computation

MDPI

Article

Incorporating Time-Series Forecasting Techniques to Predict Logistics Companies' Staffing Needs and Order Volume

Ahmad Alqatawna [1], Bilal Abu-Salih [1,*], Nadim Obeid [1] and Muder Almiani [2,*]

[1] King Abdullah II School of Information Technology, The University of Jordan, Amman 11942, Jordan; ahmadaymanq@hotmail.com (A.A.); nadim@ju.edu.jo (N.O.)
[2] Management Information Systems Department, Gulf University for Science and Technology, Kuwait City 32093, Kuwait
* Correspondence: b.abusalih@ju.edu.jo (B.A.-S.); almiani.m@gust.edu.kw (M.A.)

Abstract: Time-series analysis is a widely used method for studying past data to make future predictions. This paper focuses on utilizing time-series analysis techniques to forecast the resource needs of logistics delivery companies, enabling them to meet their objectives and ensure sustained growth. The study aims to build a model that optimizes the prediction of order volume during specific time periods and determines the staffing requirements for the company. The prediction of order volume in logistics companies involves analyzing trend and seasonality components in the data. Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) are well-established and effective in capturing these patterns, providing interpretable results. Deep-learning algorithms require more data for training, which may be limited in certain logistics scenarios. In such cases, traditional models like SARIMAX, ARIMA, and AR can still deliver reliable predictions with fewer data points. Deep-learning models like LSTM can capture complex patterns but lack interpretability, which is crucial in the logistics industry. Balancing performance and practicality, our study combined SARIMAX, ARIMA, AR, and Long Short-Term Memory (LSTM) models to provide a comprehensive analysis and insights into predicting order volume in logistics companies. A real dataset from an international shipping company, consisting of the number of orders during specific time periods, was used to generate a comprehensive time-series dataset. Additionally, new features such as holidays, off days, and sales seasons were incorporated into the dataset to assess their impact on order forecasting and workforce demands. The paper compares the performance of the four different time-series analysis methods in predicting order trends for three countries: United Arab Emirates (UAE), Kingdom of Saudi Arabia (KSA), and Kuwait (KWT), as well as across all countries. By analyzing the data and applying the SARIMAX, ARIMA, LSTM, and AR models to predict future order volume and trends, it was found that the SARIMAX model outperformed the other methods. The SARIMAX model demonstrated superior accuracy in predicting order volumes and trends in the UAE (MAPE: 0.097, RMSE: 0.134), KSA (MAPE: 0.158, RMSE: 0.199), and KWT (MAPE: 0.137, RMSE: 0.215).

Keywords: SARIMAX; ARIMA; AR; LSTM; logistics; time-series; forecasting; supply chain; machine learning; neural networks; orders forecasting

Citation: Alqatawna, A.; Abu-Salih, B.; Obeid, N.; Almiani, M. Incorporating Time-Series Forecasting Techniques to Predict Logistics Companies' Staffing Needs and Order Volume. *Computation* **2023**, *11*, 141. https://doi.org/10.3390/computation11070141

Academic Editors: Minzhang Zheng and Pedro Manrique

Received: 1 June 2023
Revised: 6 July 2023
Accepted: 7 July 2023
Published: 14 July 2023

1. Introduction

Accurate prediction of order volume is crucial for logistics companies to effectively allocate resources and meet customer demand. Inadequate resources can lead to delays and dissatisfied customers, while excessive resources can result in increased costs. Therefore, developing precise prediction models for order volumes is essential for efficient resource allocation and successful business planning. However, despite the availability of various time-series analysis methods and the complexity of the data, determining the most accurate prediction model remains a challenge in this field. Time-series analysis is a powerful technique used to analyze data patterns and trends and make predictions based on historical data within a specific time period. Widely employed in business, economics, and finance, this approach enables forecasting market trends, analyzing financial data, and predicting future demand. In conjunction with time-series analysis, machine learning leverages large datasets to develop predictive algorithms capable of identifying patterns and generating accurate predictions. By harnessing the power of data, machine learning empowers businesses to make informed decisions and anticipate future outcomes.

For 3PL providers, accurate order volume prediction is particularly important for effective resource planning and management. By utilizing advanced techniques such as time-series analysis and machine learning, 3PL providers can optimize their warehouse space, transportation resources, and labor allocation. This enables them to meet the expected demand efficiently, avoid resource shortages or excess, and ultimately enhance operational efficiency and customer satisfaction. However, given the dynamic nature of the logistics industry and the diverse factors influencing order volumes, developing accurate prediction models requires ongoing research and improvement. Despite the existing challenges, the continued advancement of data analysis techniques holds great promise for improving order volume predictions and enabling more efficient utilization of resources in the 3PL sector.

This paper aims to evaluate the performance of four distinct time-series analysis methods: Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), and Long Short-Term Memory (LSTM). The current time-series analysis methods employed in the logistics industry for order volume prediction have certain limitations that affect their accuracy. The dynamic nature of the logistics sector, characterized by constantly changing market trends, customer preferences, and external factors, poses a challenge to the existing methods. These methods may struggle to capture the intricate relationships and complexities inherent in the data, resulting in suboptimal predictions. Moreover, variations in order volumes due to seasonal patterns, exogenous variables, and other influential factors further add to the complexity of accurate prediction. Recognizing the need to overcome these limitations, our study introduces and evaluates four distinct time-series analysis methods: SARIMAX, ARIMA, AR, and LSTM. By examining the effectiveness of these methods in addressing the shortcomings of existing approaches, we aim to enhance decision-making in shipping operations and contribute to the field of order volume prediction in the logistics industry. The evaluation seeks to identify the most accurate method for improving decision-making in shipping operations by predicting order volume across three different countries. By comparing the methods using order volume data obtained from a shipping company, their effectiveness in predicting order volumes in the United Arab Emirates (UAE), Kingdom of Saudi Arabia (KSA), Kuwait (KWT), and overall will be examined.

To fit the models and forecast future order volume and trends, the time-series data will be preprocessed, and the four methods will be applied. The findings of this study will shed light on the most reliable method for predicting order volumes. Additionally, the paper will discuss a formula for estimating staffing needs in organizations handling orders, acknowledging its limitations and emphasizing the need for further research to enhance its accuracy. In conclusion, accurate prediction of order volumes is crucial for logistics companies to optimize resource allocation and ensure efficient operations. Time-series analysis, combined with machine learning algorithms, offers a powerful approach to forecasting order volumes. Through evaluating different methods, this paper aims to contribute to the field by identifying the most accurate model for predicting order volume and enhancing decision-making in the shipping industry.

2. Related Works

Managing staffing levels and order volumes accurately is crucial to a logistics company's ability to remain competitive. By using time-series forecasting and machine learning, logistics demand forecasting can be improved in terms of efficiency and accuracy. The literature review summarizes recent studies on time-series forecasting models and machine-learning techniques for logistics and supply chain forecasting.

