as well as the creation of an interactive dashboard for result visualization, ensuring a reliable and interpretable framework for the steel industry.

3.1 Dataset

This study utilizes two primary data sources: internal steel sales transactions and external economic indicators that may influence steel demand. Careful collection and preparation of both datasets are essential to ensure the accuracy and reliability of the forecasting models.

3.1.1 Steel Sales Data (Internal Data)

The internal dataset consists of historical daily sales records from Chiang Mai Center Steel Co., Ltd., including fields such as document date, product code, and total transaction value. The data, available from January 2022 onward, serve as the dependent variable in the forecasting models, as shown in Figure 1.

	เอกสารวันที่	รหัสสินค้า	รวมมูลค่า
40607	2022-01-04	IC-0658	300.00

Figure 1: Example of Raw Sales Transaction Record

From these raw records, daily sales values are aggregated and synchronized with external indicators to form a structured time-series dataset suitable for modeling.

3.1.2 External Economic Indicators

The external dataset is obtained from reputable sources, including the Iron and Steel Institute of Thailand and the Office of Industrial Economics. Key variables include the steel price index, month-over-month and year-over-year changes, and economic indices related to the construction sector, as illustrated in Figure 2.

	เอกสารวันที่	ดัชนีราคา_ผู้บริโภคทั่วไป	อัตราการเปลี่ยนแปลงเทียบกับเดือนก่อนหน้า_ผู้บริโภคทั่วไป	อัตราการเปลี่ยนแปลงเทียบกับเดือนเดียวกันปีก่อนหน้า_ผู้บริโภคทั่วไป	ดัชนี_อัตราเปลี่ยนแปลง_ผู้บริโภคทั่วไป	ดัชนี_อัตราการเปลี่ยนแปลงเทียบกับเดือนเดียวกันปีก่อนหน้า_ผู้บริโภคทั่วไป
0	2012-01-01	96.64	0.28	0.91	27.06	87.94

Figure 2: Transformed Economic Indicators for Forecasting Model Input

This dataset is aligned with the daily sales data, with missing values interpolated using a linear method. Additional derived variables are also created to capture trends and market shifts over time. Finally, the internal and external datasets are merged to form the complete dataset used in the modeling process.

3.2 Research Framework

This study develops a steel sales forecasting framework by integrating both traditional time-series models and deep learning techniques to identify the economic and industrial factors that influence sales, and to generate accurate forecasts for Chiang Mai Center Steel Co., Ltd. The framework combines internal daily sales data with macroeconomic indicators such as GDP growth rates, raw material prices, and construction-related price indices, sourced from reputable institutions.

The process begins with data collection, cleaning, and feature engineering, including generating percentage change variables, moving averages, lag features, and seasonal decomposition. Correlation analysis and Variance Inflation Factor (VIF) evaluation are then performed to remove redundant features and address multicollinearity. The data are subsequently normalized or standardized and resampled to a monthly frequency to emphasize trends over noise.

Modeling is conducted in two categories: traditional approaches (ARIMAX, VARX, and ARIMAX–MLP) and deep learning (LSTM). Leave-One-Out Cross-Validation (LOOCV) is employed to ensure reliable and unbiased performance assessment. Model performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), enabling objective comparison and selection of the most accurate model.

Finally, the forecasting results are deployed through an interactive dashboard developed with Streamlit. This interface enables users to visualize historical and forecasted sales trends, compare model accuracy, and support strategic production and demand planning. The overall workflow is illustrated in Figure 3.

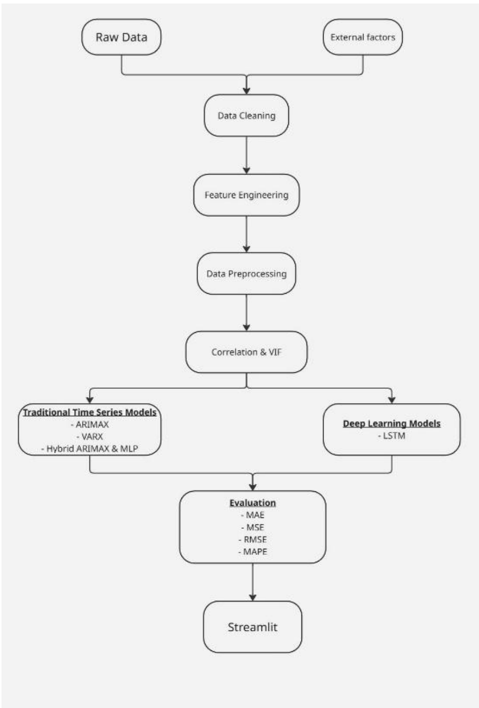

Figure 3: Workflow of Steel Sales Forecasting Model Development.

3.3 Exploratory Data Analysis and Trend Selection for Forecasting

Prior to model construction, an exploratory data analysis (EDA) was conducted to examine the underlying characteristics, temporal patterns, and potential anomalies within the dataset. This preliminary investigation aimed to establish a comprehensive understanding of the sales behavior, facilitate the identification of a representative product for in-depth analysis, and ensure that the selected data were suitable for subsequent forecasting model development.

A five-year sales dataset was analyzed to identify the top-selling product. WR-44202050 recorded the highest cumulative sales and was selected for detailed time-series analysis, as shown in Figure 4.

Figure 4: Top 5 Best-Selling Steel Products by Total Sales.

Its daily sales trend shows high variability with occasional peaks, as shown in Figure 5.

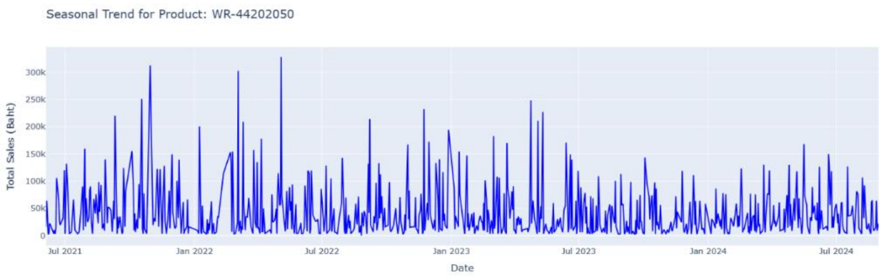

Figure 5: Daily Sales Trend of Product WR-44202050

Outlier detection via boxplot indicated that only 5.35% of data points were outliers, requiring no removal, as illustrated in Figure 6.

