Advances in Industrial Engineering, June 2026, 60(1): 165-184
DOI: 10.22059/aie.2026.406483.1962

RESEARCH PAPER

Development of an Inventory Management Strategy Model (Two-Products) Based on Demand Predicting in Digital Supply Chain Networks by Combining Data Analysis Methods

Ehsan Mardan[1 *] and Ali Qorbani[2]

[1]*Assistant Professor, Department of Industrial Engineering, Semnan University, Semnan, Iran.*
[2]*M.Sc., Department of Industrial Engineering, Semnan University, Semnan, Iran.*

Received: 17 November 2025, Revised: 27 January 2026, Accepted: 10 February 2026

Abstract

Recording and classifying commerce data electronically within the new age of information and development has many benefits for sellers and consumers in the online supply chain. We can predict customer purchase behavior patterns by analyzing this classified data. In recent years, digital stores have received additional attention due to their advantages. On the other hand, the performance of these stores is directly associated with the performance of their suppliers. Hence, supply chain management is essential during this sales system. In this investigation, the classification methods (WFRM), demand prediction analysis (Binary Logistic Regression), data classification (Discriminant Analysis), time series analysis (Trend Analysis), and mathematical modeling have been used to select suppliers to order led to the development of a management strategy to prevent shortages and reduce the average inventory and costs of sending and producing for suppliers. It ultimately covers a 29% error in sales data for assigning suppliers to orders.

Keywords:
Customer Clustering,
Data Mining,
logistic Regression,
Online Supply Chain,
Demand Forecast.

1. Introduction

The recent eruption of business monitoring and industrial automation and the genesis of digital technologies has created a massive volume of semi-structured or unstructured data. Thus, the term “Big Data” has been extensively discussed. Extracted big datasets can be integrated and converted into useful knowledge using large-scale information technologies to benefit the business in this competitive market. Organizations with big data technologies in data analytics significantly improve profitability and efficiency (Akhtar et al., 2019). Big numerical data can affect the supply chain theory. Because with a large amount of smaller data, we can describe all aspects of the supply chain, such as website traffic, customer sentiment on social media, different service levels or the development of contractual relationships, and material flows in a supply chain (Côrte-Real et al., 2019). Electronic commerce (e-commerce) in the supply chain is also undeniable (Pawar & Driva, 2000).

In general, there are many types of e-commerce. However, e-commerce is classified into three main types, including (i) company-to-company (ii) company-to-consumer, and (iii) consumer-to-consumer. A recent report shows that the predicted company-to-company e-

* Corresponding author: (Ehsan Mardan)
Email: Ehsan_Mardan@Semnan.ac.ir

commerce worldwide will reach more than $ 6.7 trillion by 2020, which is almost twice the e-commerce of company-to-consumer. E-commerce is a common range from many viewpoints and it is necessary to consider all electronic transactions between the organization and a third party (Chaffey et al., 2019). This viewpoint includes online commerce, the creation of digital value, virtual markets and stores, and new distribution intermediates (Frost & Strauss, 2016). When these viewpoints integrate other business processes into the big data environment, analytical intuition is needed to improve performance and create business value. The results can provide ideas for marketing managers on how to implement successful business strategies (Constantinides & Fountain, 2008).

In the research community, not enough attention has been paid to the sales factor of the product. Previous research shows that many key factors play a role in achieving efficiency in e-commerce, including (i) process efficiency (ii) demand compression (iii) information sharing, and (iv) risk management. When managers plan a marketing strategy to sell a product, they consider four factors in the marketing model, including (i) product (ii) cost (iii) location, and (iv) advertising (Kumar et al., 2020). One of the business strategies in e-commerce is customer segmentation. Through segmentation, predicting customer behaviors and patterns can be improved. Hence, companies can create appropriate marketing strategies according to different groups of customers. According to The Data Warehousing Institute (TDWI) survey, 41% of organizations reported that a customer-based segmentation guarantees the implementation of some types of big data analytics (Russom, 2011). The importance of customer segmentation and its positive impact has been explored by many articles since 2012 (H. Chen et al., 2012; Chen et al., 2017; Liu et al., 2015; Peker et al., 2017; Wixom et al., 2013; Wong & Wei, 2018).

As mentioned, by segmenting customers we can predict their buying behaviors. for this case we need integrated information of customers' purchases which can be relied upon so that the prediction methods and data analysis be effective. According to digital supplies chain, it is easy to record such effective information. In fact, in this research, using data mining methods and information obtained from an online store, we seek to identify and predict the customers' purchases of the considered products. But the reason for this, is to be able to determine a management strategy to reduce the storage costs in the digital supplies chain so that we can reduce the average inventory in warehouses. But how can this happen? By knowing the fact that when the customer will buy, what product, how much and with what probability, the inventory can be controlled in such a way that we neither suffer from a shortage nor increase the level of inventory. One of the most fundamental issues in the selection and optimization of inventory methods is the analysis of customer behavior and, consequently, the appropriate forecasting of their demand. It is quite clear that the more accurately we can predict customer demand based on historical data, the more accurately the manufacturer can determine the inventory status and estimate it in the future.

This research presents a comprehensive data-driven framework to address these challenges. The study begins with a Literature Review section that examines the existing research on inventory optimization and supplier allocation in the context of digital commerce. then outlines the specific challenges faced by online retailers and their suppliers in managing inventory effectively. The core of the paper is the Problem Solving Method section, which details the innovative techniques employed in this study. These include the use of WRFM (Weighted Recency, Frequency, Monetary) customer segmentation, binary logistic regression for demand prediction, and a mathematical optimization model for optimal supplier assignment. The findings and implications of this research are discussed in the Managerial Insights section, highlighting the practical benefits for online retailers and their supply chain partners. Finally, the Conclusion and Future Research section summarizes the key contributions of this work and outlines potential avenues for further exploration.

Finally in this research in part No. 2 the literature review is discussed; In part No. 3 the problem is explained; In part No. 4 the method of solving the problem is stated, which is divided into smaller parts for the convenience of expressing the solution method. In part No. 5 the numerical results of the problem solution is expressed and in the end in part No. 6 conclusions and strategies resulting from solving the problem are stated.

2. Literature Review

McCarty & Hastak, (2007) segmented the market using RFM (Recency, Frequency, and Monetary value) and logistic regression. Kim & Ahn, (2008) used genetic algorithms to segment the online market. Farajian & Mohammadi, (2010) proposed a combined clustering and dependency law method for analyzing customers behavior. Brito et al., (2015) used clustering and discovery of subgroups to segment customers in the custom fashion industry. Peker et al., (2017) proposed a new model including length, recency, frequency, monetary, and periodicity to segment customers in the retail industry. Tavakoli et al., (2018) used the combination of recency, frequency, monetary value, and k-means as clustering techniques to analyze customers patterns. Wong & Wei, (2018) proposed a combined method for analyzing the customers online purchase behavior. Dutta et al., (2015) and Tsiptsis & Chorianopoulos, (2011) have summarized the applications of customers segmentation.

According to experiences, different customers do not have the same demands. Hence, companies need to identify the specific needs of their customers and provide suitable products to satisfy different groups of customers. At the big level, market segmentation means that a continuous iterative trend from inspection and grouping of a potential factor is formed, such as identifying real buyers with similar product needs and segmenting them into groups that can then be targeted with an appropriate marketing combination, so that facilitates the goals of both sides (Mitchell & Wilson, 1998). Thus, it is an effective strategic marketing tool for marketers (Weinstein, 2006). The aim of market segmentation for companies is to reduce operating costs by allocating resources and eliminating additional activities (Dibb & Wensley, 2002). Customers classification based on behavioral classification is formed from market classification. In this case, the focus is on customers behavior and functional features. This strategy can be defined as dividing the customers club into distinct groups and compatible backend with similar features (Güçdemir & Selim, 2015). The aim of customers division is to study their purchase behavior and the determining factor of customers decisions to develop an efficient marketing strategy (Wu & Chou, 2011). In the past, the segmentation process, segmentation criteria, logic, analytical methods, and division planning have been the main topics of research in the marketing literature (Cheng et al., 2018; Ming-Chih et al., 2011).

In their scholarly investigation, Anitha and Patil (2022) deliberated upon the application of business intelligence in the detection of potential clientele by furnishing pertinent and punctual information to commercial divisions within the realm of retail. Employing the K-Means algorithm, they scrutinized present-time trading and retail data compilations. Their examination is grounded on the RFM model, and implements the tenets of data partitioning through the utilization of the K-Means algorithm. The primary objective of customer segmentation lies in ascertaining the appropriate approaches to adopt for each customer category with the aim of maximizing the profitability of each customer for the business. In their research, Tabianan et al. (2022) have specifically directed their attention towards the behavioral aspect of customers. Hence, the clustering algorithm will be employed to analyze users and determine their purchasing behavior within the e-commerce system. To process the gathered data and segment customers, a learning algorithm called K-Means clustering has been utilized.

