Int. J. Nonlinear Anal. Appl. In Press, 1–13
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2024.34189.5103

Forecasting models of demand in supply chain with high product diversification using gradient boosting machine learning methods

Mohammad Reza Fahimi[a], Ali Rajabzadeh Ghatari[b,*], Maryam Shoar[a], Maryam Khademi[c]

[a] *Department of Industrial Management, Islamic Azad University, North Tehran Branch, Tehran, Iran*

[b] *Department of Industrial Management, Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran*

[c] *Department of Applied Mathematics, Islamic Azad University, South Tehran Branch, Tehran, Iran*

(Communicated by Ali Ebadian)

Abstract

Forecasting demand in the supply chain is an essential and challenging issue for determining effective strategies and informed decisions. The related challenges also increase by raising the variety and number of products. Providing frameworks and methods with the flexibility, accuracy, and advantages necessary to forecast all product categories is crucial for managers, despite product diversity, differences in applications and features, and varying data volumes. In this regard, two supervised learning models, XGBoost Regressor (XGBR) and Gradient Boosting Regressor (GBR), have been implemented on the Global Superstore dataset on the Kaggle site. This dataset contains 3788 products in three diverse product categories, seventeen subcategories, and 51,290 orders. The limited data volume of the products prevented the possible and helpful forecasting of many products and the obtaining of appropriate results from the methods. In this experimental research, demand forecasting is used in strategic decisions. Therefore, a product aggregation approach has been proposed for this problem, which can be predicted separately according to the similarity in the subcategory products. The data of the dataset was increased using the Data Augmentation technique to investigate the effect of the amount of data on the performance of the models, and the forecasting results of the two models were compared by re-running the models. Based on the forecasting results with increased data with MSE and MAE metrics, the XGBR model achieved the lowest values of 0.12 and 0.10, respectively, and the GBR model achieved values of 0.13 and 0.15. Also, the result of the D^2 Score Metric was 0.97 in the XGBR model and 0.96 in the GBR model. The values of the error metrics decreased dramatically and by more than 80% as the data increased, and XGBR had a relative advantage in the larger data. The proposed framework and models can be used in industries with similar issues at the strategy level.

Keywords: demand forecasting, high product diversification, supply chain, machine learning, gradient boosting
2020 MSC: 68T07, 90B06

*Corresponding author
Email addresses: m.fahimi@iau-tnb.ac.ir (Mohammad Reza Fahimi), alirajabzadeh@modares.ac.ir (Ali Rajabzadeh Ghatari), m_shoar@iau-tnb.ac.ir (Maryam Shoar), Khademi@azad.ac.ir (Maryam Khademi)

Received: April 2024 *Accepted:* June 2024

1 Introduction

Research has proven the importance of supply chain management in the sustainability of many organizations, especially in today's turbulent era. The historical data of the last fifteen years indicate that many organizations were forced to shut down due to the lack of proper recognition of market signs and the inability to keep up with the rapid development of technology and the fast growth of consumer demand and expectations [1]. Forecasting is essential to support most supply chain management decisions. Time and product are the main dimensions that determine the decision-making process. Demand forecasting is the basis of most decisions in supply chain management, and forecasts at higher levels and on longer horizons are needed for strategic supply chain decisions [4]. The retail industry has seen a shift from intuition-based decision-making to data-based decision-making, with significant efforts to develop and improve predictive models over the past few decades [10]. Market and competition factors that determine retail strategy are typically dependent on forecasts [17]. Chain stores are one of the most important types of retail, consisting of several retail stores under centralized common ownership [5].

Given the nature of some chain stores in offering diverse products with different brands, applications, and features, there has always been a challenge in predicting demand to determine effective strategies. On the one hand, product diversity leads to variations in the number of orders over time and the dispersion of order distribution, compounded by insufficient time series data for low-selling products, rendering many forecasting methods impractical. On the other hand, numerous changes and the abundance of similar items make predicting each product feasible and useful. Naturally, a general demand forecast may not be beneficial, and strategic demand forecasting should be conducted at an appropriate level of product hierarchy with similar characteristics and applications.

Various studies, including Babai et al. [4], have proposed that the necessary details for forecasting should be determined based on planning and decision-making needs, and strategic supply chain decisions require forecasting at higher levels. In some cases, forecast accuracy may improve with data aggregation or forecasting at lower or higher levels. Therefore, the use of hierarchical levels in forecasting has been recommended. One of the methods of this aggregation was identified as segmental aggregation in one dimension, for example, products, making it feasible to transfer forecasting to higher or lower levels.

In this research, a different approach to demand forecasting based on subcategory aggregation has been proposed. This approach is suitable considering the research objectives for strategies and product similarities at this level. Furthermore, there are challenges in selecting and employing forecasting methods and models, considering the issues related to diverse data resulting from product differences, as described. Therefore, selecting a method that can predict all selected items with minimal error and consistent performance in the face of relatively diverse data volumes is challenging under the circumstances. While some researchers have employed various methods to achieve satisfactory results, their implementation usually imposes costs and time burdens on businesses and sometimes leads to complications. A dataset from a chain store with diverse and numerous products with the necessary characteristics was selected from the Kaggle website to address this challenge. The objective of the research was to present a combined framework and model capable of predicting appropriate demand in circumstances with limited data and product diversity. Models were proposed and tested with unchanged and augmented data using XGB and gradient-boosting machine learning methods.

2 Research literature

Numerous studies have been conducted on the use of machine learning methods, including those employed, to predict demand in supply chains with high product diversity, retail outlets, chain stores, and similar entities. In this section, some of the latest approaches, along with their details, have been reflected. Ultimately, in the conclusion section, a comprehensive comparison of the approaches will be presented, encompassing the experimental data, the framework, the demand forecasting target in the supply chain, and the employed models.

