Scholar Publishing Group
International Journal of Business Management and Economics and Trade
https://doi.org/10.38007/IJBMET.2026.070119
ISSN 2790-0959 Vol. 7, Issue 1:167-177

Research on Value Stratification and Precision Marketing of Retail Customers in Banks Based on Clustering Algorithm

Jiaqing Zhang

Small business banking, Capital one, Mclean, Virginia, 22102, US

Keywords: Retail banking customers; customer value segmentation; clustering algorithm; precision marketing; credit card customers; statistical validation

Abstract: Retail banking is shifting from scale-driven customer acquisition to refined management of existing customers. In this context, accurate customer value identification directly affects the efficiency of marketing resource allocation. To address the static nature of the traditional RFM model and the sensitivity of basic clustering algorithms to high-dimensional redundancy and extreme transaction fluctuations, this study uses an open-source credit card customer dataset from Kaggle. A customer value segmentation framework is developed by integrating Local Outlier Factor, principal component analysis, and K-Means++. Analysis of variance and post-hoc multiple comparison tests are further used for reverse validation. The results show that, based on 8950 initial samples and 18 original dimensions, 17 effective features were retained after removing the customer identifier. Local Outlier Factor identified and removed 269 abnormal samples when the neighborhood parameter was set to 20, accounting for 3.01% of the initial samples. Principal component analysis extracted 12 principal components, with a cumulative explained variance ratio of 96.53%. K-Means++ divided 8681 normal customers into four segments: the dormant long-tail segment, the basic engagement segment, the active growth segment, and the high-value behavioral segment. Further validation shows that the four segments differ significantly in activity, transaction frequency, and transaction amount. All pairwise post-hoc comparisons also reached statistical significance. These findings indicate that the proposed framework can identify interpretable customer value boundaries in high-dimensional, long-tailed, and noisy financial transaction data. The framework also provides a quantitative basis for precision marketing, resource allocation, and customer lifetime value management in retail banking.

1 Introduction

1.1 Global industry background and practical challenges

Due to competition from digital finance, the net interest margin of the entire bank group has fallen and customer acquisition costs have risen. Now the development of retail banks is not only

about increasing the number of customers. It will be based on the continuous improvement of customer lifetime value. In the service of credit cards, mobile banking and online consumption scenarios, customers generate data on purchase amount, number of transactions, frequency of purchases, credit limit, revolving balance, cash advance behavior, payment behavior, etc. The above data can be used by the bank to know its customers better. However, banks are still in a data-rich but information-poor state in reality. Banks have had more data fields than before, but they do not always contain information that can directly guide customer segmentation and resource allocation.

It is not that there is no data. It is due to the structure of actual financial transaction data. Generally speaking, the data have high dimensionality, long tails, overlap and noise. Purchase Amount, Transaction Count, Credit Limit and Payment are related in credit card data. If all the fields are directly used to build a model, some labels will be unnecessary for segmentation. At the same time, there may be extreme cases of short-term large purchases, concentrated cash advances or test transactions. These samples may cause the algorithm to think that the occasional fluctuations in account information are stable values. To boost the value of life from customers, segmenting customers should consider why some have done so more in the past. It can also help identify areas for development, reasons for customers leaving, etc. [1][2]

1.2 Research Gaps

Most of the research on customer value segmentation is still based on the traditional RFM model. The above ways can only determine the amount of contributions from current customers; they cannot observe the development process and other changes in value over time.

The first clustering algorithm is also not good. K-means has a small structure and is easy to compute. The data are of high dimension and redundant, and outliers or randomness in the initialisation of centroids will affect the results. Real banking customers are not naturally divided into several groups. They are rather spread out in terms of activities, frequency and money spent. A model that is only based on the internal clustering index may not be able to find customer groups that are far apart in geometry but have different business characteristics in this data structure. Recently, Machine Learning has been applied to the division and promotion of bank customers. However, the continuous framework that integrates noise reduction, dimension reduction, robust clustering and external value verification [5][6] has not been formed yet.

1.3 Research objective and contributions

Given the above problems, an open-source credit card customer dataset from Kaggle will be used in this paper. Local Outlier Factor, Principal Component Analysis and K-means++ were employed to divide the base of retail banking customers into various segments. Analysis of variance and post-hoc multiple comparison tests are employed in the reverse validation. Clustering is applied to the problem of customer lifetime value in this paper. The chain of research is: Anomaly Removal, Dimensionality Reduction, Robust Clustering, Value Validation and Marketing Transformation.

