


## Article

# Sales Forecasting for Fashion Products Considering Lost Sales

Dali Chen <sup>1</sup> , Wenbiao Liang <sup>2</sup>, Kelan Zhou <sup>1,\*</sup> and Fan Liu <sup>1</sup>

<sup>1</sup> School of Management and Engineering, Nanjing University, Nanjing 210093, China; dalichen@mail.nju.edu.cn (D.C.); liufan@nju.edu.cn (F.L.)

<sup>2</sup> Nari Group Co., Ltd., Nanjing 210031, China; liangwenbiao@sgepri.sgcc.com.cn

\* Correspondence: zhoukelan0303@163.com

**Abstract:** Sales forecasting for new products is significantly important for fashion retailer companies because prediction with high accuracy helps the company improve management efficiency and customer satisfaction. The low inventory strategy of fashion products and the low stock level in each brick-and-mortar store lead to serious censored demand problems, making forecasting difficult. In this regard, a two layers (TLs) model is proposed in this paper to predict the total sales of new products. In the first layer, the demand is estimated by linear regression (LR). In the second layer, sales are modeled as a function of not only the demand but also the inventory. To solve the TLs model, a gradient-boosting decision tree method (GBDT) is used for feature selection. Considering the heterogeneity in products, a mixed k-mean algorithm is applied for product clustering and a genetic algorithm for parameter estimation in each cluster. The model is tested on real-world data from a Singapore company, and the experimental results show that our model is better than LR, GBDT, support vector regression (SVR) and artificial neural network (ANN) in most cases. Furthermore, two indicators are built: the average conversion rate and the marginal conversion rate, to measure products' competitiveness and explore the optimal inventory level, respectively, which provide helpful guidance on decision-making for fashion industry managers.

**Keywords:** sales forecasting; fashion products; censored demand; inventory decision



**Citation:** Chen, D.; Liang, W.; Zhou, K.; Liu, F. Sales Forecasting for Fashion Products Considering Lost Sales. *Appl. Sci.* **2022**, *12*, 7081. <https://doi.org/10.3390/app12147081>

Academic Editor: Vincent A. Cicirello

Received: 9 June 2022

Accepted: 5 July 2022

Published: 13 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sales forecasting is a critically important task in the fashion retail industry. It is an integral component of many core fashion retail processes, including the functional areas of pricing and inventory management, marketing and production purchasing, as well as finance and accounting [1,2]. Forecasts on sales also provide the basis for regional distribution and replenishment plans [3]. Therefore, the ability to estimate the probable sales quantity could lead to improved customer satisfaction, reduced waste, increased sales revenue and more effective and efficient production plans as after-sales services [4]. Generally, accurate sales forecasting is a prerequisite to competing efficaciously in the fashion world [5].

A number of methods have been developed over the past two decades. Statistical methods are widely used in forecasting fashion products, including traditional time series techniques and discrete choice models such as the multinomial logistic model and Bayesian analysis [6,7]. Another popular method in sales forecasting is machine learning, such as artificial neural networks (ANN), evolutionary neural networks (ENN) and extreme learning machines (ELM) [8–10].

Despite the extensive research on the sales forecasting problem in fashion retail industry, there remain some practical issues that need consideration. Fashion products have the characteristic of short product life cycles [11]. Therefore, fashion products usually have a low inventory strategy in general, and every brick-and-mortar store has a low stock level, which leads to a serious problem with demand censoring [12,13]. Censored demand increases the difficulty of forecasting and the impact of inventory, and censored demand to

sales are unignorable in the fashion retail industry. Research on the total sales forecasting problem for fashion products with censored demand is relatively rare, which motivates us to design a new model to alleviate the impact of censored demand on sales considering the lost sales phenomenon. The objective of this paper is to improve the total sales forecast accuracy for fashion products and provide helpful guidance on decision-making for fashion industry managers.

In this paper, a two layers (TLs) model for forecasting the total sales of new products is proposed, considering not only the actual demand but also the effect of censored demand in brick-and-mortar stores. In the first layer, the demand is estimated by traditional linear regression (LR), while in the second layer, the sales are modeled as a function of not only the demand but also the inventory. We incorporate two key parameters in this model, with one describing a product's basic conversion rate of one inventory to sales and another one describing the customers' preference for products. To solve the model, the gradient-boosting decision tree (GBDT) method is applied for feature selection. Considering the heterogeneity, products are clustered by a mixed k-mean clustering algorithm, and parameter estimation is conducted separately in each cluster by a genetic algorithm (GA). Collaborating with a multinational fashion retailer in Singapore, experiments are conducted based on real-world data of shoes and belts. The results of experiments are compared based on different models, including our TLs model, LR, GBDT, support vector regression (SVR) and ANN. The results show that our method outperforms other sales forecasting methods in most cases. For further conclusions, we analyze the results obtained by the TLs model and define the average conversion rate when the inventory level is equal to demand as one indicator to measure product competitiveness. We also involve the marginal conversion rate as another indicator for managers to make proper inventory decisions.

Generally, our contribution is summarized in three points: (1) A TLs sales forecasting model is proposed based on the relationship between sales, inventory and demand and incorporates two parameters to describe the conversion rate of inventory to sales. The marginal sales, which is the change in total sales that arises when the inventory is increased by one unit, is modeled by the conversion rate of inventory to sales. (2) A comparison is conducted among the sales forecasting results of a fashion enterprise using the TLs model, LR, GBDT, SVR and ANN. The result of the experiments shows that our TLs model outperforms benchmarks in most cases. (3) Two different indicators, the average conversion rate and the marginal conversion rate are defined to measure the product competitiveness and decide the optimal inventory level. These could provide fashion managers with great help in product and inventory management.

The rest of this paper is organized as follows. In Section 2, we present a literature review on the related work. Section 3 introduces the proposed TLs model and gives an explanation of the parameters. Section 4 is about the results and analysis of the data experiment, introducing two indicators, the average conversion rate and the marginal conversion rate. Finally, Section 5 concludes our study and points out potential directions for future research in the subject area.

## 2. Literature Review

Our work is related to the research on the forecasting problem in the fashion industry and retail industry. The commonly adopted methods are statistical methods and machine learning methods. In the following, we review several existing methods for solving the forecasting problem.