In a supply chain study, Islam et al. [12] proposed a three-stage solution framework for addressing problems in supplier selection and order allocation planning. In the initial stage, a new modified relational deep learning prediction technique was developed for predicting product demands, and the performance of this modified technique was compared with two popular prediction techniques, SARIMA and Light Gradient Boosted Machine (LGBM). The proposed developed framework was evaluated using a real dataset from the Canadian meat industry. Based on the prediction model results, the researchers claimed that the developed deep learning network can reduce prediction errors by 55.42% compared to the SARIMA method and by 13.1% compared to the LGBM method.

Chien et al. [8] stated in the spare parts supply chain that ensuring customer satisfaction and minimizing sufficient inventory in this industry is challenging due to high variability in demand size and time intervals resulting from postsales repair and maintenance orders. A stacking ensemble approach to improve overall prediction performance was used in this research to classify demand patterns and develop relevant models through data-driven intelligent technology. The proposed composite model was based on various methods, including XGB, which creates a monitoring alert system for performance and a systematic mechanism for model retraining to maintain decision quality.

Lee et al. [16] deemed optimal pricing determination for the fast-moving consumer goods industry as one of the challenging effects of substitution in retail stores. In this study, the substitution effect necessitated examining numerous combinations of price changes and the availability of other products, for which a systematic decision support tool was proposed for price decision-making to predict demand and optimize prices in online and brick-and-mortar stores, considering the substitution effect. Researchers have introduced two methods for demand prediction and price optimization models that reflect product price changes and demand correlation structure.

The first method, i.e., the developed demand prediction method considering price changes for all products, selects the best demand prediction for the product from among time series and various machine learning methods by adjusting hyperparameters. The second method, i.e., the developed price optimization procedure, is a constrained optimization problem that operates based on a weekly time horizon and product category aggregation level capable of maximizing profit from various price combinations. Different methods, including ARIMA, RF, GBM, Extra Tree, MLR, Decision-Tree, K-Neighbours, LightGBM, HistGBM, Seasonal-Naïve, NaïveBayes, Simple-Median, and Simple-Average XGB, were compared for prediction.

Furthermore, Andrade and Cunha [3] addressed the issue of demand forecasting in retailing with diverse goods, considering segmented forecasting to ensure profitability as a support for precise decisions regarding inventory management. This study aimed to prevent inventory shortages or surpluses and associated losses and argued that the abundance of modern retail stores and products, along with complex marketing and advertising strategies, significantly influences customer demand. Given the interdependencies among products, modeling all of them becomes intricate. Researchers have proposed the XGB model, which employs a nonlinear and nonparametric composite model as the central learning algorithm. As a result of sudden changes in consumer behavior, structural change adjustment is also incorporated into these methods.

Additionally, the XGB model involves data cleansing procedures to rectify sales observations during periods of inventory shortages and logical and physical inventory discrepancies. This study, based on actual data from a large retail dataset, demonstrated that the proposed model outperforms Base-Lift model methods, a widely used benchmark for retail sales prediction. Moreover, retailers can achieve high levels of automation thanks to their improved accuracy metrics and reductions in inventory shortages and warehouse inventories.

In another study, Joseph et al. [13] considered predicting product demand in retail operators to aid in decision-making through data analysis for crafting strategies and achieving business objectives. A challenging dataset of demand prediction was utilized for a store from the Kaggle website to implement the proposed framework. The main innovation of this study lay in constructing a CNN-BiLSTM framework coupled with the Lazy Adam optimizer for accurate prediction of product demand from store items. Researchers executed advanced machine learning techniques such as SGD, XGBoost RF, KNN, SVR, CNN-LSTM Bagging, and Linear Regression for demand prediction and compared the results with the proposed model. The results using metrics including MAPE, MAE, and R-squared demonstrated that the researchers' proposed framework exhibits higher accuracy compared to traditional approaches. Apart from chain stores and retailers, the use of machine learning methods, including Gradient Boosting family methods for demand prediction in other industries and services with product diversity, has also been utilized, among which studies such as Kwon et al. [15] on South Korean blood products, Abolghasemi et al. [2] on hospital blood supply chains, and Song and Hu [22] on electric vehicle charging demand can be mentioned.

3 Theoretical foundations

3.1 Levels of decision-making and supply chain demand

Time and product are the two primary dimensions that determine the details of supply chain decisions. Stock Keeping Units (SKUs) is a product-focused decision, ranging from inventory control to overall capacity planning. From a temporal perspective, operational decisions are made at daily or weekly levels, while tactical and strategic decisions are made at monthly and yearly levels. At the strategic level, supply chain managers increasingly face uncertain capacity and increasing market and technological changes, which compels them to consider managing capacity and deciding on distribution channels (online, retail, or omnichannel) while considering the entire product portfolio offered to those markets. Alignment between strategic, tactical, and operational plans in long-term, medium-term, and short-term horizons is necessary. Thus, the level of detail in forecasting, both temporally and cross-sectionally, should be determined based on planning and decision-making requirements. Therefore, forecasting requires many details, and considerations of this nature answer the question of what should be forecasted. In response to various questions, solutions and forecasting methods can be examined at each level of the hierarchy using data or forecasts. Thus, forecasts can be derived from objectives using bottom-up or top-down methods at each level of the hierarchy [4]. There are a vast number of products that, individually or in combination, form natural hierarchical levels at each level of the supply chain. Demand should be aggregated at these hierarchical levels to determine decision-making processes across a wide range of organizational and functional levels [24].

3.2 Machine learning and its evaluation metrics

Machine learning is a subset of artificial intelligence in which machine learning algorithms can automatically learn from raw data to create predictive models based on pre-designed algorithms. Generally, there are two types of algorithms, including supervised learning and unsupervised learning. Supervised machine learning algorithms learn from labeled data, such as input and output, which are responsible for finding the relationship between input and output. Learning stops when an acceptable level of performance is achieved. Machine learning includes various algorithms such as artificial neural networks, support vector machines, random forests, regression, decision trees, and k-means algorithms, each of which has its advantages and disadvantages for implementation in any business [1]. Supervised and unsupervised learning algorithms are primarily employed for four types of tasks: regression, classification, clustering, and dependency [14]. One of the various applications of machine learning is constructing a regression or nonparametric classification model from data. Building a strong predictive model is one of the most common data-driven modeling approaches [19]. Numerous studies have been conducted on the use of machine learning in the supply chain. Efforts made in utilizing machine learning algorithms and data analysis to estimate customer demand and minimize supply chain costs can facilitate accurate (data-driven) demand prediction and align supply chain activities with these predictions to improve efficiency and customer satisfaction [21].