The three contributions of this paper are as follows: The theory in this paper links the RFM value logic to the changing behaviour of credit card users. Customer value is not only the result of historical transaction amounts. It is also shown in the following: Purchase Frequency, Transaction Count, Credit Use and Account Status over time. At the level of methods, an LOF-PCA-K-Means++ segmentation framework is constructed in this paper. It adds the internal clustering index and external statistics as well. The two paths of this Design are the stability of the model and business interpretability. Based on the results of the cluster analysis, some small groups of customers will be formed. It can put forward some quantitative plans for differential precision marketing and A/B tests given the bank's limited funds.

2 Related Work

2.1 Traditional customer value segmentation models

The RFM model is a typical old way of segmenting customers based on value. It has a small number of indicators and is easy to understand. It can provide an order of magnitude for the values of different customer groups promptly. Wilbert and others [3] have used all kinds of clustering methods on the RFM attributes. Segmentation quality was also judged based on the above and other measures. The above results show that RFM is still a good way to analyse customer behaviour. Rungruang and others [4] also used the RFM method in combination with a hierarchy. Research has helped divide the groups of customers in the past.

There are still some defects in the RFM model. Recency, frequency and monetary value are general indicators of past behaviour. They have a weak sense of the changes in value, risk, etc., and cannot identify early signs of customer churn. The value of credit card customers does not always rise steadily with their history of consumption. Some customers currently have a low amount of money, but their purchase frequency and number of transactions are increasing. Some customers have a long history of high contribution but may enter a churn period due to a drop in transaction volume. Banks only use a fixed RFM threshold and thus fail to identify the development of new customers. A small drop in the number of high-value transactions can be regarded as a stable high-value state. Therefore, RFM is better suited to be the validation reference for dynamic segmentation than to be used independently.

2.2 Machine learning for customer segmentation

Machine learning can also be employed for Customer Segmentation in this manner. Tang and Zhu [1] have studied how machine learning models can be applied to predict the needs of bank customers and optimise marketing. Based on the above data, we can better organise the customers and use different channels. Yan and others [5] have put forward an improved density-based clustering method to divide bank customers. Based on the above research, noise, parameter sensitivity and other attributes can alter the segmentation results of financial data. Vasheghani and others [6] have also studied the prediction and grouping of bank customers. Based on the above work, the construction of financial customer models should take into account behavioural factors, business interpretation and model stability. Kedar and Zhang [7-8] have also pointed out that the clustering of credit card customers should have high accuracy, interpretability and validation mechanisms.

K-means is relatively fast, and it is a simple model for unsupervised learning. DBSCAN can find noise and outliers. To see the hierarchy of the groups of customers, one can use a dendrogram. Real financial data are often correlated, have a long tail, and also contain abnormal transaction samples. A single algorithm cannot solve all the problems mentioned above. Previously, it was found in previous research that unsupervised learning could be employed for the segmentation of customer reviews, big data marketing and credit card spending pattern analysis [9-12]. Therefore, some research has started to organise bank customers. Data cleaning, dimensionality reduction, clustering robustness and external validity of the data still need to be done in the same framework. To know if the selected segments are suitable for the differences in customer value of bank operations, we need to figure this out.

3 Methodology

3.1 Dataset and experimental environment

The open-source Kaggle dataset of credit card customer information is used in this paper. The data was obtained on 2026. There are 8,950 credit card customer records and 18 original fields. The above are the fields: customer ID, balance, purchase amount, purchase frequency, cash advance, credit limit, payment behaviour, minimum payment, full-payment ratio and tenure. The purpose of the customer ID is only to differentiate individual samples. It has no model for the behaviour of customers. Therefore, this field was removed, and the remaining 17 good features were kept.

A Python Experiment will be carried out. The first stage of the data preparation is to address missing values, standardisation, outliers, principal component analysis, clustering and statistics. Impute the missing values with the median. Standardise the continuous variables. Thus, the large-scale variables of Purchase Amount and Credit Limit will not be included in the distance calculation. It can also show how much each factor has been contributed and what its level is on the same scale.

