

Article **Bi-GRU-APSO: Bi-Directional Gated Recurrent Unit with Adaptive Particle Swarm Optimization Algorithm for Sales Forecasting in Multi-Channel Retail**

Aruna Mogarala Guruvaya ¹ , Archana Kollu ² , Parameshachari Bidare Divakarachari 3,[*](https://orcid.org/0000-0002-3997-5070) , Przemysław Falkowski-Gilski [4](https://orcid.org/0000-0001-8920-6969) and Hirald Dwaraka Praveena ⁵

- ¹ Department of Artificial Intelligence and Machine Learning, Dayanand Sagar College of Engineering, Bangalore 560078, India; aruna-aiml@dayanandasagar.edu
- ² Department of Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune 412101, India; archana.kollu@pccoer.in
- ³ Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bengaluru 560064, India
- ⁴ Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland; przemyslaw.falkowski@eti.pg.edu.pl
- ⁵ Department of Electronics and Communication Engineering, School of Engineering, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati 517102, India; praveena.hd@vidyanikethan.edu
- ***** Correspondence: paramesh@nmit.ac.in

Abstract: In the present scenario, retail sales forecasting has a great significance in E-commerce companies. The precise retail sales forecasting enhances the business decision making, storage management, and product sales. Inaccurate retail sales forecasting can decrease customer satisfaction, inventory shortages, product backlog, and unsatisfied customer demands. In order to obtain a better retail sales forecasting, deep learning models are preferred. In this manuscript, an effective Bi-GRU is proposed for accurate sales forecasting related to E-commerce companies. Initially, retail sales data are acquired from two benchmark online datasets: Rossmann dataset and Walmart dataset. From the acquired datasets, the unreliable samples are eliminated by interpolating missing data, outlier's removal, normalization, and de-normalization. Then, feature engineering is carried out by implementing the Adaptive Particle Swarm Optimization (APSO) algorithm, Recursive Feature Elimination (RFE) technique, and Minimum Redundancy Maximum Relevance (MRMR) technique. Followed by that, the optimized active features from feature engineering are given to the Bi-Directional Gated Recurrent Unit (Bi-GRU) model for precise retail sales forecasting. From the result analysis, it is seen that the proposed Bi-GRU model achieves higher results in terms of an R2 value of 0.98 and 0.99, a Mean Absolute Error (MAE) of 0.05 and 0.07, and a Mean Square Error (MSE) of 0.04 and 0.03 on the Rossmann and Walmart datasets. The proposed method supports the retail sales forecasting by achieving superior results over the conventional models.

Keywords: Bi-Directional Gated Recurrent Unit; Minimum Redundancy Maximum Relevance; Particle Swarm Optimization Algorithm; Recursive Feature Elimination; retail sales forecasting

1. Introduction

Retail sales forecasting is one of the important factors in E-commerce companies due to the economic hardship and strong competition [\[1\]](#page-16-0). A good sales forecasting makes an efficient production plan, increases the revenue from sales, and improves the satisfaction of customers. The timely and effective forecasting model is an indispensable and crucial tool for handling the inventory level [\[2,](#page-16-1)[3\]](#page-16-2), whereas the inappropriate forecasting model results in insufficient or redundant stock that directly affects the competitive advantage and income [\[4\]](#page-16-3). In the recent decades, the researchers have introduced many analytical and

Citation: Mogarala Guruvaya, A.; Kollu, A.; Divakarachari, P.B.; Falkowski-Gilski, P.; Praveena, H.D. Bi-GRU-APSO: Bi-Directional Gated Recurrent Unit with Adaptive Particle Swarm Optimization Algorithm for Sales Forecasting in Multi-Channel Retail. *Telecom* **2024**, *5*, 537–555. <https://doi.org/10.3390/telecom5030028>

Academic Editor: Achilles Boursianis

Received: 10 May 2024 Revised: 9 June 2024 Accepted: 21 June 2024 Published: 1 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license [\(https://](https://creativecommons.org/licenses/by/4.0/) [creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/) $4.0/$).

statistical regression models for resolving the concerns of retail sales forecasting [\[5](#page-16-4)[,6\]](#page-16-5). The existing models are generally categorized into two types of artificial intelligence models and classical regression models. Some of the classical regression models are exponential smoothing, weighted average, linear regression, moving average, Support Vector Regression (SVR), eXtreme Gradient Boosting (XGBoost), etc. [\[7](#page-16-6)[,8\]](#page-16-7). On the other hand, the common statistical models employed in retail sales forecasting are Seasonal Auto-Regressive Moving Average $(SARIMA)$ and ARIMA [\[9](#page-16-8)[,10\]](#page-16-9).

Currently, artificial intelligence models are highly adopted for retail sales forecasting, such as neural networks, fuzzy systems, expert systems, and hybrid models [\[11,](#page-16-10)[12\]](#page-17-0). In artificial intelligence, neural networks are extensively utilized in sales forecasting because of their higher ability and flexible nature in relation to data mining. Additionally, the hybrid models integrate the benefits of dissimilar approaches for improving the performance of forecasting [\[13,](#page-17-1)[14\]](#page-17-2). Currently, most of the existing studies integrate artificial intelligence and classical models to achieve better performance in retail sales forecasting [\[15\]](#page-17-3). However, the E-commerce companies face unrivalled issues to ensure precise forecast, which can be noted as (i) complex and dynamic online sale compared to offline sale, (ii) marketing behavior greatly influencing the customers in E-business platforms, and (iii) need to collect an enormous amount of data to enhance the performance of retail sales forecasting. In order to achieve better retail sales forecasting and to reduce the time complexity, an effective automated regression model is implemented in this manuscript.

The contributions are listed below:

- Performed outlier's removal, interpolation of missing data, normalization, and denormalization in the acquired datasets. The forecasting accuracy is significantly improved by excluding outliers and interpolating missing data. The outlier's removal aims to eliminate rare or unexpected instances in datasets, and a more typical value is interpolated with the missing data to avoid problems like bias and loss of power. Additionally, data integrity is improved by performing normalization and de-normalization using the Z-score normalization technique.
- Integrated APSO algorithm, RFE, and MRMR for optimizing features in the acquired Rossmann and Walmart datasets. The feature engineering process minimizes the number of features in the pre-processed datasets and, in turn, decreases the computational complexity and processing time, alongside improving the performance of the regression model. The selection of active features decreases the complexity to linear, and the regression model consumes a minimal processing time of 20.11 s, while consuming 30.12 s in the Rossmann and Walmart datasets.
- The selected active features are passed to the Bi-GRU model for precise retail sales forecasting. When related to other regression models, the Bi-GRU model consumes limited memory and is faster in data processing. The effectiveness of the proposed regression model is analyzed based on six evaluation measures: Coefficient of determination (R2), MSE, Normalized Deviation (ND), MAE, Root Mean Square Scale Error (RMSSE), and Normalized Root Mean Square Error (NRMSE).