There are many techniques for analyzing data to divide customers and identify hidden factors. There are eight major types of data analysis techniques, including extension, continuity, classification, clustering, prediction, sequence analysis, and time series, except knowledge analysis (Cheng et al., 2018). This data analysis technique can help identify the customer, attract the customer, develop the customer, and retain the customer (Ngai et al., 2009). One of the commonly proposed techniques in customer segmentation is the RFM method. This model is used to distinguish customers with the three mentioned features: recency, frequency, and monetary value. To use this model in such a way, three features are defined. R is the distance between the last customer purchase and the next purchase. It indicates customer behavior. However, they are based on the customer perspective and do not consider the product perspective (Heldt et al., 2021).

When the time interval is shorter, R becomes larger. Frequency of purchase (frequency), F, is the number of transactions in a particular period. For example, twice a year, twice a season, or twice a month. If the number of purchases is greater, the frequency will also increase. The monetary value of purchase, M, is the amount of money that the customer pays to the seller. If this amount is higher, the monetary value will also increase (Wang et al., 2020). Khajvand and Tarokh (2011) used a weighted RFM model to estimate the customer future values. (D. Chen et al., 2012) analyzed customers in online retail by combining the RFM model and the k-means algorithm. Cheng and Chen (2009) developed a more efficient customer division method by combining RFM analysis, k-means algorithm, and Reed-Solomon (RS) error correction theory. Previous studies show that the mentioned classifications only predict customers' behavior and plan to respond to it. Prediction is the basis of planning. It means that for good planning, it is necessary to have a good prediction of what is needed to prepare a company, department, and their environments, so that we can achieve future specific developments or intermediate improvements in them (Gallego-García & García-García, 2020). Gallego-García and García-García (2020) demonstrated this well in predicting demand and delivering the product to the market on time to assist supply chain planning managers. In their research, Wu et al. (2021) employ an enhanced RFM model in order to extract distinct user attributes and subsequently employ the K-means++ clustering algorithm to achieve user classification. Chen et al. (2023) extended an RFM framework by considering the probability of purchasing ancillary services during a travel (P). They investigated four passenger groups based on the RFM-P model using unique sales data provided by Chinese airlines. Also, multi-criteria decision-making approaches have also been used in the issues of stores and digital supply chains. In their research, Beheshtinia et al. (2022) presented an approach to evaluate and rank suppliers in digital stores using a combination of two multi-criteria decision-making techniques, namely Analytic Hierarchy Process and TOPKOR.

Prediction parameters in the supply chain is of great importance. In their research, Sadeghi (2019) used an economic production quantity (EPQ) model with a periodic order quantity (POQ) policy to respond to customer demand. Abbaspour Ghadim Bonab and yousefi nejad attari (2022) in their study extracted the system information using a machine learning algorithm and presented it as a transition matrix to the Markov decision process to determine the future states of the critical spare parts inventory system. In the method presented by them, the Adaptive Neural Fuzzy Inference System (ANFIS) is used to train the model and the Markov chain is used to predict the future states of the inventory system. The following table summarizes research related to the field of customer behavior analysis based on data science techniques.

Table 1. Summary of research in customer behavior analysis with data science tools

Authors	Year	Concept	Methods
McCarty & Hastak	2007	Market segmentation	RFM/ Logistic regression
Kim & Ahn	2008	Online market segmentation	Genetic algorithm
Farajian & Mohammadi	2010	Customer behavior analysis	Combined clustering and dependency law method

Tavakoli et al.	2018	customers patterns analysis	Combination of recency, frequency, monetary value, and k-means as clustering techniques
Heldt et al.	2021	Customer segmentation	RFM
Anitha and Patil	2022	commercial divisions within the realm of retail	K-means algorithm
Tabianan et al.	2022	behavioral aspect of customers	Clustering algorithm
Chen et al.	2023	Customer behavior analysis	RFM-P

As it is known, one of the costs incurred on the supply chain is the cost of keeping the products in the warehouse. From an economic point of view, the higher the inventory level of the warehouse, the higher the amount of liquidity in the supply chain circulation, which is a negative factor on the uneconomical of the supply chain planning. Therefore, as stated earlier, we will forecast customer demand by categorizing customers and data mining methods to predict when and how much product will be purchased. In the literature review, many methods have been used to categorize and forecast demand, but not to discover the purchasing pattern of customers to control warehouse inventory. Therefore, in this research, we have used a combination of classification methods, demand prediction such as logistic regression and time series analysis, each of which has been used in the literature review, but not in a combined way.

3. Problem Description

The present study aims to adopt an effective management strategy decision for more efficient control of digital stores. The mechanism of digital stores is to respond to customer demand in order to form a supply chain. In this way, a digital store contract with many suppliers plays the role of marketing and selling the products of these suppliers. This makes it even more important to describe a successful supply chain management strategy for a digital store. Hence, a digital store plays a very important role in the sales market with multiple suppliers. As a result, predicting customer behavior becomes more important than before (Gallego-García & García-García, 2020).

Our aim in this study is to divide customers to predict their purchase behavior by combining the WRFM method and distinction analysis in the digital market to estimate the amount of demand for digital store suppliers in different periods by predicting customer behavior through binary logistic regression (two products). This research achieves the above goal by providing combined data analysis and mathematical modeling methods. On the other hand, this study reduces maintenance costs, operating costs (transportation), and the time of arrival of commodities to customers by correctly allocating orders to suppliers and estimating the amount of demand based on customer behavior in different periods based on the location of order.

Many studies have shown that a combined method, which combines many analytical approaches, can improve the effectiveness of customer segmentation (Tsai & Lu, 2009). Thus, a combination of mathematical modeling methods and data analysis methods with a completely applied approach has been used in this research. The research model is shown in Figure 1. In the first step, information was collected on the sale of a laptop and accessories store in Tehran, which had many suppliers. Then, the information about the sale of laptops, external hard drives, and internal hard drives were selected separately from the suppliers of this digital store for analysis. In the second step and WRFM analysis, customers were divided into three categories: loyal customers, regular customers, and inactive customers. In this step, distinction analysis was used to confirm the classifications. In the third step, the logistic regression model was designed, and it was predicted what kind of products each type of customer would order. In the fourth step, conditional probabilities were presented based on logistic regression results. The data trend was analyzed using time series techniques in the fifth step. Then, based on the number of sales predicted in different periods and having conditional probabilities, the number of sales in different periods was obtained separately for the customer type-product type. In the sixth step, a mathematical model for assigning orders (location of order) to suppliers was developed using primary raw data information, and the initial data error was calculated. Finally, the results of supplier allocation to orders, expressed by the mathematical model, were analyzed and combined with customer classifications methods, logistic regression, conditional probabilities and sales time series, which predicted customer's purchases. Hence, the supply chain management strategy can be expressed. Supply chain management strategy in this research means being aware of the demand constraints in different places, how to assign suppliers to orders with the lowest delivery cost and the shortest delivery time (proximity of the supplier to the customer's order) so that by knowing the amount of demand in different periods, we can reduce the stock inventory for each supplier to minimize maintenance costs, but do not face shortages. In general, this research can be divided into three phases, including data analysis, development of the mathematical model, and explanation of supply chain management strategy.

The collected data includes 2598 orders which were ordered by 255 customers. Each of these orders came from a specific location in 30 different provinces. Each order has purchased one of the laptops or hard drive products. Finally, 22 suppliers are responsible for these orders, and all locations were specified. In general, each order includes variables such as customer code, type of ordered product, order date, intended discount, order purchase volume, order location, supplier code, and allocated supplier location.

