

Review Article

Application of Data Mining Methods in Grouping Agricultural Product Customers

Tzu-Chia Chen,¹ Fouad Jameel Ibrahim Alazzawi,² Dinesh Mavaluru ,³ Trias Mahmudiono ,⁴ Yulianna Enina,⁵ Supat Chupradit ,⁶ Alim Al Ayub Ahmed ,⁷ Mohammad Haider Syed,⁸ Aras masood Ismael,⁹ and Boris Miethlich ¹⁰

¹Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City 24301, Taiwan

²Department of Computer Engineering, Al-Rafidain University College, Baghdad, Iraq

³Department of Information Technology, College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia

⁴Department of Nutrition, Faculty of Public Health, Universitas Airlangga, FKM Unair Jl. Mulyorejo Kampus C Surabaya, Surabaya City, Indonesia

⁵I.M. Sechenov First Moscow State Medical University (Sechenov University), Department of Propaedeutics of Dental Diseases, Moscow, Russia

⁶Department of Occupational Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

⁷School of Accounting, Jiujiang University, 551 Qianjindonglu, Jiujiang, Jiangxi, China

⁸College of Computing and Informatics, Saudi Electronic University, Saudi Arabia

⁹Sulaimani Polytechnic University, Technical College of Informatics, Information Technology Department, Sulaymaniyah, Iraq

¹⁰Department of Business Studies, IIC University of Technology Bldg.No. 069, Concrete Road 121206, Phnom Penh, Cambodia

Correspondence should be addressed to Boris Miethlich; boris.miethlich@gmail.com

Received 25 September 2021; Accepted 31 January 2022; Published 11 March 2022

Academic Editor: Alireza Goli

Copyright © 2022 Tzu-Chia Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The sheer complexity of the factors influencing decision-making has required organizations to use a tool to understand the relationships between data and make various appropriate decisions based on the information obtained. On the other hand, agricultural products need proper planning and decision-making, like any country's economic pillars. This is while the segmentation of customers and the analysis of their behavior in the manufacturing and distribution industries are of particular importance due to the targeted marketing activities and effective communication with customers. Customer segmentation is done using data mining techniques based on the variables of purchase volume, repeat purchase, and purchase value. This article deals with the grouping of agricultural product customers. Based on this, the K-means clustering method is used based on the Davies-Bouldin index. The results show that grouping customers into three clusters can increase their purchase value and customer lifespan.

1. Introduction

Nowadays, an essential part of doing business is identifying the variety of customers that purchase the offered products/services and establishing a relationship with them such that they remain a source of revenue for the business in the future. Indeed, retaining valuable old customers is as important as

attracting new customers [1–3]. One way to gain a deeper understanding of customers is to segment them into several groups and examine the characteristics of each group. In fiercely competitive markets where customers have many options to choose from, analyzing customer behavior and selecting the appropriate marketing method may determine whether a company can survive the competition [4].

A true understanding of customer behaviors is one of the most important aspects of customer relationship management (CRM) [5]. With the rapidly increasing use of information technology in many fields of business, which has resulted in the collection of large volumes of customer data, businesses now need more sophisticated data analyses to gain accurate knowledge and a deep understanding of customer behaviors and purchase patterns in order to adopt marketing strategies and respond to customer needs [6, 7]. As a result, the need for customer data mining and analysis tools as a prerequisite for adopting appropriate marketing and CRM approaches to satisfy and retain customers is felt more than ever. One of such data mining techniques is customer segmentation, which involves dividing heterogeneous customers into homogeneous groups with similar purchasing behaviors and patterns in order to make it easier to understand how they behave and develop the appropriate marketing and CRM strategy accordingly [8]. The success or failure in this effort can play a major role in the survival of a business [9]. In today's highly complex and competitive business environment, which demands constant business growth, customer segmentation can help improve long-term customer loyalty and customer relationships, ultimately affecting businesses' profitability [10]. Data mining can help a company identify and track customer behaviors and behavioral patterns over the course of their interaction with the company, which will ultimately lead to better customer service, higher sales, more effective distribution, and marketing strategies [11]. In this regard, it is necessary to use novel artificial intelligence to help the decision-makers in the agriculture industry [12–15].

Previous studies show that in order to determine appropriate strategies in the production and sale of agricultural products, it is necessary to perform appropriate grouping for them and their customers. This issue has many challenges. First, much information is generated in this field and which data are used has a significant impact on the research method. Second, customer lifespan has a significant impact on determining strategies. These challenges have not been seriously addressed in previous research. In this regard, the main question of this research is how to provide a framework for customer grouping by considering the lifespan and using big data analysis.

Accordingly, the machine learning approach is used to group agricultural customers in this research. Moreover, a fuzzy formulation for calculating customer lifespan is provided. In summary, the main contribution of this research and its innovation can be stated in presenting a method based on big data analysis in order to group customers by considering the impact of customer lifespan and finally presenting appropriate strategies for each group. Moreover, in this article, business-to-business (B2B) customers of a business are examined and segmented using the RFM (recency, frequency, and monetary) value model.