The related papers on the topic "sales forecasting" are reviewed in Section [2.](#page-1-0) The details about methodology, simulation analysis, and the conclusion are discussed in Sections [3–](#page-3-0)[5,](#page-15-0) respectively. In this manner, the present research manuscript is organized.

2. Literature Survey

He et al. [\[16\]](#page-17-4) integrated the PSO algorithm with a Long Short-Term Memory (LSTM) network for precise sales forecasting related to the e-commerce companies. Here, the PSO algorithm was implemented for optimizing the number of iterations and the hidden neurons in the LSTM layers. The experiments conducted on the real and online benchmark datasets demonstrated the effectiveness of the presented forecasting model over the nine comparative models. The LSTM network was effective in sales forecasting but required more training samples to learn efficiently.

Ji et al. [\[17\]](#page-17-5) and Massaro et al. [\[18\]](#page-17-6) employed the XGBoost 1.0.0 model for effective sales forecasting by using commodities sales features and data series. Here, sales forecasting was performed for optimizing the inventory management by considering the relevant factors, like transportation time, delivery time, order lead-time, and inventory cost. Generally, the XGBoost model included regularization methods for preventing overfitting, which led to erroneous prediction of new sales data. In addition, the XGBoost model was memory intensive, particularly for large real-time datasets.

Wong and Guo [\[19\]](#page-17-7) adopted an effective learning algorithm for accurate sales forecasting and incorporated Extreme Learning Machine (ELM) and Improved Harmony Search Algorithm (IHSA). The incorporation of ELM and IHSA enhanced the ability of network generalization. The experiments performed on the benchmark and real-time fashion retail datasets revealed that the presented model achieved a superior performance compared to the traditional neural networks. On the other hand, the incorporation of ELM and IHSA with the learning algorithm increased the time complexity of the system.

Loureiro et al. [\[20\]](#page-17-8) explored the usage of deep learning model (DLM) in the fashion retail sales forecasting by considering a diverse and wide set of data. Here, the DLM performance was compared with other shallow models, and the numerical analysis stated that the DLM obtained results were superior to the shallow models in terms of different evaluation measures. Nonetheless, the DLM was computationally costly and needed a higher-end system to deal with a diverse and wide set of data.

Lu [\[21\]](#page-17-9) combined the SVR model with the variable selection approach for precise sales forecasting. This study forecasted the weekly sales of computer products like display cards, hard disks, main boards, notebooks, and liquid crystal displays. In addition, Luo et al. [\[22\]](#page-17-10) integrated the LSTM network with an Extreme Deep Factorization Machine (XDeepFM) model for retail sales forecasting by effectively exploring the correlation between the sales factors. The LSTM network was employed for correcting the residuals to enhance the accuracy of the forecasting model. The numerical evaluation proved that the presented forecasting model was superior compared to other traditional forecasting models. As discussed in the earlier literature, the LSTM network was precise in sales forecasting but needed enormous training samples to learn efficiently.

Zhang et al. [\[23\]](#page-17-11) integrated a Moth–Flame Optimization (MFO) algorithm with the ELM to predict the transaction volume of e-commerce. The use of a traditional MFO algorithm significantly improved the convergence accuracy but demanded a high execution time. Weng et al. [\[24\]](#page-17-12) designed a new framework for precisely forecasting the sales of the supply chain, along with the designed framework that incorporated the LSTM network with a light-Gradient Boosting Machine (GBM) model. The designed framework's performance was validated on three supply-chain sales datasets, and the obtained results demonstrated the efficacy of the designed framework over other traditional models by means of accuracy but needed high time complexity.

Shilong [\[25\]](#page-17-13) presented an accurate and efficient sales forecasting framework based on the machine-learning models. At first, the vectors were extracted from the historical sales dataset by conducting feature engineering, and then the XGBoost model was implemented for forecasting the future sales. The experiments performed on the online Walmart retail goods dataset proved that the presented framework obtained a more precise sales prediction than the existing models with limited memory resources and computing time. Nevertheless, the XGBoost model faced two major issues in retail forecasting, namely overfitting and outliers.

Punia et al. [\[26\]](#page-17-14) integrated a random forest with the LSTM network for accurate retaildemand forecasting. The presented framework superiorly analyzed complex relationships between the regression and temporal types that resulted in better prediction than the existing forecasting models. A multivariate real-time dataset was employed for analyzing the performance of the presented framework in light of different measures, such as bias, variance, and prediction accuracy. In the multivariate real-time dataset, the random forest

resulted in overfitting with noisy regression tasks, and, additionally, the random forest faced difficulty in handling the categorical variables.

Karmy and Maldonado [\[27\]](#page-17-15) suggested the SVR model for retail sales forecasting on the hierarchical time-series data. Three models were formulated (middle-out, top-down, and bottom-up SVR) for forecasting retail sales related to the travel industry. The SVR model generally underperformed when the number of training samples was lower than the number of features in every data point.

Kilimci et al. [\[28\]](#page-17-16) integrated eleven forecasting models that comprised DLMs, SVR model, and time-series approaches for effective retail-demand forecasting. In addition, a novel decision strategy was implemented from the inspiration of the ensemble model based on the concept of boosting. The presented system's performance was tested on a real-time dataset and evaluated by means of different error rates. The integration of DLMs, the SVR model, and time-series approaches increased the system complexity, which was one of the main issues in this literature.

Kohli et al. [\[29\]](#page-17-17) introduced a new sales prediction framework based on K-Nearest Neighbor (KNN) and linear regression models to formulate commendable decisions and find potential risks in the businesses. The efficacy of the presented framework was tested on an online Rossmann dataset by means of different error rates. As mentioned in the previous literature, the integration of two machine-learning models generally increased the system complexity and processing time.

From the overall analysis of the above-stated literature, we noted that the state-of-theart models ELM, XDeepFM, light GBM, XGBoost, LSTM DLM, SVR, and KNN were applied for precise sales forecasting related to the e-commerce companies. The integration of the above-stated learning models normally maximizes the system complexity and processing time. Furthermore, aggregated forecasting assists strategic decision making on the position which frequently relates to the operational decisions at the store level. These aspects affect the demand and promotional data and increase the substantial complexity, and, subsequently, the forecasters possibly face dimensionality issue of too many constraints and not enough data. To address the aforementioned problems and to achieve better retail sales forecasting, an effective Bi-GRU model is proposed in this manuscript which is clearly described in the following subsections.

3. Methods

Future sales forecasting is the important factor in retail business for identifying the potential risks and making appropriate decisions. The precise demand forecasting is vital in organizing and planning the labor force, transportation, purchasing, and production of successful and profitable retail business [\[30\]](#page-17-18). The proposed automated regression model for retail sales forecasting includes three important steps, which are listed below and graphically specified in Figure [1:](#page-4-0)

- Retail data pre-processing: Interpolating missing data, outliers' removal, normalization, and de-normalization.
- Feature engineering: MRMR, RFE, and APSO.
- Retail sales forecasting: Bi-GRU model.

Figure 1. Workflow of the proposed automated regression model. **Figure 1.** Workflow of the proposed automated regression model.