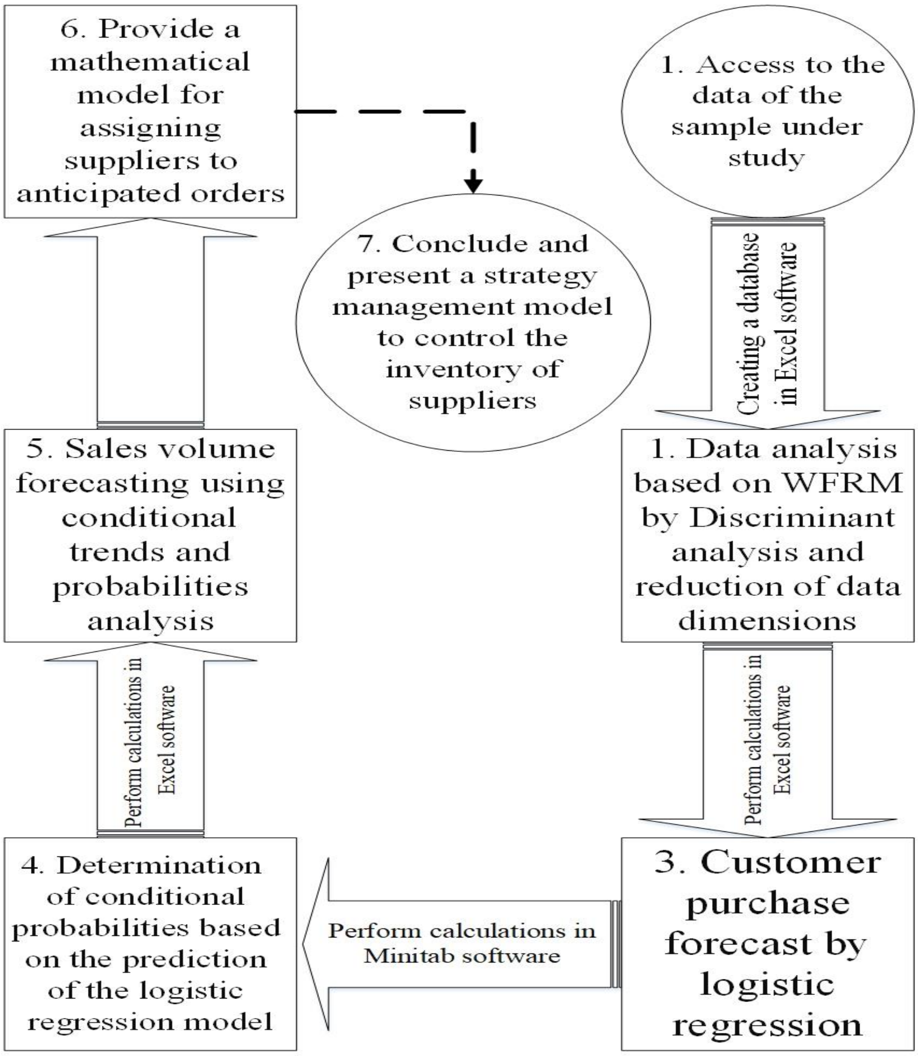

Figure 1. Research model

It is important to note that there are many tools available for analyzing and predicting customer behavior. In this study, we used classification tools to analyze customer behavior and time series techniques to predict customer behavior based on historical data.

4. Mathematical Formulatin and Problem Solving Method

Due to the large size of the problem, such as high orders, a large number of suppliers, and the large number of locations where an order is likely to occur, the mathematical model was designed to determine only which commodities from each province should be ordered and which supplier is responsible for reducing costs and delivery time to the customer. Since the purpose of this problem is to represent a strategy, we intend to compare the performance of the presented strategy with what has happened in practice. Developing this model is to respond to any order given from any province by the nearest supplier. On the other hand, this model also shows the error value of the collected data.

Table 2. Mathematical model indices

Description	**Index**
product type	i
The province from which it is ordered	j
Number assigned to suppliers	k
Set of products	$P = p1 \cup p2\ , i \in P$
Indicate the order of product 1	$p1 = 1$
Indicate the order of product 2	$p2 = 2$
Set of suppliers	$K = k1 \cup k2\ , k \in K$
Set of order points from product 1	$k1 = \{2,3,6,9,10,11,14,16,20,21\}$
Set of order points from product 2	$k2 = \{1,2,6,8,10,11,14,15,16,20,22\}$
Set of order points	$n = \{1, 2, \dots\ , 30\}\ , j \in n$

Table 3. Sets, Parameters and Decision Variables

Description	**Sets**
product type	i
province	j
supplier	k
Description	**Parameters**
The cost of sending product i, which is ordered from province j, and this delivery is done by supplier k.	C_{ijk}
Maximum available capacity of products i, which is available in the supplier warehouse k.	p_{ik}
Minimum product to be sold to supplier k.	l_k
Minimum need for product i, that we do not face shortages.	M_i
Description	**Decision Variables**
binary decision variable indicating whether supplier k is selected to supply product i from province j.	X_{ijk}
the quantity of product i supplied from province j by supplier k	y_{ijk}

The location of orders is also divided according to the province. This means that if the order is from one of the cities, Tehran province, that order will be considered for Tehran province, it is as follows table 4. The values of the parameters expressed in Table 3 are extracted from the collected data. Between 2,598 orders per year, each period allocates some of these orders. Hence, these parameters are extracted separately and based on suppliers in each period. Therefore, the mathematical model is as follows:

$$Min\ Z = \sum_{i}^{P}\sum_{j}^{n}\sum_{k}^{K} C_{ijk} \times y_{ijk} \tag{1}$$

$$\sum_{i}^{p}\sum_{j}^{n}\sum_{k}^{k} X_{ijk} = 1 \tag{2}$$

$$\sum_{i}^{p}\sum_{j}^{n}\sum_{k}^{k} y_{ijk} \leq X_{ijk} \times p_{ik} \tag{3}$$

$$\sum_{i}^{p}\sum_{j}^{n}\sum_{k}^{k} y_{ijk} \leq X_{ijk} \times p_{ik} \tag{4}$$

$$\sum_{i}^{p}\sum_{j}^{n}\sum_{k}^{k} y_{ijk} \geq\ l_k \tag{5}$$

$$\sum_{i}^{p}\sum_{j}^{n}\sum_{k}^{k} y_{ijk} \geq\ M_i \tag{6}$$

Equation (1) shows the value of the objective function of the mathematical model, which is of the cost type. Equation (2) ensures that for an order of product type 1 and is from province No. 1 (Tehran), only one of the 22 suppliers is assigned to it. Equation (3,4) shows that if product “i” is ordered from province “j” and supplier “k” is selected to respond to this order, how much product is needed to supply this order. Here, because some suppliers are single-product, some counts are not done for the indices. Equation (4) provides the suppliers who must have the minimum sales to continue working together. Equation (5) ensures that both products, codes 1 and 2, have been responded fully.

To solve this problem, First, the data was collected and then the customers were classified. The WRFM method was used for this aim. Customers were divided into three classes by Excel functions, and all customers received scores. Customers were divided into regular, loyal, and inactive customers based on this strategy. For each of the FRM factors, respectively, the number of annual purchases by each customer (F), the time interval between two customer purchases and the sum of these times (R), the total volume of purchases per year by each customer (M) were separately calculated in Excel software and then became scaleless (Normalizing Scores). Finally, the score of each customer is determined. We can classify based on each customer score (the distance between the lowest and highest scores is divided by three, and three classes are created). Based on the results, it is assumed that customers who have a minimum purchase (M) of 50,000 currency units, a minimum number of purchases (F) 20 times a year, and a maximum time interval (R) of 18 days, are considered loyal customers. Customers who have a minimum purchase of 35,000 currency units and a maximum purchase of 49,999, a minimum number of purchases of 15 times a year and have a purchase in maximum every 30 days are regular customers, and the rest of the customers who have different conditions are considered as inactive customers. The Equation for calculating customers scores is as follows equation (7).

$$Total\ point = ((\ Factor\ M\ \times\ W_M) + (Factor\ F\ \times\ W_F) - (Factor\ R\ \times\ W_R))/6 \tag{7}$$

Factor M is customer normalized score of purchase volume. As well as factor F is customer normalized score of number of annual purchases. Too, Factor R is customer normalized score of the time interval of purchase. So the assumed weights mentioned in Equation 1 are as follows.

$$\mathrm{W_M} = 3, \mathrm{W_F} = 1, \mathrm{W_R} = 2$$

After determining the types of customers, distinction analysis was used to confirm the mentioned method. These calculations were performed by MINITAB software. The results show that the linear distinction analysis correctly classified 226 out of 255 customers. Thus, a correct allocation rate of 88.8% was obtained based on the distinction function, which confirms the accuracy of the calculations by the WFRM method. The incorrectly assigned items were corrected by distinction analysis to continue the calculations. Distinction analysis is a data science tool for categorizing customers with the aim of determining customers who are behaviorally similar, so that customers within each group have the most behavioral similarity, based on which a more accurate prediction model can be provided regarding their demand.

Since the control variable data has no normal distribution and is discrete, the logistic regression method was used to predict the type of purchase of commodities by the customer. The response variable (the purchased product type) has two states. These two states indicate the type of commodity. Therefore, it is not a sequential variable and is considered a binary variable. Hence, a binary logistic regression was used to predict the purchase probability for each product type. According to the calculations in MINITAB software, the average probability for purchasing product 1 was 0.37 and for purchasing product 2 was 0.51. To ensure these results, confirmatory indices are reviewed in Table 1. According to the results, there are two confirmation indices above 0.5 and close to 1, which indicates the acceptability of the designed model. An index is close to 0.5 that, regardless of its difference from 0.5 and the strength of the other two indices, can confirm the logistics model. It means that it can predict correctly with a small amount of error. In addition, the optimal classification threshold to group observation is equal to 0.5.