The rest of the article is organized as follows. The theoretical foundation of the research area and literature review are explained in Section 2. The problem statement is provided in Section 3. In Section 4, the numerical results are presented; in Section 5, the discussion is provided; and finally, in Section 6, the article is concluded.

2. Theoretical Background

CRM begins with customer identification, which involves identifying and targeting the groups of people who may be profitable for the business. In this regard, it is important to analyze the behavior of past and present customers in order to understand their common characteristics Goli et al. [16, 17]. Businesses often analyze the data they collect from their customers and sales records through the customer segmentation process, which involves dividing customers into multiple groups and categories with similar needs, characteristics, and functional variables in order to develop appropriate customer retention and customer acquisition strategies for each group in the next stages of CRM. Customer segmentation is one of the most effective tools for customer-centric marketing (Butel). Segmentation is a data mining technique that attempts to find how data items should be clustered in order to have the greatest similarity within each cluster and the greatest differentiation between clusters [18].

The customer segmentation process has three main components: segmentation variables, method, and validation. The relationship between product decision-making information systems and big data-driven is inspired by [19–22]. Moreover, the following items are a brief description of each component and how they are chosen or performed in this study.

2.1. Segmentation Variables. Customer segmentation should be done based on variables that are measurable for every customer, such as purchase time, customer type, purchase frequency, and purchase value. In the RFM model, customers are segmented based on three variables: 1—recency (R), which indicates the time elapsed since the customer's last purchase; 2—frequency (F), which indicates the number of times the customer has made a purchase over a certain period, and 3—monetary value (M), which indicates the amount of money the customer has paid the business. The FRM model can be used to segment customers for the purpose of determining preferable marketing policies. The results of this model are also a common input for the development of marketing strategies. This model is also commonly used in customer lifetime value (CLV) calculations.

2.2. Segmentation Method. RFM is a descriptive model and cannot predict customer behavior. Thus, it initially segments customers in terms of each segmentation variable alone without considering the impacts of other variables. Clustering methods and algorithms can be divided into two categories: hierarchical methods and partitioning methods. Among clustering algorithms of the latter category, the k-means algorithm has received a lot of attention because of its sophistication, high speed, and accuracy. This algorithm tries to assign the data items to k clusters iteratively with the goal of minimizing the distance of each item from the center of the cluster to which it is assigned. The iterations of this algorithm continue until cluster centers remain the same

over a number of successive iterations. This algorithm aims at achieving maximum similarity in the data items assigned to each cluster and maximum difference between the data items assigned to different clusters.

2.3. Segmentation Validation. One input of the clustering method is the number of clusters (k), which needs to be specified and will have a notable impact on the clustering performance. One way to determine the best number of clusters is to execute the algorithm with different numbers of clusters and compare the outcomes in terms of an indicator like the Davies–Bouldin index. This index calculates the average similarity of data in each cluster and can be used to assess the lack of similarity between clusters. Lower values for this index indicate higher clustering quality.

2.4. Literature Review. In a study by Anitha et al., these researchers used the RFM model and the k -means algorithm to analyze customer behavior. In this study, the optimal number of clusters was determined with the Silhouette method [23]. Wu et al. used the RFM model and the k -means clustering method to analyze the value of customers of industrial equipment manufacturing companies. After preparing the data, customers were clustered into six segments by k -means based on RFM indicators, and customer characteristics in the segments were studied through a CLV analysis. In the end, this article also offered some suggestions for using appropriate promotion programs for different segments of customers [24]. Monalisa et al. [25] used the RFM model with the fuzzy C -means algorithm to segment the customers of a company called LWC. These researchers also calculated the CLV of the obtained segment. A research paper by Miguies [26] stated that customer segmentation is a key step for developing and maintaining customer relationships, which tend to lead to increased sales. This article proposed a method of market segmentation for retailers based on customer lifespan as indicated by exchange volume and a method of customer segmentation based on purchase history.

Rabiei et al. [27] combined the RFM model with a classification method to develop a model for estimating CLV. For this purpose, these researchers constructed an RFM model and computed the CLV of customer clusters by incorporating the CLV calculations into the $C5$ algorithm. In a study by Mohammadi et al. [28], a combination of three methods, namely, RFM, FAHP, and k -means, was used for customer segmentation. In this study, FAHP was used to weight the segmentation variables, and the resulting variables were used to cluster the customers into five segments. In a study by Christy et al. [29], customer segmentation with the RFM model was performed by using both k -means and fuzzy C -means algorithms for clustering. Taghi Livari and Zarrin Ghalam [30] segmented the customers of an insurance company with the RFM model using self-made SOM and k -means algorithms. Nguyen et al. [31] reviewed the previous studies on the relationship between customer behavior and order fulfillment in the context of marketing in online retail companies and introduced various marketing

tools for improving the level of service offered to customers. In a study by Arunachalam and Kumar [32], customer segmentation was conducted with fuzzy methods and SOM. Shokouhyar et al. [33] segmented the customers of the after-sales services of an automobile manufacturer using the RFM model. These researchers used KANO and SERVQUAL models as tools to measure customer satisfaction.

