Journal of Cases on Information Technology
Volume 26 • Issue 1

Customer Segmentation Marketing Strategy Based on Big Data Analysis and Clustering Algorithm

Xiaotong Li, School of Design, Tongmyong University, Busan, South Korea
Young Sook Lee, School of Design, Tongmyong University, Busan, South Korea*

ABSTRACT

Traditional customer segmentation methods cannot obtain more effective information from massive customer data, which affects the formulation of marketing strategies. Based on this, this study constructs a customer segmentation marketing strategy model that integrates support vector machines and clustering algorithms. This model first utilizes support vector machines to segment existing customer data, and then integrates support vector machines and clustering algorithms to construct a customer segmentation model. Finally, simulation experiments are conducted using the dataset. The results show that the model algorithm obtains the optimal solution when the quantity of iterations is 50. Meanwhile, the average error rate of the model algorithm in the customer segmentation process is 6.82%, the average recall rate is 91.28%, and the average profit predicted by the impact strategy developed by the segmentation model is 29.88%, which is 2.53% different from the true value.

KEYWORDS

Big Data Analysis, Clustering Algorithm, Customer Segmentation, Marketing Strategy, Support Vector Machine

INTRODUCTION

Due to the internet and big-data technology, enterprises need more precise customer-segmentation marketing strategies to improve market share and customer loyalty in the face of fierce market competition. Traditional customer segmentation methods cannot respond effectively and in a timely manner to a large amount of customer data, thereby affecting the braking of marketing strategies (Othayoth et al., 2022; Weking et al., 2020). Support vector machine (SVM) and K-means clustering algorithm (CA) are widely used technologies in the fields of machine learning and data mining, respectively. SVM is a nonlinear classification algorithm, while K-means algorithm (KMA) is a classic CA. To better address customer-segmentation issues, this study combined SVM and KMA to enhance the accuracy of customer classification. The method of integrating SVM and KMA can utilize the nonlinear classification ability of SVM and the data-clustering ability of K-means to better discover patterns hidden in massive data (Widyawati et al., 2020; Corpuz, 2021; Cui et al., 2021; Fakhriza, 2021).

DOI: 10.4018/JCIT.336916
*Corresponding Author

This study aims to explore customer-segmentation marketing strategies that integrate SVM and KMA and analyze their feasibility and effectiveness in practical applications. First, SVM is used to segment existing customer data, and then SVM and KMA are fused to construct a customer-segmentation model. The model is used to process and analyze these data to discover different customer groups and features. Finally, customers can be divided into different subgroups, such as high-value customers, potential customers, low-value customers, etc. The main contribution of the research is to apply SVM and KMA to customer-segmentation marketing strategies to improve the accuracy and efficiency of customer classification. This will help enterprises comprehend customer requirements and improve market competitiveness.

The first part of the study introduces the current research status of customer data dimensionality reduction, clustering, and customer segmentation. The second part studies a customer-segmentation marketing-strategy model that integrates SVM and K-means CA. The third part verifies the performance of the constructed model through simulation experiments and practical applications. The fourth part summarizes the relevant outcomes and analyzes the merits and demerits of the research methods used.

RELATED WORK

Customer-segmentation marketing strategies play an important role in improving market competitiveness, reducing marketing costs, enhancing marketing effectiveness, and strengthening customer-relationship management. By segmenting customer groups, enterprises can better meet customer needs, achieve sustained growth, and achieve sustainable competitive advantages. To better leverage customer-segmentation marketing strategies, many scholars have analyzed and studied relevant sales data. Scholars such as Sokol and Holý (2021) have utilized data clustering analysis techniques to analyze customer behavior and value in the retail industry. This study obtained information on shopping proximity, frequency, and purchasing power by segmenting customers and applied data-clustering analysis to a chain pharmacy. The results indicate that this method can bring more customer needs to the attention of merchants.

Nikaein and Abedin (2021) constructed a data-mining method based on a radio frequency machine learning model for enhancing the efficiency of marketing and reducing costs during the marketing process and applied it to the pharmaceutical industry. The results indicate that this model can help sales managers more effectively plan for each customer, improve visit efficiency, and lower costs.

To reduce customer churn in potential customer orders, Fitriani and Febrianto (2021) compared data-mining methods such as naive Bayes, random forests, and SVM. They used these to obtain data-feature information about potential customers to eliminate the problem of category imbalance in the marketing process of banks. The results indicate that random forests have high mining ability, with a maximum accuracy of 92.61%.

To better segment customers, Ahani et al. (2019) constructed a method based on machine learning for hot spring hotel segmentation and travel-choice prediction, thereby helping hotels develop more efficient marketing strategies. This method classifies and processes customer-evaluation information on hotel websites to obtain marketing-related data. The results indicate that this selection-prediction method can obtain a large amount of data from social media and help hot-spring hotels better push advertisements to target customers.

To find potential customers for online stores through customer segmentation, Nurmalasari et al. (2020) constructed a clustering method based on KMA. This method involves clustering analysis of existing customer-browsing information to find the best clustering data and prepare for customer segmentation. The results indicate that applying cluster analysis to customer segmentation can provide improved sales strategies for each group of customers in online stores.

Bekamiri et al. (2020) proposed a dynamic model to better allocate more-effective promotion strategies to customer groups using the model to calculate customer lifetime value. Banks can propose corresponding product-sales strategies by analyzing and calculating customer lifetime value. The results indicate that the model has significantly improved the product-sales process of some banks in the Middle East and North Africa.

In order to improve the performance of real-time input and collection of medical-visit data, Rodger (2015) proposed a research methodology based on the correspondence of a set of variables in relation to bodily injury. The method utilizes the relationship between three variables and the head, trunk, limbs, and abrasions of the body as a means of determining their association and analyzing them through survival, mortality, and morbidity. The results show that the methodology has a positive effect on future decision-making and planning.

