

Article

Bank Customer Segmentation and Marketing Strategies Based on Improved DBSCAN Algorithm

Xiaohua Yan ^{1,2}, Yufeng Li ^{1,*}, Fuquan Nie ^{3,*} and Rui Li ¹ 

¹ School of Electronic Information Engineering, Shenyang Aerospace University, Shenyang 110136, China; xiaohua_yanedu@163.com (X.Y.); lirui14@stu.sau.edu.cn (R.L.)

² School of Information Engineering, Zhengzhou Institute of Technology, Zhengzhou 450000, China

³ School of Mechanical and Electrical Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China

* Correspondence: liyufeng@sau.edu.cn (Y.L.); tghnkj@163.com (F.N.)

Abstract: This study conducts a case study on the characteristics of fixed deposit businesses in a Portuguese bank, analyzing the current customer data features and the limitations of marketing strategies. It also highlights the limitations of the traditional DBSCAN algorithm, including issues with parameter selection and a lack of diverse clustering metrics. Using machine learning techniques, the study explores the relationship between customer attribute features and fixed deposits. The proposed KM-DBSCAN algorithm, which combines K-means and DBSCAN, is used for customer segmentation. This method integrates both implicit and explicit customer indicators, incorporates weight factors, constructs a distance distribution matrix, and optimizes the process of selecting the neighborhood radius and density threshold parameters. As a result, the clustering accuracy of customer segmentation is improved by 15%. Based on the clustering results, customers are divided into four distinct groups, and personalized marketing strategies for customer deposits are proposed. Differentiated marketing plans are implemented, with a focus on customer relationship management and feedback. The model's performance is evaluated using silhouette coefficients, accuracy, and F1 score. The model is then applied in a real-world scenario, leading to an average business revenue growth rate of 16.08% and a 4.5% increase in customer engagement.

Keywords: KM-DBSCAN algorithm; customer segmentation; differentiated marketing strategy; evaluation metrics



Academic Editor: Francesco Zirilli

Received: 18 December 2024

Revised: 9 February 2025

Accepted: 11 February 2025

Published: 13 March 2025

Citation: Yan, X.; Li, Y.; Nie, F.; Li, R. Bank Customer Segmentation and Marketing Strategies Based on Improved DBSCAN Algorithm. *Appl. Sci.* **2025**, *15*, 3138. <https://doi.org/10.3390/app15063138>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the everchanging global economic environment and the intensified market competition [1], the bank deposit business has undergone profound changes. When bank interest rates rise, customers are more inclined to choose time deposits to obtain higher returns, and when interest rates fall, they are more inclined to choose more liquid demand deposits or other means of investment [2]. In addition, inflation also has an impact on customers' purchasing power and financial management behaviors [3]. As customers' needs are getting increasingly more diversified and personalized, customers of different ages, occupations, income and risk preferences have different expectations for banks' deposit business. Young customers tend to deposit with strong liquidity, while middle-aged and elderly customers are more inclined to time deposits with higher stability and security. Therefore, banks should segment the market [4] to understand the needs of different customer groups, and launch corresponding deposit options.

The big data technology, artificial intelligence and financial technology have shown great potential in analysis and prediction of customer behaviors, formulation of marketing strategies and personalized recommendation with their development [4]. They provide banks with powerful tools for analyzing customer behaviors, market trends and risk management. It also brings new opportunities for the application of customer segmentation theory. Customer Segmentation Theory focuses on dividing the customer base in a market into subgroups with similar characteristics or behaviors [5]. This approach allows for more precise marketing strategies and service plans tailored to each group. By discovering user preferences within vast amounts of data, banks can provide differentiated services to various customer segments, rescue customers at risk of attrition, and stand out in the competitive landscape of customer resources [6], laying a theoretical foundation for achieving precise marketing.

Current research and application of customer segmentation theory have the following shortcomings: Firstly, Simplicity of Customer Segmentation Models: Existing models tend to be overly simplistic, neglecting the diversity of customers. They often rely on single-dimensional analyses, focusing only on explicit indicators such as age, assets, or account balances, without considering the integration of these indicators. Additionally, implicit indicators—like customer attitudes, comments, feedback, and preferences—are frequently overlooked [7]. In actual banking cases, understanding implicit factors such as customer satisfaction with current services, brand recognition, risk tolerance, and online comments and feedback can provide deeper insights into customer needs and identify potential demands and preferences. Secondly, Lack of Dynamic Adaptability and Real-Time Feedback Mechanisms: Traditional tabular analysis methods require manual consolidation of periodic customer data for analysis. In today's fast-paced environment, banks need to establish flexible operational mechanisms [8] that can adjust business structures and product types based on real-time data analysis. By leveraging artificial intelligence and machine learning technologies, banks can timely acquire customer data, establish multi-channel feedback mechanisms, and analyze customer preferences to reduce churn rates. This paper proposes a bank customer segmentation strategy [9] based on the DBSCAN algorithm, which employs web scraping techniques to quickly collect real-time data and obtain customer attribute information, feedback, and comments [10]. In the process of segmenting bank customers using this data, the strategy comprehensively considers both explicit and implicit customer data indicators and adjusts service plans based on the changing needs of customers.

A well-designed banking marketing plan can enable precise customer targeting and product customization. By tailoring financial management solutions based on the business scale and needs of different types of corporate clients, and through effective customer relationship management strategies, such as timely follow-ups to understand the latest client requirements, banks can enhance clients' financial knowledge and awareness of funds management [11]. This, in turn, increases customer loyalty and satisfaction. Therefore, achieving precise marketing is key to improving banking profitability. In the current information age, major banks widely use big data technologies and machine learning methods to aggregate customer information into databases, process customer data in bulk, and employ techniques such as clustering algorithms and association rule analysis to assess customer consumption behavior and investment preferences. This approach updates the traditional manual processing methods in marketing models and operational costs [12]. However, there are several shortcomings in the current research on marketing plans: Firstly, Lack of Differentiated Marketing Approaches: When recommending banking services, there is no strict differentiation among customers, leading to the use of indiscriminate marketing methods that waste considerable human and material resources without achieving sig-

nificant results. Secondly, Lack of Awareness in Customer Relationship Maintenance: In terms of customer relationship management, there has been insufficient focus on effective communication with customers and timely handling of their feedback and comments, resulting in decreased customer loyalty and a significant loss of customers.

In recent years, machine learning-based clustering methods have gained increasing attention in customer segmentation. In particular, the DBSCAN algorithm [13] has become an important tool in customer segmentation within the financial industry due to its ability to automatically identify cluster structures without the need for pre-set cluster numbers and its capacity to handle noise points. However, despite its significant theoretical advantages, DBSCAN faces several challenges in practical applications. Firstly, the algorithm is highly sensitive to parameter selection (such as neighborhood radius and density threshold), which are often set empirically and lack a systematic optimization mechanism. Secondly, customer groups typically have complex attributes across multiple dimensions, and relying on a single clustering metric or simple parameter selection may not fully capture the diverse needs of the customer base, leading to inaccurate clustering results.

This paper aims to explore the current development status and limitations of existing bank customer segmentation methods and marketing strategies. It proposes a customer segmentation framework based on the KM-DBSCAN algorithm to address the challenges banks face in accurately segmenting their customer base. The framework aims to optimize the accuracy of customer segmentation, reduce time complexity, and enhance the application efficiency of customer segmentation in large-scale datasets. Additionally, it offers marketing strategy optimization suggestions and verifies the feasibility and effectiveness of the approach through practical applications in the financial industry.

2. Literature Review

Bank customer segmentation is a key step for banks to achieve targeted marketing, risk control, and resource optimization. In the context of increasingly fierce competition in modern banking, the importance of customer segmentation is growing. Through effective customer segmentation, banks can identify the needs and characteristics of different customer groups, offering tailored products and services to each group, thereby enhancing customer satisfaction and loyalty [14]. Accurate customer segmentation not only helps banks gain a competitive edge in the market but also supports cost control, optimizes resource allocation, and improves marketing efficiency. In terms of targeted marketing, by segmenting different customer groups, banks can push suitable financial products based on customers' consumption habits, financial needs, and behavior patterns. For example, for high-net-worth individuals, banks can offer personalized wealth management services; for younger customers, banks can promote online payment or investment products; for customers with credit risks [15], banks can provide more appropriate loan terms and risk control strategies. Through such customized services, banks can improve the match between products and services, thereby increasing customer stickiness and loyalty.

However, bank customer segmentation faces several challenges in practical applications. Firstly, with the advent of the big data era, customer attributes and behavior data have become increasingly complex and multidimensional [16]. Traditional customer segmentation methods often rely on a few simple dimensions (such as age, gender, income, etc.), which cannot fully capture the diverse needs and dynamic changes of customers. Therefore, handling large volumes of high-dimensional data and extracting effective information from it is a major challenge for bank customer segmentation. Secondly, customer data often contains a significant amount of noise and outliers [17], which can seriously impact the results of customer segmentation. Traditional segmentation methods typically lack sufficient robustness when dealing with noisy data, which can lead to inaccurate

clustering results and even misguide marketing decisions. How to handle and remove noise data while ensuring clustering accuracy becomes another key issue in solving the bank customer segmentation problem. Lastly, customer behavior patterns and needs are becoming increasingly diverse and rapidly changing. Traditional segmentation methods struggle to capture these dynamic changes. For example, with the development of financial technology, more and more customers are opting for online banking and mobile payments, and this shift in behavior has significant implications for a bank's marketing strategies and service design. Therefore, how to use flexible and efficient algorithms [18] to capture the dynamic features of customer behavior and update them in real time is another major challenge for bank customer segmentation.

In summary, bank customer segmentation not only holds significant theoretical and practical importance but also faces numerous challenges in actual operation, such as data complexity, noise interference, and changes in customer needs. Therefore, how to employ advanced technologies and methods to optimize the bank customer segmentation process, improving its accuracy and timeliness, has become a core issue that the banking industry urgently needs to address.

Market segmentation theory was first proposed by American scholar Smith in 1956. He suggested that when formulating marketing strategies, businesses should combine factors such as customer attributes, preferences, and values with product differentiation to develop targeted marketing models and service plans. Since the 1960s, numerous scholars have contributed to the development of market segmentation theory [19], continually creating new research outcomes. Customer segmentation theory mainly includes behavioral segmentation, attribute-based segmentation, value-based segmentation, and needs-based segmentation, each of which seeks to understand customer needs and behavior patterns from different perspectives, thereby enhancing customer loyalty and business profitability.