In one of the new and most related research items, Ernawati et al. [34] assessed the decision-making (DM) methods that collaborated with the RFM model and synthesized them to propose a customer segmentation framework. This study uses a comprehensive literature review published in 2015–2020. The most widely used methods are clustering and visualization from seven DM methods analyzed.

Purwanto et al. [35] proposed a clustering method for potential customer grouping in digital trade environments. They applied data mining that was collected from potential customers by online questionnaires. The authors do the data collection by distributing questionnaires online through Google form. Ciccullo et al. [36] assessed the agri-food supply chain as well as the role of food waste prevention technology in the food industry. This study reveals that adopting different technological options can represent the engine to establish vertical collaborations between the adopter of the technology and another stage in the agri-food supply chain to fight food waste. Donnet et al. [37] proposed a conceptual framework to find the effect of each critical risk index in the production and sale of agricultural products. Insights from this study can help farmers and agribusiness managers by defining and adapting their strategies within their local contexts.

Previous studies show that calculating the CLV is always fraught with complexity. Because when measuring the value of indicators, the value of its inputs must be known definitively. On the other hand, CLV depends on several factors, including customer preferences. Therefore, this study proposes using the fuzzy analytical hierarchy process method (FAHPM) approach to calculate the CLV.

3. Problem Statement

Many manufacturing and distribution companies have a large database of customer information and purchase history, including customer profile, purchase dates and frequency, and the amount of money exchanged. The owners of these large businesses are perfectly aware that they can use these data to achieve improvement in all aspects of CRM, including customer identification, acquisition, retention, and promotion. One of the most important processes of customer identification in businesses that deal with a large number of customers with massive amounts of data is customer segmentation. This process can help a business interact properly with its customers and ensure that provided products and services meet the standards demanded by customers to their satisfaction [38]. In this context, the primary measures of segmentation quality are the similarity of customers placed in the same segment and the dissimilarity of customers

placed in different segments in terms of functional variables. One of the models that are commonly used in customer segmentation and analysis is the RFM model proposed by Hughes in 1994. In this model, the functional variables for the segmentation of customers are the recency of purchases (R), their frequency (F), and their monetary value (M) [39, 40]. This study used the RFM model to segment the customers through clustering with the k-means algorithm. The Davies–Bouldin index was used to evaluate the quality of clusters (segments), and in the end, the CLV of each cluster was also estimated.

4. Methodology

In this study, the RFM model and the k-means algorithm were used to segment 12 major B2B customer groups of a food company, each group representing similar consumer markets (e.g., hospitals, universities, municipalities). The data mining and analysis process for knowledge extraction was carried out using the standard CRISP-DM methodology. This methodology comprises a number of major steps, which include understanding the business environment, data selection, data preparation, modeling, model evaluation, and evaluation of the results (Figure 1). The steps of this method are described in Figure 1.

As mentioned earlier, the PVC model is applied to manage the CRISP-DM process. The proposed PVC model is inspired by [41–43].

Step 1: understanding the business environment

Considering the importance of customer identification and retention for all businesses, in this study, we tried to identify and group the customers of the studied business in order to improve its performance in this area. Customer segmentation into similar categories is one of the major topics of CRM, which is based on the idea that not all customers should be viewed and treated the same way. In this context, the definition of a customer may differ depending on the situation. In business-to-business (B2B) enterprises, most or all customers are either a business or an organization (Butel). Given the importance of B2B customers for the studied business, this study was focused on this type of customer.

Step 2: dataset selection

In this step, the data required for conducting the research as described in the previous step were collected. The required dataset is collected in the Cambodia agriculture affair institute. The list of products with the most important customers is gathered and put in the dataset. For each customer, the purchase recently and purchase frequently is considered.

The following customer data needs to be collected to build the RFM model.

Purchase recency indicates the number of days passes since the customer's last purchase. Lower recency is more desirable.

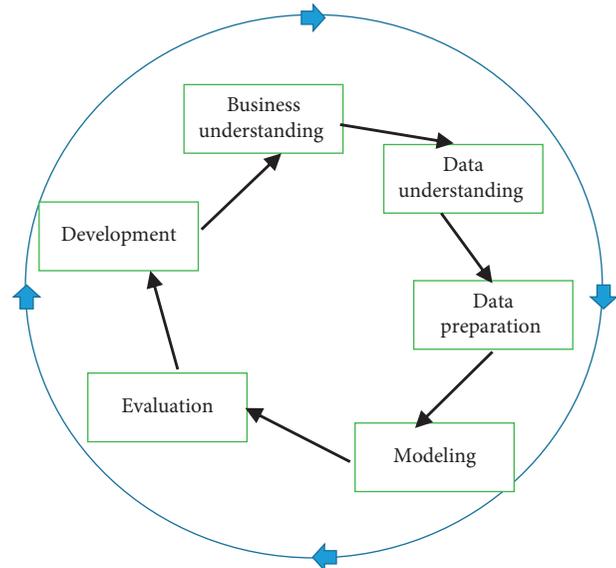


FIGURE 1: CRISP-DM process.