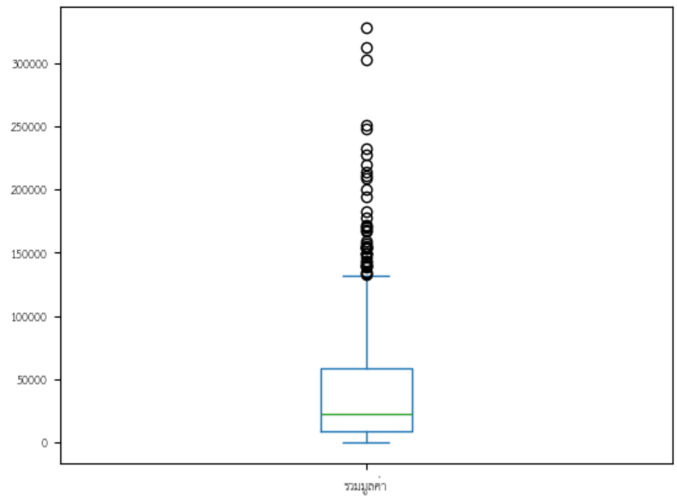

Figure 6: Boxplot of Daily Sales for Outlier Detection

Seasonal decomposition (STL) separated the data into trend, seasonality, and residual components, as presented in Figure 7.

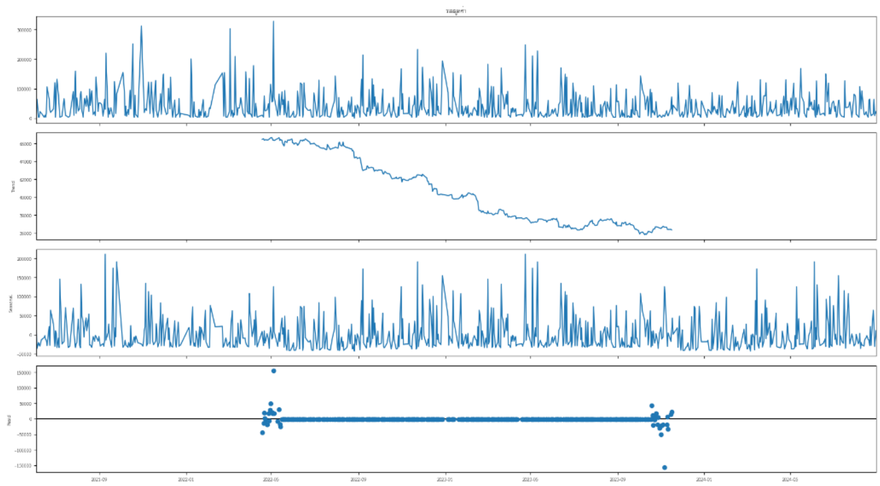

Figure 7: Seasonal Decomposition of Daily Sales Data

Testing decomposition periods of 30, 90, 180, and 365 days revealed that a 90-day period provides the best balance between smoothness and responsiveness for quarterly planning, as shown in Figure 8.

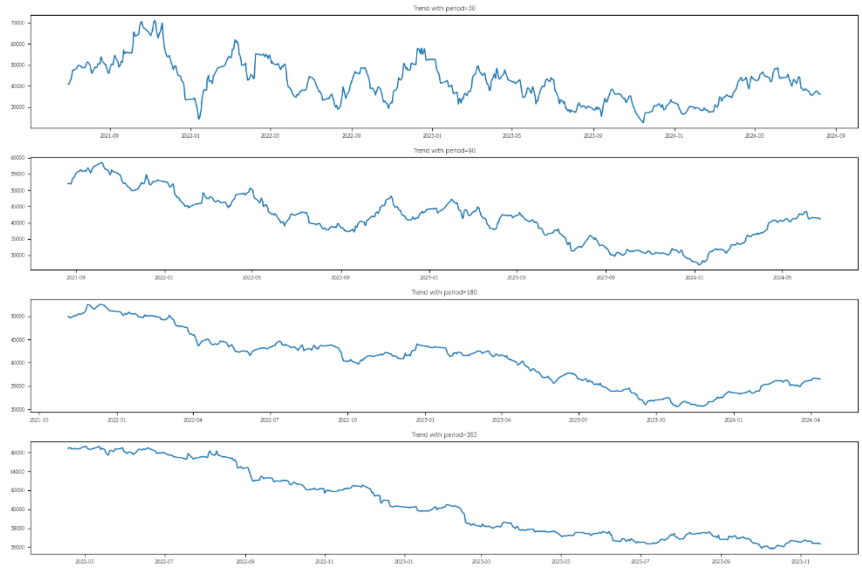

Figure 8: Comparison of Trends at Different Periods

Additionally, to further validate the presence or absence of strong seasonality, the Autocorrelation Function (ACF) was analyzed, as shown in Figure 9. The ACF plot showed that while the autocorrelation at lag = 0 was naturally equal to 1, the values at higher lags were low and fell within the confidence intervals. This indicates a lack of strong temporal autocorrelation or seasonality, especially annual seasonality. Therefore, the use of a 365-day period was ruled out, as it was inconsistent with the observed data structure.