Joung and Kim [20] proposed an interpretable machine learning method for customer segmentation. They divided customer groups by combining sentiment analysis and customer behavior data. This data-driven approach provides an effective supplement to traditional demographic-based segmentation methods, though it faces challenges when dealing with complex and noisy data. This study also adopts machine learning methods and proposes the KM-DBSCAN algorithm. Unlike traditional algorithms such as K-means, this algorithm can handle density clustering problems with noisy data, thereby better identifying potential needs within customer groups. Sun [21] and other scholars proposed a Gaussian Peak Heuristic Clustering (GPHC) method. This method identifies customer preference patterns using genetic algorithms and hierarchical clustering and optimizes the final customer segmentation results through heuristic information. This study draws on the customer segmentation ideas of the GPHC method, further combining the density clustering characteristics of the DBSCAN algorithm. It innovatively proposes a multi-dimensional clustering framework that integrates customer preferences and behavioral data. This method not only helps uncover potential patterns in customer demand but also provides data support for precise marketing strategies for bank customer groups. Although the GPHC method achieves high clustering accuracy in customer demand data, it may face challenges when dealing with customer data that is unevenly distributed or contains excessive noise. In contrast, the DBSCAN algorithm effectively addresses heterogeneity and noise in customer data through density clustering and can identify potential customer groups without explicit category labels. This makes DBSCAN particularly important in bank customer segmentation, especially when dealing with large datasets and complex, diverse customer groups. Tang et al. [22] proposed a three-dimensional joint segmentation model for the B2B market, designed to segment customers based on the behaviors and needs of direct customers as well as the needs of downstream clients. This model

provides marketing personnel with more accurate sales allocation and communication strategies. Although this research targets B2B customer segmentation, the proposed three-dimensional segmentation model offers insights for bank customer segmentation as well. In bank customer segmentation, in addition to the explicit needs and behavior data of direct customers, implicit needs (such as the latent demand for financial products) may also affect segmentation results. This study integrates explicit indicators, behavioral data, and implicit indicators to improve the accuracy and practicality of the clustering results. Tabianan et al. [23] proposed a customer segmentation method based on K-Means clustering. This method analyzes customer clustering by collecting purchase data from e-commerce customers, identifying high-profit customer groups, and promoting products accordingly. The study focuses on customer behavior characteristics and uses the K-Means clustering algorithm to analyze customer purchasing patterns, improving segmentation accuracy by optimizing intra-group similarity and inter-group differences. This algorithm demonstrates how clustering analysis can provide theoretical support for customer segmentation and offers practical experience for implementing bank customer segmentation. Although Tabianan et al. [23] used K-Means clustering, the KM-DBSCAN algorithm proposed in this study shows more advantages compared to K-Means when dealing with noisy data and irregularly distributed customer groups. Particularly in cases where bank customer groups may exhibit complex distribution patterns, KM-DBSCAN is better at identifying heterogeneity between customer groups.

Othayoth et al. [24] explored the application of various machine learning algorithms in customer segmentation. In the face of information overload related to product details in the e-commerce domain, they used techniques such as customer profiling, similarity clustering, and RFM unit classification to segment customer groups based on behavioral data, thereby achieving personalized recommendations. Sarkar et al. proposed a customer segmentation marketing strategy that combines the K-Means algorithm with RFM analysis, with an overall clustering purity evaluation of 0.95, indicating that the K-Means clustering algorithm combined with RFM analysis achieved an accuracy rate of 95% in customer attribute and feature segmentation. Hicham et al. [25] introduced a clustering ensemble-based customer segmentation technique, which integrates DBSCAN, Mini Batch K-means, and Mean Shift clustering models, and then uses spectral clustering methods to combine multiple clustering results into the final customer group segmentation. Li et al. [26] proposed a customer segmentation model that integrates Support Vector Machines (SVM) with clustering algorithms. This model first uses SVM to segment the current customer data, and then combines SVM with clustering algorithms to construct the customer segmentation model. The methods mentioned above explore the application of machine learning algorithms in customer segmentation from different perspectives. They effectively address the challenges posed by information overload in the context of big data, which can complicate decision-making and marketing strategy formulation. However, these methods often rely on a single model for customer segmentation, fail to integrate multiple customer attribute indicators [27], and do not investigate the impact of different parameters in the model algorithms on the clustering results.

Customer segmentation is also widely applied in practice. Bank of America has enhanced its retail banking marketing effectiveness by implementing customer segmentation strategies. By analyzing customer account activities, spending behaviors, and other financial data, the bank uses clustering algorithms to segment customers and provide personalized products and services such as loans, credit cards, and savings accounts to different customer groups. For example, Bank of America offers exclusive wealth management services and investment products to high-net-worth clients, flexible financial products to younger customers, and low-interest loans and preferential savings accounts to low-

income customers. Citibank utilizes big data analytics and machine learning algorithms to achieve precise customer segmentation based on customer behavioral data, transaction data, and other historical information. The bank recommends financial products, behavior predictions, and risk management tailored to each customer group. Citibank's recommendation system is based on historical customer behavior data, providing personalized financial product suggestions. It also uses Natural Language Processing (NLP) techniques to process and analyze interactions with customers, extracting insights about customer satisfaction and latent needs. China Construction Bank (CCB) uses the K-means clustering algorithm to analyze customer data, segmenting customers into groups based on characteristics such as consumption habits, credit status, and account activity. Using the RFM model, the bank further classifies customers into high-value, potential, and at-risk groups, offering personalized services and products, and devising targeted marketing strategies for each group.

Although customer segmentation has achieved some success in the banking industry, there are still limitations in practical applications. Although Citibank enhances segmentation accuracy using big data and NLP techniques, it still largely relies on explicit indicators such as transaction behavior data and historical records. Implicit characteristics (e.g., psychological needs or latent preferences) may not be sufficiently considered, which limits the depth and precision of segmentation. Banks like Citibank and China Construction Bank, despite utilizing complex machine learning algorithms and clustering techniques, often face high computational complexity with large, high-dimensional datasets. This results in slower processing speeds and challenges in responding quickly for real-time decision-making.

The KM-DBSCAN algorithm proposed in this paper addresses these issues. First, it reduces dependency on parameters by introducing K-means preprocessing, solving the difficulty of selecting global parameters in DBSCAN, especially the choice of neighborhood radius and density threshold. Second, it integrates both explicit and implicit features. By innovatively combining explicit and implicit indicators and using a weighted mechanism, the model increases sensitivity to different features, helping banks better capture customers' multidimensional needs in practice, compensating for the limitations of relying solely on explicit features. Third, it improves computational efficiency. By using K-means as a preprocessing step, the algorithm reduces the computational complexity of DBSCAN, enhancing its efficiency in big data scenarios. This is particularly advantageous when handling large amounts of customer data.

3. Algorithm Construction Process

3.1. Data Preprocessing

The sample data utilized in this study consist of survey responses from customers of a Portuguese bank. These data, sourced from the UCI Machine Learning Repository (Bank Marketing Dataset), include a total of 45,211 records [28]. The customer information includes attributes such as age, occupation, education level, account balance, whether the customer has had a credit default, whether the customer has a housing loan, and whether the customer has a personal loan. The term deposit activity data includes information such as the communication method with the customer, the result of the last marketing campaign, the number of days since the last contact, the duration of the last call, the number of contacts with the customer in the current campaign, and whether the customer subscribed to the term deposit. The experimental environment software includes Anaconda and Jupyter Notebook, and the machine learning libraries used are Scikit-learn, Pandas, Numpy, Matplotlib, and Seaborn.

(1) Data cleaning

The bank's original data set includes some dirty data, such as missing age, abnormal data, missing education background, and abnormal values of binary variable. The main tasks of data cleaning are as follows:

Missing data processing. In the sample data set, there are some missing data. For example, part of the data in the column recording the length of the last call is missing. In order to avoid large errors in the experimental results, the mean of all the data in this column is calculated, and used to fill in the missing data.

Duplicate data processing. After analysis of the basic structure of the data, it is found that there is a large amount of duplicate data in the data set. In order to avoid the error of duplicate data on the experimental results, the duplicate data is merged and only one row of data is retained.

Code value conversion of data. The string of the columns such as job type, education background, borrowing, and credit default in the data set is converted into numerical value for convenience of subsequent data analysis.

(2) Feature engineering

Standardization of variables. The binary variables such as customer's credit default, housing loans, borrowing etc. are subject to one-hot encoding, while polytomous variables such as occupation, education background, etc. are transformed into z-score to calculate the mean and standard deviation of each feature [29]. Then, the Z-score standardization formula is used to reduce the impact of outlier values on the model and make it easier to compare the importance of different eigenvalues. The principle of Z-score standardization is to subtract the data from its mean and then divide by the standard deviation. The formula is as shown in (1) and (2) below:

$$x' = \frac{x - \bar{x}}{\sigma} \quad (1)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2)$$

where, x is the customer information data before the standardization; x' is the data after the standardization. The mean of the standardized data is 0 and the standard deviation is 1, which brings the data closer to the normal distribution and eliminates the quantitative differences between the data.

(3) Overall data correlation test

The matrix model describing the correlation between the variables is shown in Figure 1:

The correlation coefficient matrix [30] shows a high correlation between whether the customer makes a fixed deposit and the length of the last call, from which it can be inferred that the length of the bank's last contact with the customer in the marketing event is crucial to the success of the marketing; at the same time, there is also a strong correlation between the variables *pdays* and *previous*, that is, the number of contacts with the customer before the current event and the time interval between the contact with the customer are related; the month of the last contact in the year is related to whether the customer has a housing loan.

