

Research on User Profiling of “Internet + Nursing Service” Platform Based on Improved RFM Model

Yiliang Xie^{1*}, Chen Chen², Rui Zhang³

¹School of Statistics, Xi’an University of Finance and Economics, Xi’an, China

²Xi’an Buzz Lightyear Software Technology Co., Ltd., Xi’an, China

³Xi’an Yanwei Hat Medical Technology Co., Ltd., Xi’an, China

Email: *xyll0315@163.com

How to cite this paper: Xie, Y. L., Chen, C., & Zhang, R. (2025). Research on User Profiling of “Internet + Nursing Service” Platform Based on Improved RFM Model. *Open Journal of Social Sciences*, 13, 184-195. <https://doi.org/10.4236/jss.2025.134012>

Received: January 5, 2025

Accepted: April 20, 2025

Published: April 23, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Under the background of aging populations, constructing user portraits for the “Internet + Nursing Service” platform helps to deeply understand the demand characteristics of the elderly groups characteristics of elderly groups, laying a foundation for providing precise services. In this study, data from the nursing platform of Company X in Xi’an, extracted from the “Smart Health and Elderly Care Products and Services Promotion Catalogue (2022 Edition)”, were analyzed. User portraits were constructed by combining an improved RFM model with a two-step clustering algorithm, while the random forest algorithm was applied for model training and user identification. The results indicated that users could be categorized into three groups: loyal consumers, potential developers, and one-time users. Based on the classification outcomes, strategies such as promoting aging-friendly transformations, implementing refined marketing, and focusing on central urban areas were proposed to enhance the platform’s high-quality development and improve user experience and well-being.

Keywords

Internet + Nursing Service, User Profiling, RFM, Two-Step Clustering, Random Forests

1. Introduction

The construction of a Healthy China was proposed in the Report to the 20th National Congress of the Communist Party of China, recognizing people’s health as a crucial indicator of national prosperity. The 2024 State Council document Opin-

ions on Developing the Silver Economy to Improve Elderly Well-being emphasizes digital empowerment in elderly care services. Against the backdrop of rapid population aging in China, the high-quality development of “Internet + Nursing Services” has become imperative to address the surging demand for medical, rehabilitative, and daily care services. Previous studies have primarily focused on the intervention effects of continuity of care on specific diseases and factors associated with participation willingness among patients and healthcare providers. However, these investigations have predominantly employed questionnaires and interviews to assess subjective needs, potentially creating discrepancies with actual service requirements.

In contrast, the analysis of operational data from “Internet + Nursing Services” platforms enables more accurate identification of core demands, pain points, and latent needs. User profiling, as an analytical methodology characterizing user attributes, behaviors, objectives, and needs, has demonstrated unique advantages in precision service delivery. A data-driven profiling model developed through tag-based methodology facilitates the identification of diversified behavioral preferences and service demands among elderly users. This approach provides substantive support for optimizing service delivery mechanisms and enhancing elderly well-being through technological innovation in nursing services.

2. Research Status

“Internet + Nursing Services” refers to a healthcare delivery model where medical institutions utilize registered nurses and integrate next-generation technologies (IoT, cloud computing, big data analytics) to provide post-discharge care and mobile health services through online-offline coordination. This model, colloquially termed “online nursing reservations” (Huang et al., 2020), has demonstrated relatively high satisfaction rates among service recipients, particularly benefiting patients with chronic conditions and disabled elderly individuals (Shi et al., 2023). However, extant literature reveals critical knowledge gaps regarding determinants of service adoption and precise identification of high-value user segments. User profiling techniques enable data-driven characterization of user attributes, demand patterns, and behavioral trajectories. This methodology empowers healthcare providers to differentiate user cohorts and develop personalized care plans, thereby enhancing service precision and satisfaction metrics. Furthermore, these analytical frameworks facilitate targeted optimization of platform services through high-value user identification.

