



Research on Brand User Profiling and Personalized Recommendation Strategies under a Multimedia Platform Data Integration Framework

Xiaoyu Cui^{1,*}, Jingjing Zhang¹

¹ Economics and Management School, Beijing Polytechnic University, Beijing, 100176, China

SUMMARY: *The consumer behavior of shopping and entertainment on multimedia platforms has more multidimensional data properties than before, which offers more information to create brand user profiles and formulate recommendation strategies in the internet age. The paper chooses Brand H as the object of study and simulates and evaluates static data and rating data of 20,000 users in China to serve as a reference model of brand user profiling. To enable the user data clustering and profiling, the K-Means clustering algorithm is chosen as the data aggregation algorithm. Simultaneously, a new recommendation model is presented, which combines the traditional stacking, but with increased dimensions of input features and weighting of various features. It results in better-stacking-based recommendation model with enhanced prediction accuracy. When building user profiles and providing personalized recommendations to consumers regarding Brand H, the recommended improved stacking recommendation model had a maximum average reciprocal hit rate of 0.281 and a maximum average precision of 0.738, indicating that it can be applied effectively in practice.*

KEYWORDS: *personalized recommendation; K-Means clustering; brand user profiling; improved stacking*

1 Introduction

With the advent of big data, user profiling and personalized recommendations have become essential tools of the internet industry due to the development and spread of multimedia platforms. With the combination of information on multimedia platforms as well as the analysis of user behavior trends, preferences, habits, and social networks, it is feasible to characterize and gain a better insight into users interests and requirements more precisely, thus allowing making personalized recommendations and accurate marketing of brands [1-4].

The process of user profiling is the combination of integration, analysis, and modeling of user data in order to come up with profiles of the essential features, behavioral tendencies, and values of users [5, 6]. The concept of user profiling goes beyond the traditional, static descriptions in the framework of multimedia platform data integration in terms of dimensions such as gender and age. Rather, it is about looking through large volumes of user data to identify correlations and trends, dividing users into various groups and attributing them with certain attributes and features [7-11]. Sources of data play a significant role in the user profiling. Brands have access to massive volumes of user data that includes browsing history, purchase history, and activities on social networks [12-14]. Such data reflects personal preferences and behavioral characteristics, which can be strongly used in user profiling.

*103067@bpu.edu.cn

<https://doi.org/10.65102/is2026198>

Eke, C.I. et al. [15] systematically brought in user profiling, discussed the most current trends, and offered some research on modeling issues including data acquisition and feature extraction. Maraj, D. et al. [16] developed further details on the research that is aimed at user profiling and approaches to safeguarding user data privacy by giving examples of popular internet services to illustrate the variety of information gathered about users. Mushtaq, M.S. [17] highlighted the significance of user perception of service quality in multimedia services. They built a testing platform to explore the factors that influence the experience quality of video streaming over wireless and wired networks. With research on user profiling, they offered an understanding of all the metrics that impact experience quality. Yan, H et al. [18] highlighted the need to emphasize user profiling to improve shopping experiences and sales revenues. They have been able to define unique customer groups based on their preference towards particular product categories that meet the same demands because of the analysis of online shopping and browsing history records to learn more about the behavior of users.

On the basis of customer profiling, personalization of brand recommendations has become one of the leading applications of the big data era [19, 20]. Conventional recommendation systems depend heavily on the similarity of content and past user actions, which is rather general in nature. But in the context of multimedia platform data fusion, recommendation systems use deep learning and machine learning technologies that help them discover better underlying interests and requirements of the users and thus make recommendations more personalized and accurate [21-24]. The main idea behind personalized recommendations is offering information and services suitable to the interests of the user. Processing data and algorithmic models are the essential components to personalized recommendations [25, 26]. It is possible to reveal latent user needs and characteristics based on user profiles and behavioral data to develop personalized recommendation models [27]. As an example, collaboration filtering algorithms consider past user behavior and preferences to suggest items that are likely to be liked by other users who are similar. Content-based recommendation algorithms consider the attributes of the items and the preferences of the user to propose items of similar content [28, 29].

Gu, Y et al. [30] suggested a hierarchical user profiling structure (HUP) to overcome multi-level user profiling of e-commerce recommendation systems. Experiments on a large scale proved that HUP performed better than any existing approach when it came to hierarchical profiling and recommendations problems. Amoretti, M. et al. [31] have introduced a recommendation generation procedure that combines user profiling with context-based data filtering. This has been applied to an intelligent travel application named UTravel to suggest points of interest to end-users and experimental assessment outcomes were presented. Huang, J. [32] have introduced a recommendation generation procedure that combines user profiling with context-based data filtering. This has been applied to an intelligent travel application named UTravel to suggest points of interest to end-users and experimental assessment outcomes were presented. Bonomo, M. et al. [33] presented a generalized model to recommend possible users by comparing online social network profiles with brand profiles and confirmed the effectiveness of the method in finding the best user group to use in targeted advertising campaigns. Lin, Q. et al. [34] analyzed personalized pricing strategy in the context of strategic user profiling, which is based on social networks, and stated that with the increasing accuracy of profiling, brands are likely to increase equilibrium uniform pricing to encourage more social engagement. Suggested an IoT-based system, which is the basis of user profiling Mohamed, S. et al. [35]. It holds all the contextual information, provides the possibility of personalizing the content and makes personalized product recommendations depending on multi-source identification of the data.

This paper first simulates and analyzes static data and rating data from H brand users, demonstrating the user information construction pathway and outlining preprocessing steps for user data. It then details the operational flow of the K-Means clustering algorithm as a method for mining user characteristics. Subsequently, it briefly describes the architecture and limitations of traditional stacking recommendation models, designs improvement methods for their primary shortcomings, and proposes a recommendation model based on enhanced stacking. Finally, using Brand H as the research subject, it visualizes user preference information, organizes user profile tags, and constructs a brand user profile. Combining the brand user profile with brand-category products, it calculates the weight matrix for brand products and categories, thereby conducting simulation experiments on the recommendation effectiveness of the proposed model.