In this research, two proposed models based on machine learning methods were used, XGBoost and Gradient Boosting, for the regression task of predicting the amount of demand in the supply chain in the time interval of the selected data set. The model input is the characteristics associated with each order, including the order date, price, and discount, which are available in the selected research data set, and the output is the predicted demand amount of each subcategory in the time axis. The model performance in demand forecasting is evaluated with the following statistical metrics:

A. The mean absolute error or MSE (Mean Squared Error) evaluation Metric is one of the error Metric calculated from Equation (3.1), and the closer its value is to zero, the more favorable it is.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i + \hat{Y}_i)^2 \tag{3.1}$$

B. The mean absolute error or MAE (Mean Absolute Error) evaluation Metric is one of the error Metric calculated from Equation (3.2), and the closer its value is to zero, the more favorable it is.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i + \hat{Y}_i|^2 \tag{3.2}$$

C. The D^2 Score Metric calculates the explained variance ratio. This score is a generalization of R^2 where the squared error is normalized and replaced by an arbitrary descriptor $dev(y, \hat{y})$ such as (tweedy, mean absolute error, and pinball). The best score for this Metric is to reach 1, like R^2, in situations where it can become negative due to bad model performance. This metric is used in some libraries dedicated to statistics and machine learning of the Python programming language, including the scikit-learn library, and is calculated from Equation (3.3) [9]. This library is evaluated in the article [20] titled "Scikit-learn: Machine Learning in Python."

$$D^2(y, \hat{y}) = 1 - \frac{dev(y, \hat{y})}{dev(y, y_{null})} \tag{3.3}$$

3.2.1 Gradient boosting method

The Enhanced Gradient Boosting Regression Tree method, proposed by Friedman [11], is a powerful machine learning technique based on gradient boosting, which has demonstrated significant success in a wide range of practical applications. In the Gradient Boosting model, the learning process sequentially adapts new models to present a more accurate estimation of the response variable. The main idea of this algorithm is to construct new base learners in a way that maximizes the correlation with the negative gradient of the loss function relevant to the entire dataset. A loss function can be arbitrary, but a least squares error loss function is sequentially fitted during the learning process for better understanding. The researcher generally chooses loss functions, and they can be trained with different loss functions to adapt to the specific requirements of each application. This method, introduced by Friedman [11] and extended by Natekin and Knoll [19], generates sequential base models from the weighted version of the training data to strategically achieve the best combination of trees. The aim of adding each new base model is to correct the errors of the previous base models. Therefore, this method has the potential to provide accurate predictions [26].

3.2.2 XGBoost method

Among the various methods of machine learning, the XGB algorithm is one of the Gradient Boosting-based approaches introduced by Chen and Guestrin [7] from the University of Washington. In this method, Gradient Boosting is a highly effective and widely used machine learning method. XGB is a scalable and comprehensive tree-boosting system used by data scientists to achieve advanced results in many machine learning challenges. In addition, the XGB algorithm provides approximate tree learning by considering sparse data and utilizing the weighted Quantile sketch technique. Most importantly, researchers have proposed insights into hidden memory access patterns, data compression, and partitioning for building a scalable tree-boosting system. XGB scales to billions of data samples with significantly fewer resources compared to existing systems by combining these insights. In general, the XGB model is among the common algorithms in prediction tasks, employing a collection of decision trees (DT) to construct a robust regressor. This machine learning method is suitable for large-scale implementation, employing automatic parallelization for accelerated execution time. Regularization penalties, such as tree depth and final node weights, are part of the XGB objective function. Thus, the iterative process is reduced, and tree construction performance is enhanced. In this approach, a level-wise decision tree growth technique is utilized to reduce model complexity.

4 Research method

4.1 Research data

The dataset of this experimental research, entitled "Global Superstore Orders 2016 [6]," was published on the Kaggle website, which pertains to Global Superstore, a renowned retail company operating globally, offering a wide range of products, including furniture, electronics, office supplies, and more. These data encompass 51,290 product orders placed by customers worldwide from January 1, 2011, to December 31, 2014.

Table 1: Number of products and orders of categories and subcategories of the dataset

Category	Sub Category	Products				Orders				Average order per product
		Quantity		Percent of total		Quantity		Percent of total		
Technology	Accessories	263	876	6.94%	23.13%	3075	10141	6.00%	19.77%	11.69
	Machines	179		4.73%		1486		2.90%		8.30
	Phones	305		8.05%		3357		6.55%		11.01
	Copiers	129		3.41%		2223		4.33%		17.23
Furniture	Bookcases	166	841	4.38%	22.20%	2411	9876	4.70%	19.26%	14.52
	Chairs	204		5.39%		3434		6.70%		16.83
	Furnishings	301		7.95%		3170		6.18%		10.53
	Tables	170		4.49%		861		1.68%		5.06
Office Supplies	Appliances	212	2071	5.60%	54.67%	1755	31273	3.42%	60.97%	8.28
	Art	272		7.18%		4883		9.52%		17.95
	Binders	326		8.61%		6152		11.99%		18.87
	Envelopes	159		4.20%		2435		4.75%		15.31
	Fasteners	127		3.35%		2420		4.72%		19.06
	Labels	185		4.88%		2606		5.08%		14.09
	Paper	392		10.35%		3538		6.90%		9.03
	Storage	247		6.52%		5059		9.86%		20.48
	Supplies	151		3.99%		2425		4.73%		16.06
Total		3788		100%		51290		100%		–

The categorization and number of products are presented in Table 1. Since the research dataset had limited data and for understanding and investigating the effect of increasing data on model results, the SMOTE (Synthetic Minority Oversampling Technique) technique was used to augment the current dataset. Ultimately, the implementation results of the models were examined in both the original data and the augmented data scenarios.