The conceptual framework of user profiling was pioneered by Alan Cooper in human-computer interaction studies (Mao et al., 2024). Initial implementations employed tag-based approaches to capture demographic attributes (Long, 2023). With advancements in machine learning algorithms, contemporary profiling systems now integrate multidimensional data layers encompassing psychographic characteristics, social networks, and real-time behavioral analytics. The integration of these multidimensional data points yields dynamic user representations,

which can have a deeper insight into the needs and expectations of users. While extensively applied in e-commerce and digital marketing, this approach remains under explored in elderly care contexts. Groundbreaking work by He et al. (2021), established methodological validity through elderly user profiling in smart pension systems. This theoretical foundation suggests methodological transferability to Internet + Nursing Services, particularly for demand prediction and service personalization in aging populations.

3. Data and Methods

3.1. Data Collection

The dataset was sourced from the Smart Health and Elderly Care Product/Service Promotion Catalog (2022 Edition) (MIIT-Electronics [2023] No. 176), containing 1600 service records generated from November 30, 2022 to July 18, 2023 on a Xi'an-based nursing platform.

During the process of selecting study subjects, the following data treatments were implemented in this research: 1) Data with order status marked as "completed" were selected; 2) Only user data receiving routine care services were retained; 3) Six characteristic parameters from users' most recent service records were extracted, including age, gender, disease condition, self-care ability, consciousness status, and past medical history; 4) The study scope was strictly restricted to orders within Shaanxi Province. After rigorous data cleansing, 778 validated records were obtained, including 332 core users receiving standard nursing services.

3.2. RFM Model

The RFM model has been widely adopted in customer relationship management as a quantitative valuation framework. This triaxial system evaluates: Recency (R): Time since last service utilization. Frequency (F): Total service engagements in observation period. Monetary (M): Cumulative consumption value. Through comparative analysis with cohort averages, users are stratified into eight value segments based on quantile distributions (He et al., 2022).

3.3. Two-Step Clustering

Two-step clustering, alternatively termed hierarchical clustering, is a pattern recognition technique employing similarity matrices for natural cluster formation. Its noise resistance and boundary detection capabilities make it particularly effective for healthcare analytics with heterogeneous datasets (Zhong et al., 2020).

3.4. Random Forest Algorithm

The random forest algorithm, a bagging-based ensemble learning method, enhances prediction stability through bootstrap aggregating of decision trees. By majority voting of constituent classifiers (typically 500+ trees), it achieves superior classification accuracy compared to single-tree approaches (Xu et al., 2024).

4. Empirical Analysis

4.1. Description of Basic Attributes of a User

Table 1. Basic characteristics of users receiving routine nursing services.

Basic characteristics	Scale (%)	Basic characteristics	Scale (%)
Sex		State of an illness	
N	(332)	N	(332)
Male	47.0	Stabilize	73.2
Female	53.0	Normal	19.6
		Be terminally ill	7.2
Age		State of consciousness	
N	(332)	N	(332)
<50	5.4	Sober	75.3
50 - 60	6.0	Muddled	19.3
61 - 70	12.7	Stun	5.4
71 - 80	22.3		
81 - 90	42.2	Regional distribution of orders	
>90	11.4	N	(751)
Ability to take care of oneself		Xincheng district	9.9
N	(332)	Beilin district	30.9
Take care of oneself	7.5	Lianhu District	14.8
Partial self-care	21.4	Baqiao District	0.8
Can't take care of themselves	71.1	Waiyang District	3.5
		Yanta District	18.6
Nursing care program		Yanliang District	13.4
N	(778)	Lintong District	0.1
Venous blood was collected	6.9	Chang'an District	1.4
Indwelling gastric tube or nasogastric tube care	41.1	Gaolin District	3.7
Nursing of indwelling catheter	27.5	Huyi District	2.9
Other	12.7		
Nursing care of peripherally inserted central catheters	11.8	Past medical history	
		N	(332)
		Have	81.9
		No	18.1

See **Table 1**, a total of 332 users received usual care services, and the ratio of male to female was almost equal. The average age of the users is 77.7 years old, and 191 of them are over 80 years old, accounting for 57.5%. In terms of self-care ability, the proportion of users who could not take care of themselves reached 71.1%. In terms of conscious state, the proportion of awake users was as high as 75.3%. Regarding the condition, 78.0% of the users were in a stable state, and 81.9% of the users had a past medical history. In terms of nursing items, the number of orders for indwelling gastric tube and nasogastric tube care and indwelling urinary tube care were more than 200, accounting for 41.1% and 27.5% of the total respectively. Notably, the reuse rates of peripherally inserted central catheter care, indwelling gastric or nasogastric tube care, and indwelling urinary catheter care were 50%, 42.3%, and 37.6%, respectively, with one user receiving 22 times of peripherally inserted central catheter care services. In terms of the geographical distribution of orders, only Xi'an's orders were analyzed, and the number of orders in Beilin District was the largest, accounting for 30.9% of the total.