Table 4. Values of confirmatory indicators of logistic regression

Pairs	Numbers	Percentages	Measurement indicators	values
Compatibility indicator	2090531	97.4	Somers'D	0.95
Incompatibility indicator	54721	2.6	Goodman-Kruskal Gamma	0.95
Nodes	612	0.0	Kendall's tau-a	0.47
Total	2145864	100		

After calculating the probabilities of purchasing products, the conditional probability can be designed. The conditional probability model is as follows equation (8).

$$\frac{P(Purchase\ of\ product\ type) \times P(customer\ type\ to\ product\ type)}{P(Purchase\ by\ customer\ type)} \tag{8}$$

Equation (2) is correct because the probability of a customer type is independent of the type of purchase. Hence, it is possible to predict the sales limits of different products by estimating sales in different periods with time series techniques. Effective predictions can be made by multiplying these conditional probabilities on the number of sales in different periods. This will be a great help in expressing the chain management strategy.

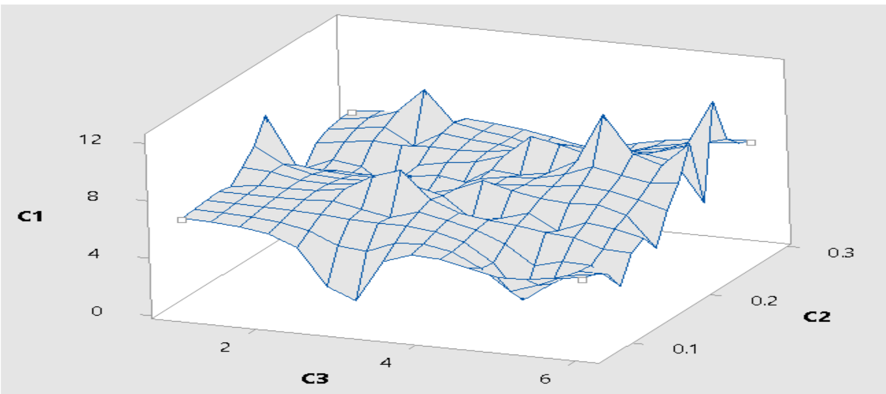

Figure 2. Conditional probability results

It should be mentioned C1 is the period, C2 is Probability value and C3 is Customer type, product type. To show the type of customer, the type of product, we have used the symbolism of Table 5.

Table 5. Classification and symbolization of customer type, product type

Representative of category (customer type, product type)	customer type	product type
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3

According to Figure 1, the probability of buying a product of type 1, if the customer is of type 1, in the first period is 0.2133, which is higher than subsequent periods. Likewise, these possibilities can be shown to other customers based on product type. The results show that in periods 1, 2, 3, 6, 7, 8, 9, 10, 11, and 12, buying a product of type 2 by a customer of type 1 is higher than in other cases. As a result, after predicting the number of customers' purchases in different periods, the highest number of sales is related to this type of state. The probability of purchasing from each province in different periods was determined by the results of data analysis.

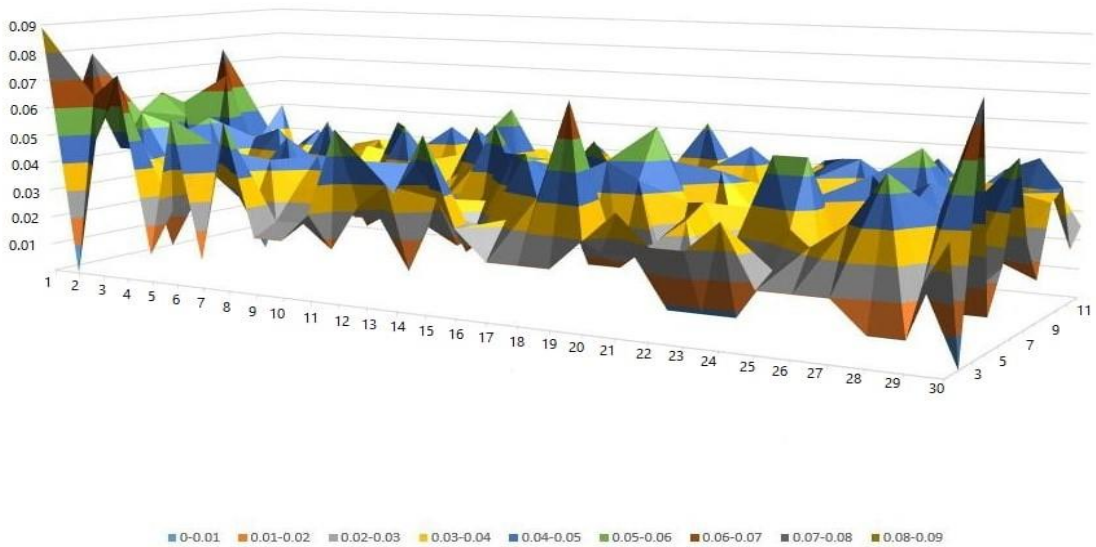

Figure 3. Sales probabilities of the total sales volume in each period by province

The numbers shown at the bottom of Figure 3 are a symbol of the provinces of the country, the details of which are shown in Figure 4. On the other hand, the horizontal numbers on the right of Figure 3 indicate different periods (12 months of a year) and the vertical numbers on the left indicate the probability.

Table 6. Symbols representing each province

province	Symbol	province	Symbol	province	Symbol
Tehran	1	Golestan	11	Semnan	21
Khorasan R	2	Mazandaran	12	Qazvin	22
North Khorasan	3	Sistan and Baluchestan	13	Qom	23
Southern Khorasan	4	Guilan	14	Kurdistan	24
Esfahan	5	Ardabil	15	Kermanshah	25

Fars	6	Elam	16	Kohgiluyeh and Boyer Ahmad	26
Eastern Azerbaijan	7	Boshehr	17	Lorestan	27
Yazd	8	Chaharmahal and Bakhtiari	18	Markazi	28
Alborz	9	Khuzestan	19	Hormozgan	29
Western Azerbaijan	10	Kerman	20	Hamedan	30

The results are shown in Figure 3. According to the results, the highest sales in period 1 belong to Tehran province and the lowest sales in period 1 belong to Khorasan Razavi province, which did not have any orders. Table 7 summarizes the data analysis after determining the type of customer to purchase the product type.

Table 7. Contingency Table

	Product Type		
	Laptop	Hard drive	Total
Customer type 1	457	661	1118
Customer type 2	366	521	887
Customer type 3	449	504	953
Total	1272	1686	2958

At present, the data process is analyzed with the probability values. The time-series data in Figure 6 show an ascending form, while no significant stagnation is observed in the mean and variance. On the other hand, the average purchase amount in the first season was the lowest, and the average level has increased in the second season, which can be seen until the last period. In December, the average level changed further. Thus, time series analysis methods can be used to analyze the above data only with the knowledge that the general trend of the data is ascending.

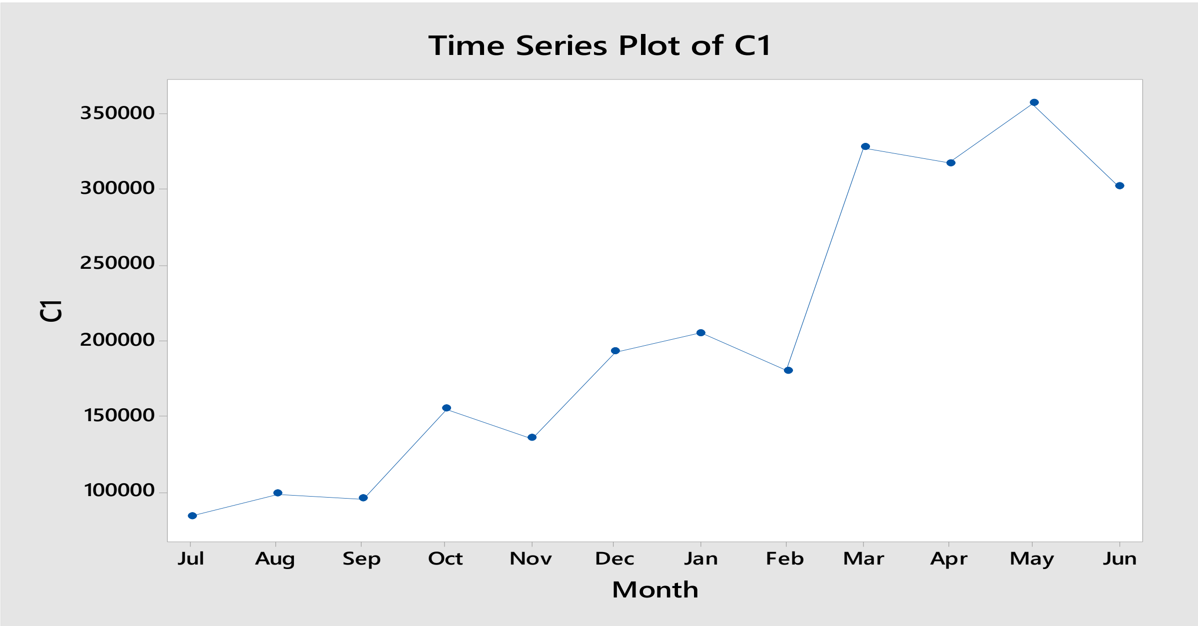

Figure 4. Monthly purchase volume trend

Therefore, single and double smoothing methods and Holt-Winters were used. For this purpose, we use MINITAB software to predict the next 12 periods of the year. Initially, a single smoothing method was used, and the results are shown in Figure 5.