4.2 Research machine learning models

The employed machine learning models are based on two methods: XGBoost Regressor (XGBR) and Gradient Boosting Regressor (GBR). These models were implemented using the Python programming language and the relevant libraries, and their execution results were evaluated in predicting demand values. The implemented models were designed and configured with the following specifications:

4.2.1 XGBR

In this study, the XGBR model, designed for regression tasks, is initialized with the following hyperparameters:

The "max_depth" parameter is set to 8, controlling the maximum depth of each tree in the ensemble, which helps control the model's complexity and prevent overfitting. The "n_estimators" parameter is set to 1000, determining the number of trees to be built in the ensemble. More trees can improve the model's performance but also increase computation time. The "subsample" parameter is set to 0.8, indicating that each tree is trained on a random subsample of 80% of the training data. This randomness in the training process helps reduce overfitting. The "Random_state" parameter is set to 42, ensuring the reproducibility of results. The XGBR model aims to construct a set of gradient-boosting regression trees that ensure stable results with the specified characteristics of these hyperparameters.

4.2.2 GBR

The GBR model, which is a group learning method for regression tasks, is initialized with the following hyperparameters:

The "max_depth" parameter is set to a value of 8, which controls the maximum depth of each regression tree in the ensemble to restrict the complexity of each tree and helps prevent overfitting. The GBR model constructs a sequence of regression trees sequentially. Each subsequent tree is trained to correct the errors made by the previous trees. This model gradually improves its predictions by minimizing a loss function, usually using gradient descent optimization. The model allows each regression tree in the ensemble to have a maximum depth of 8 levels by setting the "max_depth" parameter to 8. This complexity control helps balance between overfitting and underfitting. The GBR model with the specified hyperparameters aims to construct an ensemble of regression trees for regression tasks. The group approach and iterative training process enable the model to learn complex relationships and make accurate predictions for unseen data.

4.3 Implementation phases of modeling

The following stages were utilized for implementing the proposed models on the selected dataset in accordance with the strongest methodology in the execution and implementation of CRISP-DM data mining projects:

Phase One (Business Understanding): In this phase, based on the project objectives, the current status of the data and identified resources, requirements, and constraints therein were examined.

Phase Two (Data Understanding): This phase involved describing and analyzing the data, validating the data quality, and standardizing and formatting the data appropriately.

Phase Three (Data Preparation): Data preprocessing and cleansing were conducted in this phase, where missing, outliers, and noisy data were removed and corrected. Subsequently, the most relevant features were selected for the proposed model to improve the performance of the proposed model in the subsequent stages by creating potential and engineering new features.

Phase Four (Modeling): In this phase, the data were divided for training, testing, and validation, and implementing classification models (supervised machine learning) was carried out.

Phase Five (Evaluation and Validation): Following modeling, error metrics, accuracy, precision, recall, and F-score were calculated to evaluate the results.

Phase Six (Deployment): In this phase, a roadmap for project deployment was outlined, and finally, the project report and final results were presented in the form of tables and charts.

5 Results and findings

MSE and MAE were used as metrics to evaluate the results regarding errors after implementing the proposed models for segmenting each of the sub-categories. The results of the XGBR model execution are presented in Table 2, and the results of the GBR model are presented in Table 3. Additionally, the evaluation results of the models using the D^2 Score metric are provided in Table 4. The model results on the original data and the augmented data, along with the resulting changes, are presented.

5.1 Results and findings of XGBR model execution

The proposed XGBR model with the specified characteristics could achieve MSE values ranging from 1.72 to 4.16 on the original dataset, with only two cases exceeding 4. On the other hand, employing the model on augmented data shows a significant improvement, reducing MSE to between 0.12 and 0.57. MSE in subcategories has decreased by 2.6 on average following data augmentation, resulting in an 87% improvement. In addition, the XGBR model on the original dataset achieved MAE values ranging from 0.85 to 1.4 in predicting subcategories, while employing the model on augmented data has led to a notable improvement, reducing MAE to the range of 0.10 to 0.25. MAE in subcategories decreased on average by 0.97 with data augmentation, leading to an 83% improvement.

Moreover, the minimum MAE for predicting the demand for subcategories in the original data pertains to Machines, Chairs, and Envelopes, as well as Machines, Furnishings, and Envelopes in augmented data. In the D^2 Score metric, which is similar to the R^2 metric, values closer to 1 are preferable. In the execution of the XGBR model on the original dataset, the obtained D^2 Score ranges from 0.15 to 0.51.

Table 2: Error evaluation results of XGB Regressor model implementation

Category	**Sub Category**	**Main data results**		**Augmented data results**		**Changes with increasing data**		**Percentage changes with increasing data**	
		MSE	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
Technology	Accessories	3.5368	1.3462	0.5717	0.2472	2.9651	1.099	83.84%	81.86%
	Machines	2.5171	1.0274	0.3564	0.1667	2.1607	0.8607	85.84%	83.77%
	Phones	3.3888	1.2228	0.4801	0.2339	2.9087	0.9889	85.83%	80.87%
	Copiers	2.8948	1.2532	0.3647	0.2047	2.5301	1.0485	87.40%	83.67%
Furniture	Bookcases	3.7367	1.3354	0.5713	0.2326	3.1654	1.1028	84.71%	82.58%
	Chairs	1.8253	0.8514	0.2544	0.1784	1.5709	0.673	86.06%	79.05%
	Furnishings	2.0095	0.8893	0.1198	0.1028	1.8897	0.7865	94.04%	88.44%
	Tables	4.0837	1.4018	0.5385	0.2261	3.5452	1.1757	86.81%	83.87%
Office Supplies	Appliances	2.2817	0.9899	0.3330	0.1912	1.9487	0.7987	85.41%	80.68%
	Art	4.1645	1.3167	0.4871	0.2099	3.6774	1.1068	88.30%	84.06%
	Binders	2.7031	1.0873	0.3674	0.1938	2.3357	0.8935	86.41%	82.18%
	Envelopes	1.7150	0.8592	0.1251	0.1007	1.5899	0.7585	92.71%	88.28%
	Fasteners	3.6172	1.3344	0.4853	0.2314	3.1319	1.103	82.66%	86.58%
	Labels	3.1413	1.2531	0.5491	0.2516	2.5922	1.0015	82.52%	79.92%
	Paper	3.7736	1.2644	0.4556	0.2096	3.3180	1.0548	87.93%	83.42%
	Storage	2.3168	1.0556	0.2561	0.1698	2.0607	0.8858	88.95%	83.91%
	Supplies	3.3542	1.4092	0.4241	0.2052	0.2052	2.9301	87.36%	85.44%
Mean		3.0035	1.1704	0.3964	0.1974	2.6071	0.9730	87.10%	83.20%