3.1. Retail Data Pre-Processing 3.1. Retail Data Pre-Processing

After acquiring the retail sales data from the Rossmann and Walmart datasets, data pre-processing was carried out to eliminate unreliable samples for better forecasting. In pre-processing was carried out to eliminate unreliable samples for better forecasting. In this scenario, the data pre-processing step included the processes of interpolating missing data, outliers' removal, normalization, and de-normalization [\[31,](#page-17-19)[32\]](#page-17-20). After acquiring the retail sales data from the Rossmann and Walmart datasets, data

In the initial process, identifying outliers is important in statistics and data analysis $\frac{1}{2}$. In the initial process, identifying outliers is important in statistics and data analysis because it has a significant impact on the results of statistical analyses. Removing the because it has a significant impact on the results of statistical analyses. Removing the outliers involves excluding data points significantly deviating from the norm to enhance liers involves excluding data points significantly deviating from the norm to enhance the the model's accuracy and generalization on new data. The outliers are the instances which are highly deviated from other instances in the Rossmann and Walmart datasets. In this study, the instances, $s_{i,j}$, are considered as the outliers in the time-series data, S_j . When the following condition is satisfied, $abs(s_{i,j} - mean(S_j)) > n \times std(S_j)$, the instances $s_{i,j}$ are eliminated from the time-series data, S_j . Here, the absolute value function is $abs(.)$, mean function is represented as $mean(.)$, $n = 3$, and the standard deviation function is denoted as *std*(.). After the elimination of outliers, the missing data are interpolated because they affect the validity and generalizability of the data, which depend on the quantity, category, and pattern. The missing data are supplied by computing the mean value of two nearest the missing data are supplied by computing the missing data are supplied by computing the mean value of mean value of α

Following that, the data normalization is performed by implementing the z-score normalization technique, where the normalization process speeds up the training time of the Bi-GRU model. The key concept of normalization is to remove potential biases and distortions that occur because of the dissimilar feature scales. The z-score normalization technique normalizes the input and the output variables based on Equation (1), and the de-normalization is mathematically formulated in Equation (2).

$$
S'_{j} = \frac{S_{j} - mean(S_{j})}{std(S_{j})}
$$
\n(1)

$$
S_j = mean(S_j) + S'_j \times std(S_j)
$$
 (2)

3.2. Feature Engineering

After pre-processing the Rossmann and Walmart datasets, the feature selection is initially accomplished by employing MRMR and RFE techniques. Generally, the feature selection process decreases the feature sets by choosing the most active features or eliminating the drossy features. MRMR is chosen, as it identifies the relevant features and minimizes the redundancy, thus enhancing the classification accuracy. Therefore, the MRMR overcomes the other techniques based on the accuracy and number of supportive features. The MRMR is a feature measurement criterion; it computes the redundancy and correlation between the features based on mutual information. Here, the MRMR technique performs feature selection by following two conditions, such as maximal relevance (max_R) and minimal redundancy (min_D) , which are mathematically specified in Equations (3) and (4). Moreover, $m(x, y)$ denotes the mutual information, and f denotes a set of features. The mathematical expression of mutual information $m(x, y)$ is represented in Equation (5).

$$
max_R = \frac{1}{|f|} \sum_{i \in f} m(k, i)
$$
\n(3)

$$
min_D = \frac{1}{|f|^2} \sum_{i,j \in f} m(i,j) \tag{4}
$$

$$
m(x,y) = \iint q(x,y) \log \frac{q(x,y)}{q(x)q(y)} dx dy
$$
 (5)

On the other hand, the RFE is one of the effective feature selection techniques which fits the model and eliminates irrelevant features, until the discriminative active features are selected [\[33\]](#page-17-21). RFE is selected here because it efficiently minimizes the dimensionality of high-dimensional datasets by removing redundant features to help enhance the computational and storage necessities. The RFE technique includes three major benefits: (i) complete elimination of irrelevant information in the data, (ii) ease in performing data visualization, and (iii) required limited computational power. By considering Equations (3) and (4), Equation (6) is generated. The RFE technique efficiently selects the optimal feature subsets using Equation (6).

$$
maxQ(D,R); Q = D - R \tag{6}
$$

In addition to this, feature optimization is carried out utilizing the APSO algorithm, where it selects active features from the pre-processed Rossmann and Walmart datasets. The APSO algorithm is applied here, as it selects the algorithmic constraints at run time to enhance the exploitation efficiency, and it performs a global search over the entire search space with a higher convergence speed. As discussed in the previous sections, this process decreases the complexity of the Bi-GRU model and its processing time. The conventional PSO algorithm is one of the effective metaheuristic-based optimization algorithms [\[34\]](#page-17-22) that generally mimics the behavior of birds and fish schooling [\[35\]](#page-17-23). The velocity and the position of the particles are updated in the PSO algorithm by Equations (7) and (8).

$$
v_{id}(n+1) = I_w \times v_{id}(n) + c_1 \times r_1 \times [p_{id}(n) - x_{id}(n)] + c_2 \times r_2 \times [p_{gd}(n) - x_{id}(n)] \tag{7}
$$

$$
x_{id}(n+1) = x_{id}(n) + v_{id}(n+1) \tag{8}
$$

where the global best positions of the particles are represented as p_{gd} ; the present best positions of the particles are denoted as p_{id} ; the random numbers are indicated as r_1 and r_2 ; the acceleration coefficients are denoted as c_1 and c_2 ; the inertia weight is represented as I_w , used for balancing the local and the global searches; and the iteration number is indicated as n .

The APSO algorithm optimizes the features based on Adaptive Uniform Mutation The APSO algorithm optimizes the features based on Adaptive Uniform Mutation (AUM) function from the Human Group Optimization (HGO), where the particle's positions (features) are denoted as $p_i(n) = (p_{i,1}, p_{i,2}, p_{i,D})$. The AUM function extends the ability of feature optimization in the exploration phase. Additionally, a nonlinear function, p_m , is employed for controlling the mutation range and decision in each particle. The nonlinear function, p_m , is updated in each iteration by performing Equation (9).

$$
p_m = 0.5 \times e^{(-10 \times \frac{n}{M_I})} + 0.01 \tag{9}
$$

If the iteration increases, the nonlinear function, *pm*, tends to decrease, while the maximum number of iterations is represented as M_I . The mutation randomly selects the active features from the datasets when the nonlinear function, μ is higher than the the active features from the datasets when the nonlinear function, *pm*, is higher than the random number, which usually ranges between zero and one. The selected active features random number, which usually ranges between zero and one. The selected active features from the Rossmann and Walmart datasets are finally passed to the Bi-GRU model for retail sales forecasting. The APSO algorithm terminates when it reaches the maximum number sales forecasting. The APSO algorithm terminates when it reaches the maximum number of iterations (100). of iterations (100). If the iteration increases, the nonlinear function, , tends to decrease, while the m are negative of iterations is represented as M . The mutation randomly selected as M . The mutation randomly selected from the Rossmann and Walmart datasets are finally passed to the Bi-GRU model for retail

The parameters considered in the APSO algorithm are as follows: the cognitive constant, c_2 , is two; the social constant, c_1 , is three; the size of the population is 100; and the number of iterations is 100. The features selected in the Walmart dataset are and the number of iterations is 100. The features selected in the Walmart dataset are the date, Consumer Price Index (CPI), fuel prices, store, weekly_sales, and holiday_flag. Correspondingly, the features selected in the Rossmann dataset are the day of the week, open promo, customers, sales, and store number. The architecture of the APSO algorithm is mentioned in Figure [2.](#page-6-0) in Figure 2.