According to a study by Singha and Panse [1], when managing supply chains, forecasting customer demand is important, but forecasting errors can lead to challenges. The study examined advanced algorithms for machine learning, such as MLP, CNN, and LSTM networks, specifically for time-series forecasting. Singha and Panse conducted their analysis using Kaggle's 'Store Item Demand Forecasting' dataset. They created a comparative forecasting mechanism using the ANN approach to determine the most effective training technique for predicting demand signals. The results showed that the MLP technique outperformed other techniques in terms of accuracy estimation, although the characteristics of the data also influenced the model's accuracy. Singha and Panse recommended trying different techniques and preprocessing the data accordingly to achieve better results. By providing on-time delivery and improving forecasting accuracy, costs can be reduced, and customer satisfaction can be increased, thereby creating a powerful decision-support tool. Singha and Panse suggested that future studies should explore other types of ANN for improved prediction accuracy and even better results.

Lee et al. [2] suggested an enhanced prediction model for container volume in Busan ports, employing external variables and time-series decomposition techniques. The authors recognized that container volume data is influenced by various external factors, making it more complex and diverse than what traditional statistical methods can handle. While deep learning models excel at analyzing patterns, capturing time-series, external variables, and outliers, they have not extensively explored the overall trends of container volumes. To address this, the authors introduced a multivariate LSTM prediction approach combined with port volume time-series decomposition, resulting in improved prediction accuracy. It is anticipated that future research focusing on a more detailed approach will further enhance container volume forecasting capabilities.

Forecasts of containerized freight volumes are crucial for port terminal operators, port authorities, regulators, and governmental agencies. Ferretti et al. [3] compared multivariate regression models based on deep learning and seasonal autoregressive integrated moving averages to illustrate the potential of using deep learning models for forecasting container throughput. A mapping in latent space is proposed as an innovative representation of seasonality. According to the study, SARIMA and CNN did not perform as well as recurrent neural networks in either forecasting scenario. A more accurate forecast can be achieved by incorporating external regressors and additional available information. According to the study, adopting deep learning forecasting models is beneficial not only for container terminal operators but also for local and national government institutions responsible for port management.

The SARIMA is used in another study by Clarabelle and Gatc [4] to predict the number of passengers aboard a ship. The study aims to address the issue of inadequate safety equipment due to the fluctuating visitors to the Thousand Islands. AR (Autoregression) and MA (Moving Average) parameters were determined using ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots. The study used SARIMA because it can provide short-term predictions with seasonal elements. An analysis of SARIMA (3,0,1) (1,0,1)7 was conducted to predict 90 days in advance. According to the results of the study, the predictions obtained were based on predictions for passengers on ships without COVID-19 outbreaks.

To avoid risks and maximize benefits, a study by Li and Wei [5] examines the impact of e-commerce and the novel Coronavirus pandemic on the logistics industry. The authors proposed using freight volume as a measure of logistics needs and LSTM networks as a predictor model. Based on the Changsha logistics needs prediction index system, the authors compare LSTM results with the following: Grey Model (1, 1), linear regression, and backpropagation neural networks. Based on the results, the LSTM model has the smallest prediction errors compared to other models used to predict regional logistics needs post-epidemic. As well as noting several limitations, the study suggested that the

LSTM network could be further optimized, including the subjective selection of the index system. Research on regional logistics forecasting indexes should be extended to other angles in the future, according to the authors.

Fadda et al. [6] introduce a novel approach to address the challenges faced in managing vehicle fleets in last-mile logistics. With the rise of the on-demand economy and the increasing demand for fast and extensive parcel deliveries, last-mile operators often rely on third-party services and rented vehicles to optimize operational costs. To estimate the required quantity and types of vehicles for the next day's deliveries, the authors tackle the Tactical Capacity Planning (TCP) problem. They propose a solution that combines machine learning and optimization techniques to forecast vehicle requirements using historical data. Their approach introduces the concept of microzones in the city and solves the deterministic Variable Cost and Size Bin Packing Problem (VCSBPP). The paper provides a detailed description of the implementation, examines various forecasting methods, and compares their performance. Significantly, this study represents the first attempt to incorporate microzones in demand forecasting for the TCP problem. The paper concludes by discussing related approaches, outlining the methodology, presenting the results, and suggesting future directions.

The study by Bruni et al. [7] explores the challenges and solutions in the last-mile delivery sector, which is affected by the growing urban population and increased demand for services and goods. The study focuses on the adoption of third-party logistics (3PL) as a management solution to enhance efficiency and mitigate uncertainties. The authors discuss the Variable Cost and Size Bin Packing problem with Stochastic Items (VCSBPPSI) and the limitations of traditional solvers. They propose a machine learning (ML) heuristic based on supervised classification techniques to address these limitations and improve solution accuracy. The study highlights the contributions of the proposed ML heuristic, including its application to other two-stage stochastic problems, its independence from other solution methods, and its integration into a real case study.

The growing global demand for power load forecasting has led to the development of the Informer model, which aims to predict future power loads using historical load data. Xu et al. [8] proposed a multi-step power-load forecasting model based on Informer, employing a seq2seq structure with a sparse self-attention mechanism and specific input and output modules. By effectively addressing long-range relationships in time-series data and leveraging the parallel advantages of self-attention, the Informer model improves both prediction accuracy and efficiency. The authors validate the model by training, verifying, and testing it using the power-load dataset from the Taoyuan substation in Nanchang. Comparative analysis against traditional models, including RNN, LSTM, and LSTM with attention mechanisms, demonstrates the superior performance of the Informer-based power-load forecasting model, particularly within 1440 time steps. This research contributes to the potential of the Informer model in power-load forecasting, offering more refined predictions for complex application scenarios and valuable insights for long-term grid system planning. Future research will focus on refining the model through improved experimental environments, exploring the incorporation of additional environmental factors, and investigating its application in other domains, such as photovoltaic and wind power.

3. Methodology

3.1. Data Collection

The statistical data used in this paper was collected from the database of a shipping company. The data was exported from the database using NetSuite, a comprehensive software suite that integrates various business functions such as accounting, ERP, CRM, and e-commerce. The dataset focused on the volume of orders during specific time periods, collected from different countries and clients. The dataset comprises approximately 8,505,971 records. It contains various information, including client names, order placement and receipt dates, delivery country, and shipment description. The data used in the experiment consists of the daily volume of orders, covering a time frame of 730 days from 1 January 2021 to 31 December 2022.

3.2. Preprocessing

The data collection process involved using NetSuite to gather daily order information from 2021 to 2022, including deliveries, returns, and closures. Each country's order count was separately recorded. Before proceeding with the data analysis, several preprocessing steps were implemented. This involved cleaning the data to handle missing values and removing test-generated orders that could distort the results. Empty or duplicated values were eliminated using MS Excel to ensure data accuracy and consistency. After these operations, the dataset included the following columns: date, order counts for UAE, Kuwait, and KSA, counts for delivered, closed, and returned orders, a column indicating public holidays with corresponding descriptions, and a column indicating sales seasons. The final dataset consisted of 730 rows, properly arranged and ready for further analysis. Table 1 presents the size and type of the dataset across the countries examined.