4.2. User Value Stratification

4.2.1. Enhanced RFM-L Framework

While customer lifetime value (CLV) and AIPL models exist for user segmentation, the RFM framework was prioritized due to its operational simplicity and interpretability. However, two critical limitations were identified in conventional implementations: 1) The 8-category system induces managerial complexity in healthcare contexts, and 2) The homogeneity assumption contradicts observed user heterogeneity in nursing service adoption patterns.

To address these limitations, a novel temporal dimension (Length, L) was integrated to quantify loyalty duration (Yang et al., 2021). In this study, four variables are defined as follows:

Recency (R): Days between last service date and data cutoff (July 19, 2023).

Frequency (F): Total service engagements (November 30, 2022-July 18, 2023).

Monetary (M): Cumulative expenditure during observation period.

Length (L): Temporal span between initial and final service encounters.

As demonstrated in **Table 2**, users A and B showed comparable RFM scores (R: 28 vs 17 days; F: 5 vs 5; M: ¥1680 vs ¥2080). However, longitudinal analysis revealed Case B's engagement duration (L = 175 days) exceeded Case A's (L = 108 days) by 1.6-fold. The incorporation of this novel variable enables a more granular differentiation of user value categories, effectively distinguishing users with similar RFM profiles.

Due to the limited access to length value L, users who have only received a single service are classified into a separate group in this study. After using the traditional RFM model classification method to classify users, it is found that user A and user B are still classified into the first category, and the new index cannot be effectively used to distinguish users. Therefore, it is necessary to optimize the classification method to more accurately distinguish the value of different users.

Table 2. Comparison table of four variables for two users.

	User A	User B
L	108	175
R	28	17
F	5	5
M	1680	2080

4.2.2. Classify User Value by Using the Second-Order Clustering Method

Traditional user profiling has been conducted using the RFM model with K-means clustering. Compared to traditional K-means clustering, two-step clustering automatically determines the optimal number of clusters and employs a log-likelihood distance metric, significantly reducing sensitivity to initial centroids and improving robustness to noise and outliers. This method was prioritized over K-means due to its suitability for heterogeneous healthcare datasets and its ability to avoid classification bias caused by Euclidean distance assumptions. K-means clustering was initially tested in pilot experiments but exhibited instability due to randomness in initial centroid selection (Silhouette Coefficient fluctuation range: ± 0.15). In contrast, two-step clustering, utilizing hierarchical pre-clustering and optimized merging strategies, achieved more stable classifications (Silhouette Coefficient > 0.3), justifying its final adoption. Consequently, two-step clustering was implemented to analyze the four refined RFM metrics, yielding the following analytical outcomes:

User categorization thresholds were derived from the natural groupings identified by two-step clustering, validated by a Silhouette Coefficient of 0.32 (Table 3). The boundary between Cluster 1 and Cluster 2 was determined by cluster centroid distances (> 2 standard deviations). The first group with 24 users and the second group with 110 users. The average proximity of the first group is lower than that of the second group, and the average consumption times, total amount and length of the first group are higher, indicating that the first group has a high consumption frequency, a large total amount of consumption and a higher loyalty. In contrast, the consumption level of the second group of users is slightly lower than first, but there is still a large consumption potential, so this group of users is worth further development and exploitation of the platform. After two-step clustering classification, user A is classified into the second group, while user B is classified into the first group, indicating that user B is more worthy of attention than user A. This result confirms the above analysis and also demonstrates the rationality and effectiveness of using two-step clustering method.

The users who have only received the service once are classified into the third group. Based on the above analysis, all users are divided into three groups. The first group of users is named loyal consumers, the second group is named potential development users, and the third group is named one-time experienced users.

Table 3. Table of classification results.