2 User Information Construction

2.1 User Information Simulation

Due to the rapid development of the internet, multimedia platforms are increasingly able to access users' private data, and more people are paying greater attention to protecting their own information. Therefore, this paper takes 20,000 Chinese users of Brand H as the research subjects, simulates the data information of Brand H users on multimedia platforms in the year 2024, and analyzes this simulated data, which can also yield excellent experimental results. This study selected a total of 20,000 simulated user profiles, proceeding below with the simulation of user static data and user rating data.

2.1.1 User Static Data Simulation

User static data simulation primarily involves modeling users' age, gender, province, and educational background.

The age distribution follows an empirical distribution between 16 and 70 years old. Based on China's population age structure over the past decade, the final calculations show that individuals aged 16-25 account for 13.97%, 21.01% aged 26-35, 9.46% aged 36-40, 47.83% aged 41-65, and 7.73% aged 66-70.

Gender distribution follows an empirical distribution between “male” and “female.”

The provincial distribution follows an empirical distribution across 31 provinces (excluding Hong Kong, Macau, and Taiwan regions). Provincial population proportion data is presented in Table 1, where provinces/municipalities with population shares exceeding 5.00% are listed as follows: Guangdong (8.84%), Henan (7.13%), Jiangsu (6.09%), Sichuan (5.84%), and Hebei (5.19%).

Table 1: Population proportion data of each province

Province	Population Proportion	Province	Population Proportion
BeiJing	1.46%	TianJin	1.07%
HeBei	5.19%	ShanXi	2.56%
NeiMengGu	1.61%	LiaoNing	3.11%
JiLin	1.62%	HeiLongJiang	2.35%
ShangHai	1.67%	JiangSu	6.09%
ZheJiang	4.48%	AnHui	4.41%
FuJian	2.85%	JiangXi	3.29%
ShanDong	7.11%	HeNan	7.13%
HuBei	3.98%	HuNan	4.8%
GuangDong	8.84%	GuangXi	3.64%
HaiNan	0.62%	ChongQing	2.36%
SiChuan	5.84%	GuiZhou	2.82%
YunNan	3.25%	XiZang	0.35%
ShanXi	2.71%	GanSu	1.86%
QingHai	0.33%	NingXia	0.6%
XinJiang	1.74%		

Educational attainment distribution follows an empirical distribution across the following categories: (E1) illiterate, (E2) elementary school, (E3) junior high school, (E4) senior high school, (E5) college (associate or bachelor's degree), and (E6) master's or doctoral degree. Specifically, (E1) illiterate accounts for 1.182%, (E2) elementary school for 13.340%, (E3) junior high school for 41.797%, (E4) high school for 20.135%, (E5) college (including junior college and undergraduate) for 22.332%, and (E6) master's or doctoral degrees for 1.207%. The static data simulation for 10 randomly selected users is shown in Table 2. For example, User 00562 is 36 years old, female, registered in Beijing, and holds a master's or doctoral degree.

Table 2: Static data simulation of 10 random users

User ID	Age	Gender	Province	Education background
00562	36	Woman	BeiJing	E6
00331	25	Man	NingXia	E5
00458	43	Man	ChongQing	E4
13679	29	Woman	ShangHai	E6
17008	49	Man	GuangDong	E4
13745	24	Woman	XinJiang	E5
08078	31	Woman	AnHui	E5
01984	68	Man	HuNan	E2
01366	59	Man	TianJin	E3
01404	37	Man	FuJian	E5

2.1.2 User Rating Data Simulation

(1) User Rating Simulation

This simulation models user behavior when rating 100 different brand products. It assumes a 50% probability that users may choose not to rate a specific product subcategory, resulting in a rating value of 0 for that category. In other words, if a user's rating data for a category is 0, it indicates that user did not rate products within that category. There is also a 50% chance that a user will rate products in this category, with scores ranging from 20 to 100 in increments of 20. This means user ratings will be randomly distributed among 20, 40, 60, 80, and 100. This study uses five users with IDs 00562, 00331, 00458, 13679, and 17008 as samples. Their predicted ratings for five randomly selected brand products are shown in Table 3. Users 00331 and 00458

did not rate laptops or mechanical keyboards, but gave high ratings to phones, water bottles, and necklaces, indicating high satisfaction with the branded products.

Table 3: Example of user rating prediction

User ID	Branded merchandise				
	Notebook computer	Mechanical keyboard	Mobile phone	Water glass	Necklace
00562	80	100	80	0	0
00331	0	0	80	100	0
00458	0	0	100	80	60
13679	40	60	60	20	0
17008	20	20	40	60	0

(2) User Rating Timing Simulation

This paper also simulated three key time points: the first rating time within the statistical period, the last rating time within the statistical period, and the platform's statistical time. First, the platform's statistical time was set to December 31, 2024, with the statistical period spanning January 1, 2024, to December 31, 2024. This period was used to calculate the first and last rating times for users within the platform during this timeframe. Text selection randomly generates two points within this timeframe. The earlier point is then designated as the user's first rating time within the statistical period, while the later point serves as the last rating time. The predicted rating times for five users (IDs: 00562, 00331, 00458, 13679, and 17008) are shown in Table 4.

Table 4: Example of predicting user rating time

User ID	The first scoring time	The last scoring time	Platform statistics time
00562	2024/2/29	2024/3/25	2024/4/9
00331	2024/1/21	2024/2/6	2024/2/12
00458	2024/8/8	2024/10/19	2024/10/27
13679	2024/4/8	2024/6/14	2024/6/20
17008	2024/2/1	2024/9/27	2024/10/3

According to Table 4, User 17008 has a first rating that was made quite early (February 1, 2024) and a last rating that was made fairly late (September 27, 2024), indicating that this user could be a frequent rater of products. Conversely, User 00331 had both its first and last ratings less than half a month apart, with the final rating being in February. Since then, no more rating activity has been recorded, and it is possible that this user does not often rate products.