Being widely applied, statistical methods include regression [14], time series techniques, Bayesian analysis and discrete choice model. Time-series methods are traditional univariate forecasting models, ranging from simpler moving averages, the exponential smoothing family, to Auto Regression-Integrated Moving Average approach (ARIMA) [15]. Green and Harrison [16] apply a Bayesian approach to conduct forecasting for a mail-order company, which sells ladies' dresses, while Lee et al. [17] propose a hierarchical Bayesian model over prelaunch sales forecasting of recorded music. The discrete choice

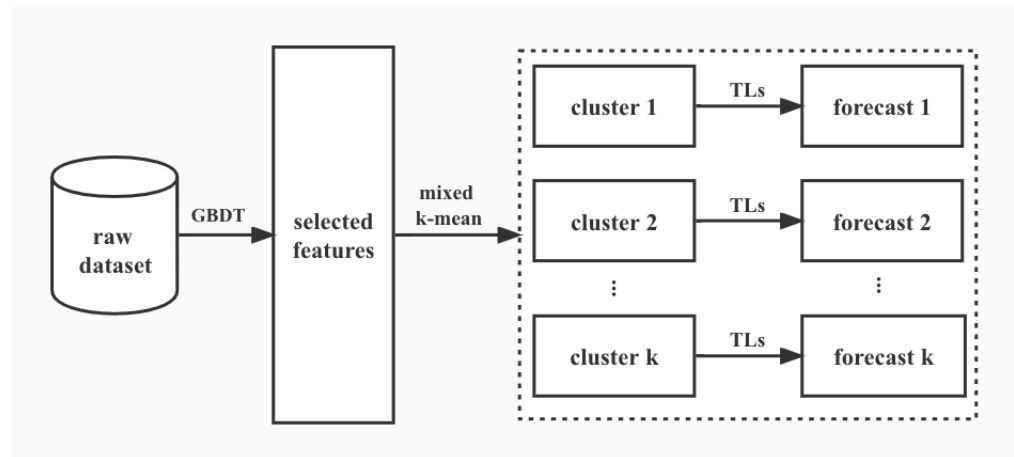
model (DCM), for example, the multinomial logistic model (MNL), is also popular in prediction. DCM conducts retail forecasting by modeling consumers' probability of choice. The literature explores the application of DCM to a forecasting problem in the retail industry [7,18]. Despite their popularity, statistical methods present several limitations. For example, statistical methods require relatively restrictive assumptions to perform well, such as the sample distribution of products and the independently identical distribution of samples. Since sales of fashion products exhibit a high degree of randomness, these hypotheses might not correspond with the actual situation and have a negative impact on the forecasting performance.

Another stream of studies on machine learning methods emerged with the advance computer technology. Popular methods such as artificial neural networks (ANNs), evolutionary neural networks (ENNs) and extreme learning machines (ELM) are widely used in the literature. Frank et al. [19] conduct fashion retail sales forecasting by using ANN due to its promising performance in the areas of control and pattern recognition. Chawla et al. [20] display a case study of an American retail corporation on forecasting by using ANN. ENN is a hybrid combination of evolutionary computation and neural networks. Au et al. [8] develop an optimized neural network structure for the forecasting of apparel sales using ENN. ELM learns faster than the conventional gradient-based learning methods and is usually adopted together with other methods by many researchers. Yu et al. [21] employ both the ELM and the traditional statistical methods to forecast sales, and Wong and Guo [22] propose a hybrid intelligent model by using ELM and the harmony search algorithm. Further, Choi et al. [23] achieve real-time sales forecasting by combining the ELM with the grey model. Machine learning performs well in identifying and modeling data patterns that are not easily discernible by traditional statistical methods. Some research studies show that machine learning methods outperform traditional statistical methods [24,25]. Machine learning methods are sometimes unstable; in other words, there are cases in which they might not work well. Moreover, these methods lack interpretability and need a substantial amount of time to conduct the prediction.

For the problem of censored demand, there have been research studies on inventory management in recent years. Nahmias and Smith [26] explore the optimal inventory levels in a retailer system with lost sales, while Heese and Swaminathan [27] provide insights regarding optimal inventory and sales effort management in the presence of unobserved sales. To the best of our knowledge, no previous literature studies the sales forecasting problem of fashion products in the case of censored demand.

### 3. Methodology

In this section, we propose a two layers (TLs) model for forecasting the total sales of new fashion products accounting for censored demand. We then develop an optimization method and establish the optimal sales plan after the introduction of the proposed model. The flow of the solution is shown in Figure 1, where the whole process of how the raw dataset transforms to forecast results is displayed from left to right. After the raw dataset enters the system, a gradient-boosting decision tree model is involved to conduct feature selection. After selecting the important features for training, we cluster the dataset by using a mixed k-mean method with the consideration of heterogeneity in different products. When finishing the steps above, a genetic algorithm is applied for parameter estimation in each cluster to give the output of forecasting in the TLs model.



**Figure 1.** Fluxogram of Methodology.

### 3.1. The Proposed Sales Forecasting Model

Sales are transformed from the potential demand, and the transformation process is limited by the scale of inventory of a fashion product. When the inventory is less than the potential demand, some demand will not be met. However, even if the inventory is sufficient, some customers may not purchase the product because products are not placed in the correct store, which leads to demand loss. Given the demand, we define marginal sales as the change in the total sales that arises with the increment in inventory by one unit. In an ideal state, as long as there exists unsatisfied potential demand, the increased inventory will be sold out, and the marginal sales are equal to one. However, in reality, marginal sales are always less than one. As the sales increase, the number of potential customers decreases, and the chance of the product being discovered by the customer drops, leading to a decrease in marginal sales as well. In this manner, we model the marginal sales by the following equations:

$$M(I) = q \left( 1 - \left( \frac{S(I-1)}{D} \right)^\alpha \right), \quad I \in N^+, \quad (1)$$

where  $S(I)$  and  $M(I) = S(I) - S(I-1)$  is sales and the marginal sales when inventory is  $I$ . The sales iterative formula describes the relationship between sales, inventory and demand. We already have the marginal sales equation when the inventory is  $I$ , and the product sales can be calculated by the iterative formula  $S(I) = \sum_{i=1}^I M(i)$ .

The right side of Equation (1) consists of two parts, and the first part is  $q$ . Since  $S(0) = 0$ , then  $M(1) = q$  describes the marginal sales of the initial inventory. Therefore,  $q \in [0, 1]$  is defined as the basic conversion rate, which is determined by a product's inherent properties.

Another part of Equation (1) actually describes the chance of products being found in brick-and-mortar stores, where  $\frac{S(I)}{D}$  is the proportion of the demand being satisfied. When the proportion is small, there are still a lot of unsatisfied demands, so the marginal sales brought by one-unit increased inventory is relatively large. When the proportion approaches one, most of the demand is satisfied, and increments in inventory can hardly transform into sales, resulting in small marginal sales. In general, marginal sales that decline as sales increase are described by  $1 - \left( \frac{S(I-1)}{D} \right)^\alpha$ . For some products, customers have a strong willingness to look for stores with stocks. Thus, the chance of these products being discovered is higher, and the decline rate of marginal sales with an increase in sales would be relatively small. We use  $\alpha$  to control the decline rate. When  $\alpha$  approaches infinity, the marginal sales will remain at the degree of the basic conversion rate.

Combining the influence of both products' inherent properties and potential customers on sales, the final increased unit of sales when one unit of inventory increases is determined by  $q \left( 1 - \left( \frac{S(I)}{D} \right)^\alpha \right)$ .

According to different  $q$  and  $\alpha$ , products are divided into three categories: major commodity, common commodity and targeted commodity.

Products that have large  $q$  and  $\alpha$  are defined as major commodities. Major commodities are usually products keeping up with the fashion trend in season. Their targeted customers have high stickiness and enthusiasm and are willing to make an effort to find them.

Products that have large  $q$  and small  $\alpha$  are defined as common commodities. Common commodities are products with basic or classic styles. Their demands are stable, but customers give up finding them when the difficulty of finding them increases. For example, when the brick-and-mortar store the customer enters is out of stock in a common commodity, the possibility is small that the customer will go to another store specifically just to find the commodity.