The model on augmented data led to a significant improvement, increasing the D^2 Score to the range of 0.92 to 0.97. The average D^2 Score within Sub Categories increased by 0.63 on average with the augmentation of data, resulting in an average improvement of 242% percent. This increase in D^2 Score ranges from a minimum of 96% to a maximum of 523%. Furthermore, the highest level of the D^2 Score Metric obtained from model execution for predicting the demand of Sub Categories in the original data in each Category related to Machines, Furnishings, and Envelopes and in augmented data remains the same.

5.2 Results and findings of GBR model execution

The proposed GBR model with the specified characteristics achieved MSE values ranging from 1.64 to 3.76 on the original dataset. However, employing the model on augmented data resulted in a significant improvement, reducing

MSE to the range of 0.13 to 0.82. The average MSE within Subcategories decreased by 2.29 on average with the increase in data, resulting in an improvement of as much as 83%. Moreover, the GBR model on the original dataset achieved MAE between 0.85 and 1.37 in predicting subcategories, and employing the model on augmented data led to a notable improvement, reducing MAE to the range of 0.15 to 0.53. With the augmentation of data, the average MAE within subcategories decreased by 0.79 on average, resulting in a 70% improvement.

The lowest MSE Metric obtained from model execution for predicting the demand of subcategories in the original data in each Category is related to Machines, Chairs, and Envelopes, and in augmented data, Machines, Furnishings, and Envelopes, respectively. Likewise, the lowest MAE criterion obtained from model execution for predicting the demand of subcategories in the original data in each Category is related to Machines, Chairs, and Envelopes, and in augmented data, Machines, Furnishings, and Envelopes respectively.

Evaluation of GBR model results with the D^2 Score Metric on the original dataset indicates the attainment of values ranging from 0.21 to 0.52. Employing the model on augmented data led to a significant improvement, increasing the D^2 Score to the range of 0.85 to 0.96. The average D^2 Score within subcategories increased by 0.56 on average with the augmentation of data, resulting in an average improvement of 190%. This increase in D^2 Score ranges from a minimum of 79% to a maximum of 349%. Furthermore, the highest level of the D^2 Score Metric obtained from model execution for predicting the demand of subcategories in the original data in each Category is related to Machines, Furnishings, and Envelopes, and in augmented data, Accessories, Furnishings, and Envelopes.

Table 3: Error evaluation results of Gradient Boosting Regressor model implementation

Category	**Sub Category**	**Main data results**		**Augmented data results**		**Changes with increasing data**		**Percentage changes with increasing data**	
		MSE	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
Technology	Accessories	3.2199	1.2519	0.5114	0.3071	2.7085	0.9448	84.12%	75.47%
	Machines	2.4467	1.0366	0.4181	0.3031	2.0286	0.7335	82.91%	70.76%
	Phones	3.0903	1.1991	0.7457	0.5028	2.3446	0.6963	75.87%	58.07%
	Copiers	2.459	1.1423	0.4451	0.3414	2.0139	0.8009	81.90%	70.11%
Furniture	Bookcases	3.327	1.2591	0.7199	0.4503	2.6071	0.8088	78.36%	64.24%
	Chairs	1.7455	0.7981	0.2664	0.2555	1.4791	0.5426	84.74%	67.99%
	Furnishings	1.9526	0.8713	0.1412	0.1776	1.8114	0.6937	92.77%	79.62%
	Tables	3.6471	1.3375	0.6033	0.3852	3.0438	0.9523	83.46%	71.20%
Office Supplies	Appliances	2.1704	1.0091	0.4729	0.3847	1.6975	0.6244	78.21%	61.88%
	Art	3.7652	1.3066	0.8263	0.5337	2.9389	0.7729	78.05%	59.15%
	Binders	2.3446	0.9797	0.4237	0.2898	1.9209	0.6899	81.93%	70.42%
	Envelopes	1.6475	0.8532	0.1305	0.1549	1.517	0.6983	92.08%	81.84%
	Fasteners	3.4591	1.2992	0.5023	0.3893	2.9568	0.9099	85.48%	70.04%
	Labels	3.0788	1.2088	0.6378	0.3342	2.441	0.8746	79.28%	72.35%
	Paper	3.4286	1.2187	0.6264	0.4012	2.8022	0.8175	81.73%	67.08%
	Storage	2.0521	0.998	0.2724	0.2597	1.7797	0.7383	86.73%	73.98%
	Supplies	3.2743	1.3678	0.4267	0.2285	2.8476	1.1393	86.97%	83.29%
Mean		2.7711	1.1257	0.4806	0.3352	2.2905	0.7905	83.21%	70.44%

5.3 Comparing the results of XGBR and GBR models

Comparing the results of the XGBR and GBR models demonstrates that the prediction of Sub Categories in the GBR model, with an average MSE of 2.77 and MAE of 1.12, exhibits lower error compared to the XGBR model, which has an average MSE of 3.00 and MAE of 1.17. However, the results of the XGBR model show a significant reduction in average MSE and MAE to 0.3964 and 0.1974, respectively, with the augmentation of data using the Data Augmentation technique outperforming the GBR method with values of 0.4806 and 0.3352. The rate and percentage of reduction in both metrics in the XGBR method are more significant than in the GBR method with the increase in data. Additionally, comparing the D^2 Score results of the two models indicates that predicting subcategories in the GBR model, with an average of 0.34, is slightly better than the XGBR model, which averages 0.31. Although both methods have shown a remarkable increase and improvement in this metric, the results of the XGBR method, with an average D^2 Score of 0.94, outperform the GBR method with values of 0.90 with the augmentation of data. The rate and percentage of increase in this metric in the XGBR method are higher than in the GBR method with the increase in data.