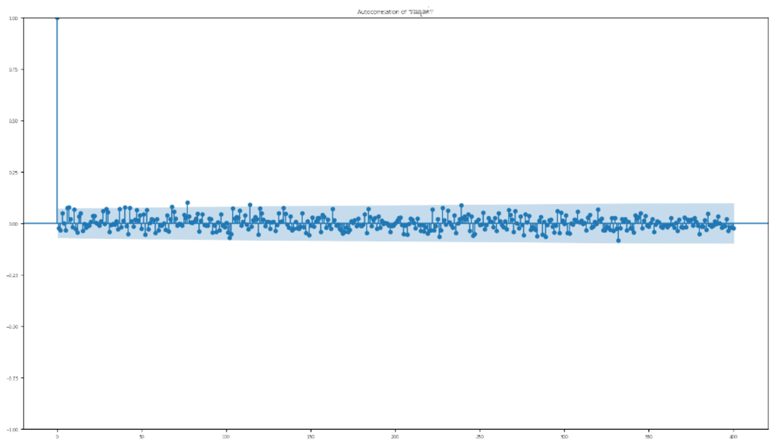

Figure 9: Autocorrelation Plot of Daily Sales Data.

Feature selection was performed using the Pearson correlation coefficient with a threshold of $|r| \geq 0.4$, resulting in several macroeconomic indicators, mainly related to raw material prices and industrial indices, meeting the criteria. To further refine the selection and ensure model stability, a Variance Inflation Factor (VIF) analysis was conducted to reduce multicollinearity among the variables. The results of the VIF analysis are summarized in Table 1, which lists the selected external variables along with their respective VIF values.

No.	Feature Name	VIF Value
1	Rate of Change Index	1.93
2	MoM Change: Steel Import Price	3.83
3	MoM Change: Metal Ore Import	5.16
4	YoY Change: Steel and Steel Products	5.58
5	MoM Change: Steel Export	4.65
6	Change Index: Steel Export	4.10
7	Change Index: Construction Materials	1.54
8	YoY Change Index: Construction Materials	2.10
9	Change Index: Basic Iron & Steel Manufacturing	3.35
10	YoY Change Index: Basic Iron & Steel Manufacturing	8.59
11	Iron Ore Price (CFR China)	2.65

Table 1: VIF Values of Selected External Variables

This combined correlation and VIF-based selection process ensured that chosen features were statistically significant, contextually relevant, and well-suited for developing a robust steel sales forecasting model.

3.4 Model Development

3.4.1 Model Development Using ARIMAX

The ARIMAX (AutoRegressive Integrated Moving Average with Exogenous Variables) model was employed as the initial approach for steel sales forecasting in this study. ARIMAX is a classical statistical time series model that allows the integration of exogenous variables to enhance predictive accuracy. Prior to model construction, the stationarity of the target time series (steel sales) was tested using the Augmented Dickey-Fuller (ADF) test. The initial test yielded a test statistic of -1.51 and a p-value of 0.53, exceeding the 5% critical value (-2.87), indicating that the series was non-stationary. To address this, first-order differencing was applied, and a subsequent ADF test produced a p-value of 5.24e-06, confirming that the differenced series was stationary and suitable for time series modeling.