3.2 Feature engineering and variable definition

The 17 good indicators in this paper are divided into four categories: account relationship indicators, credit use indicators, consumption behavior indicators and customer maintenance indicators. The main element of the account relationship variable is tenure, which is the length of time a customer has been with a credit card account. Credit use variables are Balance, Balance Frequency, Credit Limit, Cash Advance and Cash Advance Frequency. The above are indicators of how much credit resources are being utilised and funded, and the corresponding risk.

Consumption behaviour variables are at the bottom of customer value identification. The above are: Purchase Amount, One-off Purchase Amount, Installment Purchase Amount, Purchase Frequency, One-off Purchase Frequency, Installment Purchase Frequency and Transaction Count. Purchase frequency can serve as an indicator of customer behaviour. Transaction Count shows how often customers use the credit card service in recent days. Purchase Amount corresponds to the contribution of behaviour. This paper will not use the term "high-net-worth customers" because the dataset does not contain asset information. It is called a "high-value behaviour segment", and these customers have made many purchases in a short period of time and have high spending power.

Customer maintenance variables are Cash Advance Transactions, Payments, Minimum Payments and Full Payment Ratio. The above can be used to find the loan repayment situation and fund pressure. The original data fields can be turned into the behaviour dimensions of retail banking by means of variable translation. It can help to organise data and plan marketing.

3.3 Algorithmic workflow and parameter settings

LOF-PCA-KMeans++ workflow for customer value segmentation is used in this paper. Anomaly Detection is the start of the model. A small number of customers in the real credit card transaction data may have made large purchases in a short period, concentrated cash advances, or other abnormal credit use. If these samples are added to the cluster directly, they will shift the centroid away from the usual behaviour of the customers. Local Outlier Factor is used to find the abnormal samples in this paper. The neighbourhood parameter is 20. LOF is based on the density of a sample and its neighboring samples. If the customer's amount spent, number of transactions, cash advance and credit use are far from that of typical customers, then the anomaly score will rise. This step is to correct the irregular fluctuation of accounts. It will not be considered one of the company's core customers.

Dimensionality reduction is the middle part of the model. Purchase Amount, Purchase Frequency, Transaction Count, Balance, Payments and Credit Use are all related to each other in the credit card data. All 17 features can be directly used; they may be too redundant. Principal Component Analysis will be employed in this paper. The Total proportion explained is 95%. PCA is to reduce the number of correlated variables. Reduce the number of dimensions and keep the main information. No rotation was carried out on the principal components. This will be taken as the input for the Standard PCA. The extracted number of principal components is 12. The proportion of explained variance is 96.53%. This shows that the reduced feature space still has the primary behaviour information of the customers.

Clustering is the first stage of the model. K-Means++ selects the starting centroids more reasonably and is less susceptible to the randomness problem of the basic K-Means. The numbers of the candidate clusters are 2 to 8. The within-cluster sum of squares and the silhouette coefficient are shown together. Business interpretability is also used to select the final number of clusters. Add 80 per cent of the sub-samples from the repeated clustering test to the list as well. The centre of the cluster should move only slightly when different samples are taken. To know if the model will still have a relatively stable structure of customer value after the perturbation of the sample, the above method was adopted.