2.2 Preprocessing of User Information Data

The user data that has been gathered in this paper needs to be first preprocessed to enable further analysis and the process is subdivided into five steps.

(1) Data Cleaning: Eliminate duplicate, erroneous or incomplete records in order to guarantee the quality and consistency of data.

(2) Data Transformation: Transform unstructured data (text, images, etc.) into structured data to be analyzed numerically. As an example, the conversion of the user search queries into a list of keywords or the transformation of the user feedback into measurable ratings.

(3) Feature Selection: The choice of the most important features among the huge data set which are related to the user behavior and preference. It also decreases dimensionality and improves the effectiveness of clustering analysis.

(4) Data Standardization: The process of standardizing data to remove the influence of various units of measurement and scales. It is usually done through normalization or scaling.

(5) Handling Missing Values: In case of missing data, the choices available are deletion of the corresponding records or filling in the missing values with the help of statistical procedures (e.g., mean, median, mode). After these preprocessing measures have been implemented, there will be a neat and organized database that can be used to further process the data through the K-means clustering method, which will improve the precision and consistency of clustering outcomes.

3 User-Based Recommendation Algorithm Model

3.1 User Profile Features Based on the K-Means Clustering Algorithm

K-Means clustering algorithm is an easy and classical distance-based clustering technique. The distance is employed to determine similarity between objects whereby the assumption is that two objects that are close in distance will be more similar than objects that are further apart. Clustering involves dividing a large, unlabeled dataset into distinct categories based on latent data characteristics within the data itself, ensuring that data points within a category are relatively similar while those between categories exhibit greater dissimilarity. This process falls under unsupervised learning.

The K-Means algorithm, as an iterative clustering analysis method, hinges primarily on the setting of the K parameter. The main workflow of this algorithm is illustrated in Figure 1.

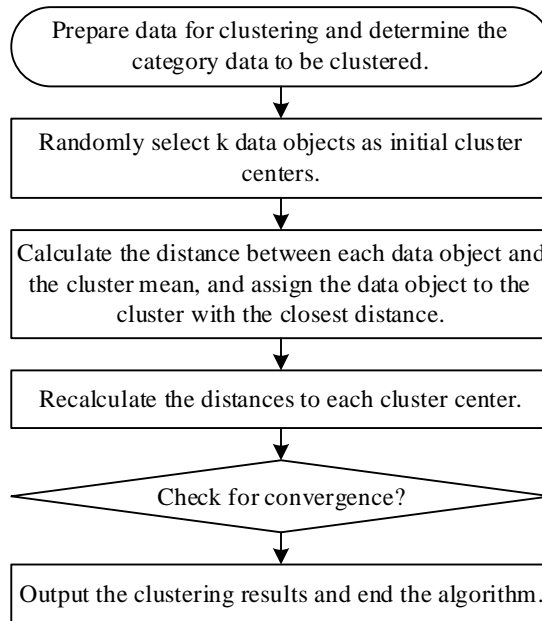


Figure 1: The process of the K-Means algorithm

The primary steps of the K-Means clustering algorithm are as follows:

(1) Randomly select K objects as initial cluster centers. The value of K is initially determined based on empirical judgment. If unfamiliar with the algorithm, cross-validation should be used to determine the optimal K value.

(2) For each point in the dataset, calculate its distance to every cluster center. The point is assigned to the cluster whose center it is closest to. The distance metric used is the squared Euclidean distance, where x represents a sample point within a cluster and μ denotes the center of that cluster. The formula is shown in Equation (1):

$$[d(x, \mu)]^2 = \sum_{i=0}^n (x_i - \mu_i)^2 \quad (1)$$

Using Euclidean distance, the sum of the squares of the distances from all sample points in a cluster to the centroid is given by Equation (2):

$$CSS = \sum_{j=0}^m \sum_{i=0}^n (x_i - \mu_i)^2 \quad (2)$$

where μ denotes the centroid of the cluster, m represents the number of samples in a cluster, and j is the index of each sample. For a cluster, the smaller the sum of distances from all sample points to the centroid, the more similar the samples within that cluster are considered to be, indicating less variation within the cluster and thus better clustering performance.

(3) Recalculate the center of each cluster by taking the average (or median) of all objects within that cluster.

(4) Repeat steps (2) and (3) until a termination condition is met: no (or minimal) objects are reassigned to different clusters, no (or minimal) cluster centers change, and the sum of squared errors reaches a local minimum. At this point, clustering is considered satisfactory, and the algorithm terminates.

K-Means algorithm has a variety of benefits including the simplicity of its principles, ease in its implementation, rapid rate of convergence and high level of interpretability. It will work especially well in situations where the resulting clusters are dense with high levels of separation between them. K-Means clustering also has its disadvantages. As an example, the value of K cannot be easily determined and is usually selected using empirical reasoning. Moreover, it is only applicable to clustered data types and works badly with other data types.

3.2 Personalized Recommendations Based on User Profiles

3.2.1 Recommendation Models Based on Traditional Stacking

When forecasting using the model prediction methodology, it has been observed that depending on one model in such a case is usually associated with some downsides. One model can be inefficient at adjusting to various types and features of data and thus perform poorly in particular situations. Moreover, multimedia platform data is naturally complex and the use of a single model to predict it has the danger of overfitting leading to poor performance in the training of new datasets. In order to overcome the problem of single-model prediction, this paper uses the stacking model ensemble approach where different models are combined in the forecasting process, increasing its precision and generalization properties.