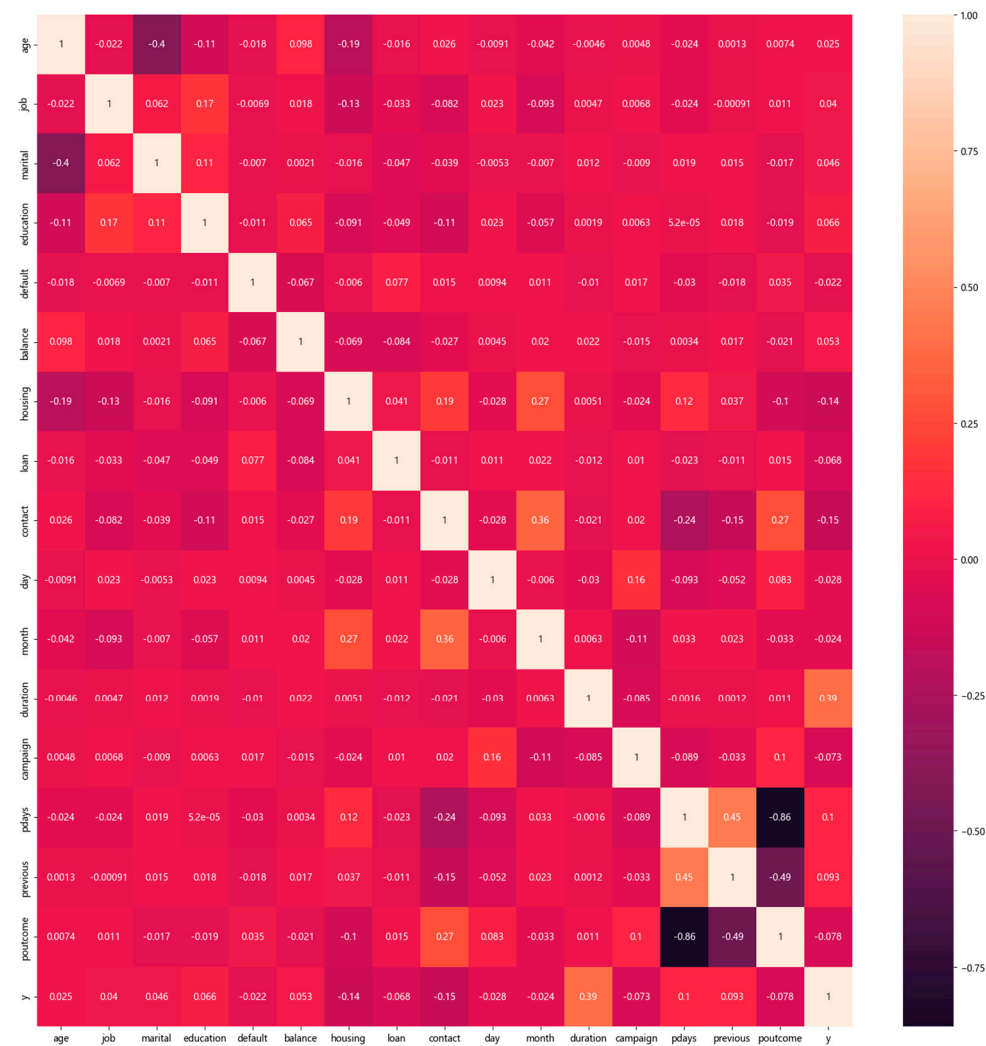


Figure 1. Overall Pearson correlation coefficient matrix.

3.2. Attribute Analysis

(1) Analysis of the relationship between related attribute columns and target columns

Some attribute columns with a high degree of correlation are filtered out from the overall correlation matrix to discuss their relationship with the target column (y column), followed by analysis of the influencing factors of customers' time deposits. The correlation coefficient matrix is shown in Figure 2.

As can be seen from the figure, whether the user makes a fixed deposit is highly related to the length of the last call (duration column) with a correlation coefficient of 0.39, which indicates that the last call with the user is critical to customers' decision-making. To increase the ratio of customers with fixed deposit, it is necessary to eliminate the obstacles to the customer's decision-making as much as possible during this call time.

In order to further explore the distribution of call duration and time deposit columns, the kdeplot function in the Seaborn library is called to draw a kernel density estimation diagram to show the distribution relationship between the two variables. As can be seen from Figure 3, the peak value is within a range of 200s–350 s. After the call of the describe() function, the average value of the call duration column is 258 s, and the 75% median value is 319 s. Therefore, the bank staff's call duration with customers should be appropriate within this interval.

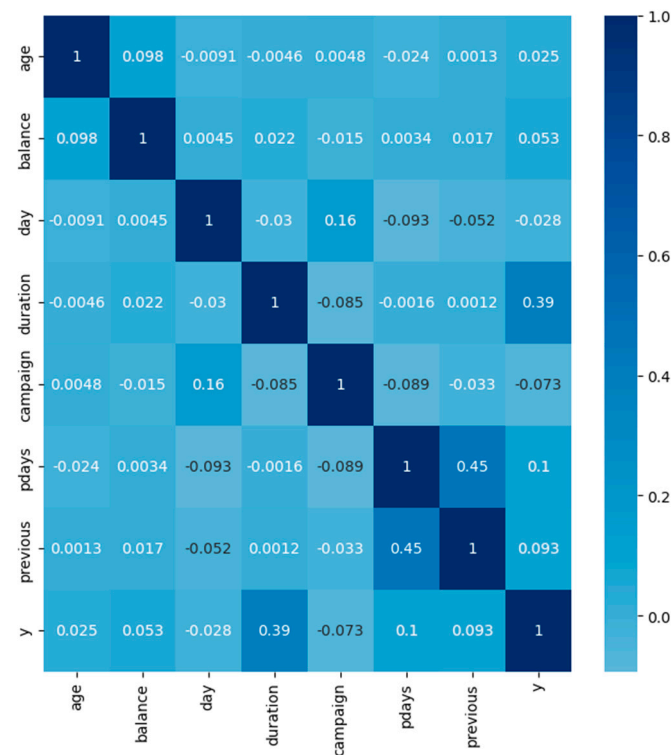


Figure 2. Coefficient matrix of feature columns and target column.

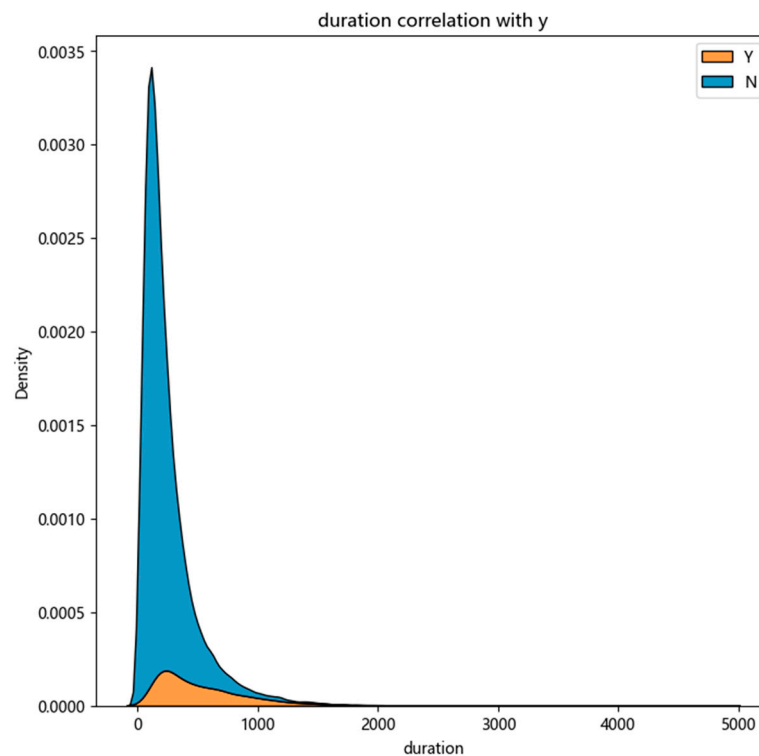


Figure 3. Call duration and fixed deposit kernel density estimation plot.

(2) Analysis of customer features

The influencing factors of customers' time deposits are analyzed from the perspective of customer features.

- (a) Explore the relationship between the customers' occupation and time deposits.

The mean() function is used to calculate the mean of the amount of fixed deposits (column y) in each occupation (column x), and the analysis of the proportion of fixed deposits in different occupations can help banks understand the financial situation and financial management behaviors of customer groups with different occupations and launch different deposit strategies and service plans in a targeted manner.

As can be seen from the figure, among the different occupations, the top five customers who choose time deposits are managers, technicians, company blue-collar, administrative personnel, and retirees. From Figures 4 and 5, it can be seen that among different occupations, the proportion of fixed deposits is higher among students and retirees. The illustration suggests that 28% of students and 23% of retirees choose fixed deposits, which indicates that although the total number of these two categories of customers with fixed deposits is not the largest, the proportion of choosing time deposit is higher in respective category. Therefore, such customers are also the customer groups that should be maintained by banks.

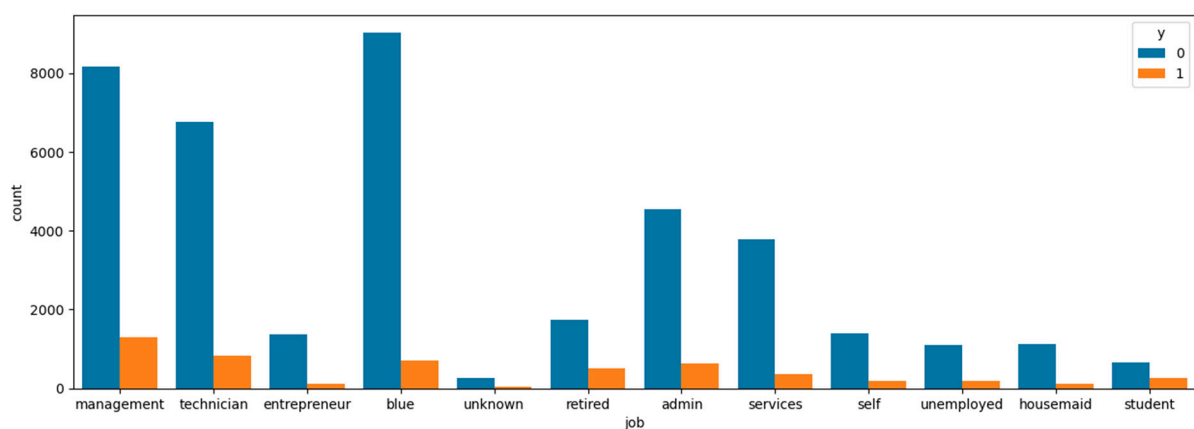


Figure 4. Bar chart of occupation categories and fixed deposits.