Table 1. Data set size and type across studied countries.

Country	Years	Type of Data	Columns Prior-Preprocessing	Columns Post-Preprocessing	Number of Orders
UAE	2021–2022	Daily Data	12	16	4,285,173
KSA	2021–2022	Daily Data	12	16	3,555,205
KWT	2021–2022	Daily Data	12	16	678,903
All Countries	2021–2022	Daily Data	12	16	8,519,281

Data size prior to preprocessing: 7.55 GB.

3.3. Models and Procedure

In this experiment, we evaluated the performance of four time-series analysis models, namely SARIMAX, ARIMA, AR, and LSTM, for predicting future order volumes in a shipping company. SARIMAX is a classic model that combines autoregressive, moving average, and seasonal factors for data modeling. ARIMA utilizes autoregressive and moving average components, while AR solely relies on autoregressive factors. In contrast, LSTM is a deep learning model capable of capturing long-term dependencies in time-series data. These models were chosen due to their widespread use in time-series analysis and their proven effectiveness in predicting future values. The accuracy of predicting future order volumes served as the evaluation criterion for each model's performance.

SARIMAX: A SARIMAX model incorporates extra external predictors, or exogenous variables, to extend the traditional ARIMA model. Unlike ARIMA, this model takes seasonality and external factors into account and assumes that the time series is stationary. SARIMAX can provide more accurate forecasts by capturing the effects of external factors on the time series [4]. The application of seasonal difference to time-series data can remove seasonal fluctuations, as indicated by the following equation for SARIMA [9]:

$$\Phi_p(B^s)\phi_p(B)(1-B^s)^D(1-B)^d y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \tag{1}$$

ARIMA: An ARIMA, a combination of an autoregressive model and a moving average model, is an effective technique for analyzing and stabilizing non-stationary time-series data. As a result, we obtain the following equation [9]:

$$\phi_p(B)(1-B^s)^d y_t = \theta_q(B)\varepsilon_t \tag{2}$$

AR: the autoregressive model (AR) assumes that a variable's current value is a linear combination of its past values plus a random error term. Economic, finance, and engineering fields use this model to forecast future values of a time series by analyzing its past behavior. The AR model is mathematically expressed by the following equation [10].

$$y\_t = c + \Sigma\_(i=1)\hat{}p\varphi\_iy\_(t-i) + \varepsilon\_t \quad (3)$$

LSTM: LSTM is an innovative and powerful architecture for deep learning based on recurrent neural networks (RNN) [11]. It was developed to overcome the problem of vanishing gradients that are encountered by traditional RNNs during training. Speech recognition and handwriting recognition are two of the many tasks in which LSTM networks are used today. As this architecture captures underlying information between events with unknown time lags, it is particularly suitable for analyzing time-series data. A typical LSTM unit or neuron consists of various components, as depicted in Figure 1, such as a cell, an input gate, an output gate, and a forget gate [11].

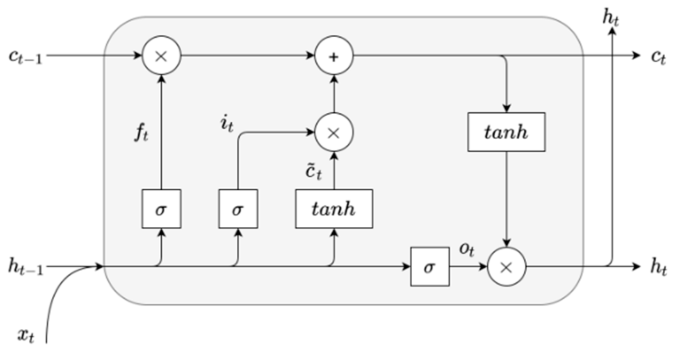

Figure 1. The organization of an LSTM unit.

The diagram illustrates an activation vector for each of the gates in memory, input, output, and forget as f_t, i_t, and o_t, respectively. The LSTM unit takes input vector x_t and generates a vector representing the hidden state h_t while preserving the cell state vector c_t and cell input activation vector $\tilde{c}_t$. These gates operate collaboratively using the following equations, where this equation involves weight matrices and bias vectors denoted by W, U, and b. And σ_g, σ_c, and σ_h represent distinct activation functions. The operator $\circ$ denotes element-wise multiplication [11].

$$\begin{gathered} f_t = \sigma_g\left(W_f x_t + U_f h_t - 1 + b_f\right) \\ i_t = \sigma_g(W_i x_t + U_i h_t - 1 + b_i) \\ o_t = \sigma_g(W_o x_t + U_o h_t - 1 + b_o) \\ \tilde{c}_t = \sigma_c(W_c x_t + U_c h_t - 1 + b_c) \\ c_t = f_t \circ c_t - 1 + i_t \circ \tilde{c}_t \\ h_t = o_t \circ \sigma_h(c_t) \end{gathered} \quad (4)$$

The LSTM (Long Short-Term Memory) model receives an input, x(t), which can either be the output of a CNN (Convolutional Neural Network) or the input sequence itself. The model takes into account the inputs from the previous time step, h_{t-1} and c_{t-1}. It generates an output, o_t, for the current time step and produces the values c_t and h_t to be utilized by the next time step LSTM.

The LSTM equations additionally compute f_t, i_t, and $\tilde{c}_t$, which are internal values utilized by the LSTM to generate c_t and h_t [11].

The aforementioned equations are specific to a single time step, requiring recomputation for subsequent time steps. For instance, if there is a sequence of 10 time steps, these equations will be executed 10 times, once for each respective time step. Notably, the weight matrices (W_f, W_i, W_o, W_c, U_f, U_i, U_o, U_c) and biases (b_f, b_i, b_o, b_c) remain constant over time. Therefore, the same weight matrices are employed to calculate outputs for different time steps, ensuring consistency in the model's computations [11].

4. Experiments and Results

These experiments describe how time-series forecasting techniques can be used to predict order volumes in logistics companies. Identifying reliable and accurate forecasting methods for these critical business metrics was the objective of this study. Furthermore, these techniques can be incorporated into logistics operations to achieve potential benefits. Our predictions were based on data collected from a shipping company. We then compared these predictions with actual order volume data to predict future order volume in the logistics industry. Figure 2 illustrates the process flow for the study.