Bangar and Chaudhary (2023) have proposed a method for optical coherence layer analysis in order to effectively identify the macular region in malignant diabetes. The method provides effective segmentation of the graph through a K-means clustering algorithm and uses a wavelet transform to enhance the image feature extraction capability on the basis of segmentation. The results show that the recognition accuracy of this method in the macular region can reach 94.46%. Song et al. (2023) proposed a classification model based on K-means with a PCA algorithm in order to improve the performance of the classification model. The study first utilizes a clustering algorithm to effectively classify the research object, and then PCA is utilized for data-dimensionality reduction. The results show that this method can significantly improve the robustness and performance of the classification model.

Hu et al. (2023) proposed a Buckling Restrained Brace (BRB) structure based on extreme gradient augmentation feature selection in order to improve the performance of the BRB system for solving an imbalance classification task. The study processes the imbalance task through different structural layers, first utilizing the main anti-buckling support, then utilizing extreme gradient augmentation for feature selection of different submodels, and finally utilizing simulation experiments to verify the model performance. The results show that the structure is able to quasi-transform the multiclassification problem into a biclassification problem, thus alleviating the imbalance of the data, which enhances the solving ability of the imbalance problem.

In summary, customer-segmentation marketing strategies are crucial for a company's market competitiveness and customer satisfaction. By accurately positioning the target market, improving marketing effectiveness, optimizing resource allocation, and enhancing customer satisfaction and loyalty, enterprises can achieve better market performance. Due to the challenges of large data volume and diversity, traditional customer-segmentation methods often cannot effectively handle them. Therefore, based on existing technologies, this study integrates SVM and K-means CA for constructing customer-segmentation marketing-strategy models with the goal of providing better technical support for the development of enterprises.

CONSTRUCTING A CUSTOMER-SEGMENTATION MARKETING-STRATEGY MODEL THROUGH THE FUSION OF SVM AND K-MEANS CA

The customer-segmentation marketing strategy that integrates SVM and K-means CA is a method of using machine-learning algorithms to more accurately divide different customer groups and implement personalized marketing strategies. This customer-segmentation marketing strategy that integrates SVM and K-means CA can help enterprises better understand customer needs and behavior patterns, thereby allowing them to formulate personalized marketing strategies to improve marketing effectiveness and customer satisfaction.

Research on Customer-Segmentation Marketing Based on Support Vector Machine

There are many factors to consider in customer-classification issues. Customer value is an important classification standard, which can be determined by the net present value evaluation system (NPVES). The NPVES is the sum of the gross profit of each transaction minus the cost, which is used to evaluate the value of customers. However, this system has some shortcomings, such as neglecting the intangible contributions brought by sales volume and customers. To improve customer-classification methods, sales volume and net present value can be considered as influencing factors and the experience of sales personnel can be referenced. Sales volume can be measured by the number of units purchased by customers, while net present value can be measured by purchase price. Taking into account both sales volume and net present value can serve as the basis for customer-value classification (Burhan et al., 2020; Bilgic et al., 2021; Monil et al., 2020). The sales volume indicator in customer value can be calculated using (1).

$$X = P_1 * W_1 + (1 - P_1) * W_2 \tag{1}$$

In (1), P_1 serves as the ratio of purchase quantity, W_1 serves as the coefficient greater than the threshold, and W_2 represents the coefficient corresponding to the threshold. The net present value indicator in customer value can be represented by (2).

$$Y = P_2 * W_3 + (1 - P_2) * W_4 \tag{2}$$

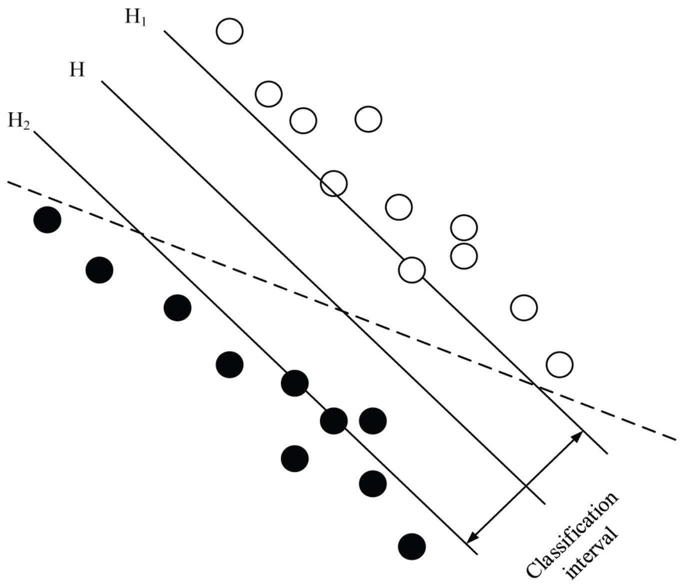
In (2), P_2 represents the ratio of the purchase unit price of the order, W_3 represents the ratio greater than the purchase unit price, and W_4 represents the ratio equal to the purchase unit price.

In customer-segmentation marketing, enterprises need to divide customers into different segmentation groups based on their characteristics and behaviors and carry out targeted marketing strategies for different segmentation groups. Traditional subdivision methods rely mainly on statistical analysis and CA, but these methods have certain limitations when dealing with high-dimensional, nonlinear, and unstructured data (Jain et al., 2023; Paranjape et al., 2021). In research on customer-segmentation marketing based on SVM, it is first necessary to preprocess and select customer-feature data, such as selecting appropriate feature variables, data cleaning, and data normalization. Then the SVM algorithm is used to classify and predict customers, dividing them into different subgroups on the basis of their characteristics and behaviors. The SVM algorithm is based on the concept of support vectors and reaches classification by finding an optimal hyperplane in the feature space. It can handle nonlinear problems through the selection of kernel functions (KFs) and achieve good generalization ability in small datasets. Customer-segmentation marketing research could help enterprises comprehend customer requirements and behaviors, formulate targeted marketing strategies, and improve marketing effectiveness and customer satisfaction. The linear separable optimal interface of SVM is shown in Fig. 1.