For parameter selection and evaluation, the dataset's limited size of 747 observations motivated the use of Leave-One-Out Cross-Validation (LOOCV), ensuring unbiased model assessment while minimizing the randomness associated with train-test splits. The differencing order d was fixed at 1 (as determined from the stationarity test), and a grid search was conducted for p and q values ranging from 0 to 3. This range was deemed appropriate for daily-level data with a relatively small sample size, avoiding overfitting and keeping the model parsimonious. The optimal configuration was identified as (p=0, d=1, q=1) with an AIC of 8393.56, which was used to fit the final ARIMAX model, as shown in Figure 10.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:               Trend_90   No. Observations:                  747
Model:               SARIMAX(0, 1, 1)   Log Likelihood               -4756.075
Date:                Mon, 24 Feb 2025   AIC                           9538.150
Time:                        15:51:15   BIC                           9598.142
Sample:                             0   HQIC                          9561.272
                                - 747
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
x1           -38.7905     31.392     -1.236      0.217    -100.317      22.736
x2           302.7390     96.146      3.149      0.002     114.297     491.181
x3           153.9798     53.577      2.874      0.004      48.971     258.988
x4          -110.6476    121.240     -0.913      0.361    -348.273     126.978
x5          -346.3181     93.338     -3.710      0.000    -529.257    -163.379
x6            16.2926     10.566      1.542      0.123      -4.417      37.002
x7            -4.5569     14.558     -0.313      0.754     -33.090      23.977
x8            16.3732     29.942      0.547      0.584     -42.312      75.059
x9            10.3308     13.914      0.742      0.458     -16.940      37.602
x10          -77.9248     37.063     -2.102      0.036    -150.568      -5.282
x11          -56.0994     24.641     -2.277      0.023    -104.395      -7.804
ma.L1          0.9322      0.023     40.490      0.000       0.887       0.977
sigma2      1.773e+04    884.997     20.032      0.000     1.6e+04    1.95e+04
===================================================================================
Ljung-Box (L1) (Q):                 369.21   Jarque-Bera (JB):                 2.22
Prob(Q):                              0.00   Prob(JB):                         0.33
Heteroskedasticity (H):               0.45   Skew:                            -0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.18
===================================================================================
```

Figure 10: Parameter Estimation Results of ARIMAX Model

Several exogenous variables, such as x2 and x3, had p-values below 0.05, indicating statistically significant contributions, while others though not statistically significant were retained for their contextual business relevance.

To assess model adequacy, residual diagnostics were performed using the plot_diagnostics() function. The output, shown in Figure 11, includes four primary components

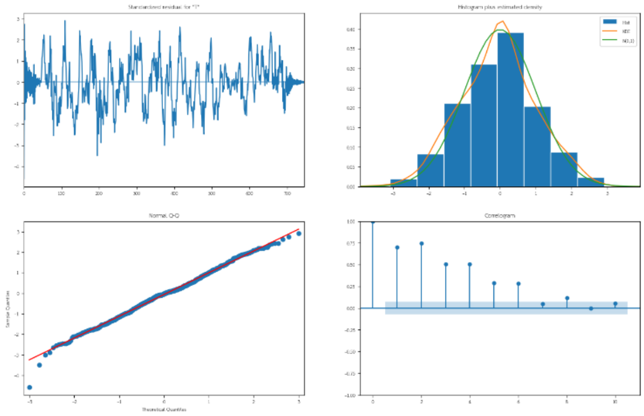

Figure 11: Residual Analysis of ARIMAX Model Using Plot Diagnostics

- **Plot 1: Standardized Residuals (Top Left)**

The standardized residual plot over time is used to assess whether the residuals are randomly distributed around zero, which is a core assumption of the ARIMAX model. Specifically, residuals should exhibit characteristics of white noise, meaning they should be independent and identically distributed without any discernible pattern. If the residuals display upward or downward trends or show clear clustering, it may indicate that the model has not fully captured the underlying structure in the data. Conversely, if the residuals are scattered consistently around the zero line without any clear sequence or wave-like structure, it suggests the model is time-appropriate and that no remaining bias exists in the residuals.

The plot shows that the residuals are evenly distributed above and below the zero line without any systematic patterns or noticeable changes in amplitude over time. This implies no remaining bias and that the residuals have an average close to zero, thereby meeting the theoretical assumption.

- **Plot 2: Histogram with Estimated Density (Top Right)**

The histogram of residuals is used to evaluate the distributional properties of the errors. The key assumption here is that residuals should follow a normal distribution. This would typically manifest as a bell-shaped curve. If the histogram is skewed or has heavy tails, it may suggest a departure from normality, which could affect the validity of statistical inference, such as the construction of confidence intervals or hypothesis testing.

The histogram appears symmetric with a shape that closely resembles the normal distribution. The KDE (orange) and the fitted normal distribution (green) curves largely overlap, indicating that the residuals closely approximate a normal distribution, which satisfies the assumption of normality required by ARIMAX.

- **Plot 3: Normal Q-Q Plot (Bottom Left)**

The Q-Q (quantile-quantile) plot compares the quantiles of the residuals with those of a theoretical normal distribution. The assumption is that the residuals should approximately follow a normal distribution. If the residuals are normally distributed, the plotted points should fall close to the diagonal reference line (y = x). Deviations, particularly at the tails, may indicate skewness or kurtosis, suggesting that the residuals deviate from normality and potentially impacting model performance and inference.