Figure 2. Architecture of the APSO algorithm. **Figure 2.** Architecture of the APSO algorithm.

The step-by-step procedure of APSO algorithm is specified as follows. The step-by-step procedure of APSO algorithm is specified as follows.

Step 1: The swarm particles, size, location, objective, number of iterations, and save Step 1: The swarm particles, size, location, objective, number of iterations, and save non-dominated solution are set into the archive. non-dominated solution are set into the archive.

Step 2: To update the p_{id} , pareto domination connection is applied.

Step 3: Due to the multiplicity of solution, p_{gd} is chosen from the archives. In the beginning, crowding distance is estimated, and, formerly, binary tournament is utilized for choosing p_{gd} .

Step 4: At that time, the decision value is reset, which depends on p_{gd} . Each value of feature vector is considered as a binary value.

Step 5: Depending on step 5, the particle's position and velocity is updated.

Step 6: Uniform mutation is accomplished.

Step 7: Then, the external archive is updated by means of crowding distance.

Step 8: Termination process: If the proposed method achieves the maximum iteration, the process is stopped; otherwise, step 2 is repeated. Therefore, the worst particles are removed by HGO. After choosing the optimal features from the Rossmann dataset (day of week, open promo, customers, sales, and store number) and Walmart dataset (date, CPI, fuel prices, store, weekly_sales, and holiday_flag) using MRMR, RFE, and APSO, forecasting is processed using Bi-GRU, which is described in the following section.

3.3. Retail Sales Forecasting

The optimal features selected from MRMR, RFE, and APSO on the Rossmann dataset and Walmart dataset are given as input to the Bi-GRU model for effective forecasting of retail sales. The Bi-GRU model has update and reset gates for performing sales forecasting, which also decreases the gradient dispersion and computational loss, while enabling the capability for shorter- and longer-term memory [\[36,](#page-17-24)[37\]](#page-17-25). Also, Bi-GRU comprises a smaller number of constraints because it does not have a forget gate, which makes it computationally efficient, less prone to overfitting, a suitable option for a smaller-type dataset.

In the Bi-GRU model, the input and the forget gates of the LSTM network are replaced by the update gate, *dTS*. The update gate helps the model in determining the past information, which needs to be passed along with the future information. This process reduces the vanishing-gradient problem in the Bi-GRU model. The update gate, *dTS*, is mathematically specified in Equation (10).

$$
d_{TS} = \sigma(W_d \times [h_{TS-1}, f_{TS}] + b_d)
$$
\n(10)

where the weight matrix is represented as W_d ; the bias matrix is denoted as b_d ; the input matrix (selected features) at the time step, *TS*, is indicated as f_{TS} ; the sigmoid activation function is denoted as σ ; and the hidden state at the previous time step, $TS - 1$, is indicated as *hTS*−¹ . In the Bi-GRU model, the reset gate, *pTS*, is utilized for controlling the historical time-series data and is responsible for the network's shorter-term memory in the hidden state. The reset gate, *pTS*, is numerically expressed in Equation (11).

$$
p_{TS} = \sigma(W_p \times [h_{TS-1}, f_{TS}] + b_p)
$$
\n(11)

where the bias matrix and the weight matrix of the reset gate (p_{TS}) are denoted as b_p and *Wp*. Then, the candidate of the hidden state, ∼ *hTS*, is specified in Equation (12).

$$
\widetilde{h}_{TS} = tanh(W_h \times [h_{TS-1} \odot p_{TS}, f_{TS}] + b_h)
$$
\n(12)

where the tangent activation function is represented as *tanh*, the dot multiplication operation is denoted as ⊙, and the bias matrix and weight matrix of the memory cell state are correspondingly denoted as b_h and W_h . The output, h_{TS} , is obtained by linearly interpolating ∼ *h*_{TS} and *h*_{TS−1}, and this process is indicated in Equation (13). The Bi-GRU model′s architecture is mentioned in Figure [3.](#page-8-0)

$$
h_{TS} = (1 - d_{TS}) \odot h_{TS-1} + d_{TS} \odot \widetilde{h}_{TS}
$$
\n(13)

Appropriate feature engineering for the Bi-GRU model is needed for retail sales forecasting to extract the implicit vectors and complex variances in the historical sequence data. The traditional Bi-GRU model extracts only feature information in the forward direction, and it automatically rejects the backward historical time-series data. So, an adaptive Bi-GRU model was implemented in this study for precise retail sales forecasting. The proposed BiGRU has the capability to process several inputs proficiently over the conventional models because of its facility to study the input from both the directions concurrently.

The proposed regression model extracts the knowledge between the variables from the forward and backward directions, as mentioned in Figure [3.](#page-8-0) In the Bi-GRU model, the forward GRU extracts prior information in the historical time-series data, and the backward GRU extracts future information in the historical time-series data. The numerical expression of the Bi-GRU O_{TS} model is specified in Equation (14).

$$
O_{TS} = A\left(\stackrel{\rightarrow}{h}_{TS}, \stackrel{\leftarrow}{h}_{TS}\right) \tag{14}
$$

where the output of the backward and forward directions is represented as *A*, and it where the output of the backward and forward difections is represented as A, and it performs operations like multiplication function, average function, summation function, ← forms operations like multiplication function, average function, summation function, \leftarrow

etc. In addition, the hidden states of backward and forward GRUs are denoted as etc. In addition, the hidden states of backward and forward GRUs are denoted as h_{TS} and \rightarrow and \hat{h}_{TS} .

The parameters considered in the BI-GRU model are as follows: the look-back is eight, The parameters considered in the BI-GRU model are as follows: the look-back is the number of neurons is 80 , the dropout rate is 0.5 , the batch size is 50 , the loss function is MSE loss, the optimizer is Adam, and the learning rate is 0.0001. The numerical results of the proposed regression model are specified [in](#page-9-0) Section 4.

Figure 3. Architecture of the Bi-GRU model. **Figure 3.** Architecture of the Bi-GRU model.