Table 4: D^2 Score metric results of XGBR and GBR models

Category	Sub Category	Main data results		Augmented data results		Changes with increasing data		Percentage changes with increasing data	
		XGBR	GBR	XGBR	GBR	XGBR	GBR	XGBR	GBR
Technology	Accessories	0.1479	0.2093	0.9162	0.9326	0.7070	0.7829	523%	338%
	Machines	0.4153	0.4100	0.9105	0.9508	0.5005	0.5355	129%	122%
	Phones	0.2697	0.2838	0.8562	0.9331	0.5724	0.6634	246%	202%
	Copiers	0.2690	0.3336	0.9023	0.9414	0.5687	0.6725	250%	170%
Furniture	Bookcases	0.2089	0.2541	0.8660	0.9308	0.6119	0.7219	346%	241%
	Chairs	0.4838	0.5161	0.9243	0.9472	0.4082	0.4634	96%	79%
	Furnishings	0.5148	0.5246	0.9474	0.9696	0.4228	0.4548	88%	81%
	Tables	0.1997	0.2364	0.8899	0.9353	0.6534	0.7356	368%	276%
Office Supplies	Appliances	0.4301	0.4191	0.8913	0.9460	0.4722	0.5158	120%	113%
	Art	0.2644	0.2701	0.8509	0.9414	0.5808	0.6770	256%	215%
	Binders	0.3351	0.4009	0.9102	0.9400	0.5094	0.6049	181%	127%
	Envelopes	0.4953	0.4988	0.9551	0.9708	0.4562	0.4755	96%	91%
	Fasteners	0.2281	0.2485	0.8834	0.9307	0.6349	0.7025	308%	255%
	Labels	0.2295	0.2567	0.9004	0.9250	0.6437	0.6955	303%	251%
	Paper	0.2730	0.2993	0.8855	0.9402	0.5862	0.6672	244%	196%
	Storage	0.3737	0.4078	0.9291	0.9537	0.5213	0.5800	155%	128%
	Supplies	0.1839	0.2079	0.9339	0.9406	0.7260	0.7567	411%	349%
Mean		0.3132	0.3398	0.9031	0.9429	0.5633	0.6297	242%	190%

5.4 Overall demand prediction results with model execution

Forecasting the demand for products in supply chains with diverse, numerous, and various applications, including chain stores and retailers, always faces challenges in selecting an appropriate supply chain framework for determining prediction levels and details, as well as choosing a suitable method and model. This study focuses on strategic planning at the subcategory level, which includes products with similar characteristics, with the aim of determining the issues and details of the supply chain for the prediction level.

In this study, demand forecasting quantities for products in the supply chain with diverse and numerous items but limited demand data were carried out using the "Global Superstore Orders 2016" dataset pertaining to a retail company with global operations. The tested dataset comprises 51,290 recorded orders of 3,788 different products, excluding any advertising data and customer feedback, over approximately four years from 2011 to 2014, encompassing 1,461 days or 48 months. These products, with various applications, are categorized into three main categories: electronics, office supplies, and furniture, and 17 Subcategories, demonstrating high diversity. Despite the limited number of orders within the dataset timeframe, the average number of orders per product ranges from only 5 to 20, which is very low. In addition, the order number for Subcategories ranges from 127 to 392, with an average of 177 orders. On average, each subcategory had only 44 orders per year, 6.3 orders per month, and not even one order per week. As a general rule, this is the average order quantity, and it has certainly been lower in some cases. This data limitation and restricted features practically hindered the successful implementation of many prediction methods and algorithms. Up to the time of writing this research, no other study has been observed utilizing this dataset for demand prediction. Although some researchers in different studies have attempted to achieve satisfactory results by adding complexity to the models and algorithms and combining various methods, their implementation usually requires equipment that imposes additional costs and time on businesses and sometimes leads to delays and disruptions in operations and reduces decision-making flexibility and agility. Moreover, data and products or specific periods are typically selected in their research. Therefore, this study encountered a set of challenges and issues arising from working with actual data without selection.

The prediction method should also align with the needs and constraints of the research data in addition to the framework and level of prediction previously described. Based on the insights gained from exploratory data analysis, specific methods are not feasible to implement or achieve satisfactory results, given the limitations involved in selecting methods. Therefore, XGBoost Regressor and Gradient Boosting Regressor were ultimately selected based on their suitability for the requirements after reviewing and experimenting with a set of machine learning algorithms, deep learning, and statistical and time series methods. These models, with specified architectures and settings, underwent a series of data analysis and feature engineering stages for demand prediction. Data augmentation techniques were employed to assess the impact of data augmentation on model performance, and the results of the models were compared with both original and augmented data.

The proposed models demonstrated acceptable outcomes despite the diversity of product subcategories and variations in their applications and features, leading to data constraints. This aligns with the characteristics mentioned by the developers Chen and Guestrin [7] regarding XGB's compatibility with sparse data. Additionally, the XGBR model exhibited significant improvement with data augmentation, confirming the scalability properties stated by Chen and Guestrin [7]. The success of the XGB method in dealing with limited data is further supported by a study by Sugiharti et al. [23], where researchers proposed a combined transfer learning method based on XGB and CNN to enhance prediction accuracy in the face of challenges posed by deep learning methods with sparse data.

Furthermore, despite differences in the number, quantity, timing, and density of orders across various subcategories of the three main categories of Technology, Furniture, and Supplies, the proposed models yielded satisfactory results with the advantage of suitable flexibility, providing relatively uniform prediction errors for all Sub Categories. Despite the mentioned data constraints, the models mentioned above performed acceptably compared to other studies and proposed models by other researchers, which may involve complexity, achieving acceptable performance in predicting subcategory products of a retail chain with high diversity.