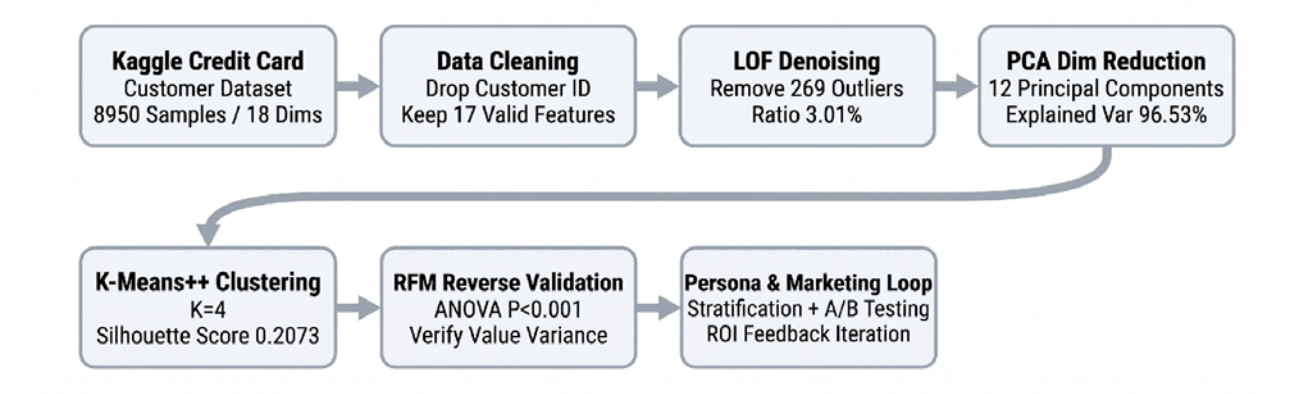

Fig. 1 Technical Workflow of This Study

3.4 Internal and external validation design

Unsupervised clustering does not have any actual labels. Therefore, only one internal indicator is not used as the sole basis for the validity of the model in this study. For the first step of verification, the silhouette coefficient is used to evaluate the coherence and separation of clusters. It is a type of structure. The final silhouette coefficient is 0.2073. The results will be discussed in the Discussion part of this paper, and the characteristics of financial transaction data and external statistics will be introduced then.

Frequency of purchase, number of transactions and amount spent are RFM proxy indicators for external validation. One-way Analysis of Variance is used to determine if the four groups of customers differ in terms of core values. Post-hoc multiple comparisons are used to find out if there are any differences between pairs. It makes sense. The silhouette coefficient is relatively small, but if the clusters of customers have similar patterns in the differences of external financial value indicators, we can still conduct clustering. ANOVA can determine if the mean of the entire group is significantly different. Post-hoc multiple comparisons are used to find the particular sections. Together, these two tests are the double verification methods in this paper.

4 Results

4.1 Data preprocessing and dimensionality reduction

Based on the 8950 initial samples and the 18 original fields, the customer identifier was removed. Seventeen good characteristics were kept. LOF found 269 outliers and removed them based on the neighbourhood parameter of 20. The Removal Rate was 3.01%. A total of 8681 normal customer samples have been saved. Thus, the real credit card data have some extreme fluctuations in account balance or irregular transaction patterns. To avoid a small number of large purchases or short-term funds being considered stable long-term values by the bank, anomaly removal is carried out. It will solve the problem of inefficient allocation of marketing resources.

PCA has been employed to obtain the first 12 principal components that explain about 95% of the total variance. Actually, the cumulative explained variance ratio is 96.53%. Therefore, it can be concluded that the original 17-dimensional feature space has some redundancy. After dimension reduction, the model still has most of the information about transaction behaviour, credit use and account relationship status. Therefore, the above model can remove outliers and reduce the dimension of the information prior to clustering.

4.2 Model optimization and segment profiles

To avoid a subjective assignment of the number of clusters, this paper tests candidate cluster numbers from 2 to 8. Within-cluster sum of squares, silhouette coefficient and business interpretability are compared together. As shown in Table 1, increasing the number of clusters from 2 to 4 significantly reduces the within-cluster sum of squares. It will still be lower than that, but the increase will be relatively small. When the number of clusters is 4, the silhouette coefficient is 0.2073. It is not the highest score of all the candidates. Not all types of customers need to be separated out of the bank's data. An Increase in the number of clusters will increase the geometric separation. It may also be a small group that is difficult to manage and operate. In order to reduce the within-cluster error and improve the value gradient and marketing feasibility of the customers, the four clusters were chosen as the final results.

Table 1 Model Optimization Indicators

Number of clusters	Within-cluster sum of squares	Silhouette coefficient
2	102304.33	0.2176
3	88560.72	0.2122
4	77575.18	0.2073
5	71125.16	0.2076
6	65425.25	0.2203
7	60229.71	0.2342
8	56925.17	0.2370

According to the clustering results, the four groups of customers have been named the dormant long-tail group, the basic-engagement group, the active-growth group and the high-value-behaviour group. As shown in Table 2, the proportion of the dormant long-tail segment is the highest at 43.30%. However, its activity is only 0.17, its transaction count is 2.91, and the amount of transactions is 268.72. This part has low activity, low frequency and a small amount of money. The proportion of the high-value behaviour group is only 6.42%. However, its activity is 0.96; the number of transactions is 74.31; and the amount of transactions is 5962.02. It is clearly in the lead across all three values. The rate of expansion is 36.38%. The activity is 0.88, but the transaction amount of this group is still far lower than that in the high-value behaviour group. It shows an

expanding state with strong activity but insufficient money supply. Fig. 2 shows the differences among the four segments in activity, Transaction Count and transaction amount.