3.4. Complexity and Convergence Analysis 3.4. Complexity and Convergence Analysis

The complexity of the method for creating a composite particle is calculated as per The complexity of the method for creating a composite particle is calculated as per the number of assessments of Euclidean distances $(T(N))$ among the particles. It is assumed that *N* particles are arranged in a list, which is denoted as L_I . Consequently, the estimation of Funkline and the consequently of Γ of Euclidean distance is essential for this process: $T(N)$ is formulated as Equation (15).

$$
T(N) = \sum_{i=1}^{\lfloor \frac{N}{3} \rfloor} (N - 3i + 2)
$$
\n(15)

The worst time complexity of the procedure is $O(N^2)$. The PSO contains a complexity of $O(N^2)$ that is at the expense of estimating $T(N)$. In spite of this, the partition technique intrinsically deliberates the distance and fitness, which delivers an efficient system to contribute a complete employment of the included particles. Hence, it is worthwhile to improve the adaptivity of PSO by incorporating other swarm intelligence algorithm for solving the optimization problems. To address the complexity, $O(N)$, and convergence speed of the algorithm, a Human Group Optimization (HGO) algorithm is integrated with the PSO algorithm to influence the particle's positions. The HGO algorithm employs an AUM function, d_i , to improve the convergence speed, which eases the implementation process of PSO.

4. Results and Discussion

The proposed regression model was simulated by a Python 3.7 software tool with libraries of SciPy, Keras, TensorFlow, SciKit Learn, Matplotlib, Numpy, Polars, and Pandas. The proposed regression model was analyzed on a system with 8TB hard disk, Linux operating system, 128GB RAM, and Intel core i7 10th generation processor. This equipment was sourced from SATA manufacturer, Mumbai, INDIA. The efficacy of the Bi-GRU model was analyzed on the Rossmann and Walmart datasets by means of NRMSE, ND, RMSSE, MAE, R2, and MSE.

4.1. Evaluation Measures

The proposed Bi-GRU model's effectiveness was validated by using six evaluation measures, namely NRMSE, ND, RMSSE, MAE, R2, and MSE. The evaluation measure NRMSE is one of the crucial measures in analyzing the proposed Bi-GRU model for sales forecasting; particularly, it is preferred while performing outliers' removal. The spread is considered as the difference between the minimum, *ymin*, and the maximum, *ymax*, values in the training dataset in NRMSE. The error measure, ND, accounts for the scale difference between the actual and predicted values. The evaluation measures NRMSE and ND are mathematically formulated in Equations (16) and (17).

NRMSE
$$
(y, \hat{y}) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}}{y_{max} - y_{min}}
$$
 (16)

$$
ND (y, \hat{y}) = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{\sum_{i=1}^{N} |y_i|}
$$
(17)

where *N* represents the number of data points, and y and \hat{y} are denoted as the actual and predicted values. The error measures, i.e., MSE, MAE, and RMSSE, determine the average square difference, mean of absolution difference, and average square scale error difference between the actual and the predicted values. The error measures, MSE, MAE, and RMSSE, are mathematically specified in Equations (18)–(20), where q is denoted as period. The R^2 is the number that ranging between zero and one; it measures how well the proposed Bi-GRU model predicts the outcomes and is numerically specified in Equation (21).

$$
MSE = \frac{1}{N} \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}
$$
(18)

$$
MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|
$$
\n(19)

RMSSE =
$$
\sqrt{\frac{1}{N} \frac{\sum_{i=q+1}^{q+N} (y_i - \hat{y}_i)^2}{\frac{1}{q-1} \sum_{i=2}^{q} (y_i - y_{i+1})^2}}
$$
(20)

$$
R^{2} = 1 - \frac{\sum(\hat{y}_{i} - y_{i})^{2}}{\sum(\overline{y}_{i} - y_{i})^{2}}
$$
(21)

4.2. Dataset Description

The proposed Bi-GRU model's efficacy was validated on the Rossmann and Walmart datasets. The Walmart dataset was released in 2014, and it contains the weekly sales data of 77 departments and 45 stores. The Walmart dataset includes features like unemployment (prevailing unemployment rate), temperature (temperature on a sale day), date (week of sales), CPI (prevailing CPI), fuel prices (fuel cost in a region), store (store number), weekly_sales (sales of a given store), and holiday_flag (whether the week is a non-holiday week, holiday week, or special holiday week). This historical dataset covers sales from

5 February 2010 to 1 November 2012, and it is available from [https://www.kaggle.com/](https://www.kaggle.com/datasets/yasserh/walmart-dataset) [datasets/yasserh/walmart-dataset](https://www.kaggle.com/datasets/yasserh/walmart-dataset) (accessed on 12 December 2023).

On the other hand, the Rossmann dataset was released in 2015, and it has the sales data of 1115 stores. This dataset includes seven features: day of the week, school holiday, state holiday, open promo, customers, sales, and store number. The Rossmann dataset is available from <https://www.kaggle.com/competitions/rossmann-store-sales/data> (accessed on 12 December 2023). The data statistics of Rossmann and Walmart datasets are mentioned in Table [1.](#page-10-0)

Table 1. Data statistics of Rossmann and Walmart datasets.

4.3. Quantitative Analysis

When managing complex sales prediction issues, several analyses are established through a combination of techniques to complement the benefits of deep learning models and to enhance the accuracy. The deep learning models are always selected because of their high precision in solving the complex problems. Consequently, in the proposed Bi-GRU model, the forward GRU extracts prior information in the historical time-series data, and the backward GRU extracts future information in the historical time-series data to minimize the error performances (NRMSE, ND, RMSSE, MAE, MSE, and R2) for sales forecasting, which is shown in the below tables.

Therefore, the efficacy of the Bi-GRU model was validated on the Rossmann and Walmart datasets. The numerical analysis of different models on the Rossmann and Walmart datasets is denoted in Tables [2](#page-10-1) and [3.](#page-11-0) As specified in Tables [2](#page-10-1) and [3,](#page-11-0) the efficacy of the Bi-GRU model was analyzed by comparing its results with other regression models, like linear regression, LSTM, Bi-Directional LSTM (Bi-LSTM), and GRU by means of NRMSE, ND, RMSSE, MAE, MSE, and R2. By viewing Table [2,](#page-10-1) we can see that the proposed Bi-GRU model obtains a high R2 value of 0.98 and a minimum NRMSE, ND, RMSSE, MAE, and MSE of 0.08, 0.07, 0.08, 0.05, and 0.04, respectively, on the Rossmann dataset.

Table 2. Numerical analysis of different regression models on the Rossmann dataset.

Table 2. *Cont.*

On the other hand, as seen in Table [3,](#page-11-0) the Bi-GRU model achieves a high R2 value of 0.99 and a minimum NRMSE, ND, RMSSE, MAE, and MSE of 0.01, 0.03, 0.08, 0.07, and 0.03 on the Walmart dataset, respectively. As compared to the other regression models—the linear regression, LSTM, Bi-LSTM, and GRU—the Bi-GRU model utilizes limited memory and is faster in regard to data processing. The Bi-GRU model is more precise while utilizing the datasets with longer sequences. In addition to this, the Bi-GRU model effectively addresses the problems of overfitting and vanishing gradient.

Table 3. Numerical analysis of different regression models on the Walmart dataset.

Table 3. *Cont.*

Table [4](#page-12-0) displays the numerical analysis of different models on the citadel POS dataset using NRMSE, ND, RMSSE, MAE, and MSE. The proposed Bi-GRU model uses less memory compared to other regression models. Table [4](#page-12-0) clearly shows that Bi-GRU obtained a high R2 value of 0.96 and an NRMSE of 0.09, an ND of 0.08, an RMSSE of 0.08, and MAE of 0.07, and an MSE of 0.06.