	Clustering of clusters	
	1	2
Mean length (L)	181.4	63.2
The mean value of nearness (R)	21.3	80.7
Mean frequency (F)	9.5	3.2
Mean value degree (M)	3527.2	1053.8
Number of cases	24	110

4.3. User Value Prediction

4.3.1. Random Forest Modeling

In the domain of user value prediction, machine learning classification models including Random Forest, BP Neural Network, Decision Tree, and Naive Bayes constitute critical technological foundations. Particularly, the Random Forest algorithm demonstrates distinctive capability in mitigating overfitting risks when processing high-dimensional data. Through ensemble learning mechanisms, this algorithm generates feature importance rankings that serve as vital decision-making references for subsequent model optimization. Considering these technical attributes, this research selects the Random Forest model as the core methodology for user value prediction.

Regarding six fundamental user characteristics (age, gender, disease condition, self-care capability, wakeful state, and medical history), age is incorporated as a continuous variable in model computation. The remaining categorical and ordinal variables, due to their non-equidistant properties, undergo unified vector space mapping processing. The model output, user value, is quantified through a three-level characterization system derived from two-step clustering grouping. All datasets underwent standardized preprocessing before being input into the Random Forest model, with 100 datasets retained as the test benchmark for validating model performance.

The feature importance distribution diagram (Figure 1) generated post-training reveals that gender feature exhibits an importance value below the zero threshold, signifying that all other features contribute positively to model fitting. Building upon this discovery, this study proposes to conduct feature elimination experiments targeting gender characteristics, aiming to investigate their potential impact on prediction accuracy enhancement.

Although gender exhibited negative importance (Figure 1), it was initially retained to assess potential confounding effects. Subsequent Pearson correlation analysis revealed no significant associations between gender and age ($r = 0.08$, $p = 0.12$) or self-care ability ($r = 0.05$, $p = 0.28$), suggesting its negative importance stemmed from sample distribution bias rather than predictive relevance. Consequently, gender was excluded from the final model to enhance performance.