Table 2 Profiles of the Four Customer Segments

Segment	Sample size	Proportion	Activity	Transaction count	Transaction amount	Segment name
Cluster 0	3759	43.30%	0.17	2.91	268.72	Dormant long-tail segment
Cluster 1	1207	13.90%	0.27	7.00	459.55	Basic engagement segment
Cluster 2	3158	36.38%	0.88	20.31	1094.08	Active growth segment
Cluster 3	557	6.42%	0.96	74.31	5962.02	High-value behavioral segment

Figure 2: Radar Profile of the Four Customer Segments. Min-Max Normalisation is used for the indicators

4.3 Reverse cross-validation and post-hoc tests

To know whether the clustering results are suitable for banking, Purchase Frequency, Transaction Count and Purchase Amount will serve as RFM proxy indicators. One-way Analysis of Variance is used for the four groups of customers. The F statistic for activity is 8485.55 according to the above results. The F statistic of transaction frequency is 4257.44. The F-statistic for the transaction amount is 3121.87. The p values of all three tests are less than 0.001. Therefore, the four parts are relatively different in terms of their primary source of financial value.

Based on the results of the post-hoc multiple comparisons, all pairwise comparisons of activity, transaction frequency and transaction amount are significant. All the adjusted p values are less than 0.001. The null hypothesis that there is no difference in groups is rejected in all cases. Some of the main results have better business interpretations. The mean difference in transaction amount between Cluster 0 and Cluster 3 is 5693.30. The mean difference in Transaction Count of Cluster 1 and Cluster 2 is 13.31. Based on the above results, the boundaries of the external value indicators

for all customer groups are relatively the same. Therefore, the validity of the model cannot be judged solely based on one internal index. Clustering Structure, Value Indicators and Post-hoc Comparisons are used together.

5 Discussion

5.1 Main findings and data-based comparison

The main result of this paper is that after LOF eliminates 3.01% of abnormal samples and PCA reduces the dimensionality of the data to 12 principal components, K-Means++ can still distinguish four distinct groups of customers with clear value gradients. The dormant long-tail group is about 43.30% of the customers. It has had only 0.17 activities and a transaction count of 2.91, with a transaction amount of 268.72. The proportion of customers in the high-value behaviour group is only 6.42%. Its activity is 0.96, the number of transactions is 74.31, and the amount of transactions is 5962.02. The other difference between the two groups is not based on money. It is evident from the activity, frequency and amount.

The active-growth part is also a section with some business functions. It makes up 36.38 per cent of the customers. The activity is 0.88; it is close to that of the high-value behaviour group. However, the amount of its transactions is only 1094.08. Thus, these customers have formed a habit of using the cards but have not fully released their spending power. Banks rank customers only by the amount of past transactions and thus give more weight to high-spending customers. They may also not be paying attention to customers who are still in a value-growth phase. Based on post-hoc comparisons, it can be concluded that all pairs of the four segments differ significantly for the three main value indicators. That is to say, the boundary division of the model is supported by external business reasons. It is also possible that there is a problem with the silhouette coefficient of 0.2073. Actual financial data are continuous and not scattered. A small number of internal geometric separations will not reduce the differentiation of external value.

5.2 Resource allocation and managerial implications

The Practical Value of Customer Segmentation is to Direct Resources. There are 3,759 dormant long-tail customers, making up 43.30 per cent of the sample. It is the largest section, but it has the lowest behaviour score. Banks should not invest too much in high-cost manual maintenance for this group. They should not be called often for instalments or wealth management. A more practical plan is low-cost and automatic, but it will have little impact. Billing reminders, small cash-back promotions and dormant benefit activation are some examples. The goal is not to immediately raise a large sum. To determine if the output is above the minimum point of the marginal cost curve.

The first group of engaged users consists of 1,207 people and constitutes about 13.9%. The number of transactions is 7.00, and the amount of these transactions is 459.55. These customers have begun to use the card to some extent, but their sense of habit is still relatively weak. Based on the daily life of the people, banks can develop inclusive activation plans. Advantages of local life, small-scale trial installations, points, etc. The resources required for this section will be relatively low or medium. Digital channels should be the main way to contact people so as not to waste marketing funds.