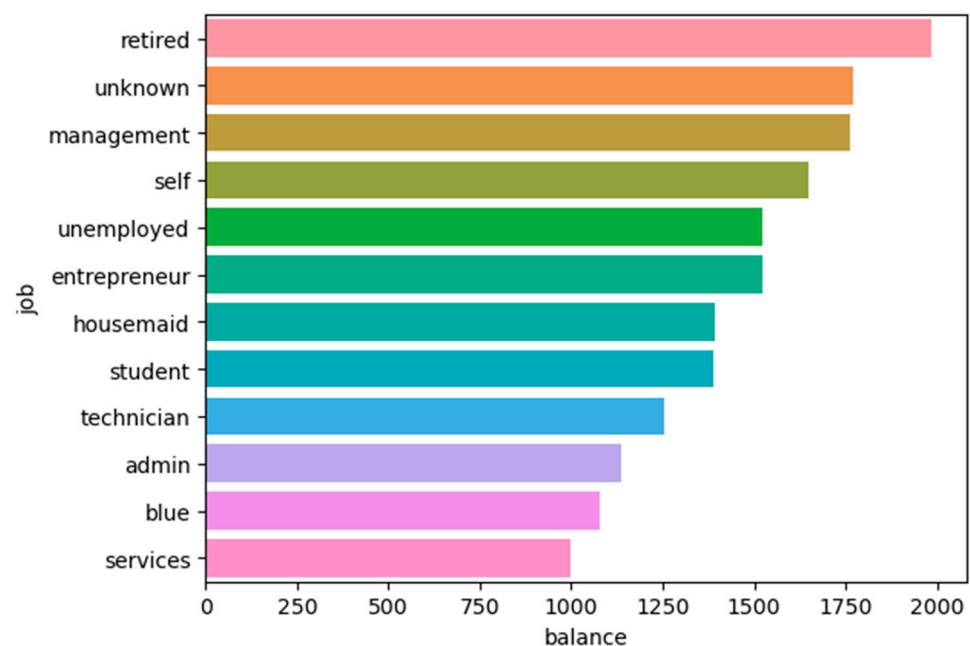


Figure 5. Ratio Chart of Occupational Category and Fixed Deposit.

- (b) Explore the different reactions of customer with different jobs to the previous marketing event.

In order to understand the different attitudes of customers with different occupations to marketing activities and provide a theoretical basis for the formulation of the next marketing plan, the Seaborn library was used to generate a histogram as shown in Figure 6. It can be seen from the figure that a high success rate of the previous marketing event was achieved among managers, technicians, and blue-collar. Therefore, the key contact groups for the next campaign are these three customer groups.

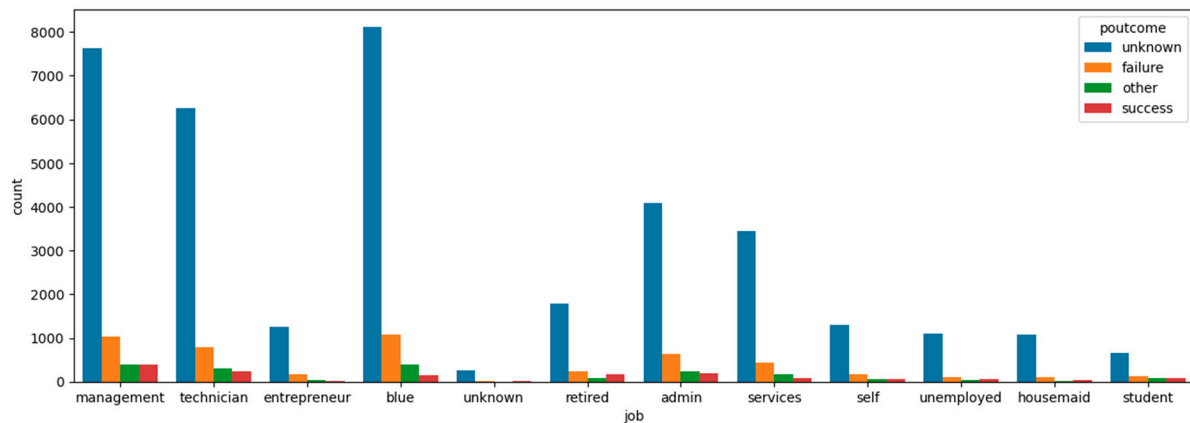


Figure 6. The bar chart of different jobs and previous marketing event.

3.3. Customer Segmentation Based on Improved DBSCAN Algorithm

3.3.1. Overview of Relevant Theories

Clustering algorithms are classic unsupervised learning methods that do not require predefined category labels [31]. They group data points with high similarity into the same cluster, ensuring that the intra-cluster similarity is high, while the inter-cluster similarity is low. Clustering algorithms are widely used in fields such as data mining, information retrieval, and customer segmentation. In the clustering analysis process, the basic dataset is first preprocessed to extract key features, then similarity measures such as Euclidean distance and cosine similarity are used to measure the similarity between data objects, and finally, appropriate partitioning methods are applied to divide the data objects into multiple clusters or groups.

K-Means and DBSCAN are both typical clustering algorithms [32]. The K-Means algorithm first initializes the centroids, then calculates the distance of each data point from its centroid. The data points are repeatedly assigned to the cluster represented by the centroid to which they are closest, and the centroids are updated. This process continues until the centroids no longer change.

The DBSCAN algorithm performs clustering based on density. It defines the clustering characteristics using two parameters: neighborhood radius (eps) and density threshold (MinPts). This algorithm can find clusters of arbitrary shapes and does not require the number of clusters to be specified in advance [33]. The process of forming clusters using DBSCAN is as follows: First, an unvisited data point pp is chosen and marked as visited. All points within the neighborhood radius of pp are identified. If the number of points within this neighborhood exceeds the density threshold, point pp is considered a core point, and a cluster CC is created. If the number of points within the neighborhood is less than the density threshold, point pp is labeled as a noise point. For each point qq in the neighborhood of pp , if the number of points within the neighborhood of qq is greater than or equal to the density threshold, all these points are added to pp 's neighborhood, and any points not yet assigned to a cluster are added to cluster CC . This process repeats until all points have been visited.

3.3.2. Innovations of the Algorithm

(1) In the traditional DBSCAN algorithm, the selection of parameters plays a crucial role in the clustering results. These parameters are often chosen based on experience and have global uniqueness. This paper introduces the K-means algorithm to improve the DBSCAN algorithm, forming the KM-DBSCAN algorithm, which helps determine the optimal parameters and achieve the best customer group feature clustering results.

(2) In the process of obtaining customer clusters, customer attribute features and behavioral information have a significant impact on the clustering results. Therefore, this paper integrates both explicit indicators (such as occupation and education level) and implicit indicators (such as customer preferences) into the KM-DBSCAN clustering algorithm, adding weight factors to assign different weights to various attributes, thereby generating more accurate customer clusters.

(3) In the traditional DBSCAN algorithm, during density-based clustering, it is necessary to traverse the entire dataset to identify core points that meet the neighborhood radius and density threshold, which significantly increases the algorithm's time complexity. This paper uses the K-means algorithm to partition the dataset into regions and applies DBSCAN clustering on the initial clusters. Finally, the initial clustering results are merged, reducing the complexity of performing DBSCAN calculations each time.

3.3.3. Algorithm Process

The customer segmentation process using the KM-DBSCAN algorithm is as follows:

(1) Construction of Explicit and Implicit Indicators

The explicit indicators integrated into this algorithm include occupation category and education level. For occupation categories, one-hot encoding is used to convert each occupation into a binary vector. For example, the occupation categories "management", "technician", "blue-collar", "admin", and "retired" are processed as follows after one-hot encoding: "management" $\rightarrow [1, 0, 0, 0, 0]$, technician $\rightarrow [0, 1, 0, 0, 0]$, "blue" $\rightarrow [0, 0, 1, 0, 0]$, "admin" $\rightarrow [0, 0, 0, 1, 0]$, "retired" $\rightarrow [0, 0, 0, 0, 1]$.

For education level, which includes "tertiary", "secondary", and "primary", label encoding is applied, converting them to [1, 2, 3]. The implicit indicator integrated into the algorithm is customer preference information. Customer preferences are derived from customer behavioral data and are presented numerically. These preferences are categorized into a range of 1–5, with higher values indicating stronger customer intent to deposit.

Standardization of Explicit and Implicit Indicators: Before clustering, the explicit and implicit indicators need to be standardized into feature vectors suitable for clustering [34]. The standardization method used in this paper is Z-score normalization. The explicit and implicit indicators are combined to form a unified feature matrix X , which contains all customers and their attribute features.

Weighting Different Features: Since different features have varying impacts on customer clustering results, this paper introduces a weight factor ω_i to assess the importance of different features. The Euclidean distance is then adjusted based on the weighted matrix to calculate the distance between each sample in the dataset. Let the customer attribute feature dataset be $y = \{y_1, y_2, y_3 \dots y_n\}$, where y_i and y_j are any two data points in the dataset:

$$d_{i,j} = \sqrt{\sum_{i,j=1}^n \omega_i (y_i - y_j)^2}$$

(2) Using K-Means Clustering Algorithm to Partition Regions

Initializing the centroids: Select data points randomly from the dataset as the initial centroids, denoted as $\{c_1, c_2, c_3\}$.

Assigning data points to the nearest centroid: Use the Euclidean formula to calculate the distance between each data point $y_i \in y$ and each centroid, and assign the data point y_i to the nearest centroid c_j . The formula is as follows:

$$d(y_i, c_j) = \|y_i - c_j\| = \sqrt{(y_{i1} - c_{j1})^2 + (y_{i2} - c_{j2})^2 + \dots + (y_{in} - c_{jn})^2}$$

Updating the centroids: For each cluster c_j , calculate the mean of the data points y_i in the cluster and use this mean as the new centroid.

$$c_j^{new} = \frac{1}{|c_j|} \sum_{y_i \in c_j} y_i$$

Repeat the above steps until the maximum number of iterations is reached, and the dataset will be finally divided into k clusters, $\{c_1, c_2, c_3 \dots c_n\}$.

(3) DBSCAN Clustering within Regions

After the regions are divided into k clusters, DBSCAN is applied to further subdivide each cluster c_j . The DBSCAN algorithm will perform density-based expansion within each cluster defined by K-Means to identify sub-clusters with higher density. The steps of the DBSCAN clustering algorithm are as follows:

Neighborhood Query: For each point y_i in cluster c_j , the algorithm performs a Euclidean distance calculation to find all the points that belong to its neighborhood.

$$N_{\epsilon}(y_i) = \{y_m \in C_j \mid \|y_i - y_m\| \leq \epsilon\}$$

Core Point Identification: If a point y_i has at least $MinPts$ points within its neighborhood, then y_i is considered a core point.

Cluster Expansion: If y_i is a core point, then y_i and all points in its neighborhood $N_{\epsilon}(y_i)$ are labeled as belonging to the same cluster. If a point is a border point, it will be assigned to the cluster of the nearest core point.

Noise Point Labeling: If a point is neither a core point nor a border point, it is labeled as noise.