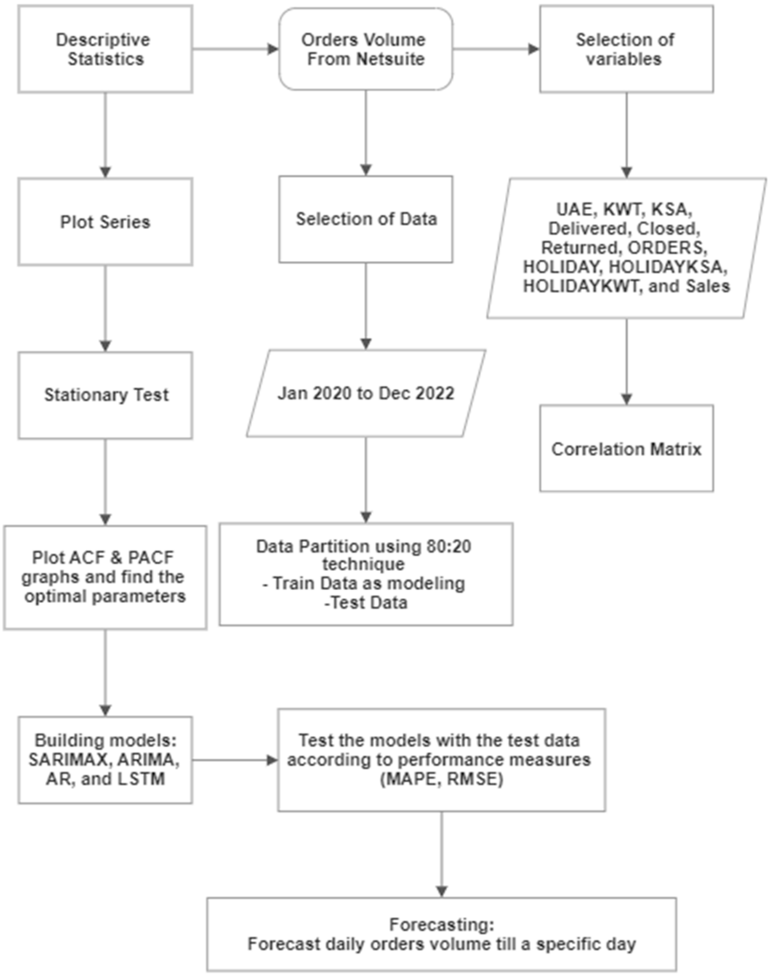

Figure 2. The experiment process flow.

The experiments in this study were conducted on Intel® Core™ i7-8550U and running Microsoft Windows 11 Pro 64-bit. The system is implemented with version 3.10 of the Python programming language, which is one of the most popular data science and DM languages; it is portable and open-source.

4.1. Descriptive Statistics

The order volume of the company varies across different operating countries, with some countries experiencing high volumes while others have relatively low volumes. Unique patterns and trends specific to the company and its operating countries were observed in the data. Certain countries exhibit higher demand for logistics services, and there are consistent variations in demand throughout the year. By analyzing the data characteristics in detail, we gain insights into the factors influencing demand and how they differ across countries.

To explore the relationships between variables in the time-series data, a heatmap was used to visualize correlations. The heatmap revealed a strong correlation between variables such as “Delivered”, “KSA”, and “Orders”. However, since these variables are part of the “Orders” column, they cannot be used as exogenous variables. On the other hand, the heatmap indicated a strong association between the “Sales” and “Holidays” columns with the “Orders” column. Based on this observation, the decision was made to include the “Sales” and “Holidays” columns as exogenous variables in the SARIMAX model. This inclusion improved the model’s performance and enabled better capturing of the underlying structure of the data. Figure 3 illustrates the correlation between the incorporated features.

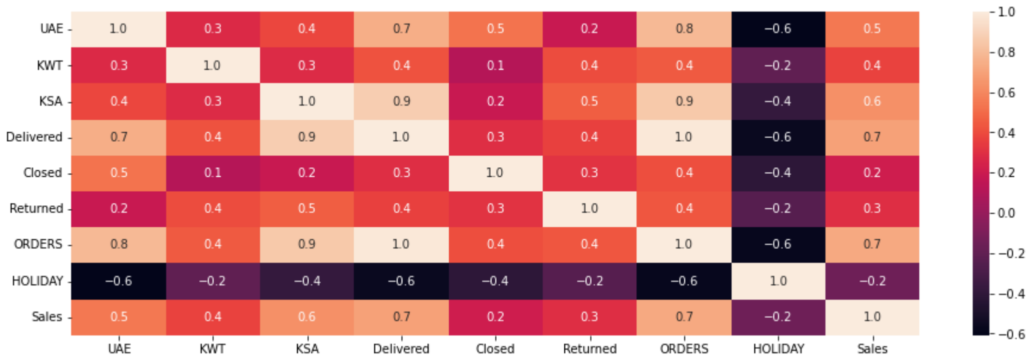

Figure 3. The correlation between the incorporated features.

4.2. Stationarity Test

In time-series analysis, determining stationarity is crucial as it assesses whether the statistical properties of a time series remain constant over time. Stationary time series exhibit consistent mean, variance, and autocorrelation, while non-stationary time series undergo changes in these properties. To test for stationarity, the Dickey–Fuller test was employed in this study [12]. This statistical test involves a null hypothesis that assumes the data is non-stationary. If the null hypothesis is rejected, it indicates stationarity. The adfuller() function from the statsmodels.tsa.stattools library was utilized to calculate the test statistic and p-value, using a significance level of 0.05. By comparing the test statistic to the critical value, it was determined whether to reject the null hypothesis. If the test statistic was lower than the critical value, the data was considered stationary; otherwise, it was deemed non-stationary. Based on the results of the Dickey–Fuller test, the time-series data used in the study was found to be stationary.

4.3. Data Splitting

We applied the 80:20 split method, a commonly used technique in time-series analysis and machine learning. This method divides the dataset into a training set (80%) and a test set (20%), which are used to train the model and evaluate its performance on unseen data, respectively. The advantage of this method is that it allows us to assess the model’s performance in a realistic setting and to detect overfitting or underfitting issues. This approach is crucial in ensuring that our model will perform well on new data in the future.

4.4. Model Implementation

SARIMAX: It is essential to determine the parameters of the SARIMAX model, i.e., p, d, q, P, D, and Q, in advance. By using a partial autocorrelation function (PACF) and an autocorrelation function (ACF), one can examine the presence of stationary and autocorrelation structures or the SARIMAX model can be used in some packages to identify the characteristics of the data and set parameters automatically [13].

In order to identify the appropriate parameters for the SARIMAX model, we combined two methods. To identify significant lags that had a correlation with the time series, we first examined partial autocorrelation functions (PACF) and autocorrelation functions

(ACF). As a result, we were able to estimate the orders of the seasonal and non-seasonal components of the model, namely p, d, q, P, D, and Q. The plots in Figure 4 illustrate partial autocorrelation functions (PACFs) and autocorrelation functions (ACFs). In addition, we applied a hyperparameter tuning procedure to maximize the model's performance. Various combinations of parameters were systematically tested in the code, and the parameter combination that produced the lowest mean squared error (MSE) was selected.