Purchase frequency: this indicates the number of times the customer had made a purchase. Higher purchase frequency rates are generally more desirable.

The monetary value of purchases: this indicates the total amount of money exchanged between the customer and the business. The higher values of this index are more desirable.

Step 3: data preparation

This step involves preparing and preprocessing data to facilitate the extraction of the knowledge it contains. For this purpose, incomplete, invalid, and inaccurate data must be discarded, and all of the remaining data must be converted into the format that suits the software (in this study, the software was RapidMiner).

Step 4: measurement of segmentation variables

As mentioned earlier, the common RFM model uses the three variables of recency, frequency, and monetary value to segment customers. In this study, these variables for each customer group were determined from the data collected over a one-year period. Table 1 shows the description and type of these variables.

Step 5: customer clustering with k-means

The results of a clustering algorithm for a dataset can strongly depend on the choice of algorithm parameters. The purpose of cluster validation is to determine how well clusters fit the data. The two basic criteria for evaluating clusters in this respect are compactness and separation.

Compactness

The data items assigned to the same cluster should be as similar to each other as possible. The most widely used criterion for measuring the compactness of clusters is variance.

Separation

TABLE 1: Type of RFM variables.

No.	Variable name	Description	Variable type	
1	ID	Customer code	Integer	Numeric
2	Monetary	Total monetary value of all purchases made by the customer	Numeric	Continuous numeric
3	Date	Purchase dates	Date	Date
4	Recency	Time passed since the last purchase	Date	Date
5	Frequency	Number of times the customer made a purchase	Integer	Numeric

The obtained clusters must be sufficiently separated (differentiable) from each other. There are three criteria for measuring the separation of clusters: the distance between the closest data in two clusters, the distance between the farthest data in two clusters, and the distance between the centers of two clusters.

In this study, the comparisons needed to determine the optimal number of clusters were performed based on the Davies–Bouldin index. Since these assessments showed that using three clusters results in the best (lowest) value for this index, it was decided to perform the customer segmentation with the parameter k in the k -means algorithm set to 3.

Step 6: determining the optimal number of clusters based on the Davies–Bouldin index

To determine the optimal number of clusters, the Davies–Bouldin index was calculated using the following formula:

$$DB = \frac{1}{n} \sum_{\substack{i=1 \\ i \neq j}}^n \max \left\{ \frac{S_n(Q_i) + S(Q_j)}{S(Q_i, Q_j)} \right\}, \quad (1)$$

where n is the number of clusters, S_n is the average distance of each record from the cluster center, and $S(Q_i, Q_j)$ is the distance between cluster centers. The best number of clusters is the one giving the lowest DB value.

The results of this step are presented in Table 2.

Since the lowest DB value was obtained with three clusters, it was determined that the optimal number of clusters according to this index is 3. Therefore, customers were clustered into three segments.

Step 7: comparison and evaluation of the results

The results of customer segmentation are shown in Table 3.

It can be seen, customers were classified into three clusters: customers in Cluster 0 are highly valuable for the business, customers in Cluster 1 are low-value, and customers in Cluster 2 have moderate value for the business. Table 4 shows the type of customers placed in each cluster.

Next, we labeled the clusters based on their value for the business. Customers with higher value creation were labeled golden, those with medium value were labeled

TABLE 2: Davis–Bouldin index values calculated for different numbers of clusters.

Index	2	3	4	5	6
DB	0.621	0.495	0.586	0.595	0.501

TABLE 3: Customer segmentation results.

No of records	2	5	5
Cluster	Cluster 0	Cluster 1	Cluster 2
Monetary value	1889969500	328290299	887774385
Frequency	18	10	14
Regency	27	104	46

TABLE 4: Placement of customers in clusters.

cluster_2	Hospitals
cluster_2	Universities
cluster_1	Automobile manufacturers
cluster_1	Ministry of energy
cluster_2	Municipality

silver, and those with lower value creation were labeled bronze customers (Table 5).

Step 8CLV calculations

n marketing, the CLV of a customer is the estimated profitability of future interactions with that customer. The higher the CLV of a customer is, the more valuable the customer is for the business. After clustering customers, the CLV of each customer was calculated by weighting the RFM variables (recency, frequency, and monetary value) as formulated in the following formula [25, 27]:

$$clv = w_m * nm + w_r * nr + w_f * nf, \quad (2)$$

where w_r , w_f , and w_m are the weights of recency, frequency, and monetary value, respectively. Based on an analysis of expert opinions with the help of ExpertChoice software, these weights were set to $w_m = 0.497$, $w_r = 0.225$, and $w_f = 0.278$. In (2), Nr , Nf , and Nm are the normalized values of model variables, which were obtained by fuzzy nondimensionalization as shown in Table 6. Having these values, CLV values of customers were determined as shown in Table 7.