The active growth group has 3,158 customers and makes up 36.38 per cent of the sample. Its activity is 0.88, and the Transaction Count is 20.31. This group of people will be used for cross-selling and the management of value migration. These customers have a large number of frequent interactions but have not fully demonstrated their purchasing power. Banks can offer different marketing activities based on billing dates, peak usage times for transactions and credit, etc. They

can suggest some off-plan plans, improve credit lines, etc. The main purpose is to increase the approval rate and retain funds. The purpose is not to raise the number of times.

The high-value behaviour group has only 557 people and makes up 6.42% of the whole group. However, the number of transactions and the amount of money are much higher in this part. Do not activate the above group. The Strategy is to keep the upper-tier group and boost their experience. Banks will offer more premium advantages, special services, adjust the amount of credit freely according to need, etc. As the activity of this section has already reached 0.96, too much promotion may harm customer experience. Therefore, communication should be restrained and the service stable.

5.3 Quantitative evaluation and A/B testing loop

Precision Marketing Strategy should be included in the measurement cycle. It is proposed in this paper that the four divisions should be covered by an A/B test system. Return on Investment, approval conversion rate, response rate, average revenue per customer and risk exposure should be taken into account together. Therefore, the bank will randomly split its customers and carry out two groups of experiments. The circumstances of the experimental group will be according to the cluster profiles. The control group will be shown a general message or script. After the marketing cycle, the bank will compare the response rate, approval conversion rate, installment income, complaint rate, overdue rate and return on investment.

The two are the return and the risk. If the experimental group has a high conversion and return on investment but also a high complaint and overdue rate, then the strategy will have short-term gains and long-term risk mismatch. If there is an increase in the conversion rate, revenue and customer experience for the experimental group, then clustering profile can be used to strengthen the marketing funnel. Banks will then use the test results to update the customer label system. It is now in operation continuously.

5.4 Limitations and future work

This paper has some defects. The data can be found in the public credit card dataset on Kaggle. It is reproducible, but it does not have regional differences, changes in interest rate environment, customers' income changes and marketing response labels from actual banks. Before the model is used in a particular commercial bank, it needs to be retrained and validated with that bank's own business data. LOF-PCA-K-Means++ will also be used in this paper. It is not a generalisation of the comparisons of density-based clustering, hierarchical clustering and spectral clustering. DBSCAN finds the shape and noise. Hierarchical clustering is to be employed to explain the structure of customer groups. More cross-model comparisons will be carried out in the future.

The new workflow is in a good state at the engineering level. K-means++ is relatively simple in terms of scale when the number of clusters, feature dimensions and iterations are fixed. Linear Algebra optimisation or incremental calculation can be used to handle a large number of samples in PCA. LOF is a nearest-neighbour method, so in order to reduce the cost of deployment, approximate nearest neighbours can be employed; at the same time, batch processing and distributed computation will also be used. Therefore, after engineering optimisation, the proposed framework can be employed to solve the problem of large-scale customer data environments in banks. In the future, Graph Neural Networks will also be employed. In line with the above, compliance, external credit information, consumption circumstances and risk events will be added to expand the contents of the customer profile.

6 Conclusion

This paper will solve the problem of having rich but sparse data on retail bank customers' value segmentation. LOF-PCA-KMeans++ is employed to conduct the segmentation of customer value. According to the above experiments, LOF has been used to find and remove 269 outliers, accounting for about 3.01% of the total data. PCA had obtained 12 principal components, and they accounted for about 96.53% of the total variance. K-means++ divided the 8681 normal customers into four groups: the dormant long-tail group, the basic-engagement group, the active-growth group and the high-value-behaviour group. The division was made according to the combined evaluation of within-cluster sum of squares, silhouette coefficient and business interpretability.

According to the analysis of variance, the four segments have different levels of activity, transaction count and transaction amount. The p values of all post hoc multiple comparisons are less than 0.05. Both of the above are valid; thus, the proposed framework can find the upper bound of interpretable customer value. It can offer some statistics to help plan the operation of the customer.

The Division of Management will use the division results to decide how to allocate resources reasonably. Low-cost automatic restart is suitable for the idle long-tail group. Inclusive Activation is for the group of general participation. Cross-selling and Value Migration Management are suitable for the active growth group. Premium Retention is for the group of high-value behaviour. In short, this paper will offer a good quantitative guide to help retail banks improve the management of customer lifetime value, optimise the allocation of precision marketing resources, and build an A/B testing loop for the digital economy.