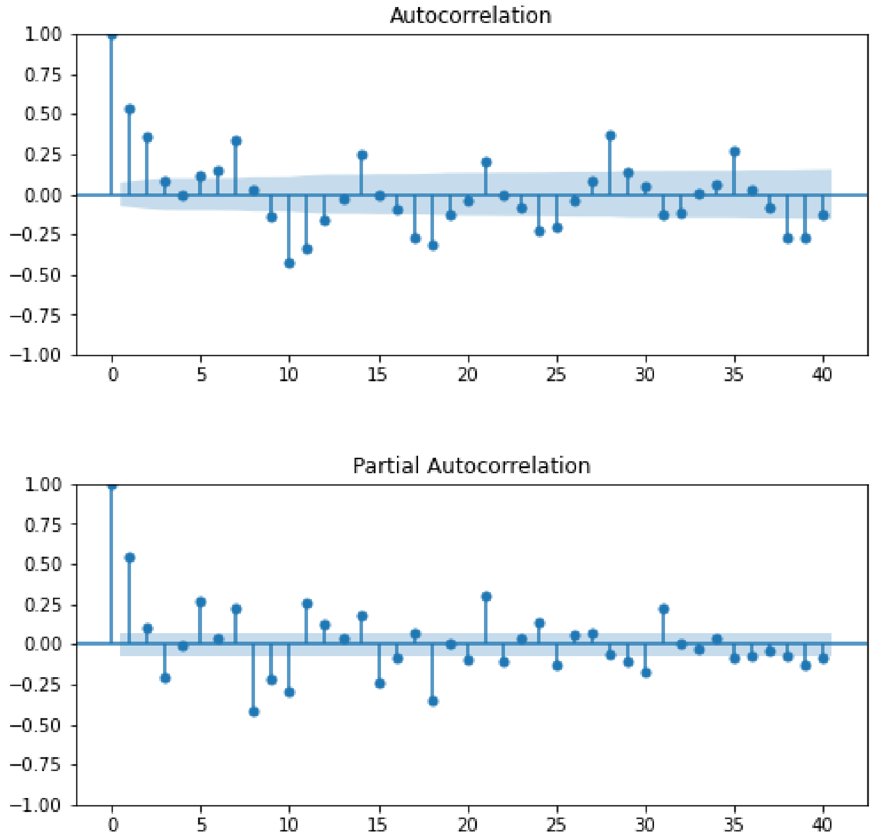

Figure 4. PACF and ACF plots.

Non-seasonal component orders:

By identifying the lag at which the PACF value drops to zero or close to zero, we can determine p, the order of the autoregressive (AR) component [14].
By identifying the lag at which the ACF value drops to zero and near zero, we were able to determine q, which is the order of the MA component [14].
Using the ACF plot, we determined the value of d, which represents the degree of differencing. It is necessary to differentiate until the ACF plot shows a gradual decline [14].

Seasonal component orders:

In order to obtain the seasonal autoregressive (SAR) component order, the seasonal differences of time series were plotted using the PACF statistic [14].
Using the ACF plot of seasonal differences, we determined the order of the seasonal moving average (SMA) component [14].
Using the seasonal ACF plot, we were able to determine the value of D, which is the degree of seasonal differentiation. ACF plots must differ until they show a gradual decline over time [14].

The results of the finest parameters analysis can be seen in Table 2. Using these parameters, we will create SARIMAX, predicting future trends in order volumes in the countries listed.

Table 2. SARIMAX inputs.

Country/Region	Seasonal	p	d	q	P	D	Q	m
UAE	Yes	4	0	4	2	1	2	7
KSA	Yes	4	0	4	2	1	3	7
KWT	Yes	3	0	3	2	1	1	7
All Countries	Yes	3	0	3	3	1	3	7

ARIMA: The values for the parameters p, d, and q of the ARIMA model are determined by analyzing the autocorrelation and partial autocorrelation plots. Once the parameters are determined, the ARIMA model is trained on the training set and then tested on the testing set. The efficacy of the model is evaluated using root mean squared error (RMSE) and mean squared error (MSE).

To ensure that the residuals of the model follow a normal distribution, diagnostic plots are created for the model's predictions and actual values. These plots help assess the goodness of fit and validate the assumptions of the ARIMA model. The results of the finest parameters analysis can be seen in Table 3.

Table 3. ARIMA inputs.

Country/Region	Seasonal	p	d	q
UAE	Yes	10	2	7
KSA	Yes	11	1	9
KWT	Yes	10	1	7
All Countries	Yes	14	1	8

AR: To implement the AR model in Python, we used the pandas [12] library to load the dataset and set the index to the date column. We then used the statsmodels library [12] to fit the AR model and predict future values based on the best number of lags determined by the AIC value. To evaluate the model's performance, we calculated the mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The actual and predicted values were also plotted to visualize the model's accuracy. The implementation of the AR model does not require stationarity testing or differencing, as it is designed to work with stationary time-series data.

LSTM: In order to build the model, we utilized the Python programming language along with the pandas, Keras, NumPy, and matplotlib libraries [12]. The dataset was loaded using pandas and preprocessed by dropping the date column and normalizing the features and target. We then created input and target sequences using a sliding window approach with a window size of 7. There were two sets of data, training and testing, with 80 percent of the data used for training. We then constructed an LSTM model using Keras with a single hidden layer containing 100 neurons and compiled it using mean squared error as the loss function and Adam optimizer. The training process involved training the model on the training set using a batch size of 32 for 30 epochs. After training, we made predictions on the test data and denormalized both the predictions and test data to calculate the root mean squared error (RMSE) and mean absolute percentage error (MAPE) using NumPy.

4.5. Data Analysis

In order to thoroughly evaluate the performance of our time-series forecasting model, the residuals were visually analyzed using the plot_diagnostics() function to produce histograms, normal Q-Q plots, and correlogram plots. These diagnostic plots helped us to understand the distribution of the residuals, identify any patterns or biases, and assess the relationship between the residuals and their lags. Figure 5 illustrates the plot diagnostics results for the All Courtiers dataset.