In Fig. 1, solid points and hollow points serve as two types of samples. H is the classification hyperplane (CH), and H_1 and H_2 are the planes closest to and parallel to the CH in each type of sample. The distance between them is called the classification interval (CI). The optimal classification surface not only needs to correctly separate the two types of samples, but also needs to maximize the CI. Although dashed lines can also separate the two types of samples, the CI is smaller than H. Support vectors are the vectors closest to the optimal CH. To better classify customers, this study used *n* indicators to participate in the classification work with *m* training samples. It predicts classification rules on the basis of these samples and applies them to the classification prediction of new data samples. This study can consider *m* as points in an *n*-dimensional space. If these points can be separated, the type of new data can be determined on the basis of the sign function, which can be expressed by (3).

$$f(x) = \mathrm{sgn}(\bar{\omega} X + b) \tag{3}$$

Figure 1. Linearly differentiable optimal interface graphs for SVM



In (3), $\bar{\omega}$ represents the corresponding coefficient value in the hyperplane, X represents the corresponding value in the real number set (RNS), and b represents the coefficient value in the RNS. The optimal hyperplane not only separates the training data but also maximizes the distance between the vector closest to the hyperplane and the hyperplane to enhance the generalization ability of classification methods. The research needs to solve the optimal hyperplane, that is, the constrained optimization problem with the minimum Lagrange function, which can be represented by (4).

$$\varphi(\bar{\omega}) = \frac{1}{2}(\bar{\omega} \cdot \bar{\omega}) - \sum_{i=1}^{n} a_i (y_i (\bar{\omega} \cdot x_i + b) - 1) \tag{4}$$

In (4), a_i serves as the support-vector value, y_i represents the customer-classification category, and x_i represents the classification index value. When a_i is greater than the coefficient value of the Lagrange function, the obtained value is the minimum value (MVA) of the function. After obtaining the MVA of the Lagrange function, customer classification is transformed into a relatively simple dual problem (DP), which can be represented by (5).

$$\max_{ai} Q(a) = \sum_{i=1}^{m} a_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i y_i x_i \tag{5}$$

After obtaining the value of the DP, the optimal classification surface function value can be calculated using (6).

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{m} a_i^* y_i (x_i \cdot x) + b^*) \tag{6}$$

In (6), $a_i^* \neq 0$ represents the support vector of the sample point and b^* represents the classification threshold in the RNS. In customer classification, if the data is linearly nonseparable, nonlinear transformations are needed to address the linear partitioning issue of hyperplanes in high-dimensional spaces. When calculating the classification function, if there is a KF, the dot product in the original space can be used to avoid interference factors. Different KFs can generate different SVMs, and commonly used KFs include linear KF, polynomial KF, radial basis KF, and sigmoid KF. Due to the superior applicability and classification performance of the radial basis function kernel compared to the other three functions, this study chose the radial basis function kernel as the function of the SVM, which can be represented by (7).

$$K(x,y) = \exp\left|\frac{\left\|x-y\right\|^2}{2\sigma^2}\right| \tag{7}$$

In (7), σ serves as the coefficient value of the radial basis KF. The specific dimension diagram in customer segmentation is shown in Fig. 2.

Construction of a Marketing-Strategy Model Integrating Support Vector Machine and CA

Combined with the above research, support vector machine in customer segmentation marketing there are slightly lower accuracy and model update difficulty, the research in the support vector machine-based customer segmentation marketing strategy research based on the introduction of K-means clustering algorithm, fusion of the two used in the construction of the model. The customer-segmentation marketing-strategy model that integrates SVM and K-means CA combines their merits, which can improve segmentation accuracy and enhance the robustness of the model. K-means CA is a commonly used unsupervised learning algorithm that can divide a dataset into K non-overlapping clusters, each with similar features within the data points. In the research on customer-segmentation marketing strategies, the K-means CA can be used to divide customers into different groups to design targeted marketing strategies for each group. Fig. 3 showcases the clustering of the K-means CA.

Figure 2. Customer-Segmentation dimension chart

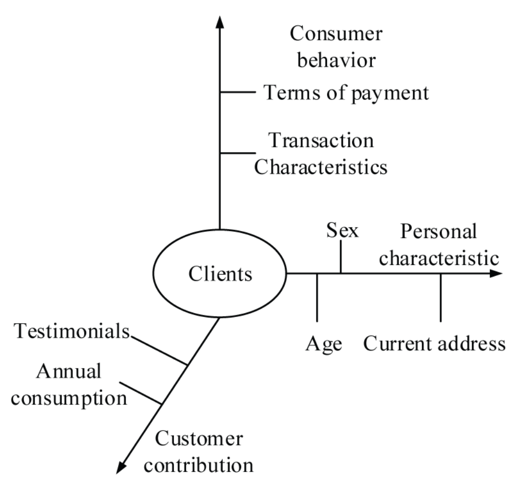
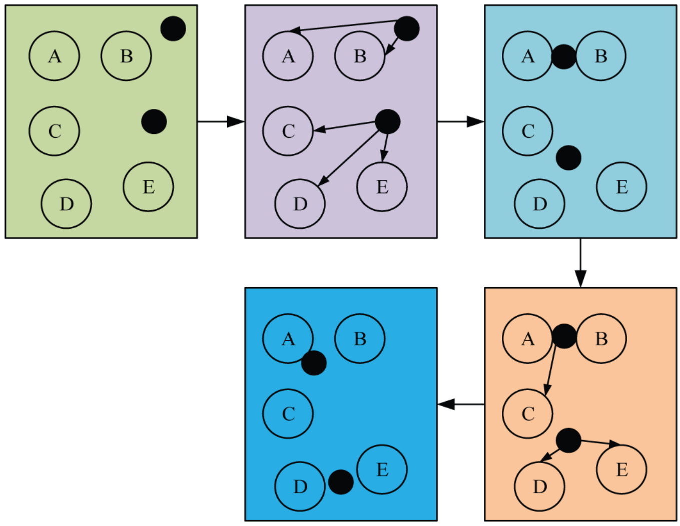


Figure 3. K-Means CA clustering process diagram



Clustering is a challenging research field, and its potential applications pose unique requirements. CA is an unsupervised learning method that discovers the internal structure of data by grouping similar data points. Among them, KMA is a commonly used CA that calculates the distance between data points for clustering. However, the ability of KMA to handle different types of attributes depends on the calculation method of distance and the processing of different types of data. For numerical data, the commonly used distance calculation method is Euclidean distance. However, the KMA also has some demerits. One is that the KMA is very sensitive to outliers. If there are outliers in the dataset, they may have a significant impact on the clustering results, leading to inaccurate clustering results. The second issue is that the KMA cannot handle nonlinear separable data. When the distribution of the dataset is complex, the KMA may not be able to find a suitable clustering center (CCE), resulting in unsatisfactory clustering results.