Table 4. Numerical analysis of different regression models on the Citadel POS dataset.

Table 4. *Cont.*

In addition to the assessment of various classifiers, a statistical test was used to assess the performance of all the classifiers. For that reason, the Friedman test, also known as the nonparametric statistical test, was applied in this research. The Friedman test uses the data ranks rather than the data themselves and tests the null hypothesis in which the column properties are all identical. In some other way, all classifiers have a comparable impact on the classification process. In this statistical test, the probability of gaining the detected sample outcomes (*p*-value) is computed as a scalar value in the range of [0, 1]. Small values of " p " decline the null hypothesis. The Friedman test is calculated as $p = 0.0071$, which shows that the proposed Bi-GRU has a superior performance compared to other classifiers in a significant manner.

In addition to this, the numerical analysis of different feature engineering techniques with the Bi-GRU model on the Rossmann and Walmart datasets is presented in Table [5.](#page-13-0) By analyzing Table [5,](#page-13-0) we can see that the Bi-GRU model with feature engineering techniques (APSO+RFE+MRMR) obtains superior retail sales forecasting results compared to the other combinations with the Bi-GRU model by means of the NRMSE, ND, RMSSE, MAE, R2, and MSE.

The active features selected from APSO+RFE+MRMR includes four major benefits: (i) improves Bi-GRU model's performance by learning meaningful patterns in the acquired datasets; (ii) reduces overfitting risk; (iii) improves computational efficiency, where the transformed features need fewer computational resources; and (iv) improves Bi-GRU model's interpretability. As mentioned in the earlier sections, the selection of active features diminishes the processing time of the Bi-GRU model to 20.11 s and 30.12 s, respectively, in the Rossmann and Walmart datasets and reduces the model's complexity to linear.

Table 5. Numerical analysis of different feature engineering techniques on the Rossmann and Walmart dataset.

Table 5. *Cont.*

4.4. Comparative Analysis

The comparative study between the existing regression models and the proposed Bi-GRU model is detailed in this section. Ozyegen et al. [\[38\]](#page-17-26) analyzed the performance of three existing regression models in retail sales forecasting, namely LSTM, Gradient Boosted Regressor (GBR), and Time Delay Neural Network (TDNN). In this study, the existing regression model's efficacy was analyzed by applying it to the Rossmann and Walmart datasets. The experiments carried out on the Rossmann and Walmart datasets state that the GBR model obtains minimal mean NRMSE and ND values when related to other two regression models. The GBR model has a mean NRMSE of 0.12 and 0.01 and ND of 0.20 and 0.10 on the Rossmann and Walmart datasets, respectively. The comparative results between the existing regression models (GBR, TDNN, and LSTM) and the proposed Bi-GRU model are clearly stated in Table [6](#page-14-0) and Figure [4.](#page-15-1)

MSE 0.04 0.03 R2 0.98 0.99

Table 6. Comparative results between the existing regression models and the proposed Bi-GRU model by means of NRMSE and ND.

On the other hand, Yang [\[39\]](#page-18-0) analyzed the sales prediction performance of XGBoost, random forest, and Ordinary Least Squares (OLS) models on the Walmart dataset. According to the numerical investigation, the XGBoost model has a minimal MAE of 0.12 and MSE of 0.06, and a high R2 of 0.98 on the Walmart dataset. The obtained numerical results are superior to the random forest and OLS models.

In addition, Niu [\[40\]](#page-18-1) performed feature engineering with the XGBoost model for precise sales forecasting on the Walmart dataset. The presented XGBoost model has a minimal RMSSE of 0.65, which is superior to other comparative regression models, like ridge and logistic regression. When related to these existing regression models, the presented Bi-GRU model has a minimal error value and high R2 value on the Rossmann and Walmart datasets. The comparative results between the XGBoost model and the proposed Bi-GRU model are specified in Table [7.](#page-15-2)

Figure 4. Graphical comparison between the proposed and existing regression models. **Figure 4.** Graphical comparison between the proposed and existing regression models.

Table 7. Comparative results between the existing regression model and the proposed Bi-GRU model
' by means of R2, MAE, MSE, and RMSSE.

Dataset	Models	R2	MAE	MSE	RMSSE
Walmart	XGBoost [39]	0.98	0.12	0.06	-
	Random forest [39]	0.97	0.14	0.08	-
	OLS [39]	0.72	0.67	1.11	-
	Bi-GRU	0.99	0.07	0.03	-
	XGBoost [40]	$\overline{}$	$\overline{}$	$\overline{}$	0.65
	Bi-GRU	-	-	-	0.08

4.5. Discussion

In this study, the integration of Bi-GRU with the hybrid feature engineering techniques achieved a higher forecasting performance on the Rossmann and Walmart datasets when compared with Ozyegen et al. [\[38\]](#page-17-26), Yang [\[39\]](#page-18-0), and Niu [\[40\]](#page-18-1). The results are superior to the those of the existing regression models, like linear regression, LSTM, Bi-LSTM, and GRU. The proposed Bi-GRU model has minimal NRMSEs of 0.08 and 0.01, NDs of 0.07 and 0.03, RMSSEs of 0.08 and 0.08, MAEs of 0.05 and 0.07, and MSEs of 0.04 and 0.03, as well as high R2 values of 0.98 and 0.99, on the Rossmann and Walmart datasets, which are shown in Tables [2](#page-10-1) and [3,](#page-11-0) respectively.

The practical implications of the findings in the context of E-commerce companies are as follows: (i) customer satisfaction by producing timely delivery of the product, and (ii) long-term strategy plan for reliable future growth. Furthermore, the precise retail sales forecasting using the Bi-GRU-APSO model improves their operations by allowing the E-commerce business to effectively allocate resources for managing cash flow and future. Additionally, the retail sales forecasting results in an accurate estimation of revenue and costs based on their long- and short-term performance. The major five benefits of timely retail sales forecasting are (i) eliminates the chances of panic sales, (ii) leverages real time data, (iii) offers simple financial planning, (iv) stabilizes inventory management, and (v) speeds up the process of product delivery.