The stacking method primarily consists of two layers of learners: the base learners and the meta-learner. This paper employs logistic regression, random forest, and XGboost as base learners. Since the base learners utilize different types of classifiers, the meta-learner selects the simpler KNN as the training model to reduce the risk of overfitting. The specific steps for constructing the recommendation model based on traditional stacking are as follows:

Step 1: Input the near-seven-day multimedia platform dataset as the training set into models such as logistic regression, random forest, and XGboost. Let the classifiers be denoted as $C_i (i = 1, 2, 3)$. Divide the training set into five portions: D_1, D_2, D_3, D_4 , and D_5 . For each training iteration, one dataset serves as the test set while the remaining four form the training set. Each test set generates a prediction label denoted as P_{ij} . Ultimately, five prediction labels are compiled into the prediction matrix $P_i = [P_{i1}, P_{i2}, P_{i3}, P_{i4}, P_{i5}]^T (i = 1, 2, 3)$, where i denotes different models.

Step 2: Input the training sets into models such as logistic regression, random forest, and XGboost, yielding prediction results P_1 , P_2 , and P_3 . Combine the prediction results from different models into a prediction matrix as the training set for the meta-learner: $P = [P_1, P_2, P_3]$.

Step 3: Input the prediction matrix P from Step 2 and the true labels of user behavior into the meta-learner. Train the meta-learner using KNN to obtain the final meta-learner model.

Step 4: Input the test set into the trained logistic regression, random forest, XGboost, and other models to obtain prediction results M_1 , M_2 , M_3 . Combine the prediction results from the three models into a matrix as the meta-learner's test set, denoted as $M = [M_1, M_2, M_3]$. Here, M_i is the result matrix obtained by taking the mode from five prediction result matrices generated via five-fold cross-validation.

Step 5: Input the result matrix M from Step 4 into the meta-learner trained in Step 3 to obtain the final prediction results. Compare these with the true labels to calculate the prediction accuracy.

The specific workflow for stacking model integration is illustrated in Figure 2.

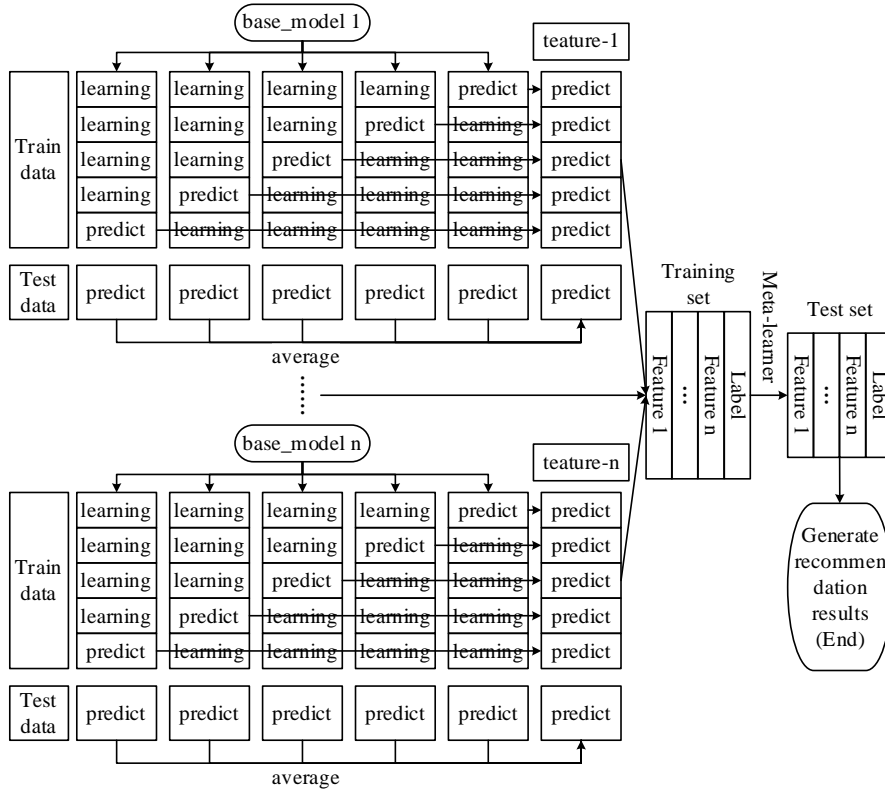


Figure 2: Stacking model fusion flowchart

3.2.2 Recommendation Model Based on Enhanced Stacking

Based on the fusion steps of the traditional stacking model discussed in the previous section, two significant issues are evident: First, the input feature dimension of the meta-learner in traditional stacking models is limited. The feature dimension of the meta-learner depends on the number of base learners. In the previous section, the base learners used were logistic regression, random forest, and XGboost, resulting in an input feature dimension of 3 for the meta-learner. This limitation may prevent the model from achieving optimal prediction performance. Second, the input feature weights for the meta-learner in traditional stacking

models are uniform. The input features for the meta-learner are composed of predictions from different base learners, yet the learning capabilities and predictive performance of these models vary. If the predictions from different base learners are concatenated directly as input features for the meta-learner, the meta-learner cannot distinguish between high- and low-quality features.

To address the issues mentioned above, this section proposes corresponding solutions. For Problem 1, where the meta-learners have low feature input dimensions, we employ a method to augment the feature dimensions. Traditional stacking model fusion utilizes the base model's predicted labels as features. However, model outputs can also include predicted probabilities and confidence values. Therefore, this section outputs the model's predicted labels, predicted probabilities, and confidence values simultaneously as input features for the meta-learner. This increases the input features to the meta-learner from three to nine, significantly enhancing the feature dimension. To address the issue of consistent feature weights in the meta-learner input, which prevents distinguishing feature quality, a weighting factor is assigned to each feature. Assuming the known training sample size is N , the base models are denoted as $C_i (i = 1, 2, 3)$, where different values of i represent models such as logistic regression, random forest, and XGboost. The prediction matrices generated by different base models are P_i , where P_{ij} denotes the prediction result of base model C_i for the j th data point, and y_j represents the true label of the j th data point in the training set. The calculation formulas are given by Equations (3)–(5):

$$\varepsilon_i = \frac{1}{N} \sum_{j=1}^N [(p_{ij} \neq y_j)], i = 1, 2, 3 \quad (3)$$

$$a_i = \frac{1}{\varepsilon_i} - 1, i = 1, 2, 3 \quad (4)$$

$$\omega_i = \frac{a_i}{\sum_{k=1}^3 a_k}, i = 1, 2, 3 \quad (5)$$

Here, ε_i denotes the prediction error rate of base model i , a_i represents the confidence of base model i , and ω_i indicates the weight of base model i . Building upon traditional stacking model fusion, the model integration steps from the previous section have been modified as follows:

Step 2: Input the training set into models such as logistic regression, random forest, and XGboost. Each model outputs prediction labels and prediction probabilities. Calculate the confidence levels according to the above formula, denoted as P_1 , P_2 , and P_3 . Here, p_i is an N -row, 3-column matrix, where N represents the sample size of the training set. Combine the predictions from different models into a prediction matrix to form the training set for the meta-learner, denoted as $P = [P_1, P_2, P_3]$.

Step 4: Input the test set into the trained logistic regression, random forest, XGboost, and other models to obtain prediction results M_1 , M_2 , M_3 . Combine the prediction structures of the three models into a matrix as the test set for the meta-learner, denoted as $M = [M_1, M_2, M_3]$. Here, M_i is the result matrix obtained by taking the mode from five prediction result matrices generated via five-fold cross-validation. Subsequently, the weights for each base model are

calculated according to the aforementioned weighting formula. These weights are incorporated into the corresponding test set, resulting in the updated matrix $M = [\omega_1 M_1, \omega_2 M_2, \omega_3 M_3]$.

4 Building Brand User Profiles and Personalized Recommendations

4.1 Brand User Preference Information Visualization

This study examines the large retail brand H, categorizing its primary merchandise segments based on operational performance as follows: (H1) Luxury Goods, (H2) Beauty, (H3) Jewelry, (H4) Watches, (H5) Men's Shoes, (H6) Women's Shoes, (H7) Young Women's Apparel, (H8) Mature Women's Apparel, (H9) Fashion Casual Wear, (H10) Men's Apparel, (H11) Men's Accessories, and (H12) Basic Sportswear. Statistics on H Brand users' average monthly clicks across different product categories on multimedia platforms are shown in Figure 3. Among these, two categories—(H2) Cosmetics and (H8) Mature Women's Wear—achieve an average of 50 or more clicks per month. This suggests that H Brand users are predominantly female.

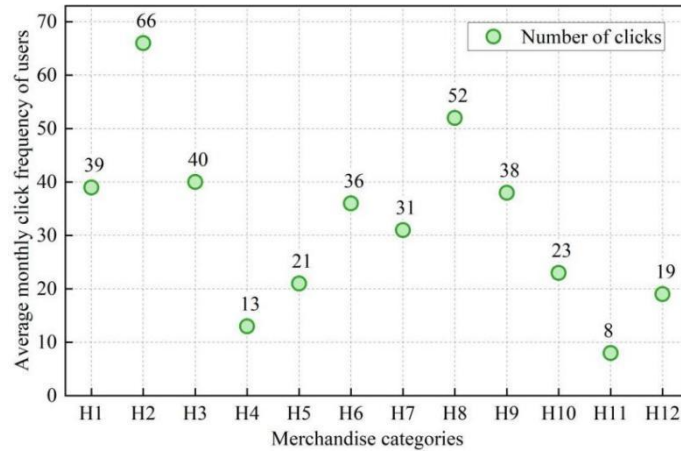


Figure 3: Analysis and display of user brand preferences

4.2 Building Brand User Personas

4.2.1 Development of User Profile Tags

To closely align with the distribution characteristics of H Brand's business and user data, thereby enabling precise development of brand user personas, this section analyzes the characteristics of membership users. Based on H Brand's consumer data, membership users can be primarily categorized into two types: (M1) dormant members and (M2) single-category spending members. (M1) Dormant members refer to those who have not made purchases across seasons, or who made purchases in the current quarter but not in the same quarter of the previous year. (M2) Single-category consumers refer to members who have only purchased products from one category since January 1st of the current year.

The table 5 demonstrates statistics related to the dormant members of Brand H (M1) based on (L1) age, (L2) active membership status, (L3) spending amount (yuan), (L4) average spending amount (yuan), (L5) spending frequency, (L6) average spending frequency, and (L7) average order value. The figures in brackets represent the percentage of this data in its particular characteristic. It is notable that most of the dormant members of Brand H (M1) belong to the age group 26-45 years old, and their spending level (>8,000 yuan) is much higher when

compared to other age groups.

Table 5: Analysis of Dormant Members

L1	L2	L3	L4	L5	L6	L7
Under 18 years old	347 (0.55%)	297 (1.00%)	5527	28 (0.56%)	1.63	3396
Aged 18 to 25	2419 (3.83%)	1343 (4.52%)	5312	182 (3.63%)	1.51	3375
Aged 26 to 35	19882 (31.48%)	9099 (30.65%)	4758	1534 (30.69%)	1.56	3041
Aged 36 to 45	16945 (26.83%)	8122 (27.36%)	4966	1319 (26.38%)	1.57	3145
Aged 46 to 55	13556 (21.47%)	6440 (21.69%)	4904	1082 (21.63%)	1.61	3030
Over 55 years old	10003 (15.84%)	4389 (14.78%)	4499	855 (17.10%)	1.73	2848

Table 6 shows the features of H Brand (M2) single-category consumer members in 12 product categories, such as (L8) number of members, (L9) total spending (yuan), (L10) number of transactions, and (L11) average spending (yuan). The single-category consumer members of H Brand (M2) are mostly made up of (H2) beauty (26.93%), (H8) mature women's clothing (4.02%) and (H10) men's clothing (8.92%). Within them, the members of the (H2) beauty category have an extraordinarily large (L11) average spending of 4,264 yuan compared to the members of the (H8) mature women's apparel category who have a (L11) average spending of 1,971 yuan, which is higher than the expenditure of the (H10) men's apparel members.