TABLE 5: Cluster labeling.

Cluster	Label
Cluster 0	Gold customers
Cluster 1	Bronze customers
Cluster 2	Silver customers

TABLE 6: Normalization results.

Cluster	Normalized R	Normalized F	Normalized M
Cluster 0	0.95	0.91	0.91
Cluster 1	0.218	0.21	0.1
Cluster 2	0.64	0.55	0.515

TABLE 7: CLV calculations for clusters.

Label	$w_m^* nm$	$w_r^* nr$	$w_f^* nf$	CLV	Rank
Gold	0.4522	0.2137	0.2529	0.919	1
Bronze	0.0497	0.0490	0.0583	0.1571	3
Silver	0.2559	0.144	0.1529	0.5528	2

$$N = \frac{x - \mu}{\sqrt{\sigma^2}}. \quad (3)$$

The higher the CLV value of a cluster is, the more valuable its customers are for the business. As expected, the CLV value of customers in the Golden Cluster was higher than that in other clusters.

5. Discussion

The numerical results showed that the customers with the highest value creation for the studied business are the Cambodian agriculture industry and “Prisons and Security and Corrective Measures Organization,” which belonged to Cluster 0 (golden customers). These customers have an average CLV of 0.919 and constitute only 17% of all customers. To retain this group of customers and maintain their loyalty, it is recommended to adopt expansion strategies.

The businesses and organizations allocated to Cluster 1 or the low-value cluster (bronze customers) had an average CLV of 0.15, which indicates limited room for improvement in cooperation and limited potential for further profitable interactions. The recommended strategy for this group of customers is negative retention. The customers placed in Cluster 2 (silver customers) include hospitals, law enforcement, the Cambodian agriculture industry, and municipalities. The average CLV of these customers is 0.55, and they make up 41% of all customers of the studied business. With an appropriate strategy, the businesses and organizations placed in this cluster can potentially be added to the list of golden customers. Using the variety of strategies listed in Francis Butel’s book “Customer Relationship Management” as the reference, in the following, we offer a number of marketing and customer loyalty strategies that can be expected to fit the studied company and its customers.

One viable strategy for customer retention is to create commitment. There are multiple verities of this strategy,

each of which can be tailored to the characteristics of customers in each segment. In particular, this strategy can be utilized to retain the silver and gold customers of the studied business.

5.1. Instrumental Commitment. The loyalty of university customers can be increased through the instrumental commitment strategy. For example, this can be done by holding at the university campus a cooking competition with the company’s products to create instrumental commitment among customers and encourage them to buy more of the products.

5.2. Relational Commitment. An improvement in work relationships can help improve customer loyalty and retention. This strategy can be applied to hospital customers. For example, the quality control agents of the production can visit these customers to ensure them of the quality and safety of offered products. Granting customers support cards can also help improve their relationship commitment.

5.3. Value-Based Commitment. Making the customers feel that the company’s values are aligned with their own values will increase their loyalty. This strategy is more suitable for hospital customers that were placed in the silver cluster. This can be done by offering customers organic products that improve their health and accelerate their recovery.

Another group of viable strategies for customer retention is bonding strategies [Butel], which involves creating a variety of structural bonds to make it more difficult for the customer to cut its relationship with the business.

5.4. Financial Bonds. This strategy could be very effective on B2B customers because most public and private organizations have periodic purchase budgets, and this makes the strategy function as both a negative and positive customer retention strategy. For example, this strategy can be implemented by giving valuable customers a credit line for their purchases.

5.5. Process Bonds. This strategy can be applied to school vendor customers. For example, the sales share of these customers can be increased by offering single-person and student-specific products.

5.6. Multiproduct Bonds. When multiple products that a customer uses are produced by a single company, their bond becomes more challenging to break. Therefore, this strategy can be very effective for maintaining and improving the loyalty of “Cambodia universities” and “Prisons and Security and Corrective Measures Organization,” as they are among the company’s loyal customers with substantial orders.

5.7. Customer Termination Strategy. The business may need to break its relationship with those customers that are not profitable and cannot be made profitable by any strategy.

TABLE 8: Summary of adaptable strategies for each cluster of customers.

Description	Customer	Strategy type
Dispatching quality control agents to check the quality of products	Hospitals	Relational commitment
Offering organic products	Hospitals	Values-based commitment
Offering single-person and student-specific products	School vendors	Process bonds
Building stronger bonds by offering more products	Prisons	Multiproduct bonds
Offering credit arrangements	Loyal customers	Financial bonds
Holding conferences and competitions	Universities	Instrumental commitment
Setting a minimum purchase volume	Customers in cluster 2	Negative retention

Bronze customers are good candidates for this strategy. For these customers, it is best to adopt a negative customer retention strategy based on setting a minimum order size to see whether there will be an increase in their purchase volume. After identifying the customers that are loyal to the company through this process, customers that do not meet the minimum purchase criteria need to be terminated.