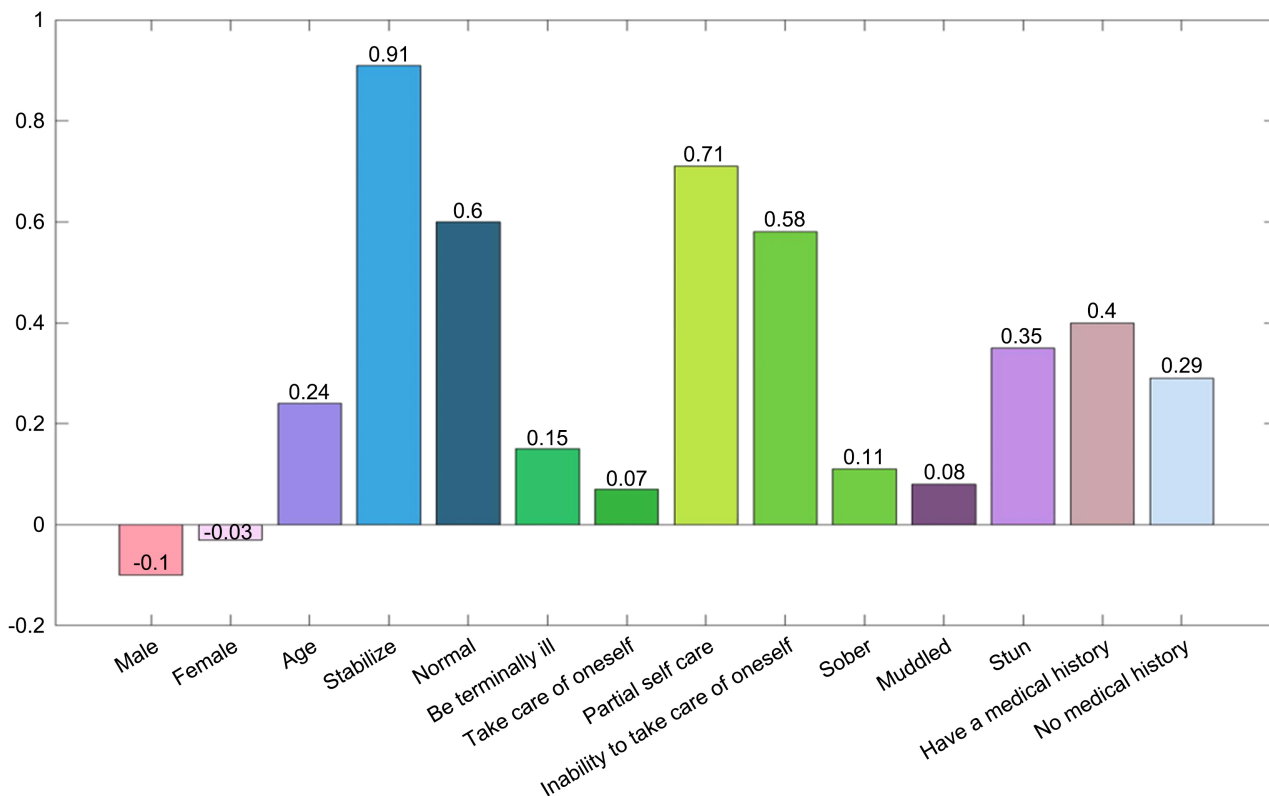


Figure 1. Importance map of input features.

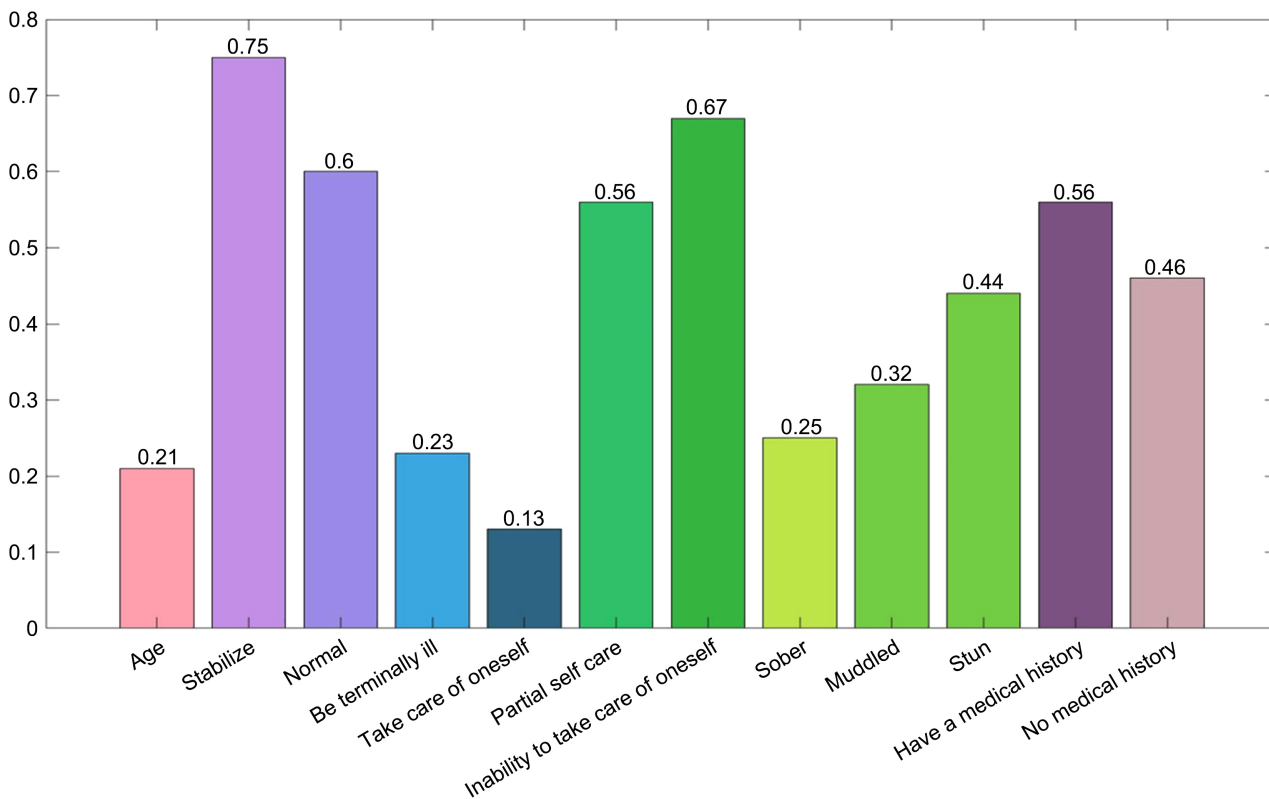


Figure 2. Importance map of input features after improvement.