5. Conclusions

A Bi-GRU model was implemented in this study with effective feature engineering techniques for precise retail sales forecasting. In the initial phase, the unreliable samples in the Rossmann and Walmart datasets were eliminated by removing outliers, interpolating missing data, normalization, and de-normalization. Next, the feature engineering was performed using the APSO algorithm, RFE, and MRMR to select the active features from the pre-processed datasets, which were finally given to the Bi-GRU model for precise retail sales forecasting. The efficacy of the proposed regression model was analyzed using six evaluation measures: NRMSE, ND, RMSSE, MAE, MSE, and R2. The proposed BiGRU model with effective feature engineering achieved a superior performance in sales forecasting in relation to the traditional regression models: linear regression, LSTM, Bi-LSTM, and GRU. As mentioned in the numerical evaluation section, the proposed Bi-GRU model has minimum MAEs of 0.05 and 0.07 and MSEs of 0.04 and 0.03 on the Rossmann and Walmart datasets. In addition, the Bi-GRU model has a high R2 value of 0.98 and 0.99 on the Rossmann and Walmart datasets, which are better than the comparative regression models. On the other hand, the selection of active features decreases the processing time to 20.11 s and 30.12 s on the Rossmann and Walmart datasets. Still, the Bi-GRU model has issues like low-learning efficiency and a slow convergence rate. Furthermore, the results analysis of this sales forecasting is completely dependent on a specified dataset. In case of changes in the dataset region, the result also varies and cannot be compared with some other different datasets. Therefore, as a future extension, a novel hyperparameter-tuning technique will be integrated with the proposed regression model which will be suitable for improving the performance of retail sales forecasting under various datasets. Additionally, this research will be analyzed with real-time data for better visualization results.

Author Contributions: The paper investigation, resources, data curation, writing—original draft preparation, writing—review and editing, and visualization were performed by A.M.G. and H.D.P. The paper conceptualization and software were conducted by A.K. The validation and formal analysis, methodology, supervision, project administration, and funding acquisition of the version to be published were conducted by P.F.-G. and P.B.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated during and/or analzsed during the current study are available in the [Rossmann dataset] and [Walmart dataset] repository: Rossmann dataset link = [https://www.kaggle.com/competitions/rossmann-store-sales/data,](https://www.kaggle.com/competitions/rossmann-store-sales/data) and Walmart dataset link = [https://www.kaggle.com/datasets/yasserh/walmart-dataset.](https://www.kaggle.com/datasets/yasserh/walmart-dataset)