Based on the present tabular research data, which includes product orders with hierarchical categorizations, previous research has examined the selection of the methods mentioned above for supply chain demand prediction with tabular data structure, which has also been utilized in similar studies. Among these studies, Yang et al. [25] stated that deep neural networks prove their efficacy in processing sensory data such as images and sounds. However, tree-based models are more popular for tabular data. One of the good features of tree-based models is their inherent interpretability. Additionally, Malioutov et al. [18] argue that these models repeatedly partition the input vector space and assign scores to the final nodes. Tree-based models not only enhance performance in tabular data but also increase the interpretability of the system, thus improving its applicability in commercial scenarios.

Based on information from the KDD Cup competition in Chen and Guestrin [7], an examination of the challenges presented by the Kaggle machine learning competition platform indicates that among the 29 winning solutions to the challenges hosted on the Kaggle blog throughout 2015, 17 solutions utilized XGB. Among these solutions, eight solely employed XGB for model training, while most other solutions combined XGB with neural networks in ensembles. Furthermore, the comparison reveals that deep neural networks were utilized in 11 solutions in the second popular method. The success of this approach in KDD Cup 2015 was also evident, where XGB was used by every winning team among the top 10 teams. In addition, the winning teams reported that ensemble methods with only a small portion of XGB, configured appropriately, outperformed. The issues addressed by these solutions include sales forecasting, web text classification, customer behavior prediction, motion detection, advertisement click-through rate prediction, malware classification, product categorization, risk prediction, and online course dropout rate prediction. Although data analysis based on domain and feature engineering plays a crucial role in these solutions, XGB is a consensus learner that highlights the impact and importance of boosting tree systems. The most significant factor in XGB's success is its scalability in all scenarios, running more than ten times faster than popular existing solutions on a single machine and scaling to billions of samples in distributed or memory-limited settings. XGB's scalability is due to several critical systems and algorithm optimizations.

6 Conclusion and recommendation

6.1 Discussion and conclusion

The current research is comparable to previous studies cited in the research literature from various perspectives.

The study by Islam et al. [12] relating to three types of red meat supply chains in Canada with a vast amount of data for a relatively small number of products differs significantly from the present research in terms of diversity, order volume, and product data volume, and there are no prediction challenges in their study under conditions of product diversity and limited data. The framework includes three stages for problem-solving: process optimization, supplier selection, and order allocation planning, with only the first step being the prediction of final product demand. The ensemble-LSTM, modified ARIMA, LSTM, and LGBM models were compared with each other. Given the massive volume of data, the execution of the combined modified LSTM deep learning model was successful, as the proposed method could reduce the prediction error compared to the LGBM method. However, in the present study, with limited data volume, the execution of multiple models based on LSTM and other similar deep-learning approaches did not yield satisfactory results, and consequently, XGBR and GBR methods were employed.

Lee et al. [16] conducted research on a three-year dataset from 2009 to 2011 of product sales transactions in US supermarket chains. Cold cereals were the only category of products considered for reasons such as frequent purchases, adequate subset sizes, a wide range of discounts, consumable products, and less impact of time series. There have been many challenges, such as product diversity and variation in order frequency, which have been overlooked, and only 12 specific products with high transaction volumes were considered. Therefore, they did not face challenges similar to those in the present study. The selection actions were not hierarchical or aggregative and aimed to investigate the relationship between price and demand. A method was sought to be designed for selecting the best demand forecasting method among different time series and machine learning techniques. A variety of methods was employed to predict product demand. The research results using the RMSPE Metric indicated that the Extra Trees, Gradient Boosting, and XGB models achieved the best results with slight differences, respectively. Considering the similarity between their study and ours, the successful outcomes of the research methods in the study testify to the correct choice of method in the present study.

The research by Andrade and Cunha [3] on a real dataset of an ample food retail store in Ecuador comprises 54 large stores and 4036 items from 2013 to 2017. This dataset also includes comprehensive information on product advertising and marketing. However, products with the lowest sales volume were removed through selection, and ultimately, 1623 products were selected using the Pareto law, while products with short sales duration and deficient sales volume were eliminated. As a result, the present study presents a prediction framework based on subcategories without removing products and marketing data. Additionally, considering the scope and purpose of their research, product dimensions, supply chains, and time, such as SKU, store, and weekly, respectively, were defined, and marketing advertising data was used to make predictions. Thus, considering the aggregation levels, the selection of the entire supply chain, and the time range for strategic purposes in the present study, these two studies differ from each other. This research used significant data and utilized the XGB method, which is similar to the present study.

Chien et al. [8] classified demand patterns in an automotive parts manufacturer, testing their proposed models on four demand patterns comprising 849 experimental products. The company provides diverse services and products worldwide. The company's data has been anonymized for confidentiality reasons. Hence, the current research faces challenges in dealing with real and limited data issues. Considering the difficulties of maintenance and repair in the spare parts industry, their study aims to classify demand patterns and develop relevant models through a stacking ensemble approach to enhance overall prediction performance.

Additionally, the UNISON six-step framework was proposed based on combinatorial learning, presenting a different approach from the current research to empower an integrated data-driven value chain and develop a demand forecasting framework to enhance supply chain resilience. The proposed model, in the form of an XGB combination with specific and intricate designs, demonstrated lower error rates compared to other individual methods used. However, in the current research, models XGBR and GBR were evaluated separately without combining or using different methods, showing results with low and uniform errors, high correlation, and acceptable performance flexibility.

Joseph et al. [13] focused on predicting demand for retail operators' products using the prediction challenge dataset of a store from Kaggle. This dataset includes sales data for 50 items for five years (2013 to 2017) in 10 stores. The objective of this challenge was to predict sales for the next three months. This dataset only contained four features and 913,000 rows of sales data, and only the training dataset provided in the Kaggle competition was considered. The dataset structure was modified efficiently to include features that represent sales data for each store and item over the five years. A series of changes were made to the data, and the research, in terms of data volume, had better conditions and could achieve better results than traditional methods using the CNN-BiLSTM deep learning model.