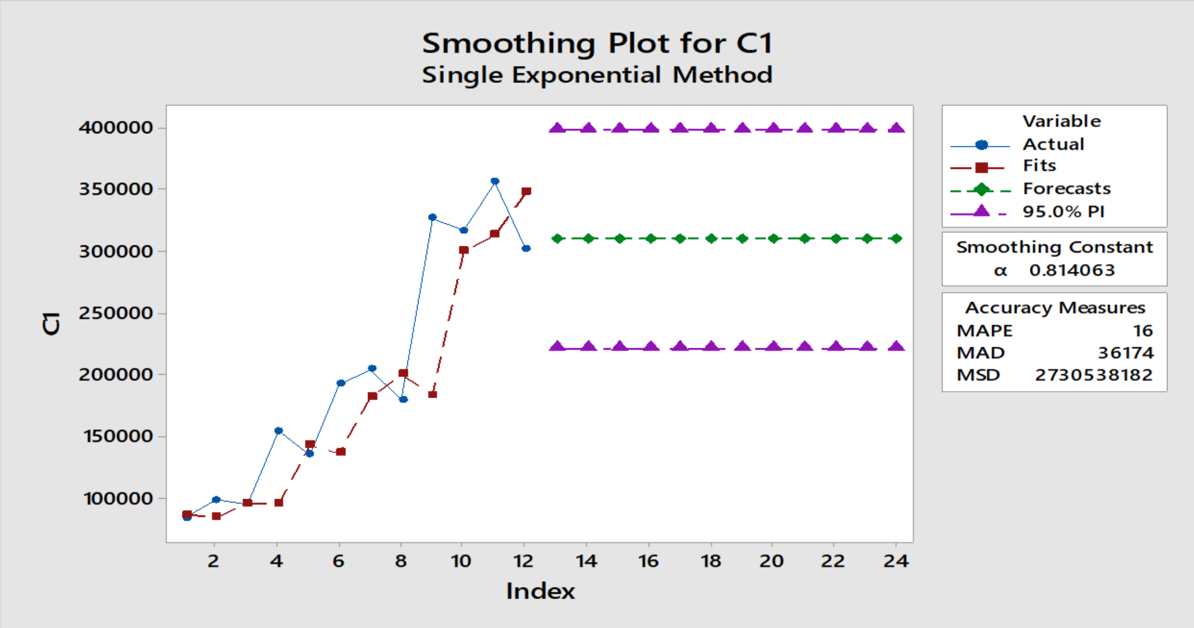

Figure 5. Sales forecast for the next 12 periods using the single average smoothing method

Due to the calculated average changes and predicted periods in the graph, this method is not suitable for predicting. We use the double smoothing method to use the trend component in the next step. It should be noted that the reason for choosing the dual time series is due to the high adequacy of this model compared to the triple time series model.

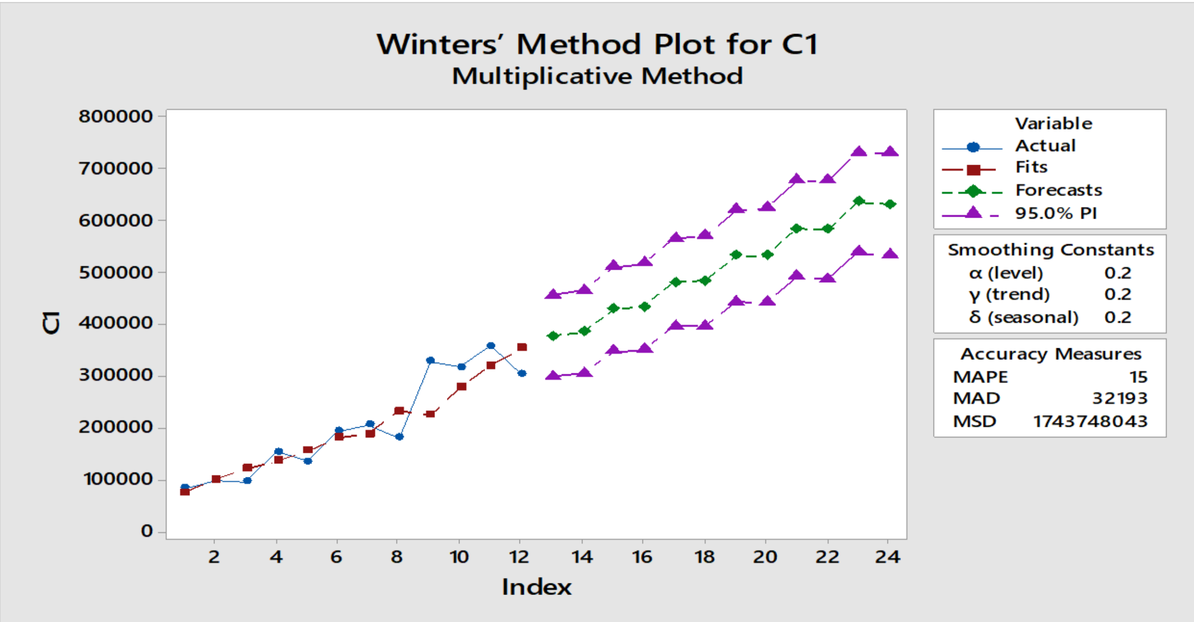

Figure 6. Predict the next 12 periods with the dual mean smoothing method

According to Figure 6, the prediction for the 12 monthly periods of the next year with a 95% confidence interval is given in Table 8.

Table 8. Predicted values from dual mean smoothing method

Predicted value ($)	the period	Predicted value ($)	the period	Predicted value ($)	the period
370735	1	473977	5	577219	9
396545	2	499787	6	603030	10
422356	3	525598	7	624651	11
448166	4	551409	8	654651	12

According to the time series in Figure 4, the data also have a seasonal trend. For example, the average level increases every three months. Thus, we will perform the Holt-Winters method to ensure the correct analysis of the time series trend. The results of this method are shown in Figure 7.