Prior to the elimination of gender features, the baseline model attained a peak accuracy of 76.0%. Post-removal experimental results demonstrated a significant performance enhancement, achieving an elevated maximum accuracy of 83.0% accompanied by an out-of-bag error of 0.17, thereby demonstrating comparatively optimal fitting performance. The refined feature importance analysis in the gender-excluded model (Figure 2) highlights two clinically relevant parameters—stable disease status and impaired self-care capacity—as predominant determinants. These findings align with preliminary data characterization showing substantial sample representation for these attributes, establishing their critical discriminative value in user stratification.

To ensure optimal prediction performance, multiple models were systematically evaluated for data fitting (Table 4). Through extensive training iterations and comparative analysis of each model's best outcomes, empirical confirmation was obtained that the random forest model consistently demonstrated the highest fitting quality. All models (Random Forest, BP Neural Network, Decision Tree, Naive Bayes) underwent hyperparameter optimization via grid search (Random Forest: $n_estimators = 500$, $max_depth = 10$; Naive Bayes: Gaussian kernel). Naive Bayes was included as a baseline model due to its computational efficiency and suitability for testing data distribution assumptions. Its low accuracy further highlighted Random Forest's superiority in handling complex feature interactions.

Table 4. Comparison of predicted performance.

Model	Accuracy
Random forest	83.0%
BP neural network	80.0%
Decision number	60.0%
Naive bayes	38.0%

4.3.2. Forecast Result

For a novel user scenario, consider an individual with the following profile: aged 80 years, exhibiting stable medical conditions yet demonstrating compromised self-care capacity, maintaining cognitive awareness, and possessing a documented medical history. Upon inputting these demographic and clinical parameters into the predictive model, the system classifies this subject as a high-value loyal consumer, thereby designating them as a prioritized entity requiring dedicated platform attention and resource allocation.

5. Discussion

From a group classification perspective, this study systematically analyzes the demographic and service utilization patterns of routine nursing care recipients at Company X in Xi'an, classifying them into three distinct segments: loyal customers, users with development potential, and one-time trial users. Building on these

classification outcomes and integrating user value attributes, the following analytical insights and strategic optimization proposals are formulated.

5.1. Loyal Consumer

In the loyal customer segment, nearly 70% of users are female, with 70.8% aged over 80 years old and the majority in stable health conditions. Clinically, this cohort predominantly presents with stable medical conditions, while frequently requiring specialized nursing interventions—primarily indwelling gastric/nasogastric tube care (58.3% prevalence) and urinary catheter maintenance. Within this group, 58.3% of users are elderly individuals with disabilities, highlighting the urgent need for long-term care services and confirming the positive correlation between declining self-care ability and increased nursing demands among the elderly population (Liu et al., 2023).

To enhance service efficacy for this demographic, the platform should prioritize three strategic optimizations. Primarily, developing condition-specific care packages for stable yet functionally impaired elderly users, incorporating age-friendly interface adaptations. This includes creating simplified digital platforms with intuitive navigation to address technological accessibility barriers. Subsequently, streamlining the matching process for nursing staff to ensure professional care delivery aligned with individual disability requirements. Ultimately, introducing subscription-based long-term care packages (e.g., monthly/annual service plans) to meet chronic care needs while improving user retention.

5.2. Potential Development Users

In the potential development user segment, 78.2% maintain normal consciousness, 89.1% have documented pre-existing medical conditions, with 66.4% of elderly users demonstrating both characteristics. Despite exhibiting lower immediate expenditure metrics compared to loyal consumers, this demographic possesses a substantial population base and latent untapped consumption potential. Existing research indicates that overemphasis on high-net-worth users may result in homogenized customer acquisition channels, escalating costs, and increased vulnerability to price competition (Li et al. 2019).

To effectively engage this segment, the platform should prioritize three optimization strategies: In the first place, formulating personalized health counseling, dietary guidance, and rehabilitation programs based on users' medical histories. Furthermore, strengthening health education initiatives and psychological support services to enhance user engagement and retention. Additionally, establishing a service quality evaluation system to continuously refine user experiences through structured feedback mechanisms.