To solve the problems of KMA, SVM can be combined with KMA. The data can be divided into multiple groups, and the CCE of each group can be determined to minimize the objective function (OFU) of non-similar indicators. Then Euclidean distance can be used as the non-similar indicator to obtain the optimal CCE by minimizing the OFU (Nurfalah et al., 2021; Huang et al., 2022; Barman & Chowdhury, 2019). By combining the nonlinear classification ability of SVM with the clustering ability of KMA, more accurate and robust clustering results can be obtained. The OFU can be defined by (8).

$$J = \sum_{i=1}^{c} Ji = \sum_{i=1}^{c} \left(\sum_{xk \in Gi} \left\| x_k - c_i \right\|^2 \right) \tag{8}$$

In (8), $Ji = \sum_{i=1}^{c} \left(\sum_{xk \in Gi} \left\| x_k - c_i \right\|^2 \right)$ represents the value function within group i. The value of the value function depends on the characteristics and location of *Gi*. After obtaining the value function, using the distance function to replace the vectors in the group can obtain the corresponding total value function value, which can be represented by (9).

$$J_z = \sum_{i=1}^{c} Ji = \sum_{i=1}^{c} (\sum_{xk \in Gi} d(x_k - c_i)^2) \tag{9}$$

In (9), x_k represents the horizontal value of the distance function and c_i denotes the vertical value of i in the distance function. Define the divided customer information groups using a two-dimensional membership matrix, so that the jth data value x_j belongs to the i group. Therefore, the value of the element in the matrix is 1; otherwise it is 0. If c_i is the nearest CCE of xj, then xj belongs to group i. Due to the fact that a given customer can only belong to one group, the membership matrix can be defined using (10).

$$\begin{cases} \sum_{i=1}^{c} u_{ij} = 1, \quad \forall j = 1, \cdots, n \\ \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} = n \end{cases} \tag{10}$$

In Equation (10), u_{ij} denotes the j-th element value corresponding to the i-th group in the affiliation matrix u. If u_{ij} is fixed, the smallest optimal CCE in (8) is the average of all vectors in group i, which can be represented by (11).

$$c_i = \frac{1}{\left|G_i\right|} \sum_{xk \in Gi} x_k \tag{11}$$

Due to the iterative nature of the KMA, it cannot be guaranteed that it converges to the optimal solution (OSO). To ensure that the OSO can be obtained, a good initial CCE is usually obtained before stacking. This requires the use of SVM to segment customer data, providing a different initial CCE and obtaining the OSO through multiple operations. On the basis of SVM, the nearest CCE obtained using the KMA can be corrected using (12).

$$\Delta c_i = \eta (x - c_i) \tag{12}$$

In (12), η represents the correction coefficient. The process of customer segmentation integrating SVM and KMA is shown in Fig. 4.

Fig. 4 shows that after clustering the training results of SVM using the KMA, they are further sparsification into new support vectors. At this time, the compression ratio of the model is smaller, the number of support vectors is also smaller, and the time consumed in the customer-segmentation process is also less, resulting in faster classification time. The overall process of customer segmentation is shown in Fig. 5.

APPLICATION ANALYSIS OF CUSTOMER-SEGMENTATION MARKETING-STRATEGY MODEL

For testing the constructed customer-segmentation model, this study obtained customer data from an e-commerce company through a market sales platform and used it to construct a dataset. The main indicators in this dataset include customer ID, gender, age, income, purchase frequency, and purchase amount. By constructing a customer-segmentation model, customers can be segmented

Figure 4. Flowchart of customer classification by fusing SVM and KMA

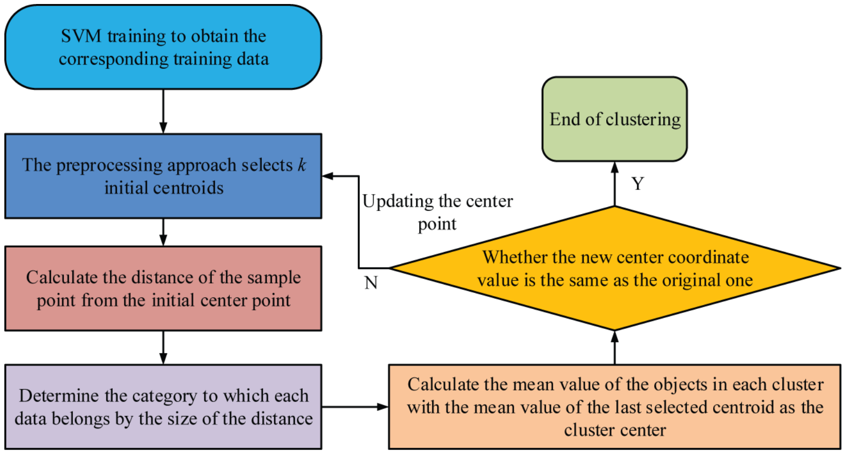
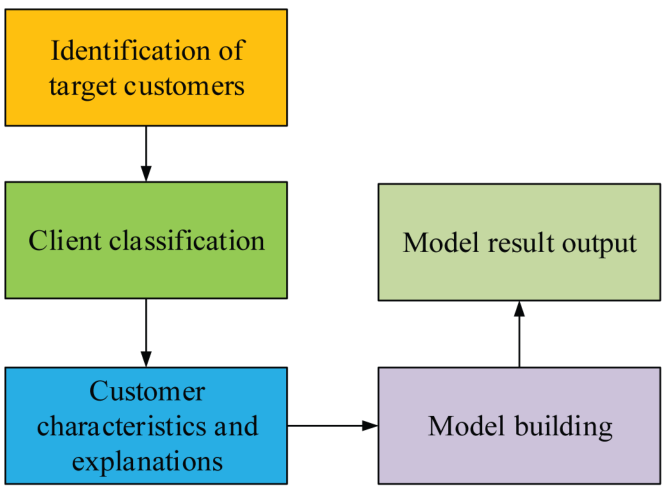