Products that have the smallest  $q$  are defined as targeted commodities. Targeted commodities are designed for minority groups. Their basic rate of conversion is the lowest. Enterprises produce targeted commodities to improve competitiveness by increasing their product variety.

The total demand of a product is denoted by  $D$ . We use multiple linear regression to estimate demand by using a product's features. The linear model can be written as:

$$D_i = \beta_0 + \beta_1^T X_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where  $n$  is the number of products,  $D_i$  refers to the demand for product  $i$  and  $\beta_1$  specifies the impact of the features on demand.  $X_i$  are features of product  $i$ . Since the TLs model is used to predict the total sales of new products and demand is decided by the product's inherent properties, such as price, factory cost and color, sales are not involved in features  $X_i$  when estimating demand. Finally, the TLs model is described in two layers. The first layer consists of Equation (2), while the second layer consists of Equation (1).

The following theorem states the fundamental properties of the TLs model.

**Theorem 1.** Suppose  $D \geq \alpha \cdot q$  holds. The TLs model has two attributes: (1)  $S(I)$  increases monotonously with  $I$  increasing; (2)  $S(I) \rightarrow D$  when  $I \rightarrow \infty$ .

**Proof of Theorem 1.** We start by proving that  $S(I)$  satisfies  $0 \leq S(I) \leq D$ , for any  $I \in \mathbb{N}^+$  using mathematical induction. First, when  $I = 1$ , from Formula (1) and given conditions we have

$$\begin{cases} S(1) = q, \\ \alpha \geq 1, \\ D \geq \alpha \cdot q, \end{cases}$$

obviously,  $0 \leq S(1) \leq D$  holds.

When  $I \geq 2$ , suppose  $0 \leq S(I) \leq D$  holds, there always exists a  $p \in [0, 1]$  such that  $S(I) = D \cdot p$  holds. From the TLs model, we have

$$\begin{cases} S(I+1) = D \cdot p + q(1 - p^\alpha), \\ S(I+1) \geq 0, \end{cases} \quad (3)$$

Define  $f(p) = D - S(I+1)$ , and bring (3) into  $f(p)$ , then we get

$$f(p) = D \cdot (1 - p) + (p^\alpha - 1) \cdot q,$$

and its derivative function

$$f'(p) = -D + (\alpha \cdot q) \cdot p^{\alpha-1}. \quad (4)$$

Since  $p^{\alpha-1} \leq 1$ ,  $D \geq \alpha \cdot q$ , (4) satisfied the following inequality:

$$f'(p) \leq -D + \alpha \cdot q \leq 0.$$

Note that  $f(p)$  is a decreasing function and  $f(1) = 0$ , we obtain

$$f(p) \geq 0, \text{ for any } p \in [0, 1]. \quad (5)$$

Therefore, according to  $f(p) = D - S(I + 1)$ , (3) and (5),  $0 \leq S(I + 1) \leq D$  holds. Through mathematical induction, we now have

$$0 \leq S(I) \leq D, \text{ for any } I \in \mathbb{N}^+. \quad (6)$$

From Formulas (1) and (6), we have sequence  $\{S(I)\}$  that is monotonically increasing and bounded with respect to  $D$ . According to the monotone bounded convergence theorem, we know that the limit of  $\{S(I)\}$  exists. Suppose  $\lim_{I \rightarrow \infty} S(I) = S^*$ , when  $I \rightarrow \infty$ , from the TLs model we have

$$S^* - S^* = q \left( 1 - \left( \frac{S^*}{D} \right)^\alpha \right),$$

It can be obtained by the above formula that  $S^* = D$ , which completes the proof.  $\square$

### 3.2. Feature Selection

Each product has a great number of category features, and category features are processed by one-hot encoding. This generates hundreds of features to be handled, while most might be unimportant. The model is too complex with all features, and a large number of features lead to the over-fitting problem. To alleviate this problem, a GBDT algorithm is used to select features by sorting the importance of features in descending order.

GBDT is a popular machine learning method constructed by an ensemble of weak CARTs (classification and regression trees). During training, the selected feature splits and their relative influence defined in [28] will be recorded. Let  $m_{org}$  denote the initial number of features, and the importance of the  $j$ -th feature in a single decision tree is expressed as:

$$\hat{f}_j^2(T) = \sum_{t=1}^{L-1} \hat{f}_t^2 \mathbb{I}(v_t = j), \quad j = 1, 2, \dots, m_{org},$$

where  $L - 1$  is the number of non-leaf nodes of tree  $T$ ,  $\mathbb{I}$  is the indicator function,  $v_t$  is the feature associated with node  $t$  and  $\hat{f}_t^2$  is the corresponding reduction of square loss as a result of the split. The global importance of  $j$ -th feature is measured by the average importance of  $j$ -th feature in a single tree. The formula is given by:

$$\hat{f}_j^2 = \frac{1}{H} \sum_{h=1}^H \hat{f}_j^2(T_h), \quad j = 1, 2, \dots, m_{org},$$

where  $H$  is the number of trees,  $T_h$  refers to the  $h$ -th decision tree.

Then the five-fold cross-validation method is applied to select important features based on the greedy strategy. By defining the dataset into the training set and validation set randomly, the method adds features based on the features' importance. Through computing the MAPE of the result, the features that have the smallest MAPE are selected to conduct the following clustering and forecasting operations.

### 3.3. Parameter Estimation

#### 3.3.1. Clustering

Products show heterogeneity with each other and have different  $\alpha$  and  $q$ . It is natural to cluster these products and conduct parameter estimation separately. Each cluster shares the same  $\alpha$ ,  $q$ ,  $\beta_0$  and  $\beta_1$ . Considering that these products usually have both continuous and classified attributes, this paper uses the mixed k-mean algorithm proposed by Ahmad and Dey [29] to cluster. Let  $X_1 = (X_{11}, X_{12}, \dots, X_{1m})^T$  and  $X_2 = (X_{21}, X_{22}, \dots, X_{2m})^T$  denote



the attribute vectors of two products, where the first  $m_r$  attributes are numeric, and the remaining  $m - m_r$  attributes are category. The distance between  $X_1$  and  $X_2$  is as follows:

$$\text{Dist}(X_1, X_2) = \sum_{j=1}^{m_r} (w(X_{1j} - X_{2j}))^2 + \sum_{j=m_r+1}^m (\delta(X_{1j}, X_{2j}))^2,$$

where  $w$  is the significance function for numeric attributes, and  $\delta$  is the distance function for category attributes defined in [29].

If the number of clusters is too small, it could not reflect the difference between products; also, if the number is too big, samples will be over-dispersed. Therefore, the number of clusters is limited to 3 to 5, and we employ the popular measure silhouette coefficient in [30] to determine the best number of clusters. The silhouette coefficient for dataset  $\{X_1, X_2, \dots, X_n\}$  is defined as

$$s = \frac{1}{n} \sum_{i=1}^n \frac{b(X_i) - a(X_i)}{\max\{a(X_i), b(X_i)\}},$$

where  $a(X_i)$  is the average distance of sample  $X_i$  to other samples in  $X_i$ 's cluster,  $b(X_i)$  is the average distance of sample  $X_i$  to all samples in the nearest cluster.