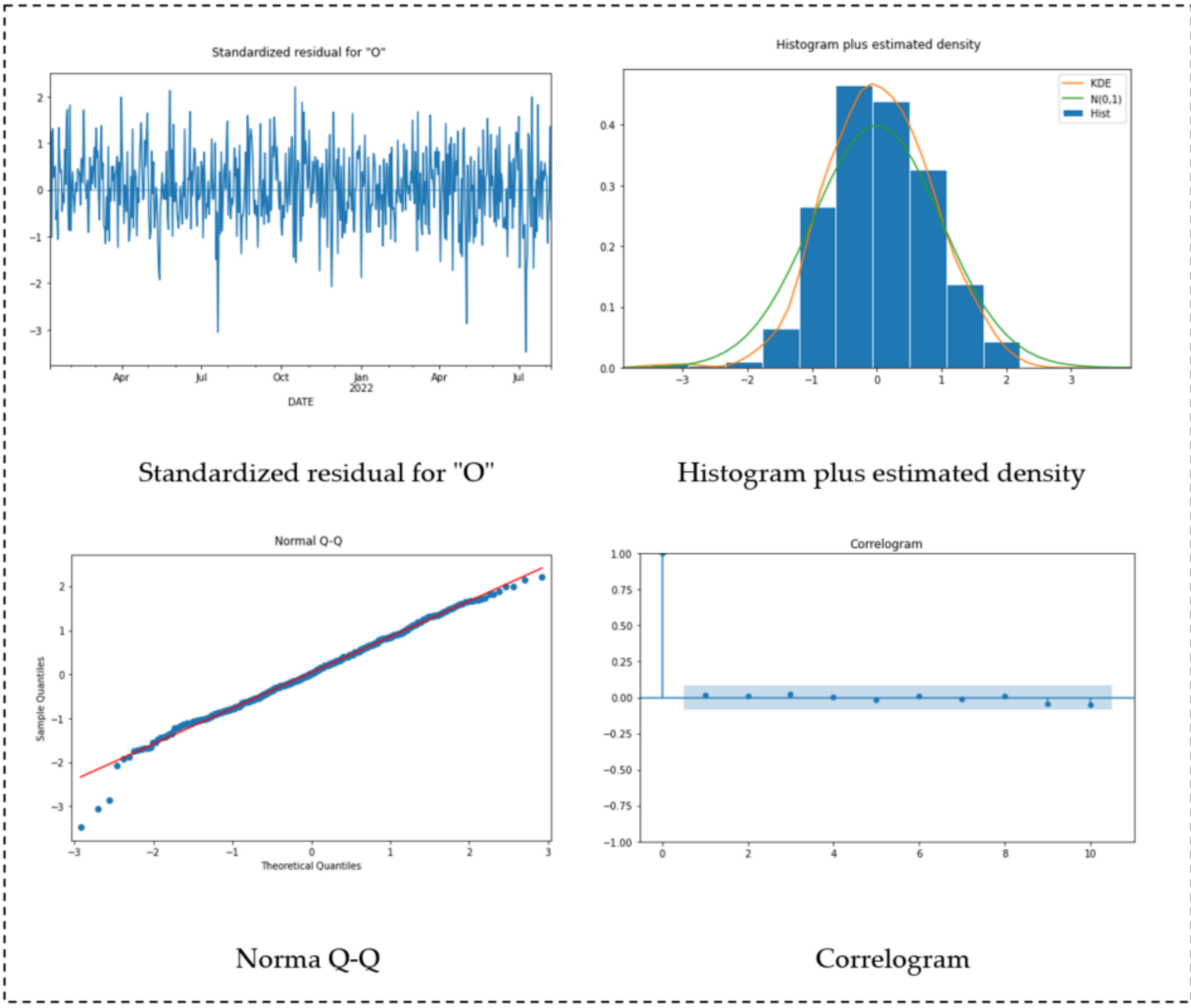

Figure 5. Plot diagnostics for 80:20 split across all countries.

Diagnostic Plots for Time-Series Models

Standardized Residual Plot: The residuals are shown in this plot standardized by their estimated standard deviations. Normally distributed residuals should be scattered randomly around the zero line without any pattern, and their values should fall within the [−2, 2] range when the model has been correctly specified. In the plot, if there are any patterns or trends, it may indicate that the model needs to be fine-tuned [14].

Histogram plus Estimated Density: A histogram and a kernel density estimate are used to illustrate the residual distribution. Normally distributed residuals should be represented by a normal histogram and density plot. Using the plot, we can identify any patterns in the residual data that seem unusual, such as skewness or bimodality [14].

Normal Q-Q Plot: In this plot, the residual distribution is compared with a theoretical normal distribution. It would be expected that the residuals would follow a straight line if the distribution were normally distributed. When the residuals deviate from the line, it means that the residuals are not normally distributed. We can use the plot to identify outliers or other unusual patterns in the residuals [14].

Correlogram: In this plot, the residuals are autocorrelated at various lags. It would be expected that the autocorrelation values for all lags would be close to zero if the residuals were uncorrelated. A significant autocorrelation value would be expected at some lags if the residuals are correlated. We can use the plot to identify any patterns or cycles in the residuals that were missed by the model [14].

4.6. Evaluation Metrics

The evaluation of the models in this study was carried out using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) shown in Equations (5) and (6), as given below:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{5}$$

The division of the discrepancy between the actual value (A_t) and the forecasted value (F_t) by the actual value $\mathrm{A_t}$ yields a ratio. By taking the absolute value of this ratio for each forecasted point in time and summing them, we obtain a total. Finally, this total is divided by the number of fitted points (n) to obtain a measure of accuracy [15].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \| y(i) - \hat{y}(i) \|^2}{N}} \tag{6}$$

In the given scenario, N represents the total count of data points. Each measurement is denoted as $y(i)$, while its corresponding prediction is indicated as $\hat{y}(i)$ [16].

It is crucial to note that RMSE (Root Mean Square Error) is not insensitive to the scale of the data. Consequently, when comparing models using this evaluation metric, the scale of the data can impact the outcomes. To address this concern, it is commonly recommended to compute RMSE on standardized data, where the scale has been adjusted or transformed to eliminate inherent dissimilarities [16].

By employing MAPE and RMSE in this thesis, we leveraged the most common and widely accepted techniques in the field to evaluate the performance of our predictive models. These evaluation metrics have been extensively used and acknowledged in the literature, enabling us to compare and assess the accuracy of our predictions against established benchmarks and prior research findings.

4.7. Results

The outcomes are evaluated using the suggested criteria and presented in tabular form to demonstrate the predictive capability of order volume trends. The study utilized the SARIMAX, ARIMA, LSTM, and AR models, and the predictions were made for a period of 146 days. The performance of the model was evaluated using the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) metrics. The results are depicted through figures, which compare the predicted trend with the actual trend and provide insight into the accuracy of the model. Table 4 shows the MAPE and RMSE metrics for the models in this project.

Table 4. Appraisal of every model.

Model	Metrics/Region	All Countries	UAE	KSA	KWT
SARIMAX	MAPE	0.087	0.097	0.158	0.137
	RMSE	0.127	0.134	0.199	0.215
ARIMA	MAPE	0.143	0.157	0.196	0.192
	RMSE	0.287	1.25	0.309	0.414
LSTM	MAPE	0.108	0.156	0.172	0.185
	RMSE	0.169	0.259	0.214	0.367
AR	MAPE	0.153	0.206	0.218	0.194
	RMSE	0.287	0.3772	0.302	0.4097

According to the results, SARIMAX can provide an effective tool for predicting order volume for shipping companies, but its performance depends on the country or region. Shipping companies may find these findings useful in planning and managing their logistics operations more efficiently. Figure 6 illustrates how the SARIMAX model fares according to 80:20.