Most of the plotted points fall close to the theoretical diagonal line, especially in the central portion of the distribution. Although there are minor deviations at the tails, the overall pattern supports the assumption of normality, indicating that the residuals are well-approximated by a normal distribution.

- **Plot 4: Correlogram (Autocorrelation Plot, Bottom Right)**

The ACF (Autocorrelation Function) plot of residuals is used to determine whether residuals are time independent. The assumption is that residuals should exhibit no significant autocorrelation meaning the prediction errors at one time point should not be systematically related to errors at another time. Most autocorrelation bars should fall within the 95% confidence bounds (typically shaded in light blue). If any lag shows significant autocorrelation, it implies the model has not fully accounted for all time dependences, and AR or MA parameters may need further adjustment.

The autocorrelation values drop off quickly after lag 1, and most bars fall within the confidence interval. This confirms that there is no significant autocorrelation remaining, and the residuals behave as white noise fulfilling one of the key statistical assumptions of the ARIMAX model.

Based on these analyses, it can be concluded that the ARIMAX (0,1,1) model is statistically sound and serves as an effective baseline model for comparison with more complex approaches in the subsequent sections.

3.4.2 Modeling with LSTM

As an alternative forecasting method, this study applied a Long Short-Term Memory (LSTM) neural network, a deep learning architecture well-suited for capturing long-term dependencies and non-linear relationships in sequential data. Unlike traditional statistical models such as ARIMA or ARIMAX, LSTM does not require data stationarity, enabling the direct use of raw time series data. To ensure numerical stability during training, all features and the target variable were standardized using StandardScaler.

A look-back window of 120 time steps (≈4 months) was used, incorporating 11 external variables selected through correlation and VIF analysis as explanatory features (X), with the 90-day moving average sales trend (Trend_90) as the target (y). The model architecture comprised three stacked LSTM layers (256, 128, and 64 units), each followed by Batch Normalization and Dropout (0.3), and concluded with two dense layers for regression output. Training used the Adam optimizer (learning rate=0.0001) with mean_squared_error loss, alongside EarlyStopping (patience=100) and ReduceLROnPlateau. Leave-One-Out Cross-Validation (LOOCV) was applied due to the limited dataset of 747 records, and the convergence behavior during training is illustrated in Figure 12.

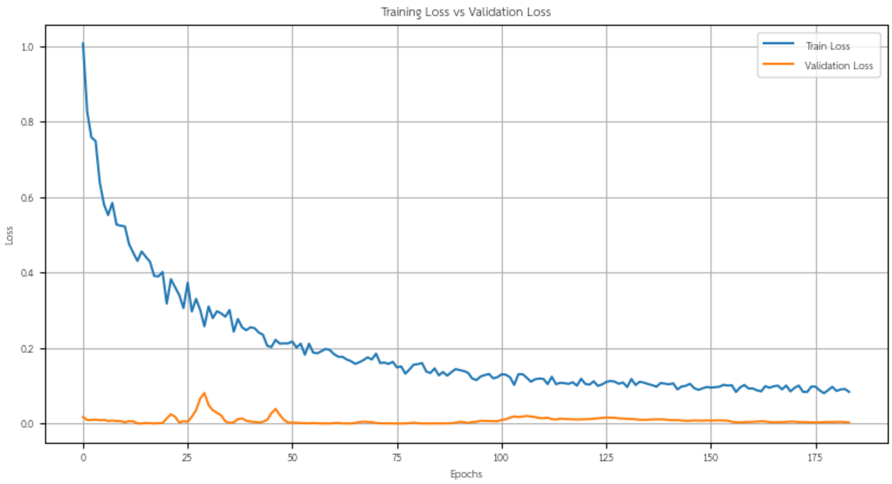

Figure 12: Training and Validation Loss Curve of LSTM Model

The loss curves showed a steady decline in training loss and consistently low validation loss, with occasional spikes corrected promptly, indicating good generalization and effective regularization.

In summary, the LSTM model effectively forecasted steel sales from raw, non-stationary time series, capturing both short-term fluctuations and long-term trends without prior differencing, making it a robust and flexible alternative to traditional models.

3.4.3 Hybrid ARIMAX & MLP Modeling

This hybrid forecasting approach combines the statistical rigor of ARIMAX, which captures linear trends and integrates exogenous variables, with the flexibility of an MLP neural network to model nonlinear residual patterns. First, an ARIMAX (3,1,3) model was trained on first-order differenced and standardized data, selected via grid search to minimize AIC (−4819.96). The model's residuals representing unexplained variance were then used as the target for the MLP.

The MLP architecture included a 32-neuron input layer (ReLU), a hidden layer at half size with 0.2 dropout, and a single output neuron. Optimized via grid search, the configuration achieved a low MSE of 0.0020, indicating effective residual learning.

Overall, this two-stage framework successfully addressed both linear and nonlinear patterns in steel sales data, yielding a high-performing and interpretable forecasting model suitable for complex, economically driven time series.

3.4.4 Forecasting Model Construction Using VARX

The Vector AutoRegression with eXogenous variables (VARX) model is a multivariate time series approach designed to capture the dynamic relationships among multiple dependent variables while integrating external influencing factors. In this study, the VARX model was applied to jointly forecast Total Sales Value and Trend_90, the latter representing a smoothed sales trend over a 90-day period. The exogenous variables, primarily economic indicators and steel-related indices, were selected through prior feature selection steps combining correlation filtering and multicollinearity reduction. The dataset was divided into a training set of the first 746 observations and a testing set comprising the remainder, with missing values imputed using median values. Standardization was applied separately to target and exogenous variables to maintain interpretability.

Lag order selection was conducted using statistical criteria from the VAR framework, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan–Quinn Information Criterion (HQIC), with the search extended up to 90 lags to capture long-term dependencies. All three criteria suggested a lag of 90, indicating extended autocorrelation structures likely driven by economic cycles or seasonal trends in the steel industry.

The VARX model was then trained using the selected lag and standardized data, achieving strong performance metrics as summarized in Table 2.