- (i) Future studies are recommended to try using other models such as LRFM for customer segmentation
- (ii) It is recommended to use association rules to discover the relationships between the products purchased by customers
- (iii) It is recommended to implement customer loyalty-building strategies and assess the outcome by using customer satisfaction measurement tools like the Kano model and gathering feedback from customers

Table 8 provides a summary of the strategies that can be adopted for the customers in each cluster.

A deep comparison of the provided results with some important previous works like [33, 40] indicates that the implemented machine learning method to conduct the strategies for the customers in the agriculture industry can lead to an appropriate plan for this industry and enhance the profitability and efficiency.

6. Conclusions and Future Directions

The agricultural sector in developing countries, as well as industrial-based countries, is the main engine of economic growth and development. In order to overcome the crisis of underdevelopment, they should go to their agricultural sector, and while trying to expand agricultural production, they should think of combining this sector with advanced technologies in order to make their products efficient [44]. Due to its extensive connections with other economic sectors, it can provide market creation, currency generation, and industry growth [45]. Therefore, in order to overcome the economic crisis, all countries should consider the agricultural sector as one of the main pillars of economic development because of its important role in the food supply, social welfare, GDP, and ultimately economic growth.

On the other hand, considering that this sector includes a large number of people who are directly or indirectly involved in the various stages of production to supply agricultural products and work, it can be said that the development of agriculture and related industries can increase GDP, job creation, and currency exchange as well as the self-sufficiency of the country in importing agricultural products. Therefore, increasing attention to the agricultural industry should become one of the main goals of any country for economic and social development [46].

In developing countries, the agricultural sector is usually large and of special importance, because agriculture plays an important role in the economy of such countries and can be used in various ways such as labor supply and capital, supply of raw materials, and cheap food, creating a market for manufactured goods in the industrial sector, and foreign exchange to economic development [47].

Industrialization can have a positive effect on agricultural growth in various ways. During industrialization, incomes increase rapidly, which in turn increases the demand for agricultural products, especially food. Moreover, it will increase employment in rural areas [30].

Industrialization increases the volume of capital in the agricultural industry. This also helps to modernize agriculture and thus increase production. Regarding the relationship between the agricultural sector and services, it can be mentioned that the two subsectors of transportation and communications have a direct relationship with the agricultural industry; The relationship between the amount of transportation and total costs is increasing. The ability to maintain agricultural products in the face of technological changes in the world is rapidly moving towards the growth of trade and globalization.

One of the most important managerial applications of this research is the issue of improvement strategies in the agricultural industry. Implementing a machine learning approach was able to identify several effective strategies. This approach can be applied to other challenges of the agricultural industry as well. For example, this approach can be used to supply industrial equipment and agricultural growth and development to determine the best relevant strategies.

The most important limitation of this research is its high dependence on input data. Lack of all necessary data availability causes the lack of achieving the desired outputs. Moreover, in order to improve this research, it is