The higher the value of the silhouette coefficient is, the better the clustering effect is. The number of clusters with the largest silhouette coefficient is selected to conduct clustering and establish the TLs model.

### 3.3.2. Optimization Tasks

The parameters in our model are estimated by minimizing the gap between actual sales and forecast results during the training process. The optimization task is defined as:

$$(\alpha_k^*, q_k^*, \beta_{0k}^*, \beta_{1k}^*) = \underset{\alpha_k, q_k, \beta_{0k}, \beta_{1k}}{\text{argmin}} \sum_{i=1}^{N_k} |S_i - \hat{S}_i|, \quad k = 1, 2, \dots, K, \quad (7)$$

where  $\alpha_k \geq 1, 0 \leq q_k \leq 1$ ,  $K$  is the number of clusters,  $N_k$  is the number of products in cluster  $k$ ,  $S_i$  is the actual sales of product  $i$  and  $\hat{S}_i$  is the predicted sales of product  $i$  from the TLs model. Here, in cluster  $k$ ,  $\beta_{0k}$  and  $\beta_{1k}$  is in the first layer for demand estimation and  $\alpha_k, q_k$  is in the second layer for sales prediction.

Actually, the optimization task of this problem is difficult for the TLs model to handle; it is an iterative formula. In order to solve the problem, a genetic algorithm (GA) is applied. GAs compute the optimal parameters iteratively. In each iteration, the parameters are tended to optimization by mutation, crossover and selection. GA is a popular technique for solving complex optimization problems [31,32], and has been widely used in the field of sales forecasting [33,34].

## 4. Numerical Experiments

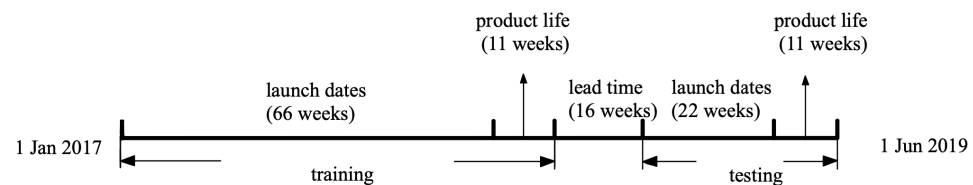
### 4.1. Data Description

The data are provided by our industry partner, a multinational fast fashion enterprise founded in Singapore. The company's products are mainly sold offline and are popular around the world. The company has many stores distributed in different large shopping malls and commercial pedestrian streets in Singapore.

Shoes and belts are selected as the forecast objects, for they are popular products that fashion managers pay much attention to. The original data includes the sales record in 30 stores from 1 January 2017 to 1 June 2019; a total of 126 weeks. After field research, we found that it usually takes 16 weeks for a company to receive the finished products from the factory. It is also known that fashion products have a short product life cycle. Fashion products would be off the shelves or have discounts 11 weeks after being released, which has a negative influence on the profit. Thus, managers of fashion companies only care

about the sales in the first 11 weeks. Under this definition, the product life period is set to 11 weeks, and the lead time for fashion products is set to 16 weeks.

The training set and testing set is divided by time (see Figure 2). Because the forecasting model is about the total sales of new products, we have to make sure that all the products in the training set and the testing set should be released into the market for at least 11 weeks. On the left side of Figure 2, the fashion products' launch time was from 1 January 2017 to 7 April 2018; a total of 66 weeks, with a product life of 11 weeks. Here, the sales records in the first 77 weeks make up the training set, and a lead time of 16 weeks is considered when dividing the testing set. On the right side of Figure 2, the products' launch time was from 14 October 2018 to 16 March 2019, with a product life of 11 weeks. Here, the sales records in the last 33 weeks are defined as the testing set. This ensures that every product in these sets has a complete life cycle.



**Figure 2.** Partition of datasets into training and testing sets.

Three statistics for each category are computed: the size of data, average and standard deviation in sales and inventory. These numerical characteristics are summarized in Table 1. From Table 1, it is noted that sales are close to inventory. This is because fashion products have a short life cycle and companies usually take a short inventory strategy. The difference between sales and inventory is thus generally small, and a serious censored demand problem is caused.

**Table 1.** Description statistics of the data sample in the dataset on shoes and belts.

Category	Dataset	Sample Size	Sales		Inventory	
			Avg	Std	Avg	Std
Shoes	Total set	5944	42.25	35.90	58.18	46.24
	Training set	4456	42.30	33.89	53.68	42.98
	Testing set	1488	42.10	41.35	55.69	54.87
Belts	Total set	868	32.74	21.11	42.68	25.70
	Training set	742	31.20	18.99	40.94	23.10
	Testing set	126	41.87	29.23	52.91	35.96

Notes; the columns 'Avg' and 'Std' are the average and standard deviation for each dataset, correspondingly.

#### 4.2. Preprocessing

The original data is incomplete and has a lot of noise data, so the data are preprocessed first. The steps are as follows:

- (1) Missing value processing. Delete data with more than 50% missing value. For continuous data, fill the missing value with the mean value. For discrete data, treat the missing values as a separate category.
- (2) Feature normalization on continuous data. Through normalization, the unit limit of data is removed, which is convenient for the comparison and weighting of indicators. The min-max normalization scheme is expressed as:

$$Y_{ij} = \frac{X_{ij} - X_{\min,j}}{X_{\max,j} - X_{\min,j}}, \quad i = 1, \dots, n; j = 1, \dots, m_r,$$

where  $X_{\min,j}$  and  $X_{\max,j}$  denote the maximum value and minimum value of continuous feature  $j$ , respectively.



After preprocessing the data, one-hot is conducted to handle categorical features, and the GBDT algorithm is applied before feature selection. Clustering and parameter estimation are both conducted based on the processed dataset after feature selection.

#### 4.3. Comparison Models

Statistical methods have been widely used in previous research on the forecasting problem in the retail industry and often require relatively restrictive assumptions or otherwise will lead to problems [7]. Nevertheless, the research problem of this paper is about the total sales forecasting in the fashion industry, especially for those fashion enterprises that have brick-and-mortar stores, while statistical methods in existing research on prediction in the retail industry usually give periodical results of forecasting rather than total sales [15,17]. Since research on the sales forecasting problem with censored demand is relatively rare, we conduct field research and find that the industry prefers to use traditional machine learning methods when facing the forecasting problem. Therefore, we introduce four more comparison models as follows:

- Linear regression (LR)
- Gradient-boosting decision tree (GBDT)
- Support vector regression (SVR)
- Artificial neural network (ANN)

Table 2 reports the base forecasters and their software tools in this experiment. The data for the benchmarks are processed data after feature selection and clustering. Moreover, inventory is also used as a feature for benchmarks, considering the fairness of comparison. When forecasting, if the result of sales is negative, sales take zero; if it is greater than the inventory, take the quantity of inventory as sales.

All experiments were run in Matlab 2019b and Python 3.8.10. In our computational work, we used a Quad-core Intel Core i5 pc clocked at 2.4 GHz processor, equipped with 16 GB RAM.