From the initial clustering, the class with the most points is selected to calculate the average, which is then used as the neighborhood radius (ϵ). The nearest distance value is selected as the density threshold ($MinPts$), and this value is used to calculate the density as follows:

$$Density = \frac{MinPts}{\pi \times \epsilon^2}$$

This density value selection process is used for determining the parameters in the DBSCAN algorithm [35]. When the number of clusters remains the same for five consecutive iterations, it indicates that the number of clusters has stabilized. At this point, the number of clusters represents the optimal clustering result for the customer groups' characteristics.

(4) Merging Cluster Results to Obtain Final Clusters

Through the above steps, the K-Means algorithm divides the dataset into k clusters, and the DBSCAN algorithm further refines each cluster. Next, the clustering results from both algorithms need to be merged. For each cluster c_j generated by K-Means, DBSCAN may create new sub-clusters or noise points within c_j . If DBSCAN identifies multiple sub-clusters within a cluster c_j , these sub-clusters are treated as new sub-clusters [36]. If DBSCAN labels some points as noise within a cluster, these noise points do not participate in the final clustering.

Final Cluster Division: If DBSCAN does not find noise or new sub-clusters within a particular cluster, the original K-Means cluster division is retained. If DBSCAN refines a cluster, it is divided into new sub-clusters.

4. Experimental Process and Result Analysis

4.1. Experimental Process

In this experiment, the customer attribute columns include the customer's age, occupation, marital status, education background, asset balance, housing loan, borrowing, and activity information. The process of customer analysis with the KM-DBSCAN algorithm is as shown in Figure 7.

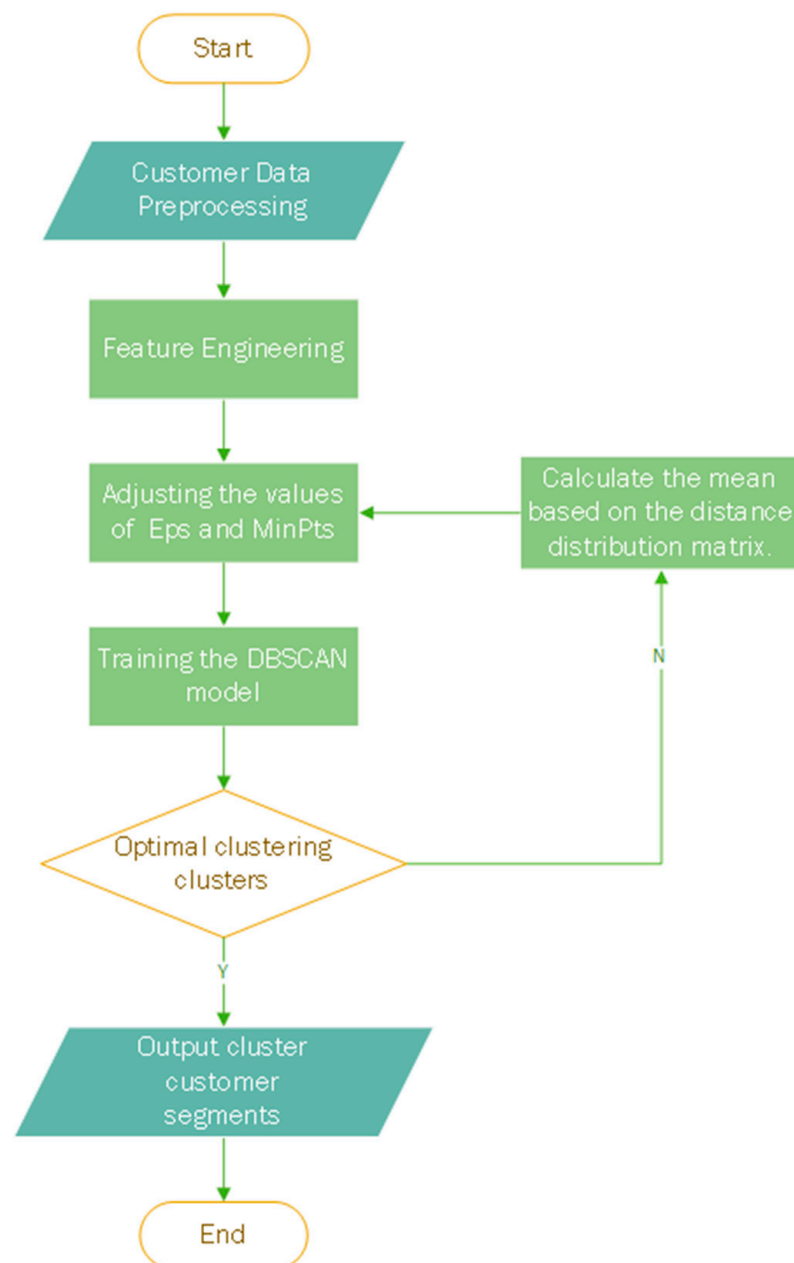


Figure 7. Flow chart of the KM-DBSCAN algorithm.

First, the attribute feature column was subjected to PCA (principal component analysis) to reduce the data dimension to improve the operating efficiency of the model. The optimal parameters of neighborhood radius (Eps) and density threshold (MinPts) were selected

for division of the clustering effect. After verification, the Eps value of 0.6 and the minPts value of 5 were selected in this experiment to obtain the optimal clustering effect.

Based on the above algorithm model and process, the following clustering results were obtained. The dataset was divided into two parts: the training set and the test set, with 80% of the data used as the training set and 20% as the test set. A comparative experiment was conducted, and the results show that the clustering accuracy of the proposed model in this study improved by 15% compared to the traditional DBSCAN algorithm.

(1) Experimental Environment

The environment configuration for this experiment is shown in Table 1:

Table 1. The environment configuration for this experiment.

Experimental Environment	Specific Configuration
Operating System	Windows 11 64-bit
CPU	Intel(R) Core(TM) i7-10700
GPU	NVIDIA GTX 1080 Ti
RAM	12 GB
Programming Language	Python 3.8
Programming Environment	Jupyter Notebook

(2) Selection of Weight Factors

In this experiment, the weight factors for the explicit and implicit indicators in the customer dataset are adjusted to explore the impact of different weight combinations on the clustering results. The optimal weight factors are determined based on the clustering performance. The Davies-Bouldin Index (DBI) is used as the evaluation metric [37]. A smaller DBI value indicates higher separation between clusters, better compactness within clusters, and thus better clustering performance. The formula for calculating the DBI index is as follows:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max_{i \neq j} \left(\frac{\delta_i + \delta_j}{d_{ij}} \right)$$

where δ_i is the compactness of the i -th cluster, representing the average distance between the samples within the cluster and the cluster center. d_{ij} is the separation between the i -th and j -th clusters, representing the distance between the centers of the two clusters.

As shown in Table 2, when the weight ratio of explicit indicators to implicit indicators is 0.7:0.3, the DBI value is the smallest, indicating that the clustering effect of customers is optimal at this point.

Table 2. Comparison of DBI Values under Different Weight Factors.

Weight Configuration	DBI Value
Explicit Weight:Implicit Weight = 0.2:0.8	1.54
Explicit Weight:Implicit Weight = 0.3:0.7	1.46
Explicit Weight:Implicit Weight = 0.4:0.6	1.42
Explicit Weight:Implicit Weight = 0.5:0.5	1.30
Explicit Weight:Implicit Weight = 0.6:0.4	1.21
Explicit Weight:Implicit Weight = 0.7:0.3	1.03
Explicit Weight:Implicit Weight = 0.8:0.2	1.25

(3) DBSCAN Parameter Selection

When using the DBSCAN algorithm for clustering within a region, in order to obtain the optimal clustering results, the optimal values for the neighborhood radius and density threshold are determined [38]. This paper sets up the following comparative experiments, and the experimental data shows that when Eps = 0.6 and MinPts = 5, the number of clusters tends to stabilize, resulting in the optimal clustering outcome. The results are shown in Table 3.

Table 3. The number of clusters for different values of Eps and MinPts.

Clustering Iterations		1	2	3	4	5
Value Selection	Number of Clusters					
Eps = 0.4	MinPts = 3	4	3	2	2	3
Eps = 0.4	MinPts = 4	3	3	2	2	2
Eps = 0.4	MinPts = 5	4	3	4	2	3
Eps = 0.5	MinPts = 3	3	3	2	2	3
Eps = 0.5	MinPts = 4	4	3	2	2	2
Eps = 0.5	MinPts = 5	4	3	3	2	3
Eps = 0.6	MinPts = 3	3	3	2	2	3
Eps = 0.6	MinPts = 4	4	3	2	2	3
Eps = 0.6	MinPts = 5	4	4	4	4	4
Eps = 0.7	MinPts = 3	4	3	3	3	2
Eps = 0.7	MinPts = 4	4	2	2	2	3
Eps = 0.7	MinPts = 5	4	3	2	2	2

Key code as follows:

```
kmeans_labels = kmeans.fit_predict(X)
#Visualize the K-Means clustering results
plt.scatter(X[:, 0], X[:, 1], c = kmeans_labels, cmap = 'viridis', s = 10)
plt.title('K-Means Clustering')
plt.show()
# Apply DBSCAN within each K-Means cluster
epsilon = 0.6 # Neighborhood radius
min_samples = 5 # Minimum number of samples
#Refine the results of K-Means
final_labels = np.copy(kmeans_labels)
for i in range(4): # DBSCAN clustering for each cluster
    # Get the points of the current cluster
    cluster_points = X[kmeans_labels == i]
    # DBSCAN clustering
    dbscan = DBSCAN(eps = epsilon, min_samples = min_samples)
    dbscan_labels = dbscan.fit_predict(cluster_points)
    # Merge results: DBSCAN refines clustering within the cluster, noise points are marked
    as -1 mask = kmeans_labels == i
    final_labels[mask] = final_labels[mask] × 100 + dbscan_labels
Experimental results as follows in Figure 8:
```

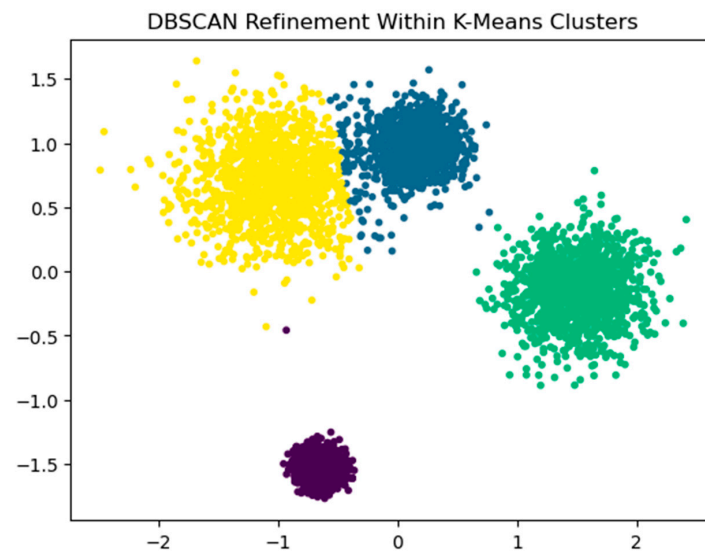


Figure 8. clustering result of KM-DBSCAN algorithm. Colors represent different clusters: yellow (Cluster 1), green (Cluster 2), blue (Cluster 3), and purple (Cluster 4).