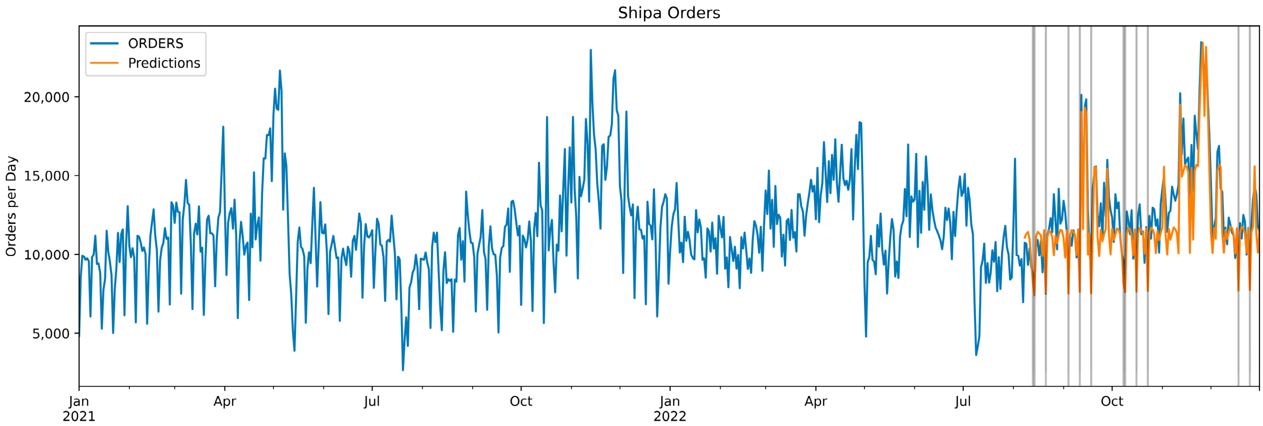

Figure 6. SARIMAX model 80:20 split (All Countries) forecasting outcomes.

Figure 6 illustrates a comparison between the predicted and actual values across all countries obtained from the SARIMAX model, which incorporates Seasonal Autoregressive Integrated Moving Average with Exogenous Variables. The data was split into an 80:20 ratio for training and testing purposes. The figure is color-coded, with orange representing the predicted values and blue representing the actual values, while the gray lines indicate an off or holiday on that day. Through a visual analysis of the orange and blue lines, we can assess the model's performance. Notably, the orange line closely aligns with the blue line, indicating that the model accurately predicts the observed values.

5. Conclusions

This study compared the effectiveness of four different time-series analysis methods for forecasting order volume in three countries: UAE, KSA, and KWT. The study collected daily new order data from a shipping company and preprocessed the time-series data. The four methods used for modeling and forecasting were SARIMAX, ARIMA, LSTM, and AR.

Based on the comparison of prediction outcomes, SARIMAX showed exceptional performance in predicting both order volume and trend across all countries. Therefore, the study concluded that SARIMAX is the most accurate model for predicting order volumes in the combined countries as well as individually for UAE, KSA, and KWT. These findings can help shipping companies in managing and planning their order volumes effectively. Operational efficiency plays a crucial role in helping shipping companies estimate the time and manpower required for packing based on the predicted order volume. This involves analyzing the average time taken to pack different types of orders, allowing companies to establish benchmarks and standards for packing time and manpower requirements. By utilizing this information, along with the forecasted order volume, companies can effectively estimate the necessary resources needed for packing. The combination of operational efficiency and accurate forecasts provides valuable insights into the expected workload, enabling shipping companies to optimize their manpower allocation and plan their packing activities accordingly. Future research can explore the inclusion of other factors in order volume prediction models.

Additionally, the study provided a formula for predicting staffing needs based on order volume. The formula suggests multiplying the expected order volume by the tasks per order amount (usually 0.2) to estimate the number of tasks. Then, dividing the expected tasks by the number of hours each agent is expected to work per month (typically 160) provides an estimate of the number of agents needed. It is important to consider other factors, such as agent skills, order complexity, and operational considerations, as the formula provides a rough estimate. This formula can guide businesses in determining staffing requirements, especially in industries with seasonal or fluctuating demand.

6. Discussion

In order to enhance the predictions made by a previous study [4], we worked on the data used in their study. The data used in their study was downloaded in order to apply our technique and model to determine if the predictions could be improved. A 90:10 split technique was used to study the data, which was the same technique they used, and the same sample of data was used without the outbreak of COVID-19 from 2018 to 2020. Our results were enhanced, and better predictions were obtained by using the SARIMAX model, which uses ships departing as an exogenous variable. We used this variable since the heatmap showed a high correlation between the number of passenger ships and the ships departing. Ports usually have a schedule or timetable that shows when ships are expected to arrive and depart. These times are often communicated to the port in advance by the shipping company or the ship's captain [17]. Therefore, it would be easy to obtain the number of ships departing in advance and use it for predictions. It is possible for a port's departure number to differ from its initial schedule because of unexpected delays or changes in plans. On the same dataset, our model and their model are compared in Table 5, with ours being the first model in the table.

Table 5. Number of passenger ships from Port X comparison.

#	Splitting Technique	Model	p	d	q	P	D	Q	m	X	RMSE Values
1	(90:10)	SARIMAX	5	0	1	1	0	1	7	Ship_Departs	627,763
2	(90:10)	SARIMA	3	0	1	1	0	1	7	-	645,296
3	(90:10)	SARIMA	3	0	1	2	0	1	7	-	688,982
4	(90:10)	SARIMA	3	0	1	1	0	2	7	-	694,836
5	(90:10)	SARIMA	3	0	1	2	0	0	7	-	739,864
6	(90:10)	SARIMA	3	0	1	1	0	0	7	-	739,982
7	(90:10)	SARIMA	3	0	1	0	0	2	7	-	799,582
8	(90:10)	SARIMA	3	0	1	0	0	1	7	-	808,420
9	(90:10)	SARIMA	3	0	1	0	0	0	7	-	810,832
10	(90:10)	SARIMA	3	0	1	2	0	2	7	-	888,593
11	(90:10)	SARIMA	3	0	1	2	0	2	7	-	888,593
12	(90:10)	SARIMA	3	0	1	2	0	1	7	-	988,982

In addition to employing time-series analysis techniques, both studies share a common goal of utilizing predictive models to inform decision-making and optimize resource allocation within their respective industries. While the first study focuses on forecasting the number of passengers aboard ships using SARIMA, our study extends this line of research by examining the performance of various time-series models, including SARIMAX, in predicting order volume within logistics companies. By exploring different industries and modeling approaches, both studies contribute to the growing body of knowledge on the application of time-series analysis for accurate predictions and efficient planning.