Figure 5. Flowchart for building models in customer segmentation



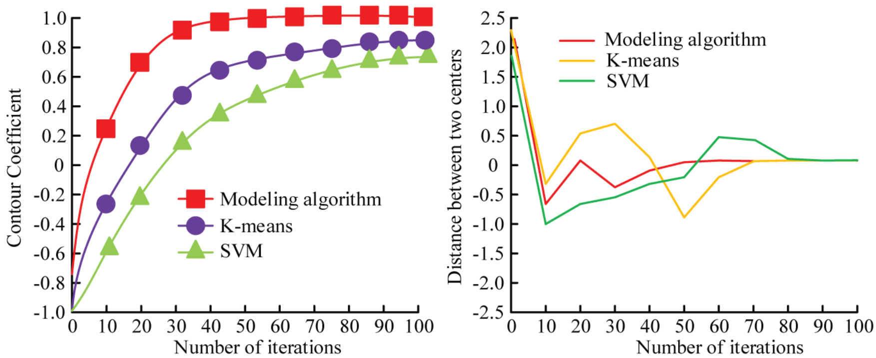
based on indicators such as gender, age, income, and purchase behavior frequency and amount, and corresponding sales strategies can be formulated.

Performance Analysis of Customer-Segmentation Marketing-Strategy Model

To test the constructed model, this study compared SVM with KMA and compared them with the model method. The contour coefficient, clustering performance, detection rate, error rate (ERA), accuracy, recall rate, and F1 value of the model for data processing were used as comparison indicators, which were used to evaluate the compactness and separation of clustering results. The value range was between [-1,1], and the closer to 1, the more excellent the clustering effect. Fig. 6 showcases the comparison results of contour coefficients and clustering performance between the model algorithm (MAL) and the comparison algorithm.

Fig. 6 (a) shows that the contour coefficient of the MAL is closer to 1, while the KMA is further away from the coefficient value of SVM by 1. This indicates that the clustering effect of the MAL on data is superior to that of the two algorithms compared. Figure 6 (b) shows that as the quantity of

Figure 6. Comparison results of profile coefficients and clustering performance of three algorithms



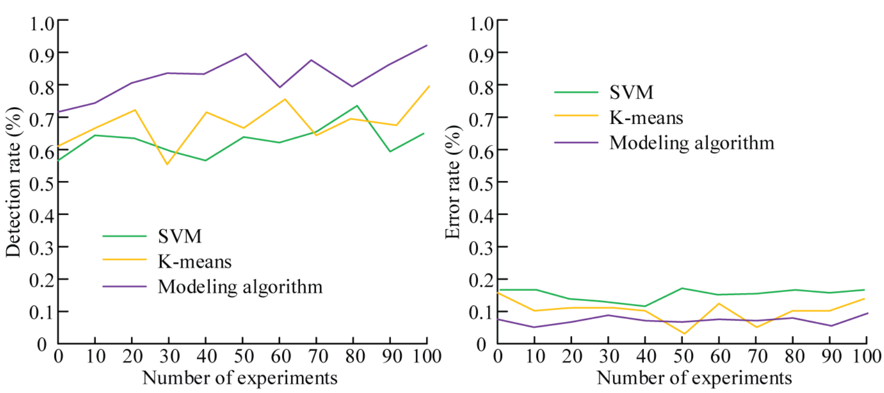
(a) Comparison of detection performance of three algorithms

(b) Clustering performance comparison of three algorithms

iterations grows, the distance between the virtual center point and the real center point of the three algorithms also changes. When the iteration number of the MAL is 50, the distance between the two centers is 0 and tends to stabilize. When the KMA and SVM have an iteration count of 68 and 79, respectively, the distance between the two centers is 0 and tends to stabilize. This indicates that the model has better performance in the process of customer data segmentation. For testing the detection ability in customer data processing, the above method was also used for comparison. The comparison indicators were detection rate and ERA, and the relevant outcomes are showcased in Fig. 7.

Fig. 7 (a) shows that the average detection rate (ADR) of the MAL is 89.22%, the ADR of K-means is 75.06%, and the ADR of SVM is 61.43%. As shown in Fig. 7 (b), the ERA of the MAL is the lowest, with an average of 6.82%. The average ERA of K-means and SVM are 10.16% and 17.35%, respectively. Compared with K-means and SVM, the average ERA of the MAL is 3.34% and 10.53% below that of K-means and SVM. This indicates that the MAL has the highest detection rate and the lowest ERA in the dataset. To further validate the performance of the model in data-dimensionality

Figure 7. Comparison of detection rate and ERA of three algorithms



(a) Detection rate

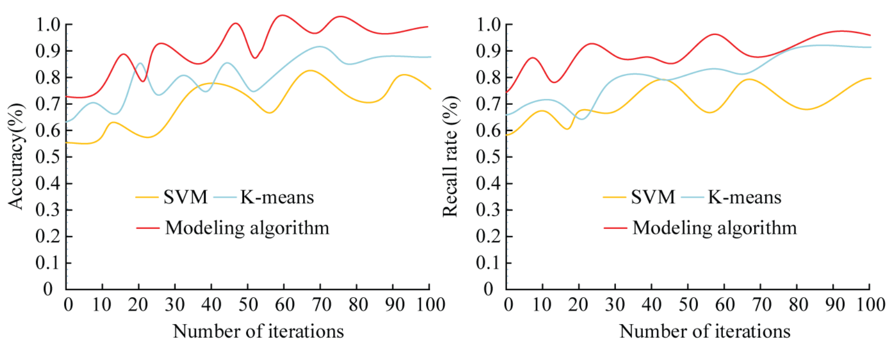
(b) Error rate

reduction and clustering, this study compared accuracy and recall as measurement indicators, and the comparison results are shown in Fig. 8.