4.2. Evaluation Metrics and Result Analysis

In order to test the effect of the algorithm, comparative experiments were conducted to compare the K-Means algorithm, DBSCAN algorithm and KM-DBSCAN algorithm with Silhouette Coefficient [39], Precision and F1 Score as evaluation indicators.

The Silhouette Coefficient, which combines the tightness inside the cluster and the degree of separation between the clusters, is commonly used to evaluate the quality of the clustering results. The calculation formula is: $SC(i) = \frac{m(i) - n(i)}{\max\{m(i), n(i)\}}$. For a data point i , the average distance n of all in-cluster points except itself is calculated as the tightness, and the average distance m of all points in the nearest neighborhood cluster is calculated as the degree of separation, thereby the contour coefficient is calculated. The value range of $SC(i)$ is $[-1, 1]$, and the closer the value of $SC(i)$ is to 1, the better the clustering effect. As shown in Figure 9, after 40 times of iterations, the number of clustering clusters of KM-DBSCAN tended to stabilize. At this time, the contour coefficient reached 0.97, which is closer to 1, and the highest contour coefficients of the K-Means algorithm and the DBSCAN algorithm were 0.85 and 0.75, respectively, which indicates that the proposed KM-DBSCAN algorithm achieves a better clustering effect.

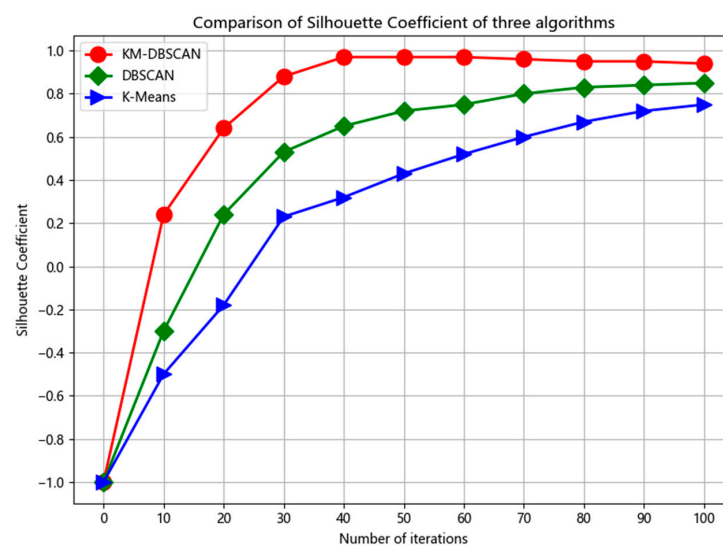


Figure 9. Comparison of Silhouette Coefficient of three algorithms.

Accuracy is commonly used to evaluate the predictive power of algorithms in different categories, and can be also used to detect the performance of model clustering. In this experiment, the customer data set was divided into training set and test set, and the accuracy is used as the evaluation index [40]. The results are shown in Figure 10. It shows that with the increased times of iterations, the accuracy of the algorithm gradually increases. The average accuracy of the KM-DBSCAN algorithm, DBSCAN algorithm and K-Means algorithm is 92.3%, 81.6%, and 71.8%, respectively, which indicates that the proposed KM-DBSCAN algorithm achieves a higher accuracy in customer segmentation.

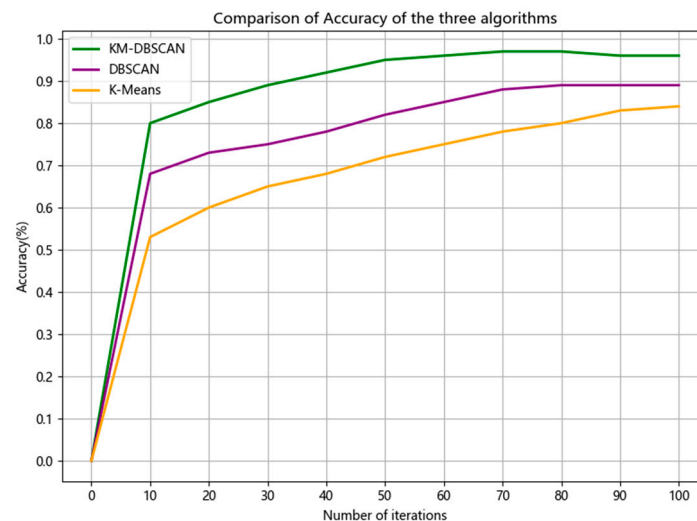


Figure 10. Comparison of Accuracy of the three algorithms.

In order to further evaluate the overall performance of the model, F1 Score is used as the evaluation index. This index is the reconciled average of accuracy and recall rate, where both the two indicators can be taken into account in optimizing the model [41]. The results are shown in Figure 11. The F1 average value of the proposed KM-DBSCAN algorithm was 0.92, and that of the DBSCAN algorithm and K-Means algorithm was 0.83 and 0.71, respectively. It can be seen that the F1 average value of the proposed algorithm is significantly higher than that of the K-Means algorithm and the DBSCAN algorithm, which verifies that the overall performance of the constructed model is higher in customer segmentation.

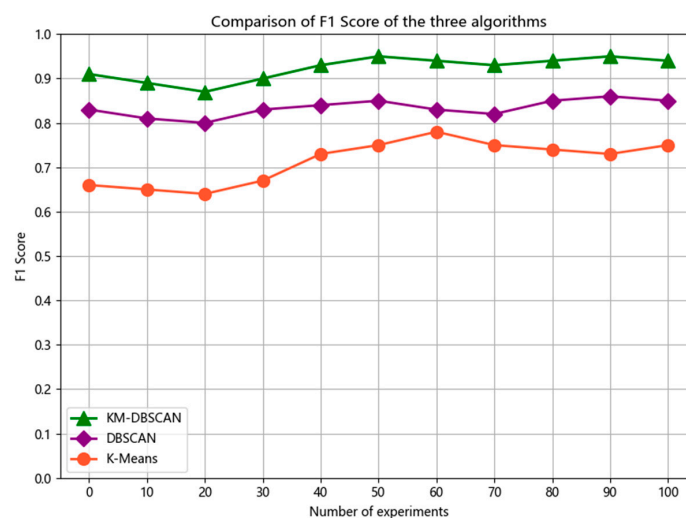


Figure 11. Comparison of F1 Score of the three algorithms.

4.3. Development of Marketing Plan

On the basis of above clustering results, the customers are subdivided into 4 groups, with the results shown in Table 1:

Customer clustering results table and feature segmentation table are shown in Table 4.

Table 4. Customer clustering results table.

Customer Segmentation	Customer Group 1	Customer Group 2	Customer Group 3	Customer Group 4
Quantity	25,432	17,997	1350	432
Average Age	42	56	32	23
Proportion of Term Deposits	42.63%	36.43%	17.56%	3.38%
Average Balance	1342	1563	700	230
Proportion Without Housing Loans	75.64%	43.25%	32.51%	13.34%
Average Call Duration	253 s	180 s	103 s	30 s
Success Rate of Previous Marketing Campaign	45%	32%	15%	5%
Customer value division	High-value customer group	medium-value customer group	low-value customer group	Potential customers

On the basis of above analysis of customer information, activity information, and customer clustering characteristics, the following marketing strategy analysis is made:

First, the characteristics of customers with fixed deposit behavior are summarized as follows:

- (1) The proportion of technical personnel, management personnel, and company blue-collar workers is high.
- (2) Customers with higher education have the habit and awareness of time deposits.
- (3) The balance of assets is high, and there is a disposable amount.
- (4) The proportion of time deposits of customers without borrowing or housing loans is high.

Second, the influencing factors of customers' time deposits under the bank's existing activities are as follows:

- (1) When customers are contacted in April, May, July, and August, they are more likely to choose the fixed deposit.
- (2) The results of the previous marketing event will directly affect customers' choice of time deposits this time.
- (3) The appropriate call duration helps with customers' decision-making on time deposits.
- (4) Proper regular contact with customers helps maintain the stability and durability of customer relationships.

Based on the results of the clustering analysis, the following marketing plan is formulated:

- (1) On the basis of clustering analysis result, the following personalized marketing plans are formulated for different customer groups:

Customer group 1: This customer group is characterized by high asset balances without borrowing or housing loans, and there is a high proportion of managers and technical personnel aged about 40 years old. This customer group should be mainly

maintained. Bank staff should continue to maintain the frequency of contact with them and recommend high-yield and flexible time deposit products.

Customer group 2: This customer group is characterized by high asset balances without borrowing or housing loans, and there is a high proportion of retirees and technical personnel aged about 50 years old. This customer group is the key customers. Bank staff should appropriately increase the frequency and duration of contact with them, and launch fixed deposit products with high security and high stability targeting the elderly population, emphasizing secured principal and fixed income. The bank should also optimize the offline service process, provide a dedicated customer service window or green channel to facilitate the business for elderly customers.

Customer group 3: This customer group have borrowing and housing loans and are about 30 years old. Key focus should be paid to this customer group. The bank should increase the frequency and duration of contact with them, take the initiative to understand their needs, and recommend short-term high-yield fixed deposit products with low required minimum deposits.

Customer group 4: The customer group have less asset balances, and students and unemployed people account for a relatively high proportion. They are less contacted by the bank, but they are key customers to be reached. In view of the characteristics of such customer group, banks should strengthen communication with them, cultivate their awareness and habits of financial management [42], and launch short-term fixed deposit products with lower required minimum deposits.

(2) Proposed customer marketing strategies for different occupational groups

Retired Group: Retired individuals receive pensions, have low material demands, and generally maintain good saving habits. This demographic represents the greatest potential for time deposit purchases. Banks should organize more communication activities targeting elderly retirees, promote awareness of financial scams, and offer gift redemption programs to attract customers to deposit their money in banks.

Student Group: This demographic typically does not have significant amounts for time deposits but possesses strong learning capabilities and a willingness to try new things. Therefore, banks can promote savings and financial management concepts to this group, increase time deposit and financial products suitable for young people, and facilitate both online and offline marketing and participation.