Table 6: Characteristics of single-category consumption membership

Category	L8	L9	L10	L11
H1	3303(7.29%)	1693(6.79%)	51	5221
H2	12209(26.93%)	6612(26.53%)	773	4264
H3	3179(7.01%)	2251(9.03%)	72	7257
H4	6104(13.46%)	677(2.72%)	24	7859
H5	769(1.70%)	971(3.90%)	91	1058
H6	1052(2.32%)	253(1.02%)	44	675
H7	4172(9.20%)	808(3.24%)	465	533
H8	1824(4.02%)	7399(29.69%)	649	1971
H9	2717(5.99%)	2517(10.10%)	261	1036
H10	4043(8.92%)	402(1.61%)	348	1105
H11	3044(6.71%)	220(0.88%)	160	1985
H12	2920(6.44%)	1117(4.48%)	276	3922

4.2.2 User Profiling Based on Clustering

The paper uses the K-Means clustering algorithm to group H brand members into three separate groups, with $K = 3$, according to the results of the analysis presented above. The figure below is a visual representation of the clustering of H brand members and the clustering results with X-axis representing age of the members, Y-axis representing frequency of consumption and Z-axis represents average transaction value per purchase.

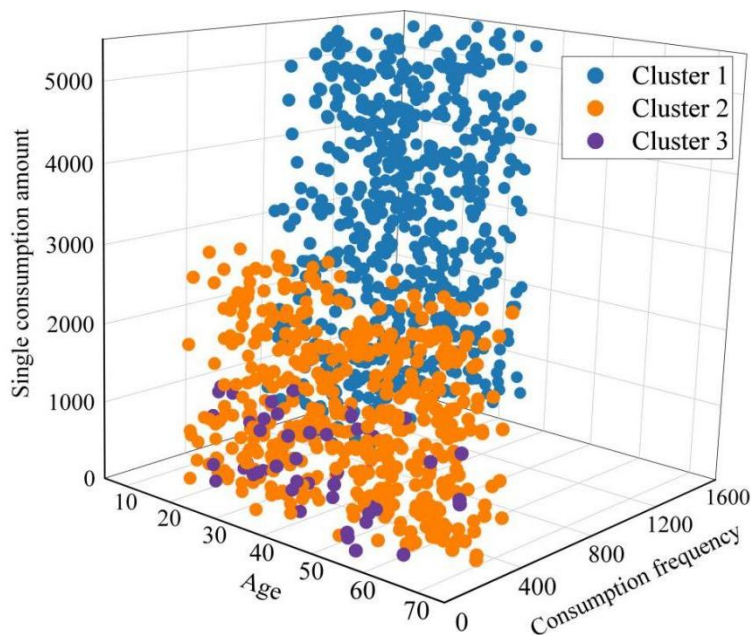


Figure 4: Member user clustering

According to Figure 4, the 18-40 age range is the most common age range occupied by Cluster 1 users with the largest user group. They have the greatest consumption frequency and mean transaction value (0-5000) of the three user groups. There are no specific age distribution features of Cluster 2 users, who form a relatively small part of the entire user population with consumption rates of 100-800. The average transaction value (0-3000) is less than in case of Cluster 1 users. Cluster 3 users do not have any distinct age distribution properties. In addition to being the smallest group, they also have the lowest consumption frequency and average transaction value of the three user groups. Namely, the users of Cluster 1 and Cluster 2 tend to purchase H brand products and have a particular amount of purchasing ability. Consequently, this paper chooses Cluster 1 and Cluster 2 users as the experimental subjects of the future personalized recommendation strategy of H brand.

4.3 Product Personalized Recommendations Based on an Improved Stacking Recommendation Model

4.3.1 Weight Matrix for Products and Categories

After collating the top 10 best-selling products across 12 categories for Brand H (numbered 1-10 within each category), the calculated weight matrix values for the 12 categories are presented in Table 7. Based on this data, the heatmap visualization of the weight matrix for all 120 products across the 12 categories is shown in Figure 5.

Table 7: The values of the weight matrix of 12 categories of goods

	1	2	3	4	5	6	7	8	9	10
H1	0.25	0.265	0.108	0.217	0.238	0.157	0.027	0.136	0.231	0.161
H2	0.013	0.195	0.287	0.13	0.099	0.273	0.032	0.217	0.175	0.094
H3	0.023	0.129	0.282	0.286	0.201	0.092	0.215	0.184	0.051	0.241
H4	0.191	0.263	0.198	0.225	0.054	0.058	0.245	0.191	0.264	0.286
H5	0.161	0.273	0.055	0.205	0.017	0.013	0.11	0.254	0.264	0.182
H6	0.084	0.087	0.092	0.058	0.133	0.135	0.177	0.028	0.171	0.034
H7	0.27	0.098	0.077	0.085	0.094	0.221	0.209	0.148	0.019	0.078
H8	0.114	0.063	0.013	0.169	0.125	0.201	0.27	0.269	0.082	0.129
H9	0.278	0.192	0.026	0.239	0.217	0.28	0.033	0.05	0.254	0.009
H10	0.236	0.25	0.247	0.077	0.239	0.035	0.185	0.037	0.031	0.11
H11	0.235	0.256	0.035	0.274	0.026	0.204	0.173	0.096	0.272	0.214
H12	0.084	0.067	0.289	0.095	0.263	0.139	0.016	0.085	0.292	0.147