The objective of our study is to enhance the predictions and to obtain better results by analyzing data correctly. In order to enhance the predictions made by authors of a previous study [1], we analyzed the data in order to obtain the separate sales amount for each store and item based on their historical data (https://www.kaggle.com/competitions/demand-forecasting-kernels-only/data), (accessed on 15 May 2023). Thus, we reduced the number of rows and created another sheet as a databank. We created a second sheet that shows the daily sales for all items across all stores. In addition, we ran our LSTM and SARIMAX models on the data. We divided the amount of the RMSE values by the mean ratio for the sales column utilizing the same technique as the study to calculate the RMSE values. We determined the sales amount for each store and item separately using the data from the databank sheet, taking advantage of the average values for each item in each store across all days. Table 6 shows the results of the studies. By emphasizing the importance of exploring different techniques and preprocessing data, our study aligns with the goal of enhancing operational efficiency and facilitating informed decision-making in supply chain and logistics management, as highlighted in the previous study. Together, these studies contribute to the advancement of forecasting accuracy and resource allocation optimization in the logistics industry.

Table 6. Summary results.

Data	Model	RMSE Values	Mean Ratio
Original Data	CNN LSTM	19.17	52.25
	LSTM	18.75	52.25
	CNN LSTM	18.75	52.25
	MLP	18.50	52.25
Analyzed data using cumulative sales	LSTM	0.056	26,125.143
	SARIMAX	0.237	26,125.143

Source: [1].

7. Limitations

Although this paper has resolved all the research issues and accomplished the objectives, it has brought to light certain constraints.

Due to the company's infancy, we were unable to use data from 2020 and earlier due to its inaccuracy, which would have impacted our projections.

The ARIMA, LSTM, SARIMAX, and AR models have multiple parameter ranges defined in Section 3.3. We may not have selected the optimal parameter range in our experiments. While we get good results from our models based on the parameter selections we provide, we do not know if they can be improved with a wider parameter selection (i.e., outside the range we provide). Since hardware resources are limited, a greater range of parameter selection will lead to greater time consumption with grid search algorithms or similar algorithms. This has resulted in a relatively small parameter selection range, and it will prevent the model from maximizing its capabilities.

Since each country has its own holidays, off days, and sales seasons, using external variables like sales season and holidays was challenging. In this case, combining data from different countries becomes difficult.

As a result of a large amount of data, it took a long time to download, analyze, and arrange it for the study. This is because Excel is not capable of showing all of the data at once, so we had to split it into quarters so we could handle it better and then collect it again.

8. Future Work

It has been demonstrated in this comparative study that different time-series analysis models are capable of accurately predicting order volume in three countries; however, further research will be necessary to enhance our understanding and improve accuracy. Future work could include further examination of the effect of external variables on order volume prediction and experimenting with different parameter ranges.

Such work should include additional examinations of the impact of external factors on order volume prediction, such as gasoline prices, labor costs, service demand, and geopolitical events.

It will also be important to find out if better results can be obtained if different parameter ranges are used for time-series analysis models.

Additionally, future research should focus on identifying alternative methods of handling large amounts of data more efficiently by examining alternative data analysis tools and techniques.

Moreover, future research should analyze how different order types affect order volume prediction and how to incorporate this information into time-series analysis models.

Our future studies will aim to increase the time period of the study to obtain more accurate predictions since it was only over two years.

Lastly, future research should explore alternative models and utilize datasets from different countries.

Author Contributions: Conceptualization: A.A.; methodology, A.A.; software, A.A.; validation, A.A., B.A.-S., N.O. and M.A.; formal analysis, A.A., B.A.-S. and N.O.; investigation, A.A.; resources, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A., B.A.-S., N.O. and M.A.; visualization, A.A.; supervision, B.A.-S. and N.O.; project administration, B.A.-S., N.O. and M.A.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This project has been supported by Gulf University for Science and Technology under project code: ISG–Case # 264060.

Data Availability Statement: The data are not publicly available due to privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Singha, D.; Panse, C. Application of different Machine Learning models for Supply Chain Demand Forecasting: Comparative Analysis. In Proceedings of the 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 23–25 February 2022; pp. 312–318. [CrossRef]
2. Lee, E.; Kim, D.; Bae, H. Container volume prediction using time-series decomposition with a long short-term memory models. *Appl. Sci.* **2021**, *11*, 8995. [CrossRef]
3. Ferretti, M.; Fiore, U.; Perla, F.; Risitano, M.; Scognamiglio, S. Deep Learning Forecasting for Supporting Terminal Operators in Port Business Development. *Future Internet* **2022**, *14*, 221. [CrossRef]
4. Clarabelle, C.; Gatc, J. Prediction Number of Passenger Ships from Port X to the Kepulauan Seribu Using SARIMA. In Proceedings of the 2022 9th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia, 25 August 2022; pp. 201–205. [CrossRef]
5. Li, Y.; Wei, Z. Regional Logistics Demand Prediction: A Long Short-Term Memory Network Method. *Sustainability* **2022**, *14*, 3478. [CrossRef]
6. Fadda, E.; Fedorov, S.; Perboli, G.; Barbosa, I.D.C. Mixing machine learning and optimization for the tactical capacity planning in last-mile delivery. In Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 12 July 2021; pp. 1291–1296. [CrossRef]
7. Bruni, M.E.; Fadda, E.; Fedorov, S.; Perboli, G. A machine learning optimization approach for last-mile delivery and third-party logistics. *Comput. Oper. Res.* **2023**, *157*, 106262. [CrossRef]
8. Xu, H.; Peng, Q.; Wang, Y.; Zhan, Z. Power-Load Forecasting Model Based on Informer and Its Application. *Energies* **2023**, *16*, 3086. [CrossRef]
9. Manigandan, P.; Alam, M.S.; Alharthi, M.; Khan, U.; Alagirisamy, K.; Pachiyappan, D.; Rehman, A. Forecasting natural gas production and consumption in united states-evidence from sarima and sarimax models. *Energies* **2021**, *14*, 6021. [CrossRef]
10. Brockwell, P.J.; Davis, R.A. Springer Texts in Statistics Introduction to Time Series and Forecasting. 2016. Available online: http://www.springer.com/series/417 (accessed on 15 May 2023).
11. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
12. Tutorials Point. Time Series Tutorial. 2019. Available online: https://www.tutorialspoint.com/time_series/index.htm (accessed on 31 January 2023).
13. Adhikari, R.; Agrawal, R.K. An Introductory Study on Time Series Modeling and Forecasting. *arXiv* **2013**, arXiv:1302.6613.
14. Peixeiro, M. *Time Series Forecasting in Python*; Manning Publications Co.: Shelter Island, NY, USA, 2022. Available online: https://www.manning.com/books/time-series-forecasting-in-python-book (accessed on 15 May 2023).
15. Wikipedia Contributors. Mean Absolute Percentage Error. Available online: https://en.wikipedia.org/w/index.php?title=Mean_absolute_percentage_error&oldid=1155495457 (accessed on 20 May 2023).

16. Krishnan, P.N. Enterprise AI and Machine Learning for Managers. 2020. Available online: https://c3.ai/glossary/data-science/root-mean-square-error-rmse/ (accessed on 20 May 2023).
17. Menon, H. What Are Liner Services and Tramp Shipping? Available online: https://www.marineinsight.com/maritime-law/what-are-liner-services-and-tramp-shipping/ (accessed on 20 August 2021).