suggested to focus on optimizing customer grouping and using the novel meta-heuristic algorithms like Gray Wolf Optimizer and Ant Lion Optimization algorithms.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] S. Chupradit, D. Yannan, H. W. Kamran, S. S. Soudagar, A. M. Shoukry, and J. A. Khader, "Measuring technical efficiency associated with environmental investment: does market competition and risk matters in banking sector," *Environmental Science and Pollution Research*, vol. 28, no. 47, pp. 66575–66588, 2021.
- [2] T. C. Dang, T. B. Nguyen, T. Y. Nguyen, T. H. Trinh, and T. T. Banh, "Factors affecting the profitability of listed commercial banks in Vietnam: does agriculture finance matter?" *AgBioforum*, vol. 23, no. 1, pp. 32–41, 2021.
- [3] P. H. Vo and T. Q. Ngo, "The role of agricultural financing and development on sustainability: evidence from ASEAN countries," *AgBioforum*, vol. 23, no. 1, pp. 22–31, 2021.
- [4] S. Vyas-Doorgapersad, "Assessing gender equality in the South African public service," *International Journal of Social Sciences and Humanity Studies*, vol. 12, no. 2, pp. 607–623, 2020.
- [5] R. Akhtar, M. M. Masud, M. S. Uddin, and Q. M. A. Hye, "Underlying drivers that influence farmers' sustainable adaptation strategies," *International Journal of Management and Sustainability*, vol. 9, no. 3, pp. 181–193, 2020.
- [6] T. T. C. Nguyen, Q. B. Tran, T. M. P. Nguyen, D. L. Tran, A. T. Nguyen, and Q. T. Dang, "The effects of risks in agricultural supply chain management on business performance: a case study in vietnam," *AgBioforum*, vol. 21, no. 3, pp. 23–34, 2019.
- [7] S. Dlamini, "Citizens' satisfaction with the South African police services and community police forums in Durban, South Africa," *International Journal of Social Sciences and Humanity Studies*, vol. 12, no. 2, pp. 593–606, 2020.
- [8] I. U. Haq, A. Maneengam, S. Chupradit, W. Suksatan, and C. Huo, "Economic policy uncertainty and cryptocurrency market as a risk management avenue: a systematic review," *Risks*, vol. 9, no. 9, p. 163, 2021.
- [9] E.-C. Chang, S.-C. Huang, and H.-H. Wu, "Using K-means method and spectral clustering technique in an outfitter's value analysis," *Quality and Quantity*, vol. 44, no. 4, pp. 807–815, 2010.
- [10] K. Khalili-damghani, F. Abdi, and S. Abolmakarem, "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case of customer-centric industries," *Applied Soft Computing*, vol. 73, pp. 816–828, 2018.
- [11] Jiawei Han, Jian Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, Amsterdam, Netherlands, 2011.
- [12] D. Adams and T. Krulicky, "Artificial intelligence-driven big data analytics, real-time sensor networks, and product decision-making information systems in sustainable manufacturing internet of things," *Economics, Management, and Financial Markets*, vol. 16, no. 3, pp. 81–93, 2021.
- [13] A. Gordon, "Internet of things-based real-time production logistics, big data-driven decision-making processes, and industrial artificial intelligence in sustainable cyber-physical manufacturing systems," *Journal of Self-Governance and Management Economics*, vol. 9, no. 3, pp. 61–73, 2021.
- [14] V. Konecny, C. Barnett, and M. Poliak, "Sensing and computing technologies, intelligent vehicular networks, and big data-driven algorithmic decision-making in smart sustainable urbanism," *Contemporary Readings in Law and Social Justice*, vol. 13, no. 1, pp. 30–39, 2021.
- [15] E. Nica and V. Stehel, "Internet of things sensing networks, artificial intelligence-based decision-making algorithms, and real-time process monitoring in sustainable industry 4.0," *Journal of Self-Governance and Management Economics*, vol. 9, no. 3, pp. 35–47, 2021.
- [16] A. Goli and B. Malmir, "A covering tour approach for disaster relief locating and routing with fuzzy demand," *International Journal of Intelligent Transportation Systems Research*, vol. 18, no. 1, pp. 140–152, 2020.
- [17] A. Goli, H. K. Zare, R. Tavakkoli-Moghaddam, and A. Sadeghieh, "Hybrid artificial intelligence and robust optimization for a multi-objective product portfolio problem Case study: The dairy products industry," *Computers & Industrial Engineering*, vol. 137, Article ID 106090, 2019.
- [18] A. Goli, H. Khademi-Zare, R. Tavakkoli-Moghaddam, A. Sadeghieh, M. Sasanian, and R. Malekalipour Kordestanizadeh, "An integrated approach based on artificial intelligence and novel meta-heuristic algorithms to predict demand for dairy products: a case study," *Network: Computation in Neural Systems*, vol. 32, no. 1, pp. 1–35, 2022.
- [19] E. Cunningham, "Artificial intelligence-based decision-making algorithms, sustainable organizational performance, and automated production systems in big data-driven smart urban economy," *Journal of Self-Governance and Management Economics*, vol. 9, no. 1, pp. 31–41, 2021.
- [20] A. Novak, D. Bennett, and T. Kliestik, "Product decision-making information systems, real-time sensor networks, and artificial intelligence-driven big data analytics in sustainable Industry 4.0," *Economics, Management, and Financial Markets*, vol. 16, no. 2, pp. 62–72, 2021.
- [21] J. Vrbka and Z. Rowland, "Using artificial intelligence in company management," in *Proceedings of the International Scientific Conference "Digital Transformation of the Economy: Challenges, Trends, New Opportunities*, pp. 422–429, Springer, Cham, Switzerland, 2019.
- [22] E. Grant, "Big data-driven innovation, deep learning-assisted smart process planning, and product decision-making information systems in sustainable industry 4.0," *Economics, Management, and Financial Markets*, vol. 16, no. 1, pp. 9–19, 2021.
- [23] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *Journal of King Saud University-Computer and Information Sciences*, 2019, In press.
- [24] H.-H. Wu, E.-C. Chang, and C.-F. Lo, "Applying RFM model and K-means method in customer value analysis of an outfitter," in *Global Perspective for Competitive Enterprise, Economy and Ecology*, pp. 665–672, Springer, London, 2009.
- [25] S. Monalisa, P. Nadya, and R. Novita, "Analysis for customer lifetime value categorization with RFM model," *Procedia Computer Science*, vol. 161, pp. 834–840, 2019.
- [26] V. L. Miguéis, A. S. Camanho, and J. F. e Cunha, "Customer data mining for lifestyle segmentation," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9359–9366, 2012.