Figure 8 (a) shows that the average accuracy of the MAL is 97.51%, while the average accuracy of K-means and SVM algorithms are 81.08% and 72.36%, respectively. Figure 8 (b) illustrates that the recall rates of all three algorithms first increase and then tend to stabilize. The MAL has the highest recall rate, with an average of 91.28%. The average recall rates of the K-means and SVM algorithms are 82.59% and 71.88%, respectively. This indicates that the performance of the MAL in customer-data segmentation is significantly superior to that of the other two algorithms. The F1 value is an indicator that comprehensively considers accuracy and recall and can be used to evaluate the comprehensive performance of classification models. The F1 value comparison results in customer-data processing are showcased in Fig. 9.

Fig. 9 demonstrates that the F1 average value of the model method is the highest, with an F1 average value of 0.92. The F1 average values of the K-means and SVM algorithms are 0.81 and 0.73, respectively. The model method is significantly higher than the F1 average values of the K-means and SVM algorithms. This indicates that from the F1 value, the performance of the model method

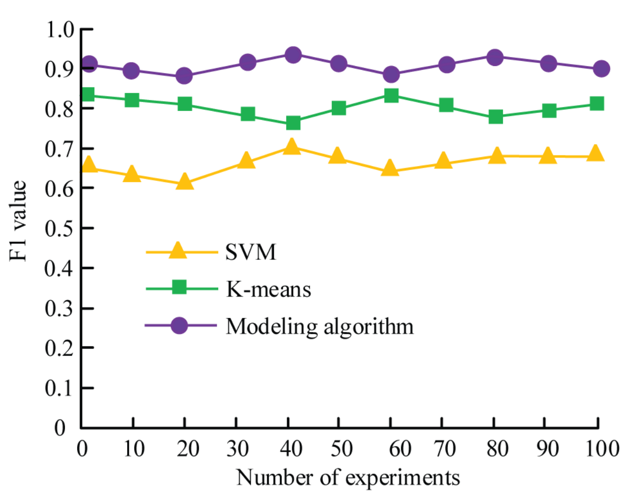
Figure 8. Comparison results of precision and recall of the three methods



(a) Comparison of the accuracy of the three algorithms

(b) Recall of the three methods

Figure 9. Comparison results of F1 values for the three algorithms



is superior to that of the K-means and SVM algorithms. This verifies the capability of the research construction model for extracting and analyzing data in customer-data segmentation.

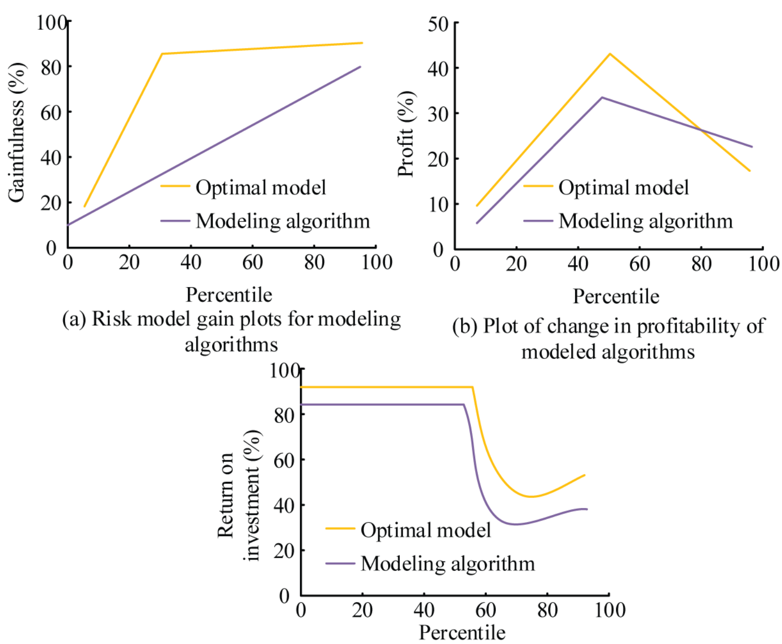
Analysis of the Application Effect of the Customer-Segmentation Marketing-Strategy Model

For testing the application effect of the customer-segmentation model in actual marketing strategies, the MAL was compared with the optimal model, and financial gain, profit, and return were used as validation indicators. The comparison results are shown in Fig. 10.

Fig. 10 (a) indicates that although the trend of the MAL is different from that of the optimal model, the difference in gain rate between the MAL and the optimal model is not significant in the end. The average gain rate of the optimal model is 94.29%, while the gain rate of the MAL is 89.79%. The difference in gain rates between the two is 4.50%, which is in line with the required range of gain rates in practical applications. Fig. 10 (b) demonstrates that the average profit margin of the optimal model reaches 32.41%, while the average profit margin of the MAL is 29.88%, with a difference of 2.53% between the two. Fig. 10 (c) demonstrates that the average return on investment of the optimal model reaches 78.91% and the average return on investment of the MAL is 69.57%. This indicates that the customer-segmentation model constructed through research can provide assistance to sales personnel in the formulation of actual marketing strategies. To further validate the application of the customer-segmentation model in actual marketing strategies, this study divided customers into high-value customers, potential customers, and retain customers. The model was used to classify the relevant data information of the three types of customers and develop marketing strategies that correspond to the customer groups. The comparison results of customer attention to the product after applying the developed marketing strategy are shown in Fig. 11.retain

Fig. 11 (a) demonstrates that in the prediction of high-value customers, the disparity between the model method and the true value in the training set is not significant. The true attention of high-value customers to the product is 88.39%, and the prediction observation of the model method is 85.02%,