```
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Wed, 19, Feb, 2025
Time:                     15:57:57
--------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                  -10.5888
Nobs:                     656.000    HQIC:                 -12.1046
Log likelihood:           2785.47    FPE:               2.18253e-06
AIC:                     -13.0644    Det(Omega_mle):    1.34066e-06
--------------------------------------------------------------------
```

Table 2: VARX Model Summary

The table reports a log-likelihood of 2785.47, an AIC of −13.0644, a BIC of −10.5888, and an HQIC of −12.1046, all of which indicate a good model fit. Furthermore, the residual correlation between Total Sales Value and Trend_90 was only 0.085, as shown in Figure 13, confirming minimal correlation and supporting the assumption of residual independence.

```
Correlation matrix of residuals
              รวมมูลค่า  Trend_90
รวมมูลค่า     1.000000  0.085045
Trend_90      0.085045  1.000000
```

Figure 13: Correlation Matrix of Residuals

Overall, the VARX model effectively captures the dynamic structure of steel sales and their relationship with external variables. Its multivariate nature allows for modeling interdependencies between multiple targets, making it particularly suitable for environments with interacting economic indicators. With evidence of low residual correlation and an optimal lag structure, the VARX model provides a statistically sound foundation for forecasting and serves as a benchmark for comparison with other methods, including deep learning-based approaches, in subsequent sections.

3.5 Forecast Evaluation Metrics

Forecast model performance was assessed using four metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

- **MAE** measures the average absolute difference between actual and predicted values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **MSE** calculates the mean of squared errors, penalizing larger deviations:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **RMSE** is the square root of MSE, expressed in the same unit as the original data:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- **MAPE** expresses forecast error as a percentage of actual values:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

These metrics provide complementary perspectives for comparing models and selecting the most accurate and reliable forecasting approach.

3.6 Visualization and Interpretation of Forecasting Results

In order to bridge the gap between advanced statistical modeling and practical business application, this study developed an interactive forecasting dashboard using the Streamlit platform. The primary objective of this system is to transform complex outputs from the ARIMAX model into clear, actionable visual insights that can be readily interpreted by non-technical stakeholders, including sales managers, production planners, and executives. By integrating multiple analytical functions into a single interface, the dashboard enhances decision-making efficiency, facilitates timely responses to market changes, and supports collaborative planning across departments.

The dashboard is structured into three main functional sections. The first section, Data Insights, presents sales performance through visualizations such as a bar chart displaying the top five best-selling products and a pie chart showing their proportion of total revenue, as illustrated in Figure 14. This design enables users to immediately identify the primary revenue drivers and understand the sales distribution across key offerings.

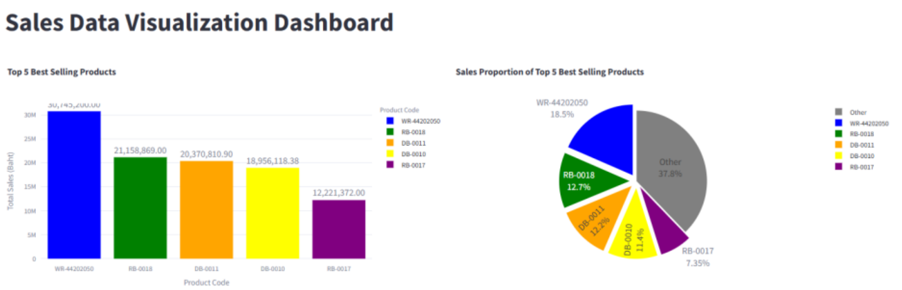

Figure 14: Sales Overview Dashboard: Top 5 Best-Selling Products and Revenue Proportion

The second section, Analysis, focuses on data exploration and statistical diagnostics. It includes a time series plot of raw sales data and a box plot for detecting potential outliers that may impact model accuracy. Seasonal trends reveal monthly fluctuations for the top five steel products between 2020 and mid-2024, as shown in Figure 15. Certain products, such as DB-0010 and RB-0017, exhibit sharp sales spikes during specific periods, possibly due to bulk orders or project-based demands, while others like RB-0018 demonstrate more consistent sales behavior.

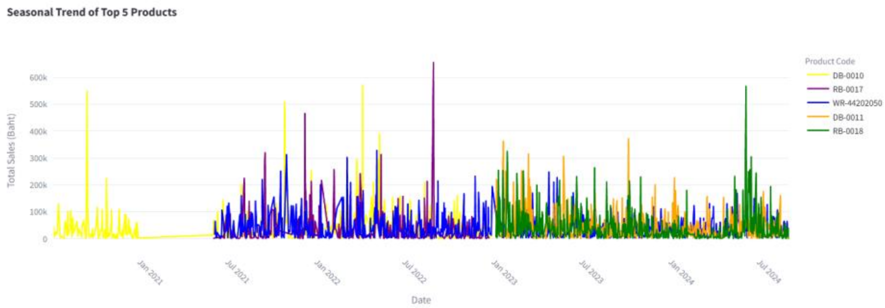

Figure 15: Seasonal Trend of Top 5 Products

Statistical analysis for outlier detection further supports data validation and quality assurance, as illustrated in Figure 16.

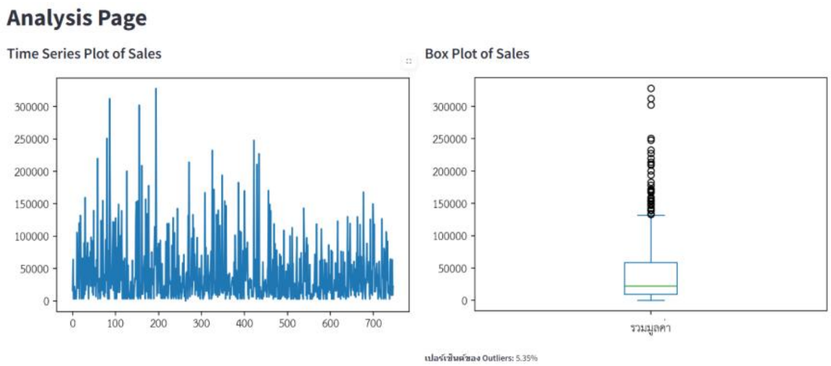

Figure 16: Statistical Analysis Page – Time Series and Outlier Detection

The third section, Forecasting, provides an interactive environment for applying and visualizing ARIMAX model predictions. Users can select forecasting horizons of 7, 30, 90, or 175 days and instantly view projected trends with associated confidence intervals. Historical forecasting comparisons show a close alignment between observed values and the predicted 90-day steel sales trend (Trend_90), indicating high model accuracy for medium- to long-term planning, as shown in Figure 17.

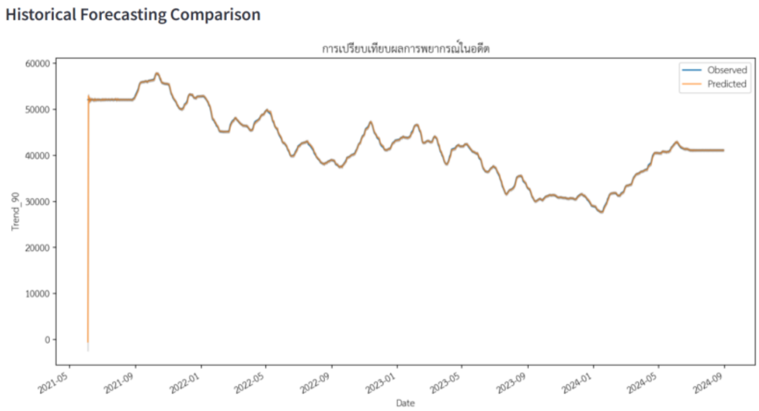

Figure 17: Historical Forecasting Comparison

The forecasted sales trend for a 175-day horizon provides extended visibility for strategic planning, as presented in Figure 18.

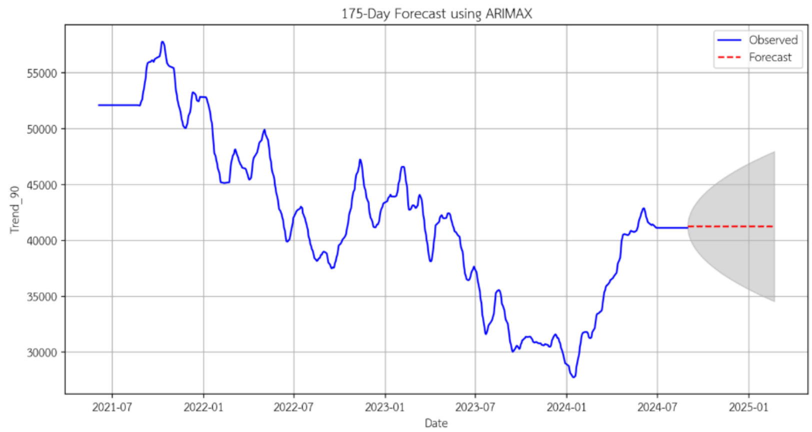

Figure 18: Forecasted Sales Trend using ARIMAX – 175-Day Horizon

Forecasted Trend_90 values from August 31 to September 9, 2024, along with the 95% confidence intervals, demonstrate stability in predictions, with values ranging between approximately 41,150 and 41,180 Baht, as shown in Figure 19. The narrow confidence intervals reflect high certainty in the model's short-term predictions, largely due to the low variability of exogenous variables such as economic indices or import indicators, which change gradually over time.

Forecast Table

Date	Forecasted_Trend_90	Lower_CI	Upper_CI
2024-08-31 00:00:00	41,162.54	40,900.81	41,424.27
2024-09-01 00:00:00	41,174.67	40,605.87	41,743.47
2024-09-02 00:00:00	41,164.24	40,403.6	41,924.88
2024-09-03 00:00:00	41,178.54	40,265.52	42,091.56
2024-09-04 00:00:00	41,157.74	40,114.36	42,201.12
2024-09-05 00:00:00	41,177.18	40,018.01	42,336.36
2024-09-06 00:00:00	41,182.48	39,918.08	42,446.89
2024-09-07 00:00:00	41,160.26	39,798.74	42,521.78
2024-09-08 00:00:00	41,182.46	39,730.29	42,634.62
2024-09-09 00:00:00	41,179.09	39,641.61	42,716.56