6.2 Suggestions for future researchers

One of the challenges in demand forecasting studies in the supply chain was the lack of suitable and sufficiently large datasets, which occurred due to the non-cooperation of many companies and the absence of dataset provision in many similar articles. In light of the limitations of the current research, the adequacy and structure of the dataset and its analysis, the discovery of suitable patterns, and the requirements and needs of prediction methods are essential topics for future research. Additionally, presenting models and frameworks for the prediction challenge under data constraints, diversity of classifications, and products can be helpful for future research.

Acknowledgment

The authors would like to express their gratitude to the esteemed Research Department of Islamic Azad University, Tehran North, for their support and cooperation in conducting the present research.

References

[1] A. Aamer, L. Eka Yani, and I. Alan Priyatna, *Data analytics in the supply chain management: Review of machine learning applications in demand forecasting*, Operat. Supply Chain Manage.: Int. J. **14** (2020), no. 1, 1–13.

[2] M. Abolghasemi, B. Abbasi, and Z. HosseiniFard, *Machine learning for satisficing operational decision making: A case study in blood supply chain*, Int. J. Forecast. In Press, (2023).

[3] L.A.C. Andrade and C.B. Cunha, *Disaggregated retail forecasting: A gradient boosting approach*, Appl. Soft Comput. **141** (2023), p. 110283.

[4] M.Z. Babai, J.E. Boylan, and B. Rostami-Tabar, *Demand forecasting in supply chains: a review of aggregation and hierarchical approaches*, Int. J. Product. Res. **60** (2022), no. 1, 324–348.

[5] B. Berman, J.R. Evans, and P. Chatterjee, *Retail Management-A Strategic Approach*, 13th eds, Pearson Education India, 2018.

[6] J. Brown, *Global superstore data of 2016*, Kaggle, https://www.kaggle.com/datasets/jamsbrown/global-superstore-data-of-2016/data, 2023.

[7] T. Chen and C. Guestrin, *Xgboost: A scalable tree boosting system*, Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discov. Data Min., 2016, pp. 785–794.

[8] C.F. Chien, C.C. Ku, and Y.Y. Lu, *Ensemble learning for demand forecast of after-market spare parts to empower data-driven value chain and an empirical study*, Comput. Ind. Eng. **185** (2023), p. 109670.

[9] S.-L. Developers, *Metrics and scoring: Quantifying the quality of predictions*, User Guide, Dispon’ıvelem: https://scikit-learn.org/stable/modules/modelevaluation.html. **26** (2021).

[10] M. Fisher and A. Raman, *Using data and big data in retailing*, Product. Operat. Manag. **27** (2018), no. 9, 1665–1669.

[11] J.H. Friedman, *Greedy function approximation: a gradient boosting machine*, Ann. Statist. **29** (2001), no. 5, 1189–1232.

[12] S. Islam, S.H. Amin, and L.J. Wardley, *A supplier selection and order allocation planning framework by integrating deep learning, principal component analysis, and optimization techniques*, Expert Syst. Appl. **235** (2024), 121121.

[13] R.V. Joseph, A. Mohanty, S. Tyagi, S. Mishra, S.K. Satapathy, and S.N. Mohanty, *A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting*, Comput. Electrical Engin. **103** (2022), 108358.

[14] E.R. Kone and M.H. Karwan, *Combining a new data classification technique and regression analysis to predict the Cost-To-Serve new customers*, Comput. Ind. Eng. **61** (2011), no. 1, 184–197.

[15] H.J. Kwon, S. Park, Y.H. Park, S.M. Baik, and D.J. Park, *Development of blood demand prediction model using artificial intelligence based on national public big data*, Digital Health **10** (2024), 20552076231224245.

[16] K.H. Lee, M. Abdollahian, S. Schreider, and S. Taheri, *Supply chain demand forecasting and price optimisation models with substitution effect*, Mathematics **11** (2023), no. 11, p. 2502.

[17] M. Levy and D. Grewal, *Retail Management*, McGraw-Hill, 2023.

[18] D.M. Malioutov, K.R. Varshney, A. Emad, and S. Dash, *Learning interpretable classification rules with boolean compressed sensing*, T. Cerquitelli, D. Quercia and F. Pasquale, Transparent data mining for big and small data, Studies in Big Data, Springer, Cham. **32** (2017), 95–121.

[19] A. Natekin and A. Knoll, *Gradient boosting machines, a tutorial*, Front. Neurorobotics **7** (2013), 21.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and É. Duchesnay, *Scikit-learn: Machine learning in Python*, J. Machine Learn. Res. **12** (2011), 2825–2830.

[21] M. Seyedan and F. Mafakheri, *Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities*, J. Big Data **7** (2020), p. 53.

[22] Y. Song and X. Hu, *Learning-based demand-supply-coupled charging station location problem for electric vehicle*

demand management, Transport. Res. Part D: Transport Envir. **125** (2023), 103975.

[23] E. Sugiharti, R. Arifudin, D.T. Wiyanti, and A.B. Susilo, *Convolutional neural Network-XGBoost for accuracy enhancement of breast cancer detection*, J. Phys.: Conf. Ser. IOP Publishing, **1918** (2021), no. 4, 042016.

[24] A.A. Syntetos, Z. Babai, J.E. Boylan, S. Kolassa, and K. Nikolopoulos, *Supply chain forecasting: Theory, practice, their gap and the future*, Eur. J. Oper. Res. **252** (2016), no. 1, 1–26.

[25] Y. Yang, I.G. Morillo, and T.M. Hospedales, *Deep neural decision trees*, arXiv preprint arXiv:1806.06988, (2018).

[26] Y. Zhang and A. Haghani, *A gradient boosting method to improve travel time prediction*, Transport. Res. Part C: Emerg. Technol. **58** (2015), 308–324.