Figure 10. Response of customer segmentation models in practical marketing strategies

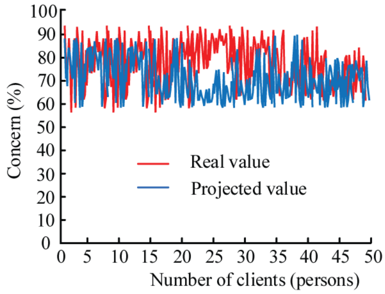


(a) Risk model gain plots for modeling algorithms

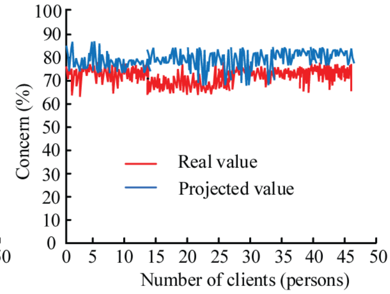
(b) Plot of change in profitability of modeled algorithms

(c) Return on investment of modeling algorithms

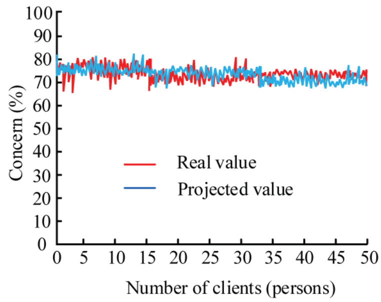
Figure 11. Comparison of the concerns of different customer groups after the implementation of marketing strategies



(a) High-value customer forecasting results



(b) Potential customer forecasting results



(c) Forecasted results of retaining lost customers

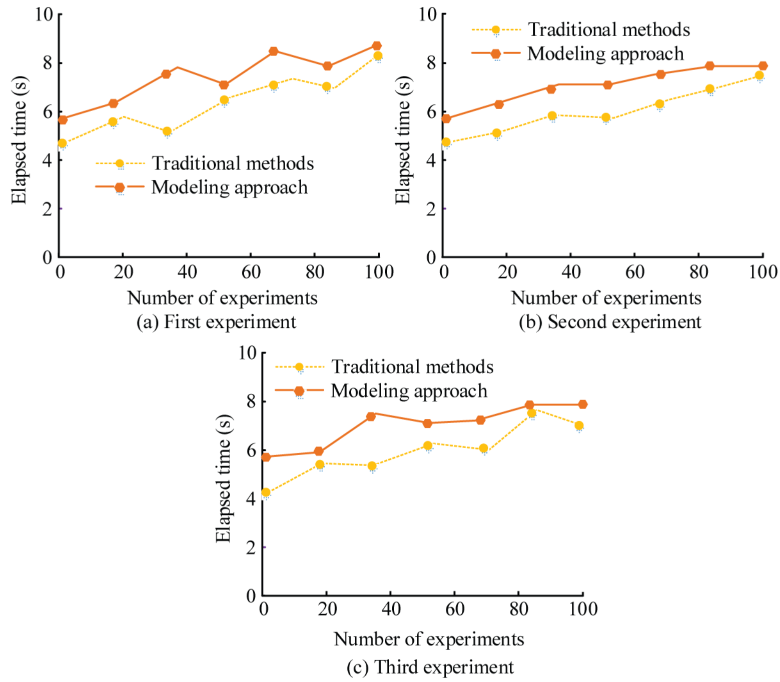
with a difference of 3.37%. Fig. 11 (b) demonstrates that the actual attention of potential customers to the product is 72.93%, and the predicted attention guided by corresponding marketing strategies is 79.57%, which has increased the attention by 6.64%. Fig. 11 (c) illustrates that the true attention level of lost customers to the product is 69.98% and the attention level guided by corresponding marketing strategies is 76.21%, an increase of 6.23%. This indicates that the marketing strategy developed through MALs for customer segmentation can increase the attention of potential customers and retain some customers who are about to be lost. To verify the operational efficiency of the MAL in customer segmentation, this study compared traditional methods with the MAL based on the model running time as an indicator. To reduce experimental errors, a total of three experiments were carried out, and the comparison results are showcased in Fig. 12.

The comparative analysis in Fig. 12 demonstrates that the average time spent on three runs of the MAL is 4.9 seconds, while the average time spent on three runs of traditional methods is 7.9 seconds. The results indicate that, from the perspective of running time, the performance of the MAL is markedly better than that of the comparison algorithm, verifying that the pattern algorithm proposed in the study has high running efficiency in customer segmentation.

CONCLUSION

This research studied customer-segmentation marketing strategies by integrating SVM and KMA. By clustering customer data, customers can be divided into different subgroups and corresponding marketing strategies can be developed based on their characteristics. First, it uses SVM to reduce the dimensionality of customer data, then uses KMA for clustering and analyzing customer data, and finally integrates SVM and KMA to construct a segmentation model. The outcomes illustrate that the contour coefficient of the MAL is closer to 1, with an ADR of 89.22% and an average accuracy of 97.51%. The gain rate of the MAL is 4.50% lower than the true value, which satisfies the design

Figure 12. Comparative results of the elapsed time



needs. On the basis of the above analysis, the research and the construction model have high credibility and robustness in customer segmentation. The research results will have an important guiding role for the marketing decision-making and strategy formulation of enterprises and will help to improve customer satisfaction and market competitiveness. However, there are still shortcomings in the research. Due to the limitations of data sources, the computational resources of this study have been limited to a certain extent. In the future, it is necessary to further increase the source path of data and use a large amount of data to alleviate the limitations, thereby improving the practicality of the strategy.

AUTHOR'S NOTE

The authors report there are no competing interests to declare.