Figure 19: Forecasted Trend_90 with 95% Confidence Intervals – Aug 31 to Sep 9, 2024

By consolidating forecasting, diagnostics, and visualization into a single accessible interface, the dashboard effectively operationalizes predictive analytics for practical business use. It empowers non-technical users to interpret forecasting results with confidence, while fostering cross-departmental collaboration by providing a shared, real-time decision-support tool that aligns with both operational needs and strategic planning objectives.

3.7 Tools and Environment

The sales forecasting models in this study were developed entirely in Python, leveraging its extensive ecosystem for time-series modeling, statistical analysis, and visualization. Google Colab served as the primary development environment, offering cloud-based execution with GPU acceleration, pre-installed libraries, and convenient sharing for collaborative research.

For deployment to non-technical stakeholders, an interactive dashboard was implemented using Streamlit, enabling intuitive exploration of forecast results without requiring programming expertise. Core libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Plotly supported data manipulation and visualization, while Statsmodels and Scikit-learn facilitated statistical modeling and evaluation. Deep learning models, including LSTM and the Hybrid MLP, were implemented with TensorFlow and Keras, complemented by keras_tuner for hyperparameter optimization.

4 Result

This chapter summarizes the results of forecasting steel sales using historical data and macroeconomic indicators. Four models ARIMAX, LSTM, Hybrid ARIMAX & MLP, and VARX were developed and evaluated using statistical metrics to identify the most accurate and reliable approach.

4.1 Forecasting Model Performance

To evaluate the accuracy of the forecasting models developed in this study, four commonly used error metrics were applied: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). These metrics provide a comprehensive assessment of model performance by capturing both the magnitude of absolute errors and the relative percentage deviations between predicted and actual values. MAE and RMSE quantify the average prediction error in absolute terms, with RMSE placing greater weight on larger errors. Meanwhile, MAPE offers insight into the forecast accuracy in percentage terms, which is particularly useful for understanding model reliability across varying scales. The combined use of these metrics enables a robust comparison of forecasting models under different assumptions and complexities. The results of these evaluations are summarized in Table 3.

Model	MAE	MSE	RMSE	MAPE
ARIMAX	4.28	18.28	4.28	0.01%
LSTM	1,099.47	1,986,197.45	1,409.33	2.83%
Hybrid ARIMAX & MLP	602.52	363,027.84	602.52	1.47%
VARX	23.86	1,009.92	31.78	0.06%

Table 3: Comparison of Forecasting Model Performance Metrics

4.2 Interpretation of Experimental Results

The ARIMAX model delivered the highest accuracy, with low RMSE and MAPE, effectively capturing linear patterns and integrating external economic variables. Although LSTM excels at learning nonlinear relationships and temporal dependencies, it performed the weakest in this study due to the limited dataset size and the discontinuous nature of daily sales data. The Hybrid ARIMAX & MLP model outperformed both LSTM and VARX by combining linear forecasting with nonlinear residual learning, but still fell short of ARIMAX's performance. The VARX model handled interdependent variables well for short-term trend forecasting (Trend_90) but performed poorly for total sales forecasting, likely due to model complexity or suboptimal parameter selection.

4.3 Summary of Key Findings

- The experimental evaluation of four forecasting models ARIMAX, LSTM, Hybrid ARIMAX & MLP, and VARX yielded the following key insights regarding forecasting accuracy and model behavior.
- ARIMAX consistently outperformed all other models, achieving the lowest RMSE (4.28) and MAPE (0.01%). Its success is attributed to its ability to model linear trends while incorporating external macroeconomic indicators such as GDP growth and construction spending.
- LSTM exhibited the weakest performance, with an RMSE of 1,409.33 and MAPE of 2.83%. The limited dataset (747 monthly records) and lack of fine-grained temporal resolution (e.g., daily data) likely hindered its ability to learn complex temporal patterns effectively.
- The Hybrid ARIMAX & MLP model showed intermediate accuracy, with RMSE of 602.52 and MAPE of 1.47%. While it improved upon LSTM by capturing non-linear residuals, it still fell short of outperforming ARIMAX. The model's sensitivity to residual structure and parameter tuning suggests it requires careful calibration.
- VARX performed well on short-term trend forecasting (Trend_90) with RMSE of 31.78 and MAPE of 0.06%, but its results on total sales were unstable due to over-parameterization and potentially inappropriate lag selection.
- ARIMAX's interpretability and low data requirements make it practical for industrial applications, especially when stakeholders require reliable forecasts for operational planning and resource allocation.
- Deep learning models like LSTM may require larger and more granular datasets to perform effectively, while hybrid models offer a promising compromise when both linear and non-linear patterns exist.
- VARX is suitable for multivariate trend analysis but can be computationally expensive and less robust in small-data scenarios due to its sensitivity to model complexity.

These findings emphasize the importance of matching the model architecture to both the structure and size of the dataset. In this study, ARIMAX proved most effective in balancing accuracy, interpretability, and computational efficiency for steel sales forecasting.

5 Conculsion and Discussion

5.1 Conculsion

This research aimed to develop and compare forecasting models for predicting future steel sales using historical sales data combined with relevant economic indicators. Four forecasting models were implemented and evaluated ARIMAX, LSTM, VARX, and a Hybrid ARIMAX-MLP model. Evaluation metrics including MAE, MSE, RMSE, and MAPE were used to assess each model's accuracy.

Among all the models, the ARIMAX model yielded the best overall performance, showing the lowest forecasting error. This result emphasizes the suitability of ARIMAX for time series data with exogenous variables, especially in economic and industrial contexts. The best-performing ARIMAX model was also deployed using the Streamlit framework, enabling users to interactively visualize and interpret forecasted trends.