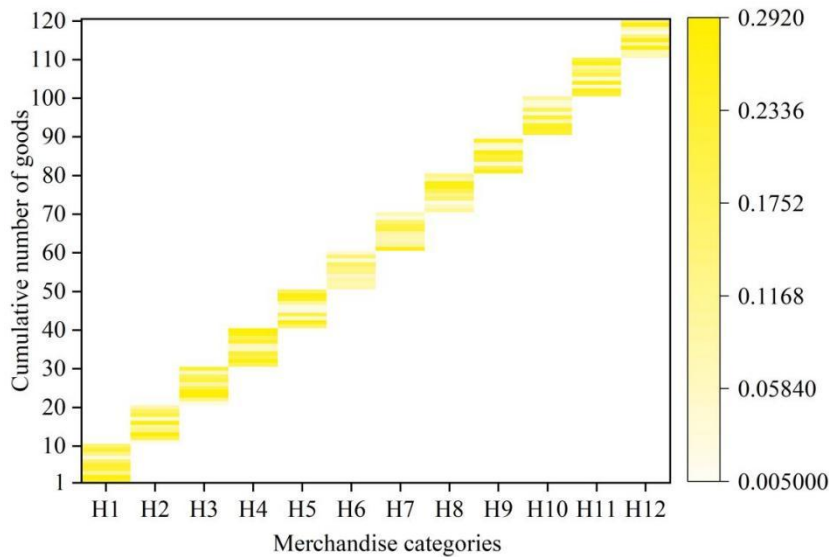


Figure 5: Category and product weight matrix heat map

4.3.2 Simulation Experiments of the Recommended Model

The current research uses Average Reciprocal Hit Rate (ARHR) as a measure of the strength of product recommendation and Average Precision (MAP) as a measure of personalized recommendation effectiveness of TOP-N products to efficiently assess the validity of the enhanced stacking recommendation model presented in this paper in terms of ranking TOP-N products recommended to users in Cluster 1 and Cluster 2 of Brand H.

The base number of user clusters is 5 and it is increased by 15 every time until a maximum of 50. At the same time, the TOP-N setting is 5 and it increases by increments of 10 until 55 is reached. Figure 6 shows how the ARHR measure of the proposed recommendation model of H-brand users changes with various sizes of user clusters. The ARHR value of the model is lower as TOP-N goes up, but the overall trend of the ARHR measure is increasing as the size of user clusters increases. It reaches its peak at 0.281 when the user cluster size is 50, which means that the model works more effectively in personalized recommendations with larger user cluster sizes.

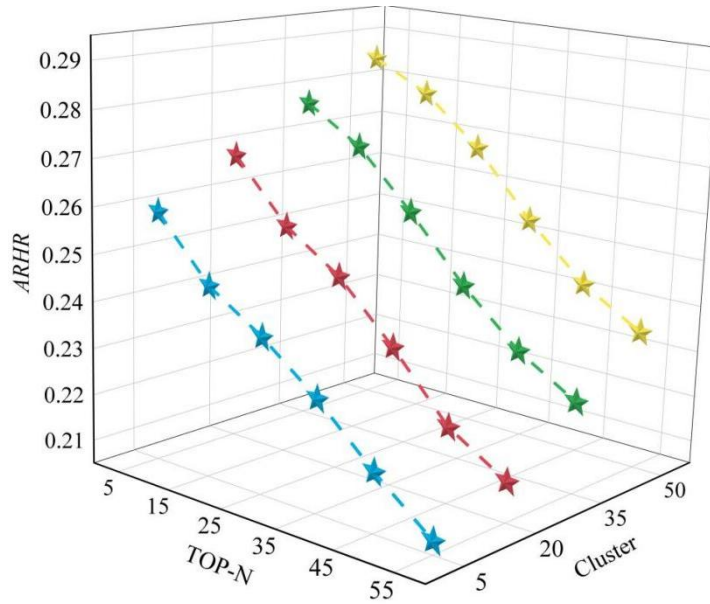


Figure 6: Change of evaluation ARHR index value

Figure 7 shows the variation in the MAP metric of the model proposed in this paper. The metric has an inverse relationship to the TOP-N metric. With an increasing number of user clusters, the MAP metric value of the model shows a general increase. It also achieves its highest value (0.738) at 50 user clusters, just like the ARHR metric. This means that the average accuracy of personalized recommendations increases as the number of user clusters grows with the model presented in this paper.

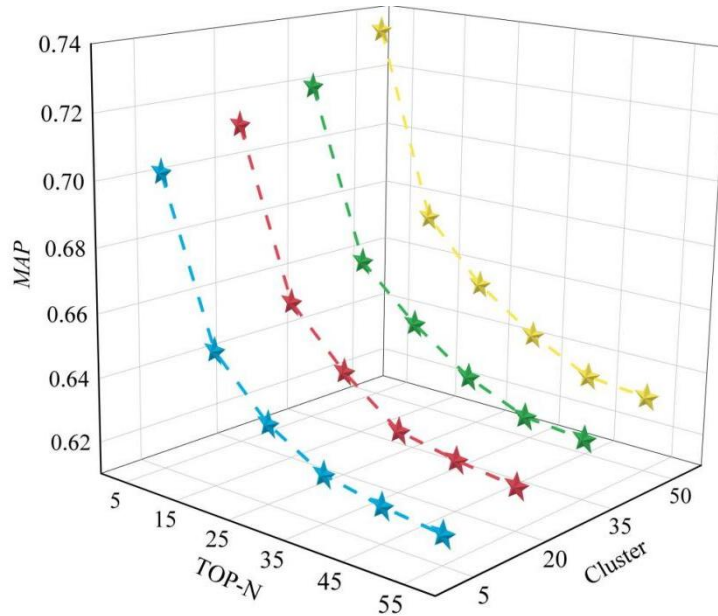


Figure 7: Change of evaluation MAP index value

5 Conclusion

This paper constructs a more comprehensive and multidimensional user profile by integrating brand users' multi-dimensional data across multimedia platforms, supported by the K-Means

clustering algorithm. The proposed recommendation model based on an improved stacking approach delivers increasingly superior recommendation results as the number of user clusters grows, providing brand enterprises with significantly more precise strategy references for personalized product recommendations tailored to different user profile types.