**Table 2.** The base forecasters for comparison.

Base Forecaster	Description	Software Tool
LR	LR is a statistical technique that uses several explanatory variables to predict the outcome of a response variable [3].	sklearn 0.22.1, linear_model. LinearRegression
GBDT	GBDT is a machine learning algorithm. It applies boosting to construct an ensemble of weak decision tree learners [28].	lightgbm 3.1.1
SVR	SVR is a neural network forecaster based on statistical learning theory and the principle of structural risk minimization [35].	sklearn 0.22.1, svm.SVR
ANN	ANN is a computing model composed of a large number of nodes connected to each other that give the output by the weight value and the activation function [36].	keras 2.4.3, tensorflow 2.4.1

#### 4.4. Measure of Forecast Accuracy

In this paper, the forecast accuracy is evaluated using the following performance measures: (1) mean absolute error (MAE); (2) mean absolute percentage error (MAPE). The definitions are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |S_i - \hat{S}_i|, \quad (8)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{S_i - \hat{S}_i}{S_i} \right|, \quad (9)$$

where  $n$  is the total number of products, and  $S_i$  and  $\hat{S}_i$  represent the actual and predicted sales of products  $i, i = 1, \dots, n$ .

#### 4.5. Results and Discussion

##### 4.5.1. Results for Parameter Estimation

A total of 20 important features are selected for clustering in the shoe category using the feature selection method in Section 3.3. The runtime of each step is shown in Table A1, and the detailed information on the selected features is given in Table A2 in Appendix A. Then, the silhouette coefficients are computed to determine the best number of clusters. For shoes, the silhouette coefficients of cluster numbers 3 to 5 are 0.2407, 0.1999 and 0.2019. For belts, the silhouette coefficients of cluster numbers 3 to 5 are 0.4240, 0.2938 and 0.3946. As a result, both shoes and belts are divided into three clusters. The statistic characteristics of each cluster are given in Table 3.

**Table 3.** Statistic characteristics of clusters.

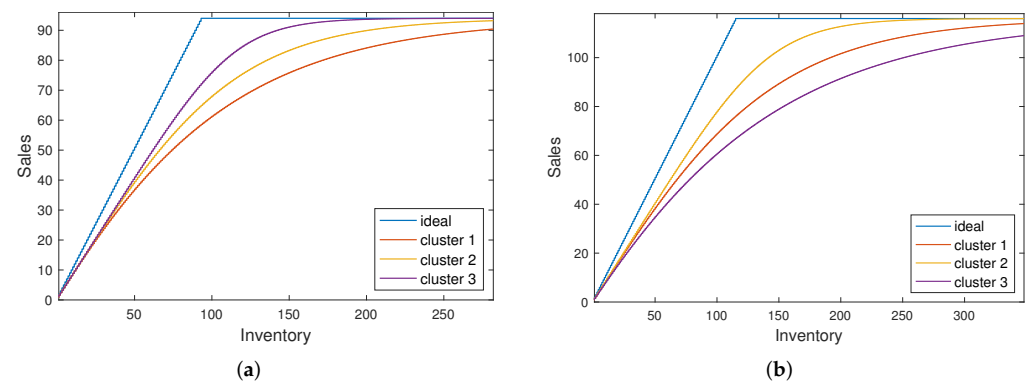
Category	Cluster	Dataset	Sample Size	Sales		Inventory	
				Avg	Std	Avg	Std
Shoes	Cluster 1	Total set	1735	48.25	41.87	62.45	53.61
		Training set	1128	47.97	38.87	60.97	48.99
		Testing set	607	48.75	46.96	65.21	61.22
	Cluster 2	Total set	2102	44.40	36.04	56.77	46.55
		Training set	1863	43.89	32.24	55.81	43.94
		Testing set	239	48.35	47.67	64.28	62.98
	Cluster 3	Total set	2107	35.17	28.53	44.78	36.81
		Training set	1465	35.91	27.80	45.35	34.66
		Testing set	642	33.48	30.09	43.49	41.29
Belts	Cluster 1	Total set	80	35.80	22.64	47.50	28.56
		Training set	73	35.42	22.76	46.51	28.19
		Testing set	7	39.71	22.63	57.86	32.60
	Cluster 2	Total set	484	36.54	24.04	46.75	29.03
		Training set	387	34.40	21.30	44.42	25.63
		Testing set	97	45.06	31.50	56.08	38.59
	Cluster 3	Total set	304	25.90	12.16	34.91	15.50
		Training set	282	25.70	12.10	34.72	15.42
		Testing set	22	28.45	12.98	37.36	16.68

After clustering, GA is used to find the optimal solution for parameters in the TLs model. The results for shoes and belts are listed in Table 4. Each cluster corresponds to a commodity classification in Section 3.1. For shoes, the difference between  $q$  is small, and sales are mainly influenced by  $\alpha$ . In the shoe dataset, cluster 3 has the highest  $\alpha$  with 5.0117 and  $q$  with 0.8065. Products in cluster 3 are recognized as major commodities. Similarly, cluster 1 is the targeted commodities and cluster 2 is the common commodities. For belts, cluster 2 is the major commodity because it has the highest  $q$  and  $\alpha$ . Cluster 1 is the common commodities, and cluster 3 is the targeted commodities. To conclude, the estimation results demonstrate our model well.

**Table 4.** Parameter estimates trained on the training set.

Category	Parameter	Cluster 1	Cluster 2	Cluster 3
Shoes	$q$	0.8159	0.8068	0.8065
	$\alpha$	1.4419	2.3391	5.0117
Belts	$q$	0.7857	0.8000	0.7513
	$\alpha$	1.9861	4.4056	1.3328

Graphs also prove our conclusion about the products category above. Each item in the dataset has its own demand estimation. In order to display the general result of our TLs model, we set the rounding of the average demand of all samples in a corresponding dataset as the demand. The maximum inventory is set to 3 times the predictive value of demand. Figure 3 illustrates the changes in the sales forecast result related to inventory compared with the ideal conversion rate of inventory to sales. Here, Figure 3a,b are shoes and belts datasets whose demand is 94 and 116, respectively. The difference between each curve is significant. The major commodity has the curve that is closest to the ideal state. The common commodity has a slower conversion rate than the major commodity, and targeted commodity is at the bottom of the three curves.



**Figure 3.** Sales forecast related to inventory of the dataset on shoes and belts compared with the ideal conversion rate of sales. (a) Shoes (mean  $D = 94$ ) (The mean  $D$  is the rounding of the average demand of all samples in the corresponding dataset). (b) Belts (mean  $D = 116$ ).

#### 4.5.2. Results for Forecast

Table 5 reports the forecasting performance of all base forecasters. The best performance is highlighted for each setting. In general, our model outperforms other models in most cases. Both shoes and belts show the best performances in cluster 3 over MAE, with, respectively, 4.5365 and 3.8445 from our model. In terms of MAPE, the proposed model achieves 0.1359 in shoe cluster 2 and 0.1365 in belt cluster 3. It is also noted that the TLs model shows the best performance over the whole shoe dataset in terms of MAE and MAPE and the belts dataset for MAPE. The reason might be that the TLs model takes inventory as a significant property of products and makes full use of it. The TLs model delves into the relationship between sales and inventory, while other methods only take inventory as an input feature and a constraint for sales. This is crucial because the existence of censored demand in fashion products makes sales closely related to inventory. As can be seen from the table, LR, GBDT, SVM and ANN did not show significant differences in the results. We note that ANN has a relatively good performance in forecasting belts' sales but performs much more poorly in the shoes' cluster. Especially in the whole belts dataset, ANN has a smaller MAE compared with the TLs model. The forecasting performance of ANN is unstable, and the possible reason is that the training sample is not big enough for ANN to extract abundant information.