High-Income Groups (Managers, Technicians, etc.): Individuals in this category have a significant amount of disposable assets, are highly educated, and possess planning awareness. They tend to have deeper thinking and foresight when approaching issues. Thus, when recommending products to this group, it is essential to present the short-term and long-term returns of deposit and financial products from a data analysis perspective.

(3) Implement cross-marketing for high-value and high-balance asset groups:

Cross-marketing refers to the practice of promoting two or more products through the analysis of customer historical data [43], targeting customers with potential needs for related marketing to expand the business scope. For example, for customers with bank loans who have higher liquidity requirements, the bank can recommend credit card applications and installment payment services. For customers with high asset balances and high-value customer groups, bank financial products such as funds, insurance, and bonds can be recommended to help customers achieve wealth appreciation.

4.4. Application Effect Analysis

In this study, the proposed customer segmentation and marketing strategy is further applied to actual scenarios, where the revenue growth rate is used as the evaluation

index to show the distribution in each percentile [44]. The result value of the proposed algorithm is compared with the optimal model result. As shown in Figure 12, the average revenue growth rate of the model algorithm was 16.08%, and that of the optimal model was 18.0%, where there is a difference of 1.92%. This is in line with the range of growth rate differences in practical applications, which indicates that the proposed customer segmentation model and marketing suggestions can help enterprises formulate marketing strategies and improve corporate efficiency.

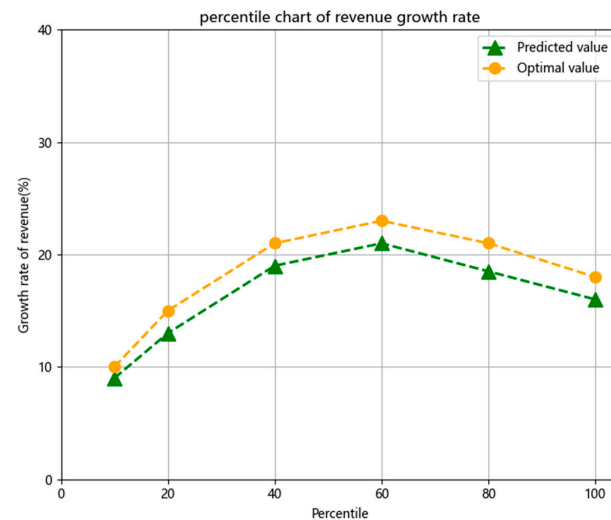


Figure 12. Percentile chart of revenue growth rate.

In this study, customers are subdivided into 4 groups. In order to further identify the application effect of customer segmentation and marketing strategy and analyze the changes in customer attention to bank time deposit products, customer group 1 and customer group 4 were selected as comparison objects. Customer group 4 is a low-value customer group in customer value theory, and should be actively reached. The results are shown in Figure 13. The average initial customer attention was 36.67%. After application of the marketing strategy, it was 41.17%, and the customers' attention to the product increased by 4.5%, which indicates that the proposed marketing strategy is effective in expanding potential customers.

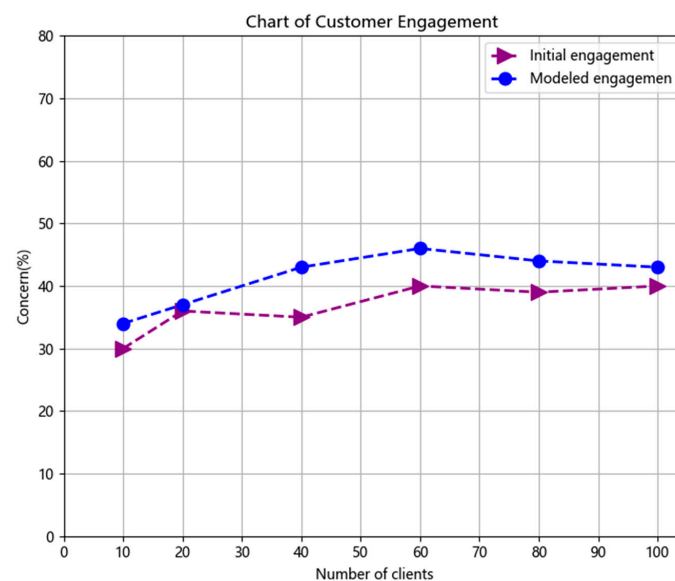


Figure 13. Chart of Customer Engagement.

5. Discussion

This study focuses on the segmentation of bank customers. To address the shortcomings of the traditional DBSCAN algorithm in terms of parameter selection, computational complexity, and adaptation to data characteristics, an improved KM-DBSCAN algorithm is proposed for customer segmentation. Based on the analysis results, marketing strategy formulation is suggested.

- (1) **Algorithm Stability:** In terms of algorithm stability, this paper introduces the K-means algorithm based on the traditional DBSCAN algorithm, improving the parameter selection method to make the clustering results more accurate and stable. Experimental results show that the KM-DBSCAN algorithm effectively avoids the clustering issues caused by the global uniqueness of parameters in the traditional DBSCAN algorithm. The silhouette coefficient reaches 0.97, indicating that the proposed algorithm has higher clustering stability.
- (2) **Consideration of Customer Attributes:** This study integrates both explicit indicators (such as occupation type, education level) and implicit indicators (such as consumption preferences) in the clustering analysis. It assigns different attribute weights by adding weight factors. The experiments show that the clustering accuracy is improved by 10.7% compared to the traditional DBSCAN algorithm. This demonstrates that the method can more accurately reflect the true characteristics of customers, improving the distinguishability of segmented groups. It also indicates that the integration of explicit and implicit indicators is an important direction in segmentation analysis.

Compared to other researchers, Dodda et al. [45] used the K-means algorithm for customer segmentation, but their approach can only effectively identify large customer groups in some cases, and is sensitive to the initial centroid selection, resulting in low clustering stability. In contrast, the KM-DBSCAN algorithm proposed in this paper can effectively identify multiple small customer groups and shows higher stability. The proposed algorithm identified four main customer segments, while the algorithm presented by Xiang et al. [46] only identified two customer groups. This suggests that the algorithm proposed in this paper is better at identifying more detailed customer characteristics.

6. Conclusions

This paper addresses the limitations of the traditional DBSCAN algorithm in customer segmentation by proposing an improved KM-DBSCAN algorithm based on K-means. The effectiveness and applicability of the algorithm are verified through experiments. The research findings show that the KM-DBSCAN algorithm can reduce computational complexity while improving clustering accuracy and stability. By incorporating a weight adjustment mechanism for both explicit and implicit indicators, this study further optimizes the customer segmentation model, providing technical support for targeted marketing and differentiated services. The results show:

- (1) With the contour coefficient, accuracy rate and F1 score as the evaluation indicators, the KM-DBSCAN algorithm achieves better overall performance than the K-Means algorithm and DBSCAN algorithm. At this time, the contour coefficient of the model was 0.97; the average accuracy rate was 92.3%; and the average F1 score was 0.92.
- (2) The proposed customer segmentation and marketing strategy is applied to real scenarios, with revenue growth rate and customer attention as evaluation indicators. The results showed that the average revenue growth rate after application of the model algorithm was 16.08%, and customer attention increased by 4.5%.

In terms of theory, this paper optimizes the traditional DBSCAN algorithm by introducing the K-means algorithm, proposing the KM-DBSCAN algorithm, which provides a new

approach for the application of density clustering methods in bank customer segmentation. This improvement not only enhances clustering accuracy and efficiency but also expands the potential of density-based clustering algorithms in real-world business scenarios. This study, through an in-depth exploration of both explicit and implicit customer characteristics, presents a customer segmentation model based on multi-dimensional information. This model not only focuses on basic customer information but also incorporates behavioral data, offering significant theoretical value, especially in the fields of customer behavior analysis and targeted marketing. This approach provides a new perspective for future academic research on consumer behavior patterns and group characteristics.

In practice, the findings of this study can assist banks in developing personalized marketing strategies based on customer group characteristics, helping banks allocate marketing and service resources more rationally, optimize resource distribution, and effectively improve customer satisfaction and operational efficiency [47].

Although the KM-DBSCAN algorithm demonstrates advantages in accuracy and efficiency, it still has certain limitations. When handling multimodal data, such as complex texts, images, and other types of data, the integration of these data types becomes more challenging, which increases the complexity of algorithm training [48]. While this paper improves customer segmentation accuracy by integrating explicit and implicit indicators, the set of integrated indicators remains limited. Customer characteristics involve more dimensions, such as social networks, which can also impact the effectiveness of customer segmentation. Future research could combine multimodal neural network methods [49] and natural language processing (NLP) techniques to process various data types (such as text, images, and speech), enhancing the model's ability to understand complex data. Fully integrating multidimensional data related to user characteristics, and utilizing social network analysis methods to segment customer groups with higher social influence, could help construct a multidimensional customer segmentation model, thereby achieving more precise customer profiling and marketing strategies. Finally, although our dataset mainly comes from a specific region, the customer behavior patterns and clustering results discovered in the study may have some similarities in other regions or countries. The algorithms and analytical framework we used also have broad applicability. In the future, we plan to incorporate datasets from more regions or countries to more comprehensively evaluate the performance of the algorithm and the generalizability of the conclusions.

Author Contributions: X.Y.: Writing-original draft, Visualization, Conceptualization, Methodology. Y.L.: Writing-review and editing, Methodology, Data curation, software and validation. F.N.: Writing—review and editing, Supervision, Software, Resources, Investigation. R.L.: Formal analysis, Conceptualization and investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported in part by the Shenyang Science and Technology Project under Grant 23-503-6-18, the Fundamental Research Funds for the Universities of Liaoning Province LJ232410143060.