REFERENCES

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Scott, W. (2019). Market segmentation and travel choice prediction in spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management*, *80*, 52–77. doi:10.1016/j.ijhm.2019.01.003

Bangar, M., & Chaudhary, P. (2023). A novel approach for the classification of diabetic maculopathy using discrete wavelet transforms and a support vector machine. *AIMS Electronics and Electrical Engineering*, *7*(1), 1–13. doi:10.3934/electreng.2023001

Barman, D., & Chowdhury, N. (2019). A novel approach for the customer segmentation using clustering through self-organizing map. [IJBAN]. *International Journal of Business Analytics*, *6*(2), 23–45. doi:10.4018/IJBAN.2019040102

Bekamiri, H., Mehraeen, M., Pooya, A., & Sharif, H. (2020). A stochastic approach for valuing customers in banking industry: A case study. *Industrial Engineering & Management Systems*, *19*(4), 744–757. doi:10.7232/iems.2020.19.4.744

Bilgic, E., Cakir, O., Kantardzic, M., Duan, Y., & Cao, G. (2021). Retail analytics: Store segmentation using rule-based purchasing behaviors analysis. *International Review of Retail, Distribution and Consumer Research*, *31*(4), 457–480. doi:10.1080/09593969.2021.1915847

Burhan, G. F., & Mansur, A. (2020). Marketing strategy planning based on customer value. *PROZIMA (Productivity, Optimization and Manufacturing System Engineering), 4*(2), 29–40. 10.21070/prozima.v4i2.1309

Corpuz, R. S. A. (2021). Categorizing natural language-based customer satisfaction: An implementation method using support vector machine and long short-term memory neural network. *International Journal of Integrated Engineering*, *13*(4), 77–91. doi:10.30880/ijie.2021.13.04.007

Cui, H., Niu, S., Li, K., Shi, C., Shao, S., & Gao, Z. (2021). A K-means++ based user classification method for social e-commerce. *Intelligent Automation & Soft Computing*, *28*(1), 277–291. doi:10.32604/iasc.2021.016408

Fakhriza, M. D. (2021). Comparison of K-means method and fuzzy clustering algorithm in determining customer satisfaction test in delivery services. [Jurnal Informatika]. *JIKA*, *5*(2), 152–158. doi:10.31000/jika.v5i2.4511

Fitriani, M. A., & Febrianto, D. C. (2021). Data mining for potential customer segmentation in the marketing bank dataset. *JUITA: Jurnal Informatika*, *9*(1), 25–32. doi:10.30595/juita.v9i1.7983

Hu, G., He, W., Sun, C., Zhu, H., Li, K., & Jiang, L. (2023). Hierarchical belief rule-based model for imbalanced multi-classification. *Expert Systems with Applications*, *216*, 119451. doi:10.1016/j.eswa.2022.119451

Huang, J. Q., Guo, W. L., & Lee, C. Y. (2022). Dynamic evaluation model of the default risk of online loan borrowers based on K-means and SVM. *Recent Advances in Computer Science and Communications*, *15*(9), 1259–1266. doi:10.2174/2666255814666210618111901

Jain, K., & Saxena, A. (2023). Simulation on supplier side bidding strategy at day-ahead electricity market using ant lion optimizer. *Journal of Computational and Cognitive Engineering*, *2*(1), 17–27. doi:10.47852/bonviewJCCE2202160

Monil, P., Darshan, P., Jecky, R., Vimarsh, C., & Bhatt, B. R. (2020). Customer segmentation using machine learning. *International Journal for Research in Applied Science and Engineering Technology*, *8*(6), 2104–2108. doi:10.22214/ijraset.2020.6344

Nikaein, N., & Abedin, E. (2021). Customers' segmentation in pharmaceutical distribution industry based on the RFML model. *International Journal of Business Information Systems*, *37*(1), 29–44. doi:10.1504/IJBIS.2021.115071

Nurfalah, R., Riana, D., & Anton (2021). Identifikasi citra beras menggunakan algoritma multi-SVM dan neural network pada segmentasi K-means. *Jurnal RESTI, 5*(1), 55–62. 10.29207/resti.v5i1.2721

Nurmalasari, M., Mukhayaroh, A., Marlina, S., Hartini, S., Muryani, S., Sinnun, A., Nurajizah, S., & Adiwihardja, C. (2020). Implementation of clustering algorithm method for customer segmentation. *Journal of Computational and Theoretical Nanoscience*, *17*(2), 1388–1395. doi:10.1166/jctn.2020.8815

Othayoth, S. P., & Muthalagu, R. (2022). Customer segmentation using various machine learning techniques. *International Journal of Business Intelligence and Data Mining*, *20*(4), 480–496. doi:10.1504/IJBIDM.2022.123218

Paranjape, P. N., Dhabu, M. M., & Deshpande, P. S. (2021). dSubSign: Classification of instance-feature data using discriminative subgraphs as class signatures. *International Journal of Software Engineering and Knowledge Engineering*, *31*(7), 917–947. doi:10.1142/S0218194021500285

Rodger, J. A. (2015). Discovery of medical big data analytics: Improving the prediction of traumatic brain injury survival rates by data mining Patient Informatics Processing Software Hybrid Hadoop Hive. *Informatics in Medicine Unlocked*, *1*, 17–26. doi:10.1016/j.imu.2016.01.002

Sokol, O., & Holý, V. (2021). The role of shopping mission in retail customer segmentation. *International Journal of Market Research*, *63*(4), 454–470. doi:10.1177/1470785320921011

Song, B., Fan, W., & Zhang, S. (2023). Research on an integrated intelligent classification algorithm based on K-means PCA-RF machine learning. *Highlights in Science*. *Engineering and Technology*, *49*(7), 20–29.

Weking, P. N. M., Sudarma, M., & Saputra, K. O. (2020). User experience analysis of “Ayooklik.com” online store using use questionnaire in determining product segmentation. *Majalah Ilmiah Teknik Elektro*, *19*(6), 19–25. doi:10.24843/MITE.2020.v19i01.P03

Widyawati, W., Saptomo, W. L. Y., & Utami, Y. R. W. (2020). Penerapan agglomerative hierarchical clustering untuk segmentasi pelanggan. *Jurnal Ilmiah SINUS*, *18*(1), 75–87. doi:10.30646/sinus.v18i1.448