To conclude, the TLs model shows the best performance in the experiment in most cases. Since the products in this dataset are typical fashion products with high demand

volatility, the TLs model is likely to give better performance in the fashion industry with censored demand and multiple brick-and-mortar stores.

**Table 5.** Forecasting performance of the five forecasters and one benchmark over two categories in the test data.

Category	Method	Cluster 1		Cluster 2		Cluster 3		Total	
		MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE
Shoes	TLs	<b>0.1531</b>	<b>6.3392</b>	<b>0.1359</b>	<b>6.9151</b>	<b>0.1481</b>	<b>4.5365</b>	<b>0.1482</b>	<b>5.6539</b>
	LR	0.2610	10.2327	0.2861	9.4790	0.2012	5.6482	0.2392	8.1336
	GBDT	0.1648	8.2280	0.1381	7.7411	0.1656	5.3156	0.1609	6.8932
	SVR	0.1631	9.2175	0.1482	10.0306	0.1530	5.8745	0.1564	7.9057
	ANN	0.2134	9.2214	0.2491	10.5849	0.1787	5.6153	0.2042	7.8846
Belts	TLs	0.1468	8.1172	<b>0.1488</b>	<b>6.3018</b>	<b>0.1365</b>	<b>3.8445</b>	<b>0.1466</b>	5.9736
	LR	0.1110	4.8785	0.1559	6.4324	0.1537	4.2487	0.1530	5.9648
	GBDT	0.1763	8.6786	0.1606	8.0955	0.1526	4.6342	0.1602	7.5236
	SVR	0.2134	9.6330	0.1934	11.8215	0.1457	4.8524	0.1862	10.4831
	ANN	<b>0.1087</b>	<b>4.1958</b>	0.1557	6.4796	0.1476	4.0534	0.1516	<b>5.9291</b>

Notes; the column ‘Total’ is the forecasting result of all samples in clusters 1, 2 and 3 in the shoes and belts datasets.

#### 4.5.3. Further Discussion

In this subsection, further conclusion from the model for the fashion managers is provided. The conclusion consists of two parts: measure of product competitiveness and auxiliary index for inventory decision.

Three important indicators are introduced first: average conversion rate  $\gamma_a$ , marginal conversion rate  $\gamma_m$  and demand satisfied proportion  $\gamma_d$ . The expression of each indicator is as follows:

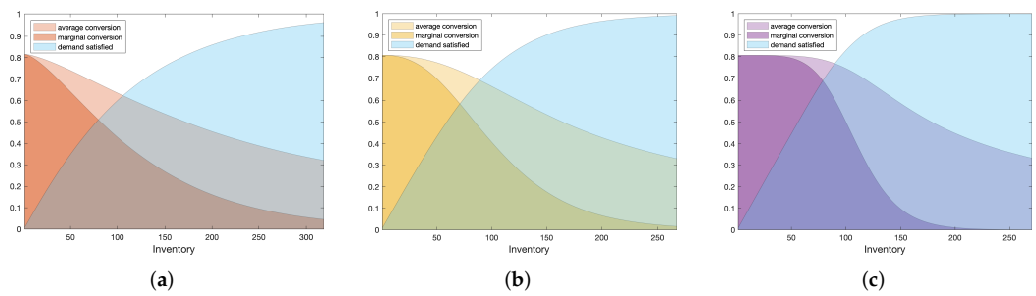
$$\gamma_a(I) = \frac{S(I)}{I}, \quad (10)$$

$$\gamma_m(I) = S(I) - S(I-1), \quad (11)$$

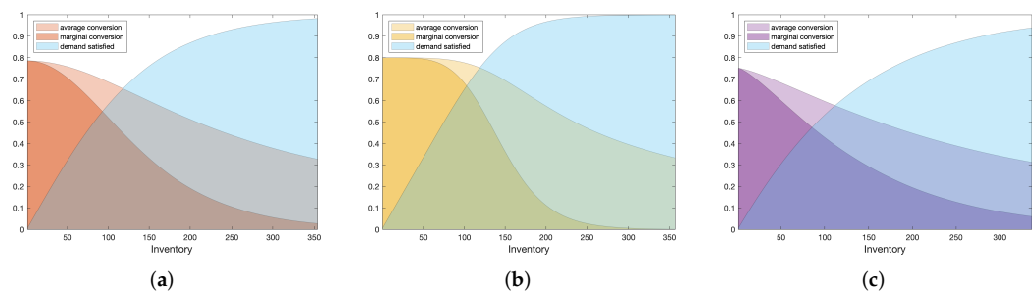
$$\gamma_d(I) = \frac{S(I)}{D}, \quad (12)$$

where  $I \in N^+$ ,  $S(0) = 0$ ,  $\gamma_a$  is the converting ratio of the total inventory to sales, and  $\gamma_m$  is the converting ratio of a unit of increased inventory to sales, also known as the marginal conversion rate.  $\gamma_d$  is defined as the proportion of satisfied demand. All ratios are related to  $q$  and  $\alpha$  in the proposed model. Figures 4 and 5 display the relationship between inventory and three ratios: the average conversion rate, marginal conversion rate and demand-satisfied proportion over the shoes and belts datasets. Figure 4a–c plots the curves of three clusters in the shoe dataset. The demand set uses the same method shown in Figure 3, where the three cluster in the shoe dataset have an average demand of 106, 89 and 90, respectively. The setting of Figure 5 is the same as Figure 4.

In general, the graph shows a similarity based on the product category. In terms of  $\gamma_m$ , major commodities, described in Figures 4c and 5b, have a relatively high  $\gamma_m$  and decline rapidly when the inventory level exceeds the demand. Common commodities and targeted commodities have smooth curves compared with major commodities, as shown in Figures 4a,b and 5a,c. Further, when the inventory level is the same, more demand is satisfied in major commodities than in the other two commodities.



**Figure 4.** The curves of average conversion rate, marginal conversion rate and demand satisfied proportion related to inventory for the shoe dataset by cluster. (a) Cluster 1 (mean  $D = 106$ ). (b) Cluster 2 (mean  $D = 89$ ). (c) Cluster 3 (mean  $D = 90$ ).



**Figure 5.** The curves of average conversion rate, marginal conversion rate and demand satisfied proportion related to inventory for the belts dataset by cluster. (a) Cluster 1 (mean  $D = 118$ ). (b) Cluster 2 (mean  $D = 119$ ). (c) Cluster 3 (mean  $D = 112$ ).