References

[1] *Wang, Z. (2026). Exploration of the Application of Resource Circular Economy in the Agricultural Industry Chain.*

[2] *Yin, J. (2026, March). An Efficient Iterative Algorithm for Calibrating Agency MBS Estimation Parameters Based on Bayesian Optimization, Implemented in Python. In 2026 IEEE Madhya Pradesh Section Conference (MPCON) (pp. 1462-1467). IEEE.*

[3] *Zheng, H. (2026, April). Research on Cloud-Edge Collaborative Elastic Computing and Cost Optimization for High-Concurrency Scenarios. In 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 1489-1496). IEEE.*

[4] *Liu, H. (2025, December). Mlops Model Deployment System for Multi-Cloud Environments and Improvement of Commercial AI Service Availability. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 1047-1052). IEEE.*

[5] *Hong, Y. (2025, December). Chain Demand Mutation Identification and Emergency Response Decision-Making Integrated With Social Media Sentiment Analysis. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 425-432). IEEE.*

[6] *Chen, M. (2026, March). Design of a Privacy-Oriented AI Compliance Hook System Based on Static Code Analysis. In 2026 IEEE Madhya Pradesh Section Conference (MPCON) (pp. 648-654). IEEE.*

[7] *Lyu, N. (2026, April). Toward Robust AI Agents: A Closed-Loop Task Planning–Execution–Feedback Framework for Open Scenarios. In 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 1182-1188). IEEE.*

[8] *Wang, Y. (2025, December). Construction of an Intelligent Recommendation System for Personalized Training Plans for Sports Rehabilitation Patients Based on Graph Neural Networks. In 2025 IEEE 1st International Conference on Recent Trends in Computing and Smart Mobility (RCSM) (pp. 1-6). IEEE.*

[9] *Zhang, C., Han, J., Zou, Y., Dong, K., Li, Y., Ding, J., & Han, X. (2024, April). Detecting the anomalies in LiDAR pointcloud. In WCX SAE World Congress Experience. SAE Technical Paper.*

[10] *Wang, Y. (2025, July). Research on Federated Learning Algorithm Design and Privacy Protection Mechanism of Sports Rehabilitation Data Based on Multimodal Fusion. In International Conference on Frontier Computing (pp. 563-574). Singapore: Springer Nature Singapore.*

[11] *Zhang, Q. (2025, July). Research on Improving the Effectiveness of Programmatic Advertising Using Multi-armed Bandit Algorithms. In International Conference on Frontier Computing (pp. 543-552). Singapore: Springer Nature Singapore.*

[12] *Zhang, K. (2025, December). Research on Cross-Selling Precision Marketing Strategy Based On Xgboost Model and SHAP Explanatory Framework. In 2025 IEEE 1st International Conference on Recent Trends in Computing and Smart Mobility (RCSM) (pp. 1-7). IEEE.*

[13] *Zhu, P. (2025, December). Construction of Multi-Scale Biostatistical Analysis Framework and its Application in Biomedical Signal Feature Recognition and Classification. In 2025 5th International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-7). IEEE.*

[14] *Yanchun Wang. (2025) Research on Enhancing ERP System Efficiency Through AI in Cross-border Supply Chain Environments. Advances in Computer and Communication, 6(5), 268-273.*

[15] *Wu, Y. (2026). Federated Learning-based Algorithm Design for Privacy Preservation in Cross-domain Data Sharing. Engineering Advances, 6(1).*

[16] *Ding, J. (2025). Exploration of Process Improvement in Automotive Manufacturing Based on Intelligent Production. International Journal of Engineering Advances, 2(2), 17-23.*

[17] *Hou, Y. (2026). Research on Heterogeneous Server Upgrade Strategies and Resource Utilization Efficiency Oriented Toward Green Computing Objectives. Advances in Computer and Communication, 7(1).*

[18] *Huang, J. (2025, September). Performance Evaluation Index System and Engineering Best Practice of Production-Level Time Series Machine Learning System. In 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT) (pp. 01-07). IEEE.*

[19] *Liu, X., & Yang, D. (2025, March). LLM Data Strategy: Improving Data Availability and Efficiency. In Doctoral Symposium on Computational Intelligence (pp. 425-437). Singapore: Springer Nature Singapore.*