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

References

- 1. Esmeli, R.; Bader-El-Den, M.; Abdullahi, H. Towards Early Purchase Intention Prediction in Online Session Based Retailing Systems. *Electron. Mark.* **2021**, *31*, 697–715. [\[CrossRef\]](https://doi.org/10.1007/s12525-020-00448-x)
- 2. Fildes, R.; Ma, S.; Kolassa, S. Retail Forecasting: Research and Practice. *Int. J. Forecast.* **2022**, *38*, 1283–1318. [\[CrossRef\]](https://doi.org/10.1016/j.ijforecast.2019.06.004)
- 3. Wang, X.S.; Ryoo, J.H.J.; Bendle, N.; Kopalle, P.K. The Role of Machine Learning Analytics and Metrics in Retailing Research. *J. Retail.* **2021**, *97*, 658–675. [\[CrossRef\]](https://doi.org/10.1016/j.jretai.2020.12.001)
- 4. Wang, J.; Liu, G.Q.; Liu, L. A Selection of Advanced Technologies for Demand Forecasting in The Retail Industry. In Proceedings of the 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), Suzhou, China, 15–18 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 317–320.
- 5. Güven, I.; Şimşir, F. Demand Forecasting with Color Parameter in Retail Apparel Industry Using Artificial Neural Networks (ANN) And Support Vector Machines (SVM) Methods. *Comput. Ind. Eng.* **2020**, *147*, 106678. [\[CrossRef\]](https://doi.org/10.1016/j.cie.2020.106678)
- 6. Cui, H.; Rajagopalan, S.; Ward, A.R. Predicting Product Return Volume Using Machine Learning Methods. *Eur. J. Oper. Res.* **2020**, *281*, 612–627. [\[CrossRef\]](https://doi.org/10.1016/j.ejor.2019.05.046)
- 7. Salinas, D.; Flunkert, V.; Gasthaus, J.; Januschowski, T. Deepar: Probabilistic Forecasting with Autoregressive Recurrent Networks. *Int. J. Forecast.* **2020**, *36*, 1181–1191. [\[CrossRef\]](https://doi.org/10.1016/j.ijforecast.2019.07.001)
- 8. Kavitha, M.; Srinivasan, R.; Kavitha, R.; Suganthy, M. Sales Demand Forecasting for Retail Marketing Using Xgboost Algorithm. In *Intelligent and Soft Computing Systems for Green Energy*; Wiley Online Library: Hoboken, NJ, USA, 2023; pp. 127–140.
- 9. Schaeffer, S.E.; Sanchez, S.V.R. Forecasting Client Retention-A Machine-Learning Approach. *J. Retail. Consum. Serv.* **2020**, *52*, 101918. [\[CrossRef\]](https://doi.org/10.1016/j.jretconser.2019.101918)
- 10. Ramos, P.; Santos, N.; Rebelo, R. Performance of State Space and ARIMA Models for Consumer Retail Sales Forecasting. *Rob. Comput. Integr. Manuf.* **2015**, *34*, 151–163. [\[CrossRef\]](https://doi.org/10.1016/j.rcim.2014.12.015)
- 11. Sharma, A.; Shafiq, M.O. Predicting Purchase Probability of Retail Items Using an Ensemble Learning Approach and Historical Data. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 14–17 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 723–728.
- 12. Yu, M.; Tian, X.; Tao, Y. Dynamic Model Selection Based on Demand Pattern Classification in Retail Sales Forecasting. *Mathematics* **2022**, *10*, 3179. [\[CrossRef\]](https://doi.org/10.3390/math10173179)
- 13. Dhote, S.; Vichoray, C.; Pais, R.; Baskar, S.; Mohamed Shakeel, P. Hybrid Geometric Sampling and adaboost Based Deep Learning Approach for Data Imbalance in E-Commerce. *Electron. Commer. Res.* **2020**, *20*, 259–274. [\[CrossRef\]](https://doi.org/10.1007/s10660-019-09383-2)
- 14. Pavlyshenko, B.M. Machine-Learning Models for Sales Time Series Forecasting. *Data* **2019**, *4*, 15. [\[CrossRef\]](https://doi.org/10.3390/data4010015)
- 15. Nosratabadi, S.; Mosavi, A.; Duan, P.; Ghamisi, P.; Filip, F.; Band, S.S.; Reuter, U.; Gama, J.; Gandomi, A.H. Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics* **2020**, *8*, 1799. [\[CrossRef\]](https://doi.org/10.3390/math8101799)
- 16. He, Q.Q.; Wu, C.; Si, Y.W. LSTM with Particle Swam Optimization for Sales Forecasting. *Electron. Commer. Res. Appl.* **2020**, *51*, 101118. [\[CrossRef\]](https://doi.org/10.1016/j.elerap.2022.101118)
- 17. Ji, S.; Wang, X.; Zhao, W.; Guo, D. An Application of a Three-Stage Xgboost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise. *Math. Probl. Eng.* **2019**, *2019*, 8503252. [\[CrossRef\]](https://doi.org/10.1155/2019/8503252)
- 18. Massaro, A.; Panarese, A.; Giannone, D.; Galiano, A. Augmented Data and XGBoost Improvement for Sales Forecasting in the Large-Scale Retail Sector. *Appl. Sci.* **2020**, *11*, 7793. [\[CrossRef\]](https://doi.org/10.3390/app11177793)
- 19. Wong, W.K.; Guo, Z.X. A Hybrid Intelligent Model for Medium-Term Sales Forecasting in Fashion Retail Supply Chains Using Extreme Learning Machine and Harmony Search Algorithm. *Int. J. Prod. Econ.* **2020**, *128*, 614–624. [\[CrossRef\]](https://doi.org/10.1016/j.ijpe.2010.07.008)
- 20. Loureiro, A.L.; Miguéis, V.L.; da Silva, L.F. Exploring the Use of Deep Neural Networks for Sales Forecasting in Fashion Retail. *Decis. Support Syst.* **2018**, *114*, 81–93. [\[CrossRef\]](https://doi.org/10.1016/j.dss.2018.08.010)
- 21. Lu, C.J. Sales Forecasting of Computer Products Based on Variable Selection Scheme and Support Vector Regression. *Neurocomputing* **2014**, *128*, 491–499. [\[CrossRef\]](https://doi.org/10.1016/j.neucom.2013.08.012)
- 22. Luo, T.; Chang, D.; Xu, Z. Research on Apparel Retail Sales Forecasting Based on Xdeepfm-LSTM Combined Forecasting Model. *Information* **2022**, *13*, 497. [\[CrossRef\]](https://doi.org/10.3390/info13100497)
- 23. Zhang, B.; Tan, R.; Lin, C.J. Forecasting of E-Commerce Transaction Volume Using a Hybrid of Extreme Learning Machine and Improved Moth-Flame Optimization Algorithm. *Appl. Intell.* **2021**, *51*, 952–965. [\[CrossRef\]](https://doi.org/10.1007/s10489-020-01840-y)
- 24. Weng, T.; Liu, W.; Xiao, J. Supply Chain Sales Forecasting Based on Lightgbm and LSTM Combination Model. *Ind. Manag. Data Syst.* **2020**, *120*, 265–279. [\[CrossRef\]](https://doi.org/10.1108/IMDS-03-2019-0170)
- 25. Shilong, Z. Machine Learning Model for Sales Forecasting by Using Xgboost. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 15–17 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 480–483.
- 26. Punia, S.; Nikolopoulos, K.; Singh, S.P.; Madaan, J.K.; Litsiou, K. Deep Learning with Long Short-Term Memory Networks and Random Forests for Demand Forecasting in Multi-Channel Retail. *Int. J. Prod. Res.* **2020**, *58*, 4964–4979. [\[CrossRef\]](https://doi.org/10.1080/00207543.2020.1735666)
- 27. Karmy, J.P.; Maldonado, S. Hierarchical Time Series Forecasting Via Support Vector Regression in The European Travel Retail Industry. *Expert Syst. Appl.* **2019**, *137*, 59–73. [\[CrossRef\]](https://doi.org/10.1016/j.eswa.2019.06.060)
- 28. Kilimci, Z.H.; Akyuz, A.O.; Uysal, M.; Akyokus, S.; Uysal, M.O.; Atak Bulbul, B.; Ekmis, M.A. An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain. *Complexity* **2019**, *2019*, 9067367. [\[CrossRef\]](https://doi.org/10.1155/2019/9067367)
- 29. Kohli, S.; Godwin, G.T.; Urolagin, S. Sales prediction using linear and KNN regression. In *Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019*; Springer: Singapore, 2021; pp. 321–329.
- 30. Ma, S.; Fildes, R. Retail Sales Forecasting with Meta-Learning. *Eur. J. Oper. Res* **2021**, *288*, 111–128. [\[CrossRef\]](https://doi.org/10.1016/j.ejor.2020.05.038)
- 31. Tang, L.; Li, J.; Du, H.; Li, L.; Wu, J.; Wang, S. Big Data in Forecasting Research: A Literature Review. *Big Data Res.* **2022**, *27*, 100289. [\[CrossRef\]](https://doi.org/10.1016/j.bdr.2021.100289)
- 32. Yang, R.; Yin, L.; Hao, X.; Liu, L.; Wang, C.; Li, X.; Liu, Q. Identifying A Suitable Model for Predicting Hourly Pollutant Concentrations by Using Low-Cost Microstation Data and Machine Learning. *Sci. Rep.* **2022**, *12*, 19949. [\[CrossRef\]](https://doi.org/10.1038/s41598-022-24470-5)
- 33. Hossain, M.M.; Swarna, R.A.; Mostafiz, R.; Shaha, P.; Pinky, L.Y.; Rahman, M.M.; Rahman, W.; Hossain, M.S.; Hossain, M.E.; Iqbal, M.S. Analysis of The Performance of Feature Optimization Techniques for The Diagnosis of Machine Learning-Based Chronic Kidney Disease. *Mach. Learn. Appl.* **2022**, *9*, 100330. [\[CrossRef\]](https://doi.org/10.1016/j.mlwa.2022.100330)
- 34. Jain, M.; Saihjpal, V.; Singh, N.; Singh, S.B. An Overview of Variants and Advancements of PSO Algorithm. *Appl. Sci.* **2022**, *12*, 8392. [\[CrossRef\]](https://doi.org/10.3390/app12178392)
- 35. Yu, C. Design of Drug Sales Forecasting Model Using Particle Swarm Optimization Neural Networks Model. *Comput. Intell. Neurosci.* **2022**, *2022*, 6836524. [\[CrossRef\]](https://doi.org/10.1155/2022/6836524)
- 36. Li, X.; Ma, X.; Xiao, F.; Xiao, C.; Wang, F.; Zhang, S. Time-Series Production Forecasting Method Based on the Integration of Bidirectional Gated Recurrent Unit (Bi-GRU) Network and Sparrow Search Algorithm (SSA). *J. Pet. Sci. Eng.* **2022**, *208*, 109309. [\[CrossRef\]](https://doi.org/10.1016/j.petrol.2021.109309)
- 37. Zhu, Z.; Dai, W.; Hu, Y.; Li, J. Speech Emotion Recognition Model Based on Bi-GRU and Focal Loss. *Pattern Recognit. Lett.* **2020**, *140*, 358–365. [\[CrossRef\]](https://doi.org/10.1016/j.patrec.2020.11.009)
- 38. Ozyegen, O.; Ilic, I.; Cevik, M. Evaluation of Interpretability Methods for Multivariate Time Series Forecasting. *Appl. Intell.* **2022**, *52*, 4727–4743. [\[CrossRef\]](https://doi.org/10.1007/s10489-021-02662-2) [\[PubMed\]](https://www.ncbi.nlm.nih.gov/pubmed/34764613)
- 39. Yang, T. Sales Prediction of Walmart Sales Based on OLS, Random Forest, and Xgboost Models. *Highlights Sci. Eng. Technol.* **2023**, *49*, 244–249. [\[CrossRef\]](https://doi.org/10.54097/hset.v49i.8513)
- 40. Niu, Y. Walmart Sales Forecasting Using Xgboost Algorithm and Feature Engineering. In Proceedings of the 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Chengdu, China, 23–25 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 458–461.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.