5.3. One-Time Experience Users

This cohort exhibits heterogeneous clinical profiles yet demonstrates epidemiological congruence with other segments through shared attributes: advanced age

(mean 78.6 ± 5.2 years), stable disease progression, functional dependency, preserved cognitive status, and prevalent medical comorbidities (82.4%). Platforms should prioritize optimizing the initial service encounter to enhance the probability of repeat care service utilization.

For this segment, the platform can implement the following initiatives: First, prioritizing initial service delivery quality and post-service follow-up care via SMS or app notifications, soliciting feedback on service satisfaction while providing follow-up recommendations or promotional offers. Second, concentrating resource allocation in high-order density areas (e.g., urban centers) with demonstrated service demand to improve accessibility. Third, implementing discounted pricing or loyalty reward programs to convert users into long-term subscribers, thereby maximizing their long-term value potential.

6. Conclusion

This study constructs user profiles for “Internet + Nursing Service” platforms using an enhanced RFM model integrated with two-step clustering and random forest algorithms, classifying users into loyal customers, potential developers, and one-time experience users. This methodological framework improves both the accuracy and practicality of user profiling, providing critical support for optimizing resource allocation, service targeting precision, and regional resource integration. Notably, the innovation lies in the integration of the user relationship length (L) metric with machine learning algorithms and big data analytics techniques, which significantly enhances classification accuracy. Future research directions include diversifying data sources, refining algorithmic models, and conducting scenario-based experimental validations to further enhance the model’s practical applicability. The proposed approach contributes to advancing high-quality development within the “Internet + Nursing Services” domain.

Acknowledgements

Thanks to the reviewers for their comments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- He, S. P., Su, H. Z., & Yao, K. F. (2022). Adaptive Comprehensive Evaluation Method for Dam Measured Behavior Based on RFM Model. *Journal of Yangtze River Scientific Research Institute*, 39, 82-88.
- He, Z. Y., Zhu, Q. H., & Bai, M. (2021). Construction of Urban Elderly User Portraits from the Perspective of Elderly Care Services. *Journal of Intelligence*, 40, 154-160.
- Huang, Y. S., Yuan, C. R., Song, X. P. et al. (2020). Development Status of Internet+ Nursing Service. *Chinese Nursing Research*, 34, 1388-1393.
- Li, P., Peng, Y. N., & Kang, J. (2019). The Impact of Platform Market Assets on Merchant

- Loyalty: The Moderating Role of Platform Competition. *Management Review*, 31, 103-118.
- Liu, X. C., Wang, X. L., & Ma, Y. Z. (2023). Practice of Extended Care for Elderly Patients with Chronic Diseases Based on Internet + Medical Alliance. *Journal of Nursing Science*, 38, 100-104.
- Long, Q. (2023). Construction and Analysis of User Portraits in University Library Multi-dimensional Spaces: Taking Wuhan University Library as an Example. *Library Journal*, 42, 120-131.
- Mao, T. T., Liu, J., & Mao, J. B. (2024). Empirical Study on the Response Behavior Portrait of Elderly Users to False Information on Mobile Social Media. *Library and Information Service*, 68, 117-128.
- Shi, H. H., Qian, W. Y., Feng, J., Ke, Z. W., & Bi, D. J. (2023). Practice and Effect Analysis of Internet+ Nursing Service in Taizhou City. *Hospital Management Forum*, 40, 91-93, 27.
- Xu, H. M., Qiu, S. Y., Wang, J. X. et al. (2024). Clinical Outcomes of Predicting Postoperative Hemorrhage Rate in Pediatric Tonsillectomy Based on Random Forest Model. *Journal of Clinical Otorhinolaryngology Head and Neck Surgery*, 38, 883-890.
- Yang, L., Kou, Y. G., Bai, Z., & Liu, H. C. (2021). Study on Civil Aviation Customer Segmentation Based on Improved RFM Model. *Mathematics in Practice and Theory*, 51, 33-39.
- Zhong, Y. Y., Chen, J., & Shao, Y. M. (2020). Study on Segmentation Algorithm of Urban Traffic Vulnerable Groups Based on Second-Order Clustering. *Journal of Computer Applications Research*, 37, 132-134.