#### (1) Measure of product competitiveness

The average conversion rate when the inventory level is equal to demand is denoted by  $\gamma_{a(D)}$ .  $\gamma_{a(D)}$  is defined as the inventory conversion level of this SKU, and also can be understood as the product competitiveness of this SKU. Through computing,  $\gamma_{a(D)}$  over each cluster in shoes is 0.6232, 0.6908 and 0.7678, respectively. For belts, the ratios are 0.6561, 0.7532 and 0.5771 for clusters 1 to 3. Obviously, cluster 3 in shoes and cluster 2 in belts have the highest product competitiveness, which means that they have the highest percentage of inventory converting into sales, corresponding with the characteristics of a major commodity. Cluster 1 in shoes and cluster 3 in belts has the lowest product competitiveness and is known as a targeted commodity. All the results correspond with the conclusion we give about commodity categories in Section 4.5.1.

#### (2) Inventory decision

The TLs decision  $I^\zeta$  is defined as an auxiliary index for inventory decision:

$$\text{TLs-decision } I^\zeta = \max\{I : \gamma_m(I) \geq \zeta, I \in N^+\}, \quad (13)$$

where  $\zeta$  is a given constant of the marginal conversion rate,  $\zeta \in (0, 1)$ .

Table 6 displays the TLs decision  $I^\zeta$  for each SKU in shoes' cluster 3 when  $\zeta$  is given. Moreover, it gives the actual inventory of each SKU. All SKUs are randomly selected from cluster 3 in the shoe dataset. By combining the data of inventory cost, sales profit and production scale, the enterprise set a benchmark  $\gamma_{m_0}$  for marginal conversion rate. The enterprise then set  $\zeta = \gamma_{m_0}$  in Formula (13) and gets the recommended inventory level. For example, if the company's minimum acceptable, marginal conversion rate is 0.7, it can be seen from Table 6 that the actual inventory level for SKU2 and SKU8 is higher than the TLs-decision  $I^{0.7}$ , which means that the marginal conversion rate for these two SKUs are relatively low and generates revenue in an inefficient way. The company might consider decreasing the inventory level as a better inventory decision. By contrast, the

actual inventory level for the rest of the SKUs is lower than the TLs decision  $I^{0.7}$ . These SKUs might have too little stock, and the company should increase some inventory instead.

**Table 6.** TLs inventory decision for the SKUs in shoes cluster 3.

Item	Inventory	TLs Inventory Decision		
		$\zeta = 0.65$	$\zeta = 0.70$	$\zeta = 0.75$
SKU1	32	67	61	54
SKU2	167	91	83	73
SKU3	43	70	64	56
SKU4	49	100	79	69
SKU5	17	37	33	29
SKU6	52	69	63	55
SKU7	31	88	81	71
SKU8	106	104	95	83
SKU9	48	78	72	62
SKU10	25	93	85	74

## 5. Conclusions, Limitations and Future Research

In this work, a TLs model is proposed for sales forecasting in the fashion industry. To the best of our knowledge, although the sales forecasting model has been explored by a great number of researchers, the TLs model proposed is brand new. The TLs model delves into the relationship between sales demand and inventory and collaborates two parameters to describe the conversion rate of inventory to sales. To solve the proposed model, feature selection and clustering are conducted to process the historical data, and a GA algorithm is used to operate the parameter estimation process. Furthermore, the proposed model is compared with the LR, GBDT, SVR and ANN models on historical data provided by a multinational fashion retailer in Singapore. The experimental result shows that the TLs model outperforms other models in most cases in terms of forecast accuracy. The four traditional methods do not show significant differences when forecasting, and the forecasting performance of ANN is unstable. Moreover, the results give a conclusion that in the shoe dataset, cluster 1 is a targeted commodity and cluster 2 is a common commodity, while cluster 3 is a major commodity. For belts, clusters 1 to 3 are common commodities, major commodities and targeted commodities, respectively.

Moreover, two indicators are built based on our TLs model: the average conversion rate and the marginal conversion rate. By defining the average conversion rate, a measure of product competitiveness is put forward for fashion managers. As for the marginal conversion rate, it is an auxiliary index for deciding on a suitable inventory decision. By setting the acceptable rate based on factors such as inventory cost, sales profit and production scale, the TLs-decision formula provides managers with the recommended inventory level.

Nevertheless, the proposed TLs model shows limitations. The TLs model forecasts the total sales of all stores, which does not consider the forecasting problem of sales in each brick-and-mortar store and periodical sales. Here, a top-down approach could be applied for further research. Moreover, the TLs model could consider alternatives for techniques in each step of the proposed algorithm to achieve better performance. A possible improvement for further work is to find an alternative method to replace the linear regression used in the first layer. Moreover, more effort could be put into extending the TLs model with more realistic constraints and mining deep information about products, such as the impact of feature crosses on sales.



**Author Contributions:** Investigation, W.L.; Methodology, D.C.; Supervision, F.L. and K.Z.; Writing original draft, K.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China of grant number 71801124.

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

## Appendix A

The runtimes of the GBDT feature selection and mixed k-mean algorithm in the shoe dataset were 488.3135 and 224.0785 s, respectively, while in the belts dataset, the runtimes were 76.2131 and 1.6885 s. Table A1 describes the training time of our TLs model and benchmark models in different clusters and in the total dataset. It is noted that the TLs model and ANN model have the longest runtime. This might be because the TLs model is an iterative formula, and ANN is relatively more complex compared with the other three benchmark models.

**Table A1.** Runtimes of the TLs model and benchmark models.

Category	Method	Cluster 1	Cluster 2	Cluster 3	Total
Shoes	TLs	34.2324	52.8239	34.1007	121.1570
	LR	0.0227	0.0199	0.0192	0.0618
	GBDT	0.5384	0.6462	0.8382	2.0228
	SVR	0.2913	0.3303	0.3230	0.9446
	ANN	36.1684	57.1717	47.2212	140.5613
Belts	TLs	1.8975	9.0857	5.1129	16.0961
	LR	0.0103	0.0110	0.0082	0.0295
	GBDT	0.1098	0.3187	0.1454	0.5739
	SVR	0.0073	0.0234	0.0150	0.0457
	ANN	7.2949	16.4111	13.2441	36.9501

Notes; the column 'Total' is the sum of training time of three clusters.

**Table A2.** Result of feature selection.

Category	Name	Attribute Type	Meaning
Shoes	price	numerical	price for sales
	factory_cost	numerical	cost of product
	size	categorical	35, 36, etc.
	heel_height_range	categorical	high, median, etc.
	toe_box	categorical	round, pointed, etc.
	color_tone	categorical	dual_tone, mono-tone, etc.
	type_ck	categorical	fashion, basic, etc.
	article_name	categorical	flats, heels, etc.
	launch	categorical	lunch time
	supplier_code	categorical	supplier
	color_group	categorical	dark, mid, light, etc.
	out_sole	categorical	synthetic, rubber, etc.
	heel_material	categorical	wrapped, chromed, etc.
	season	categorical	spring, winter, etc.
	category_name	categorical	slip on, ankle strap, etc.
	color	categorical	black, blue, etc.
	heel_type	categorical	heel, wedge, etc.
Belts	sub_class_name	categorical	open toe, covered, etc.
	design_name	categorical	jubo, yilin, etc.
	country_of_origin	categorical	China, Malaysia
	price	numerical	price for sales
	size	categorical	35, 36, etc.
	color	categorical	black, blue, etc.