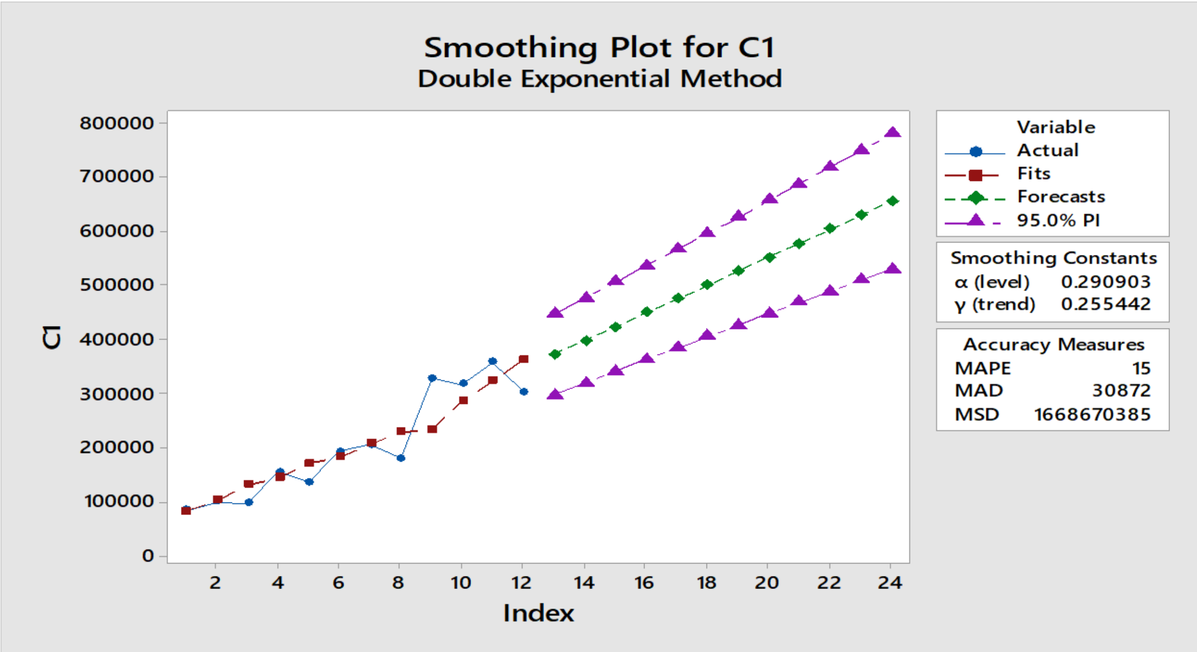

Figure 7. Sales forecast for the next 12 periods using the Health-Winters method

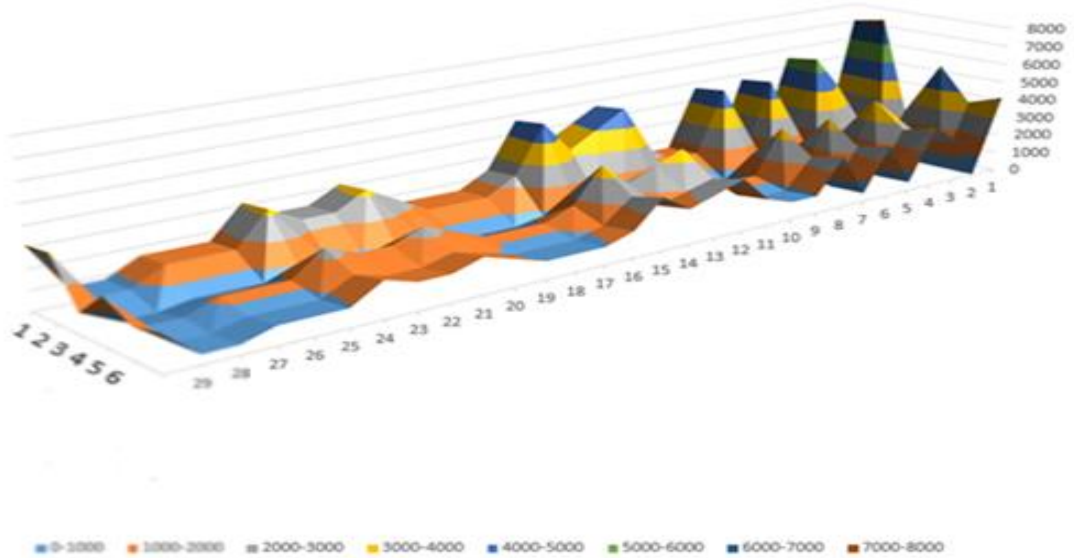

Figure 8. The amount of purchase of each product is predicted in the first period, provided that the type of customer is known separately for each province

It should be noted that the horizontal numbers to the left of Figure 8 are symbolized according to Table 5. According to Figure 8, this graph also includes the seasonal trend in the predictions. However, the Absolute Error Percentage (AEP), the Mean Absolute Deviation (MAD), and the Mean Squared Deviations (MSD) show that the best prediction method for this dataset is the double smoothing method, and its values are given in Table 5. Therefore, the values predicted by this method will be used in the following. The number of customer purchases can be predicted separately and based on the type of product and province in each period, with sales predicting probabilities and conditional probability.

According to Figure 8, the highest probability of purchases in the first period was related to Tehran province, so the share of sales of Tehran province in the first period according to Figure 4 was 33035 monetary units, which was the highest amount on provincial sales in all periods. Likewise, this analysis can be done for different periods. Then, by knowing the probabilities values, the customers purchase prediction, the sales prediction, and the desired digital store, we developed a mathematical model for the correct assignment of each order to the supplier.

Table 9. Suppliers and their capacities

Supplier	Product 1	Product 2	Minimum sales capacity	Supplier	Product 1	Product 2	Minimum sales capacity
1		160	20	12		170	50
2	140	160	20	13	140		50
3	150		40	14	50	50	25
4		100	20	15		60	25
5		200	30	16	40	50	20
6	100	50	50	17	200		30
7	130		100	18		200	70
8		100	20	19	70	150	150
9	200		20	20		100	40
10	70	75	15	21	80		20
11	80	90	40	22	150	50	30
The sum of capacities	870	935		The sum of capacities	730	830	

Table 10. Location of each supplier

Supplier	The provincial code in which the supplier is located	Supplier	The provincial code in which the supplier is located
1	1	12	5
2	1	13	5
3	1	14	6
4	1	15	6
5	1	16	10
6	1	17	10
7	1	18	10
8	2	19	10
9	2	20	19
10	2	21	19
11	5	22	19

Table 11. Assign suppliers to customer orders

product type		Product number 1			Product number 2		
Province code number	**Place of order**	**Suppliers assigned to orders in this province**	**Selected suppliers**	**Product quantity allocated to selected suppliers**	**Suppliers assigned to orders in this province**	**Selected suppliers**	**Product quantity allocated to selected suppliers**
1	Tehran	2-3-6-7	3	150	1-2-6	1-2	60-160
2	Khorasan R	9-10	9	200	8-10	8	100
3	North Khorasan	2-9-10	10	15	10	10	15
4	Southern Khorasan	9-10	10	25	1-10	1-10	60-20
5	Esfahan	11	11	80	11-12	11	90
6	Fars	4-14	4-14	9-50	15	15	46

7	Eastern Azerbaijan	2-16-17	16-17	45-25	16-18	16-18	50-60
8	Yazd	2-11	2	40	12	12	100
9	Alborz	2-3-6-7	2-6	100-50	1-12-14	1-12-14	40-70-20
10	Western Azerbaijan	16-17	17	40	4-18	4-18	91-140
11	Golestan	2-3-6-7	6	25	4-5	4-5	121-9
12	Mazandaran	6-7	6	30	18	18	40
13	Sistan and Baluchestan	9-10	10	15	10	10	15
14	Guilan	6-7	7	10	19	19	40
15	Ardabil	16-17	17	115	19	19	15
16	Elam	20-21-22	20	30	-	-	0
17	Boshehr	15-14	15	25	14	14	15
18	Chaharmahal and Bakhtiari	12-13	13	80	-	-	0
19	Khuzestan	20-21-22	22	110	20	20	80
20	Kerman	13	13	20	14	14	15
21	Semnan	2-3-6-7	7	50	6	6	20
22	Qazvin	2-3-6-7	7	5	6-15	6-15	30-14
23	Qom	2-3-6-7	7	60	19	19	50
24	Kurdistan	20-21-22	21	38	-	-	0
25	Kermanshah	20-21-22	20	40	-	-	0
26	Kohgiluyeh and Boyer Ahmad	-	-	0	20-22	20	20
27	Lorestan	20-21-22	21	30	21-22	22	30
28	Markazi	2-3-6-7-13	7-13	5-10	10-19	10-19	25-10
29	Hormozgan	21-22	21-22	12-10	22	22	15
30	Hamedan	2-3-6-7-13-17	13-17	50-10	19-22	19-22	40

According to Table 11, the number of orders the supplier should provide separately for product No. 1 & No. 2 is known. But based on the results of the mathematical model there are suppliers that are assigned to the orders of the provinces; however, they are not selected as the main supplier; this means that in case a supplier is not capable of providing the orders of that province, we can use the rest of the suppliers assigned to that province, which will definitely cost the same amount. In cases which there is only on supplier assigned and there is no reservation provider, we can say that in case of lack of assigned supplier, if another provider considers for allocation, the cost (target function) will change.

According to the first line, the mathematical model recognizes that when product No. 1 is ordered from province No. 1 (Tehran), if it is supplied by suppliers No. 2, 3, 6 and 7, the costs will be minimized. To ensure the correct response, we must first identify the suppliers who have product No. 1 according to Table 3. Suppliers who can supply product No. 1 are suppliers No. 2, 3, 6, 9, 10, 11, 14, 16, 20, and 21, respectively. The province that has sent the order is province No. 1 (Tehran). The suppliers selected by the software (suppliers No. 2, 3, and 6) are also in Tehran. So the response is valid. According to the problem data, suppliers have responded to demands regardless of the mathematical model. Here, we compare the optimal response to the problem with the actual response based on the data. According to calculations in Excel software, No. 1 is assigned to real data that had a response similar to the response of the mathematical model of the problem. The right decision has been made for 2100 orders out of 2958 orders. This means that a suitable supplier is considered to meet that demand. Accordingly, the error values of the actual data relative to the optimal response of the mathematical model are as follows equation 9.

$$Erorr = \left(1 - \left(\frac{2100}{2958}\right)\right) * 100 = 29.00608519 \quad (9)$$

According to these calculations, costs can be reduced by up to 29% if the model of assigning suppliers to orders is used.