Data Availability Statement: The datasets used in this study are publicly accessible.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Schilling, L.; Fernandez-Villaverde, J.; Uhlig, H. Central bank digital currency: When price and bank stability collide. *J. Monet. Econ.* **2024**, *145*, 103554. [[CrossRef](#)]
2. Anginer, D.; Bertay, A.C.; Cull, R.; Demirguc-Kunt, A.; Mare, D.S. Bank capital regulation and risk after the Global Financial Crisis. *J. Financ. Stab.* **2024**, *74*, 100891. [[CrossRef](#)]

3. Ranawat, N.S.; Chakraborty, A. The Impact of Third-Party Financial Products on the Consumer Loan Services Market in the Banking Sector: An Analysis of Sales Progress and Consumer Behavior. *Asia Pac. Financ. Mark.* **2024**, *31*, 367–387. [\[CrossRef\]](#)
4. Munusamy, S.; Murugesan, P. Modified Dynamic Fuzzy C-Means Clustering Algorithm—Application in Dynamic Customer Segmentation. *Appl. Intell.* **2020**, *50*, 1922–1942. [\[CrossRef\]](#)
5. Yuniningsih, Y.; Santoso, B.; Sari, I.M.; Firdausy, A.A.; Romadhon, I.C. Financial Literacy and Motivation to Stimulate Saving Behavior Intention in Form of Bank Customer Deposits. *J. Econ. Financ. Manag. Stud.* **2022**, *5*, 3334–3340. [\[CrossRef\]](#)
6. Zhan, Z.; Xu, B. Analyzing Review Sentiments and Product Images by Parallel Deep Nets for Personalized Recommendation. *Inf. Process. Manag.* **2023**, *60*, 103166. [\[CrossRef\]](#)
7. Yu, X.; Li, W.; Zhou, X.; Tang, L.; Sharma, R. Deep Learning Personalized Recommendation Based Construction Method of Hybrid Blockchain Model. *Sci. Rep.* **2023**, *13*, 17. [\[CrossRef\]](#)
8. Fu, Z.; Lian, T.; Yao, Y.; Zheng, W. Mulsimnet: A Multi-Branch Sub-Interest Matching Network for Personalized Recommendation. *Neurocomputing* **2022**, *495*, 37–50. [\[CrossRef\]](#)
9. Costa, R.; Di Pillo, F. Aligning innovative banks' sustainability strategies with customer expectations and perceptions: The CSR feedback framework. *J. Innov. Knowl.* **2024**, *9*, 100596. [\[CrossRef\]](#)
10. Kovacs, T.; Ko, A.; Asemi, A. Exploration of the investment patterns of potential retail banking customers using two-stage cluster analysis. *J. Big Data* **2021**, *8*, 141. [\[CrossRef\]](#)
11. Leclercq-Machado, L.; Alvarez-Risco, A.; Esquerre-Botton, S.; Almanza-Cruz, C.; Anderson-Seminario, M.D.L.M.; Del-Aguila-Arcentales, S.; Yanez, J.A. Effect of Corporate Social Responsibility on Consumer Satisfaction and Consumer Loyalty of Private Banking Companies in Peru. *Sustainability* **2022**, *14*, 9078. [\[CrossRef\]](#)
12. Armutcu, B.; Tan, A.; Ho, S.P.S.; Chow, M.Y.C.; Gleason, K.C. The effect of bank artificial intelligence on consumer purchase intentions. *Kybernetes* **2024**, ahead-of-print. [\[CrossRef\]](#)
13. Al-Shammari, M.; Mili, M. A fuzzy analytic hierarchy process model for customers' bank selection decision in the Kingdom of Bahrain. *Oper. Res.* **2021**, *21*, 1429–1446. [\[CrossRef\]](#)
14. John, M.J.; Shobayo, O.B. An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics* **2023**, *2*, 809–823. [\[CrossRef\]](#)
15. Chowdhury, S.; Helian, N.; de Amorim, R.C. Feature weighting in DBSCAN using reverse nearest neighbours. *Pattern Recognit.* **2023**, *137*, 109314. [\[CrossRef\]](#)
16. Kim, B.; Jang, H.-J. Genetic-Based Keyword Matching DBSCAN in IoT for Discovering Adjacent Clusters. *CMES Comput. Model. Eng. Sci.* **2023**, *135*, 1275–1294. [\[CrossRef\]](#)
17. Ouyang, T.; Shen, X. Online structural clustering based on DBSCAN extension with granular descriptors. *Inf. Sci.* **2022**, *607*, 688–704. [\[CrossRef\]](#)
18. Sun, Y.; Liu, H.; Gao, Y. Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon* **2023**, *9*, e13384. [\[CrossRef\]](#)
19. Ritter, T.; Pedersen, C.L. Is segmentation a theory? Improving the theoretical basis of a foundational concept in business-to-business marketing. *Ind. Mark. Manag.* **2024**, *116*, 82–92. [\[CrossRef\]](#)
20. Joung, J.; Kim, H. Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews. *Int. J. Inf. Manag.* **2023**, *70*, 102641. [\[CrossRef\]](#)
21. Sun, Z.-H.; Zuo, T.-Y.; Liang, D.; Ming, X.; Chen, Z.; Qiu, S. GPHC: A heuristic clustering method to customer segmentation. *Appl. Soft Comput.* **2021**, *111*, 107677. [\[CrossRef\]](#)
22. Tang, Y.E.; Mantrala, M.K. Incorporating direct customers' customer needs in a multi-dimensional B2B market segmentation approach. *Ind. Mark. Manag.* **2024**, *119*, 252–263. [\[CrossRef\]](#)
23. Tabianan, K.; Velu, S.; Ravi, V. K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability* **2022**, *14*, 7243. [\[CrossRef\]](#)
24. Othayoth, S.P.; Muthalagu, R. Customer Segmentation Using Various Machine Learning Techniques. *Int. J. Bus. Intell. Data Min.* **2022**, *4*, 20. [\[CrossRef\]](#)
25. Hicham, N.; Karim, S. Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation Using Clustering Ensemble and Spectral Clustering. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 58–65. [\[CrossRef\]](#)
26. Li, X.; Lee, Y.S. Customer segmentation marketing strategy based on big data analysis and clustering algorithm. *J. Cases Inf. Technol.* **2024**, *26*, 1–16. [\[CrossRef\]](#)
27. Tang, X.; Zhu, Y. Enhancing bank marketing strategies with ensemble learning: Empirical analysis. *PLoS ONE* **2024**, *19*, e0294759. [\[CrossRef\]](#)
28. Ziko, I.; Granger, E.; Yuan, J.; Ayed, I. Clustering with Fairness Constraints: A Flexible and Scalable Approach. *arXiv* **2019**, arXiv:1901.08223.
29. Torrens, M.; Tabakovic, A.A. Banking Platform to Leverage Data Driven Marketing with Machine Learning. *Entropy* **2022**, *24*, 347. [\[CrossRef\]](#)

30. Bohan, F. Research on the Marketing Strategy of Banking and Finance Business Given Big Data Technology. *SHS Web Conf.* **2023**, *154*, 02018.
31. Sis, B. Phenotypes and prognostic subgroups derived by the rejectclass clustering algorithm are not fully reproducible in an independent multicenter study. *Transplantation* **2024**, *108*, 1060–1061. [[CrossRef](#)]
32. Yin, L.; Hu, H.; Li, K.; Zheng, G.; Qu, Y.; Chen, H. Improvement of DBSCAN Algorithm Based on K-Dist Graph for Adaptive Determining Parameters. *Electronics* **2023**, *12*, 2537. [[CrossRef](#)]
33. Cheng, F.; Niu, G.; Zhang, Z.; Hou, C. Improved CNN-Based Indoor Localization by Using RGB Images and DBSCAN Algorithm. *Sensors* **2022**, *22*, 9531. [[CrossRef](#)]
34. Hou, G.; Wang, J.; Fan, Y. Wind power forecasting method of large-scale wind turbine clusters based on DBSCAN clustering and an enhanced hunter-prey optimization algorithm. *Energy Convers. Manag.* **2024**, *307*, 118341. [[CrossRef](#)]
35. Wenying, Z. An improved DBSCAN Algorithm for hazard recognition of obstacles in unmanned scenes. *Soft Comput.* **2023**, *27*, 18585–18604.
36. Hu, S.; Pang, Y.; He, Y. An Enhanced Version of MDDB-GC Algorithm: Multi-Density DBSCAN Based on Grid and Contribution for Data Stream. *Processes* **2023**, *11*, 1240. [[CrossRef](#)]
37. Hu, Y.; Zhang, C.; Cui, Z.; Ling, N. Construction and performance evaluation of big data prediction model based on fuzzy clustering algorithm in cloud computing environment. *J. Electr. Syst.* **2023**, *19*, 1–13.
38. Cheng, D.; Zhang, C.; Li, Y.; Xia, S.; Wang, G.; Huang, J.; Zhang, S.; Xie, J. Gb-dbscan: A fast granular-ball based dbscan clustering algorithm. *Inf. Sci.* **2024**, *674*, 120731. [[CrossRef](#)]
39. Qian, J.; Wang, Y.; Zhou, Y.H.X. Mdbscan: A multi-density dbscan based on relative density. *Neurocomputing* **2024**, *576*, 127329. [[CrossRef](#)]
40. Komatsu, H.; Kimura, O. Customer segmentation based on smart meter data analytics: Behavioral similarities with manual categorization for building types. *Energy Build.* **2023**, *110*, 112831. [[CrossRef](#)]
41. Sciascia, I. From market segmentation to customer loyalty. *Int. J. Bus. Manag.* **2023**, *21*, 132–145. [[CrossRef](#)]
42. Salaheldin, K.M.; Mohamed, H.; Islam, T.E. Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Comput. Appl.* **2024**, *36*, 4995–5005.
43. Wilbert, H.J.; Hoppe, A.F.; Sartori, A.; Stefenon, S.F.; Silva, L.A. Recency, frequency, monetary value, clustering, and internal and external indices for customer segmentation from retail data. *Algorithms* **2023**, *16*, 396. [[CrossRef](#)]
44. Li, Y.; Qi, J.; Chu, X.; Mu, W. Customer segmentation using k-means clustering and the hybrid particle swarm optimization algorithm. *Comput. J.* **2022**, *65*, 1285–1297. [[CrossRef](#)]
45. Dodda, R.; Babu, A. Text Document Clustering Using Modified Particle Swarm Optimization with k-means Model. *Int. J. Artif. Intell. Tools* **2023**, *33*, 2350061. [[CrossRef](#)]
46. Xiang, R.F. Use of n-grams and K-means clustering to classify data from free text bone marrow reports. *J. Pathol. Inform.* **2024**, *15*, 100358. [[CrossRef](#)]
47. Alblwi, A. Mdefc: Automatic recognition of human activities using modified differential evolution based fuzzy clustering method. *J. Comput. Sci.* **2024**, *81*, 102377. [[CrossRef](#)]
48. Wang, L.; Huang, Y.; Hong, Z. Digitalization as a double-edged sword: A deep learning analysis of risk management in Chinese banks. *Int. Rev. Financ. Anal.* **2024**, *94*, 103249. [[CrossRef](#)]
49. Lin, H.; Zhan, Y.; Liu, S.; Ke, X.; Chen, Y. A deep learning based bank card detection and recognition method in complex scenes. *Appl. Intell.* **2022**, *52*, 15259–15277. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.