## References

- Seaman, B. Considerations of a retail forecasting practitioner. *Int. J. Forecast.* **2018**, *34*, 822–829. [\[CrossRef\]](#)
- Zhang, G.P. Neural networks for retail sales forecasting. In *Encyclopedia of Information Science and Technology*; IGI Global: Hershey, PA, USA, 2005; pp. 2100–2104.
- Fildes, R.; Ma, S.; Kolassa, S. Retail forecasting: Research and practice. *Int. J. Forecast.* **2019**. [\[CrossRef\]](#)
- Aye, G.C.; Balcilar, M.; Gupta, R.; Majumdar, A. Forecasting aggregate retail sales: The case of south africa. *Int. J. Prod. Econ.* **2015**, *160*, 66–79. [\[CrossRef\]](#)
- Macchion, L.; Moretto, A.; Caniato, F.; Caridi, M.; Danese, P.; Vinelli, A. Production and supply network strategies within the fashion industry. *Int. J. Prod. Econ.* **2015**, *163*, 173–188. [\[CrossRef\]](#)
- Musalem, A.; Olivares, M.; Bradlow, E.T.; Terwiesch, C.; Corsten, D. Structural estimation of the effect of out-of-stocks. *Manag. Sci.* **2010**, *56*, 1180–1197. [\[CrossRef\]](#)
- Vulcano, G.; Van Ryzin, G.; Ratliff, R. Estimating primary demand for substitutable products from sales transaction data. *Oper. Res.* **2012**, *60*, 313–334. [\[CrossRef\]](#)
- Au, K.F.; Choi, T.M.; Yu, Y. Fashion retail forecasting by evolutionary neural networks. *Int. J. Prod. Econ.* **2008**, *114*, 615–630. [\[CrossRef\]](#)
- Ni, Y.; Fan, F. A two-stage dynamic sales forecasting model for the fashion retail. *Expert Syst. Appl.* **2011**, *38*, 1529–1536. [\[CrossRef\]](#)
- Sun, Z.L.; Choi, T.M.; Au, K.F.; Yu, Y. Sales forecasting using extreme learning machine with applications in fashion retailing. *Decis. Support Syst.* **2008**, *46*, 411–419. [\[CrossRef\]](#)
- Choi, T.M.; Sethi, S. Innovative quick response programs: A review. *Int. J. Prod. Econ.* **2010**, *127*, 1–12. [\[CrossRef\]](#)
- Ding, X.; Puterman, M.L.; Bisi, A. The censored newsvendor and the optimal acquisition of information. *Oper. Res.* **2002**, *50*, 517–527. [\[CrossRef\]](#)
- Besbes, O.; Muharremoglu, A. On implications of demand censoring in the newsvendor problem. *Manag. Sci.* **2013**, *59*, 1407–1424. [\[CrossRef\]](#)
- Papalexopoulos, A.D.; Hesterberg, T.C. A regression-based approach to short-term system load forecasting. *IEEE Trans. Power Syst.* **1990**, *5*, 1535–1547. [\[CrossRef\]](#)
- Babai, M.Z.; Ali, M.M.; Boylan, J.E.; Syntetos, A.A. Forecasting and inventory performance in a two-stage supply chain with arima (0, 1, 1) demand: Theory and empirical analysis. *Int. J. Prod. Econ.* **2013**, *143*, 463–471. [\[CrossRef\]](#)
- Green, M.; Harrison, P. Fashion forecasting for a mail order company using a bayesian approach. *J. Oper. Res. Soc.* **1973**, *24*, 193–205. [\[CrossRef\]](#)
- Lee, J.; Boatwright, P.; Kamakura, W.A. A bayesian model for prelaunch sales forecasting of recorded music. *Manag. Sci.* **2003**, *49*, 179–196. [\[CrossRef\]](#)
- Chintagunta, P.K.; Nair, H.S. Structural workshop paper—Discrete choice models of consumer demand in marketing. *Mark. Sci.* **2011**, *30*, 977–996. [\[CrossRef\]](#)
- Frank, C.; Garg, A.; Sztandera, L.; Raheja, A. Forecasting women's apparel sales using mathematical modeling. *Int. J. Cloth. Sci. Technol.* **2003**, *15*, 107–125. [\[CrossRef\]](#)
- Chawla, A.; Singh, A.; Lamba, A.; Gangwani, N.; Soni, U. Demand forecasting using artificial neural networks—A case study of american retail corporation. In *Applications of Artificial Intelligence Techniques in Engineering*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 79–89.
- Yu, Y.; Choi, T.M.; Hui, C.L. An intelligent fast sales forecasting model for fashion products. *Expert Syst. Appl.* **2011**, *38*, 7373–7379. [\[CrossRef\]](#)
- Wong, W.K.; Guo, Z. 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.* **2010**, *128*, 614–624. [\[CrossRef\]](#)
- Choi, T.M.; Hui, C.L.; Liu, N.; Ng, S.F.; Yu, Y. Fast fashion sales forecasting with limited data and time. *Decis. Support Syst.* **2014**, *59*, 84–92. [\[CrossRef\]](#)
- Alon, I.; Qi, M.; Sadowski, R.J. Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *J. Retail. Consum. Serv.* **2001**, *8*, 147–156. [\[CrossRef\]](#)
- Chu, C.W.; Zhang, G.P. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *Int. J. Prod. Econ.* **2003**, *86*, 217–231. [\[CrossRef\]](#)
- Nahmias, S.; Smith, S.A. Optimizing inventory levels in a two-echelon retailer system with partial lost sales. *Manag. Sci.* **1994**, *40*, 582–596. [\[CrossRef\]](#)
- Heese, H.S.; Swaminathan, J.M. Inventory and sales effort management under unobservable lost sales. *Eur. J. Oper. Res.* **2010**, *207*, 1263–1268. [\[CrossRef\]](#)
- Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
- Ahmad, A.; Dey, L. A k-mean clustering algorithm for mixed numeric and categorical data. *Data Knowl. Eng.* **2007**, *63*, 503–527. [\[CrossRef\]](#)
- Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [\[CrossRef\]](#)
- Hadavandi, E.; Shavandi, H.; Ghanbari, A. An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: Case study of printed circuit board. *Expert Syst. Appl.* **2011**, *38*, 9392–9399. [\[CrossRef\]](#)

32. Konar, A. *Computational Intelligence: Principles, Techniques and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
33. Jiménez, F.; Sánchez, G.; García, J.M.; Sciavicco, G.; Miralles, L. Multi-objective evolutionary feature selection for online sales forecasting. *Neurocomputing* **2017**, *234*, 75–92. [[CrossRef](#)]
34. Lau, R.Y.K.; Zhang, W.; Xu, W. Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Prod. Oper. Manag.* **2018**, *27*, 1775–1794. [[CrossRef](#)]
35. Lu, C.J. Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing* **2014**, *128*, 491–499. [[CrossRef](#)]
36. Ainscough, T.L.; Aronson, J.E. An empirical investigation and comparison of neural networks and regression for scanner data analysis. *J. Retail. Consum. Serv.* **1999**, *6*, 205–217. [[CrossRef](#)]