According to Figure 4, among the total sales of the first month, the share of sales in Tehran province was 33053 monetary units, of which 7049 monetary units were ordered by customers type 1 to purchase a product type 1. This type of analysis of results in supply chain management can also determine the importance of orders for supply. This means that 7049 monetary units for product type 1 plus 7342 monetary units for product type 2 have the priority for supplying orders. We can do these calculations in different periods and get the sales details of each period separately and based on the provinces.

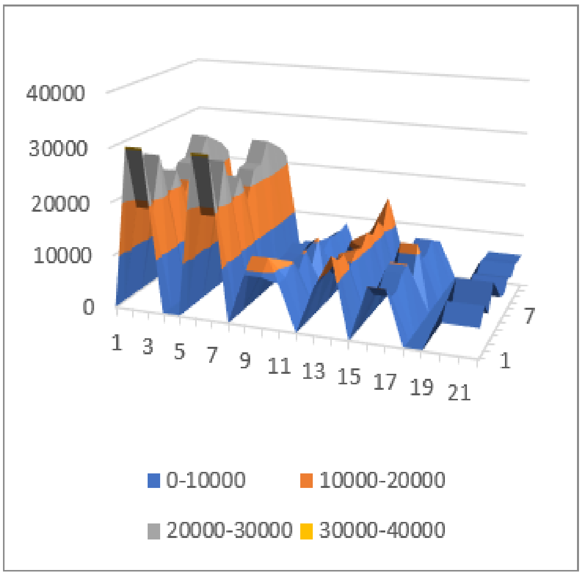

Figure 9. Details of sales forecast in different periods by province for type 1 customer and product 1.

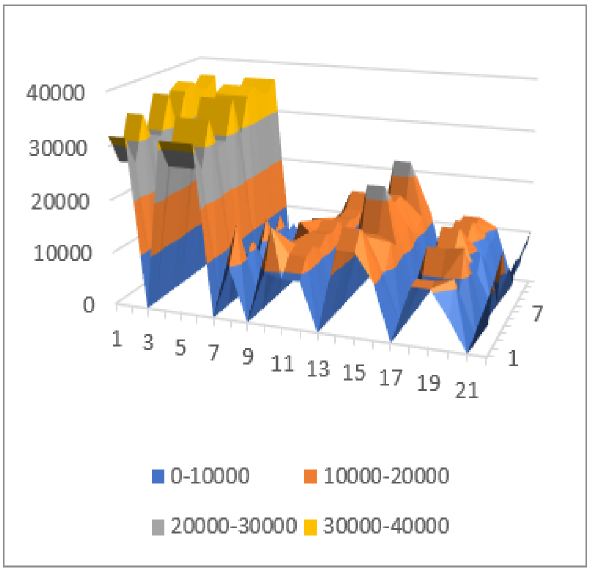

Figure 10. Details of sales forecast in different periods by province for type 1 customer and product 2

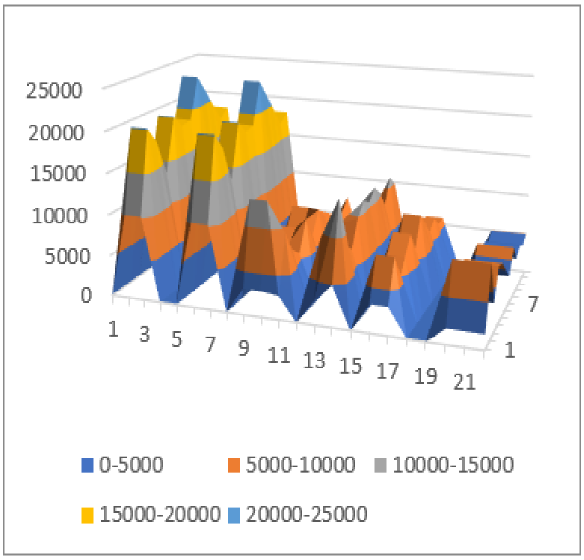

Figure 11. Details of sales forecast in different periods by province for type 2 customer and product 1.

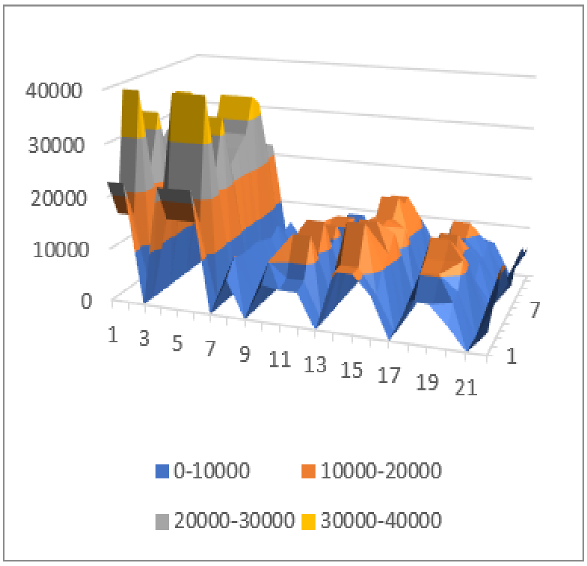

Figure 12. Details of sales forecast in different periods by province for type 2 customer and product 2.

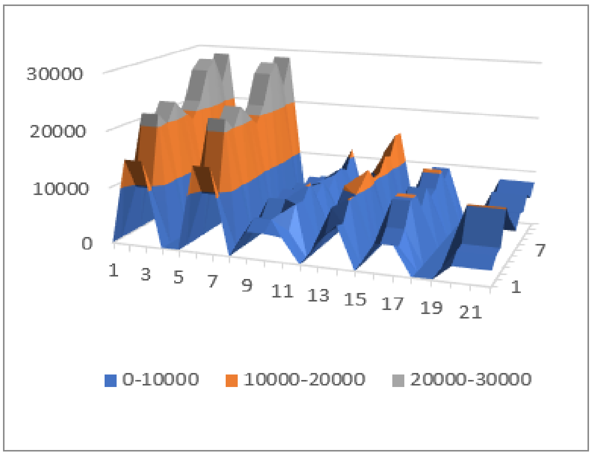

Figure 13. Details of sales forecast in different periods by province for type 3 customer and product 1.

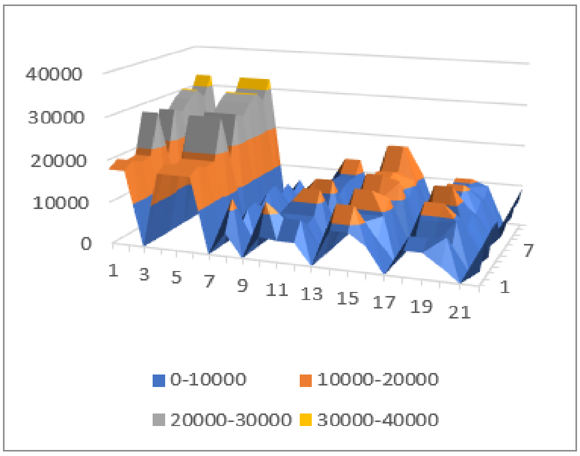

Figure 14. Details of sales forecast in different periods by province for type 3 customer and product 2.

In general, from the results, it can be concluded that:

- The model also identifies additional "assigned suppliers" in each province, beyond the initial selections.
- Provides a back-up option and ensures continuous order fulfillment in case the main supplier faces shortages or capacity constraints.
- Supplier allocation takes into account the geographical proximity between the province and suppliers and minimizes logistics costs and delivery time.

This comprehensive supplier allocation strategy, driven by advanced analytics and mathematical modeling, enables the online retailer to optimize inventory management, increase supply chain efficiency, and ultimately increase customer satisfaction through availability. Improve product reliability.

5. Managerial Insights

The study found that 29% of order assignments deviated from the optimal supplier allocation determined by the model. This was primarily due to supplier stockouts, forcing the selection of higher-cost alternatives. By predicting customer demand more accurately using the categorization and binary logistic regression methods, the model can help suppliers optimally allocate inventory to fulfill orders at the lowest cost.

The research was able to determine the predicted purchase rates of products across different time periods with a high degree of confidence. This insight can enable suppliers to proactively manage their inventory levels and replenishment cycles. By aligning inventory quantities and replenishment timing with the forecasted demand patterns, suppliers can minimize storage costs and ensure product availability.