5.2 Discussion

The findings indicate that the performance of forecasting models depends largely on the data structure and the complexity of the model applied. The ARIMAX model demonstrated the highest performance due to its strength in capturing linear trends and its ability to effectively incorporate external economic indicators to explain steel sales behavior. In contrast, although the LSTM model is designed to learn complex nonlinear patterns, its performance was limited by the relatively small training dataset (747 records), preventing it from reaching its full potential. The Hybrid ARIMAX-MLP model, which integrates both linear and nonlinear modeling approaches, also failed to outperform the standalone ARIMAX. This may be attributed to overfitting on the residuals or insufficient parameter tuning.

For the VARX model, which is inherently suited for multivariate time series forecasting, practical implementation revealed some limitations. The large number of variables and the high lag order required for optimal performance led to over-parameterization and increased computational burden, with no substantial gain in forecasting accuracy. As a result, VARX proved less efficient and not cost-effective in this context.

To present the forecasting results in a user-friendly and interactive format, a lightweight web application was developed using Streamlit. While there are other platforms such as Gradio, Dash, or Voila that can serve a similar purpose, Streamlit was chosen due to its simplicity, seamless integration with Python, and suitability for rapid prototyping. Since this study was entirely Python based from data preprocessing to model development and evaluation Streamlit offered a consistent, low-cost, and open-source solution that aligned well with the research workflow. Additionally, its interactive nature allows end users to easily explore forecasting results and gain insights for strategic decision-making.

Nonetheless, the study also identified certain internal organizational factors that may significantly impact steel sales but were not included in the models. These factors include inventory management and product aging, which can directly influence pricing and profitability. For example, steel that is kept in storage for an extended period may develop rust or degrade in quality, reducing its market value. In cases where production or procurement exceeds actual demand, companies may face a "stock overhang," prompting price reductions to liquidate inventory and maintain cash flow. These factors reflect sales behavior that cannot be fully explained by economic indicators alone.

Therefore, to enhance forecasting accuracy and better capture real-world business dynamics, future model development should consider incorporating internal logistics data such as stock levels, purchasing cycles, and inventory turnover rates. Doing so would improve not only the precision of forecasts but also support more effective planning in production, cost management, and inventory control.

6 Future Work

To build upon the current findings, future research may explore the following directions

- **Data Enrichment:** Collect more granular and long-term data, including external market and policy indicators, to enhance model training and performance.

- **Refined Variable Selection for ARIMAX:** For future research, it is essential to refine the variable selection process within the ARIMAX model. Specifically, a new model should be constructed using only exogenous variables that demonstrate statistical significance (P-value < 0.05) after the initial training. This approach aims to produce a more concise and interpretable model. However, reducing the number of explanatory variables may impact the accuracy of the forecast. Therefore, it is recommended to conduct a comprehensive evaluation of the model's performance to determine whether ARIMAX remains the most suitable method, and to reforecast using the revised set of variables that align with the updated model structure.

- **Real-time Forecasting:** Develop a data pipeline that allows real-time integration of updated indicators, enabling adaptive forecasting and timely monitoring of market dynamics.

- **Decision Support Integration:** Expand the scope of the model to be part of a comprehensive decision-support system for production planning, inventory management, and risk assessment.

References

1. Venkatachary, M., Subbaiah, V. K., Satyanarayana, B., & Rao, S. R. (2021). Econometric Modeling for Forecasting Steel Demand in India. *International Journal of Advances in Engineering and Management (IJAEM)*, 3(7), 2628-2640.
2. Sonja Strasser, Shailesh Tripathi. (2024). Forecasting Steel Demand: Comparative Analysis of Predictability across Diverse Countries and Regions. Procedia Computer Science, 232, 2740–2750.
3. Office of Industrial Economics. (2024). รายงานภาวะเศรษฐกิจอุตสาหกรรม [Industrial Economics Situation Report], June 2024. Bangkok, Thailand: Ministry of Industry.
4. Christogonus Ugoh, C. A. Uzuke (2021). Application of ARIMAX Model on Forecasting Nigeria's GDP. American Journal of Theoretical and Applied Statistics.
5. S M Ulyah (2019). Forecasting index and stock returns by considering the effect of Indonesia pre-presidential election 2019 using ARIMAX and VARX approaches. Journal of Physics: Conference Series, 1277(1).
6. Wiwik Anggraeni, Retno Aulia Vinarti, Yuni Dwi Kurniawati (2015). Performance Comparisons Between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study. Procedia Computer Science, 72, 630-637.
7. Tian Guo, Tao Lin (2018). Multi-variable LSTM neural network for autoregressive exogenous model. Journal of Computational and Applied Mathematics, 344, 305-317.
8. Thomas Shering, Eduardo Alonso, Dimitra Apostolopoulou (2024). Investigation of Load, Solar and Wind Generation as Target Variables in LSTM Time Series Forecasting, Using Exogenous Weather Variables. Energies, 17(8), 2371-2389.
9. Paulo S. G. de Mattos Neto, João Fausto Lorenzato de Oliveira, Domingos Savio de Oliveira Santos Júnior, Hugo Valadares Siqueira, Manoel Henrique Da Nóbrega Marinho, Francisco Madeiro (2020). A Hybrid Nonlinear Combination System for Monthly Wind Speed Forecasting. IEEE Transactions on Sustainable Energy, 11(4), 2041-2052.