- [27] A. Rabiei and H. Rastegari, "Integrating RFM and classification for response modeling based on customer lifetime value," *Fen Bilimleri Dergisi (CFD)*, vol. 36, no. 4, 2015.
- [28] R. Mohammadi, B. Bidabad, T. Nourasteh, and M. Sherafati, "Credit ranking of bank customers (an integrated model of RFM, FAHP and K-means)," *European Online Journal of Natural and Social Sciences*, vol. 3, no. 3, p. 564, 2014.
- [29] A. Christy, A. Joy, L. P. Umamakeswari, and A. Neyaa, "RFMr ranking an effective approach to customer segmentation," *J. king Saudi university. Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1251–1257, 2018.
- [30] R. Taghi Livari and N. Zarrin Ghalam, "Customers grouping using data mining techniques in the food distribution industry (a case study)," *SRPH Journal of Applied management and Agile Organisation*, vol. 3, no. 1, pp. 1–8, 2021.
- [31] D. H. Nguyen, S. de Leeuw, and W. E. H. Dullaert, "Consumer behaviour and order fulfilment in online retailing: a systematic review," *International Journal of Management Reviews*, vol. 20, no. 2, pp. 255–276, 2018.
- [32] D. Arunachalam and N. Kumar, "Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data-driven decision-making," *Expert Systems with Applications*, vol. 111, pp. 11–34, 2018.
- [33] S. Shokouhyar, S. Shokoohyar, and S. Safari, "Research on the influence of after-sales service quality factors on customer satisfaction," *Journal of Retailing and Consumer Services*, vol. 56, Article ID 102139, 2020.
- [34] E. Ernawati, S. S. K. Baharin, and F. Kasmin, "A review of data mining methods in RFM-based customer segmentation," *Journal of Physics: Conference Series*, vol. 1869, no. 1, Article ID 012085, 2021.
- [35] E. Purwanto, Z. A. Long, and F. Mohd, "Clustering of potential customer in digital trade transaction," *Journal of Physics: Conference Series*, vol. 1918, no. 4, Article ID 042138, 2021.
- [36] F. Ciccullo, R. Cagliano, G. Bartezzaghi, and A. Perego, "Implementing the circular economy paradigm in the agri-food supply chain: the role of food waste prevention technologies," *Resources, Conservation and Recycling*, vol. 164, Article ID 105114, 2021.
- [37] M. Donner, A. Verniquet, J. Broeze, K. Kayser, and H. De Vries, "Critical success and risk factors for circular business models valorising agricultural waste and by-products," *Resources, Conservation and Recycling*, vol. 165, Article ID 105236, 2021.
- [38] Y. Wang, X. Ma, Y. Lao, and Y. Wang, "A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization," *Expert Systems with Applications*, vol. 41, no. 2, pp. 521–534, 2014.
- [39] C. B. Pocol, V. Marinescu, D. C. Dabija, and A. Amuza, "Clustering Generation Z university students based on daily fruit and vegetable consumption: empirical research in an emerging market," *British Food Journal*, vol. 123, no. 8, pp. 2705–2727, 2021.
- [40] J.-T. Wei, S.-Y. lin, C.-C. weng, and H.-H. Wu, "A case study of applying LRFM model in market segmentation of a children's dental clinic," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5529–5533, 2012.
- [41] N. V. Florea and A. Duica, "Improving communication and relationship with customers using models to measure their value," *Valahian Journal Of Economic Studies*, vol. 8, no. 1, 2017.
- [42] V. Kumar, *Customer Lifetime Value: The Path to Profitability*, Now Publishers Inc, 2008.
- [43] L. Stanca, D. C. Dabija, and E. Păcurar, "Community of practice: converting IT graduate students into specialists via professional knowledge sharing," *Kybernetes*, vol. 51, no. 2, pp. 557–581, 2021.
- [44] S. Bratu, "Can social media influencers shape corporate brand reputation? Online followers' trust, value creation, and purchase intentions," *Review of Contemporary Philosophy*, no. 18, pp. 157–163, 2019.
- [45] A. L. Drugău-Constantin, "Is consumer cognition reducible to neurophysiological functioning?" *Economics, Management, and Financial Markets*, vol. 14, no. 1, pp. 9–15, 2019.
- [46] J. C. Hollowell, Z. Rowland, T. Kliestik, J. Kliestikova, and V. V. Dengov, "Customer loyalty in the sharing economy platforms: how digital personal reputation and feedback systems facilitate interaction and trust between strangers," *Journal of Self-Governance and Management Economics*, vol. 7, no. 1, pp. 13–18, 2019.
- [47] S. Graessley, J. Horak, M. Kovacova, K. Valaskova, and M. Poliak, "Consumer attitudes and behaviors in the technology-driven sharing economy: motivations for participating in collaborative consumption," *Journal of Self-Governance and Management Economics*, vol. 7, no. 1, pp. 25–30, 2019.