Figures 5-10 in the full paper provide the detailed results and visualizations of the inventory planning strategies recommended for suppliers based on the research findings. In summary, the data-driven approach proposed in this study empowers online retailers and their suppliers to optimize inventory allocation, enhance replenishment planning, and ultimately improve supply chain efficiency and customer service. These actionable insights can have a significant impact on the management of digital commerce operations.

5.1. Conclusion and Future Research

This study presents a comprehensive data-driven framework to optimize inventory management in the context of the growing digital commerce landscape. By integrating customer purchase patterns, advanced categorization techniques, and predictive analytics, the proposed approach enables online retailers and their suppliers to significantly enhance inventory planning and allocation.

The core innovations of this research include:

1. Leveraging the WRFM (Weighted Recency, Frequency, Monetary) customer segmentation method to gain deep insights into differentiated buying behaviors.
2. Employing binary logistic regression to accurately forecast product demand at the customer segment level.
3. Developing a mathematical model that assigns orders to the most geographically suitable supplier, while considering inventory availability.

The results demonstrate that this integrated framework can account for 71% of sales data, reducing the error rate in supplier assignment to just 29%. This represents a substantial improvement over current industry practices. It is noteworthy that the integration of different data science approaches, although it can develop data and management analyses, sometimes also leads to an increase in error. Therefore, it can be said that the total errors in classification and prediction can ultimately lead to this number. However, the approach proposed in this study has been able to reduce the decision-making error compared to previous research. In addition, it is possible to achieve a lower error rate through trial and error and a combination of other methods, which, of course, is also highly dependent on the accuracy of the initial data.

The key management insights derived from this research empower suppliers to proactively manage their inventory levels and replenishment cycles. By aligning inventory quantities and replenishment timing with the forecasted demand patterns, suppliers can minimize storage costs and ensure product availability. Furthermore, the optimal assignment of orders to geographically proximate suppliers helps to streamline the supply chain and reduce overall logistics expenses.

These findings have significant implications for the digital commerce industry. As online shopping continues to grow, the ability to leverage customer data and predictive analytics to enhance inventory management will be a critical competitive advantage. The framework presented in this study provides a robust and scalable solution that can be adopted by retailers and suppliers to improve operational efficiency, reduce costs, and enhance customer satisfaction. Future research directions may include extending the model to handle demand uncertainty, as well as exploring the application of more advanced machine learning techniques to further refine the demand prediction capabilities.

References

Abbaspour Ghadim Bonab, A., & yousefi nejad attari, m. (2022). A predictive data-driven state-dependent decision approach to determine inventory system states for critical spare parts. *Journal of Quality Engineering and Production Optimization*, *7*(2), 205-231. https://doi.org/10.22070/jqepo.2022.16129.1233

Akhtar, P., Frynas, J. G., Mellahi, K., & Ullah, S. (2019). Big data-savvy teams' skills, big data-driven actions and business performance. *British Journal of Management*, *30*(2), 252-271.

Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 1785-1792. https://doi.org/https://doi.org/10.1016/j.jksuci.2019.12.011

Beheshtinia, M. A., Falsafi, P., Qorbani, A., & Jalinouszade, H. (2022). Evaluating and ranking digital stores' suppliers using TOPKOR method. *International Journal of Engineering*, *35*(11), 2155-2163.

Brito, P. Q., Soares, C., Almeida, S., Monte, A., & Byvoet, M. (2015). Customer segmentation in a large database of an online customized fashion business. *Robotics and Computer-Integrated Manufacturing*, *36*, 93-100.

Chaffey, D., Edmundson-Bird, D., & Hemphill, T. (2019). *Digital business and e-commerce management*. Pearson UK.

Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, *19*(3), 197-208.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.

Chen, X., Fang, Y., Yang, M., Nie, F., Zhao, Z., & Huang, J. Z. (2017). Purtreeclust: A clustering algorithm for customer segmentation from massive customer transaction data. *IEEE Transactions on Knowledge and Data Engineering*, *30*(3), 559-572.

Chen, Y., Liu, L., Zheng, D., & Li, B. (2023). Estimating travellers' value when purchasing auxiliary services in the airline industry based on the RFM model. *Journal of Retailing and Consumer Services*, *74*, 103433. https://doi.org/https://doi.org/10.1016/j.jretconser.2023.103433

Cheng, C.-H., & Chen, Y.-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications*, *36*(3), 4176-4184.

Cheng, Y., Chen, K., Sun, H., Zhang, Y., & Tao, F. (2018). Data and knowledge mining with big data towards smart production. *Journal of Industrial Information Integration*, *9*, 1-13.

Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. *Journal of direct, data and digital marketing practice*, *9*(3), 231-244.

Côrte-Real, N., Ruivo, P., Oliveira, T., & Popovič, A. (2019). Unlocking the drivers of big data analytics value in firms. *Journal of business research*, *97*, 160-173.

Dibb, S., & Wensley, R. (2002). Segmentation analysis for industrial markets: Problems of integrating customer requirements into operations strategy. *European journal of marketing*.

Dutta, S., Bhattacharya, S., & Guin, K. K. (2015). Data mining in market segmentation: a literature review and suggestions. Proceedings of Fourth International Conference on Soft Computing for Problem Solving,

Farajian, M. A., & Mohammadi, S. (2010). Mining the banking customer behavior using clustering and association rules methods. *International Journal of Industrial Engineering & Production Research*, *21*(4), 239-245.

Frost, R. D., & Strauss, J. (2016). *E-marketing*. Routledge.

Gallego-García, S., & García-García, M. (2020). Predictive sales and operations planning based on a statistical treatment of demand to increase efficiency: A supply chain simulation case study. *Applied Sciences*, *11*(1), 233.

Güçdemir, H., & Selim, H. (2015). Integrating multi-criteria decision making and clustering for business customer segmentation. *Industrial Management & Data Systems*.

Heldt, R., Silveira, C. S., & Luce, F. B. (2021). Predicting customer value per product: From RFM to RFM/P. *Journal of Business Research*, *127*, 444-453. https://doi.org/https://doi.org/10.1016/j.jbusres.2019.05.001

Khajvand, M., & Tarokh, M. J. (2011). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science*, *3*, 1327-1332.

Kim, K.-j., & Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert systems with applications*, *34*(2), 1200-1209.

Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial marketing management*, *90*, 493-507.

Liu, Y., Li, H., Peng, G., Lv, B., & Zhang, C. (2015). Online purchaser segmentation and promotion strategy selection: evidence from Chinese E-commerce market. *Annals of Operations Research*, *233*(1), 263-279.

McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of business research*, *60*(6), 656-662.

Ming-Chih, T., Yi-Ting, T., & Ching-Wei, L. (2011). Generalized linear interactive model for market segmentation: The air freight market. *Industrial marketing management*, *40*(3), 439-446.

Mitchell, V.-W., & Wilson, D. F. (1998). Balancing theory and practice: A reappraisal of business-to-business segmentation. *Industrial marketing management*, *27*(5), 429-445.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, *36*(2), 2592-2602.

Pawar, K. S., & Driva, H. (2000). Electronic trading in the supply chain: a holistic implementation framework. *Logistics Information Management*.

Peker, S., Kocyigit, A., & Eren, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: a case study. *Marketing Intelligence & Planning*.

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.

Sadeghi, H. (2019). A forecasting system by considering product reliability, POQ policy, and periodic demand. *Journal of Quality Engineering and Production Optimization*, *4*(2), 133-148. https://doi.org/10.22070/jqepo.2020.5087.1123

Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, *14*(12), 7243.

Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., & Rahmani, R. (2018). Customer segmentation and strategy development based on user behavior analysis, RFM model and data mining techniques: a case study. 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE),

Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. Expert Systems with

Applications, 36(10), 12547-12553.

Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons.

Wang, S.-C., Tsai, Y.-T., & Ciou, Y.-S. (2020). A hybrid big data analytical approach for analyzing customer patterns through an integrated supply chain network. *Journal of Industrial Information Integration*, *20*, 100177.

Weinstein, A. (2006). A strategic framework for defining and segmenting markets. *Journal of Strategic Marketing*, *14*(2), 115-127.

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive*, *12*(2).

Wong, E., & Wei, Y. (2018). Customer online shopping experience data analytics: Integrated customer segmentation and customised services prediction model. *International Journal of Retail & Distribution Management*.

Wu, J., Shi, L., Yang, L., XiaxiaNiu, Li, Y., XiaodongCui, Tsai, S.-B., & Zhang, Y. (2021). User value identification based on improved RFM model and k-means++ algorithm for complex data analysis. *Wireless Communications and Mobile Computing*, *2021*, 1-8.

Wu, R.-S., & Chou, P.-H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications*, *10*(3), 331-341.