Through visualization of H-brand user data and development of profiling tags across multimedia platforms, this study categorizes users into three distinct groups, two of which exhibit high consumption propensity and purchasing power. In simulation experiments with H-brand users, the proposed improved stacking recommendation model demonstrates significant improvements in average reciprocal recall and average precision metrics as the number of user clusters increases, reaching peak values of 0.281 and 0.738 respectively.

Funding

This work was supported by University-Level Key Research Project of Beijing Polytechnic University (project number: 2025X005-SXZ).

About the Author

Xiaoyu Cui was born in Yichun, Heilongjiang, P.R. China, in 1994. She obtained a master's degree from China University of Geosciences. She is currently a lecturer of Beijing Polytechnic University, Beijing, China. Her research interests include new media marketing, internet dissemination, and user research.

References

- [1] Farseev, A., Samborskii, I., & Chua, T. S. (2016, October). bbridge: A big data platform for social multimedia analytics. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 759-761).
- [2] Wang, X., & Liu, D. F. (2025). Pattern mining and prediction techniques for user behavioral trajectories in e-commerce. *PLoS One*, 20(5), e0320772.
- [3] Liao, S. H., & Yang, C. A. (2021). Big data analytics of social network marketing and personalized recommendations. *Social Network Analysis and Mining*, 11(1), 21.
- [4] Zeb, M. A., & Fasli, M. (2012, December). Dynamically adaptive user profiling for personalized recommendations. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 604-611). IEEE.
- [5] He, W., Liu, P., & Qian, Q. (2023). User Profiles. In *Machine Learning Contests: A Guidebook* (pp. 117-135). Singapore: Springer Nature Singapore.
- [6] Rial, A., Gómez, P., Picón, E., Braña, T., & Varela, J. (2015). Identification and characterization of adolescent Internet user's profiles. *The Spanish journal of psychology*, 18, E50.
- [7] Mekouar, L., Iraqi, Y., & Damaj, I. (2024). A global user profile framework for effective recommender systems. *Multimedia Tools and Applications*, 83(17), 50711-50731.

- [8] Bensassi, I., Elyusufi, Y., & El Mokhtar, E. N. (2019, October). Smart Connection of User Profiles in a Big Data Context. In *Proceedings of the 4th International Conference on Big Data and Internet of Things* (pp. 1-8).
- [9] Cufoglu, A. (2014). User profiling-a short review. *International Journal of Computer Applications*, 108(3).
- [10] Farid, M., Elgohary, R., Moawad, I., & Roushdy, M. (2018, October). User profiling approaches, modeling, and personalization. In *Proceedings of the 11th international conference on informatics & systems (INFOS 2018)*.
- [11] Gonzalez, R., Soriente, C., & Laoutaris, N. (2016, November). User profiling in the time of https. In *Proceedings of the 2016 Internet Measurement Conference* (pp. 373-379).
- [12] Peng, J., Choo, K. K. R., & Ashman, H. (2016). User profiling in intrusion detection: A review. *Journal of Network and Computer Applications*, 72, 14-27.
- [13] Lashkari, A. H., Chen, M., & Ghorbani, A. A. (2019). A survey on user profiling model for anomaly detection in cyberspace. *Journal of Cyber Security and Mobility*, 75-112.
- [14] Sánchez, P., & Bellogín, A. (2019). Building user profiles based on sequences for content and collaborative filtering. *Information Processing & Management*, 56(1), 192-211.
- [15] Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access*, 7, 144907-144924.
- [16] Maraj, D., Vuković, M., & Hotovec, P. (2024, October). A Survey on User Profiling, Data Collection, and Privacy Issues of Internet Services. In *Telecom* (Vol. 5, No. 4, pp. 961-976). MDPI.
- [17] Mushtaq, M. S., Augustin, B., & Mellouk, A. (2014, June). QoE: User profile analysis for multimedia services. In *2014 IEEE International Conference on Communications (ICC)* (pp. 2289-2294). IEEE.
- [18] Yan, H., Wang, Z., Lin, T. H., Li, Y., & Jin, D. (2018). Profiling users by online shopping behaviors. *Multimedia Tools and Applications*, 77(17), 21935-21945.
- [19] Pricopoaia, O., Cadinoiu, M., & Matei, A. (2020). Aspects Regarding the Online Consumer Profile and the Impact on the Purchasing Behaviour. *Economics and Applied Informatics*, 2, 91-99.
- [20] Fuentes Hernández, J. A., Ruiz Santoyo, G. A., & Amarillas Urbina, V. A. (2020). Online Purchase and Consumer Profile: Case Mexico. *Global Journal of Emerging Trends in E-Business, Marketing & Consumer Psychology*, 6(1).
- [21] Zhu, Z., Li, D., Liang, J., Liu, G., & Yu, H. (2018). A dynamic personalized news recommendation system based on BAP user profiling method. *IEEE Access*, 6, 41068-41078.
- [22] Benkaddour, F. Z., Taghezout, N., Kaddour-Ahmed, F. Z., & Hammadi, I. A. (2018). An Adapted Approach for User Profiling in a Recommendation System: Application to

- Industrial Diagnosis. *Int. J. Interact. Multim. Artif. Intell.*, 5(3), 118-130.
- [23] Yang, P., Wang, H., Fang, H., & Cai, D. (2015). Opinions matter: a general approach to user profile modeling for contextual suggestion. *Information Retrieval Journal*, 18(6), 586-610.
- [24] Bekkamov, F., & Sharifov, Y. (2024, June). Extended User Profiling Approaches for Recommendation Systems. In *ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference (Vol. 2, pp. 49-54)*.
- [25] Ha, L., Abuljadail, M. H., Joa, C. Y., & Kim, K. (2022). Personalized vs non-personalized recommendations: how recommender systems, recommendation sources and recommendation platforms affect trial of YouTube videos among digital natives in Saudi Arabia. *Journal of Islamic Marketing*, 13(12), 2778-2797.
- [26] Ferrario, T., Fersini, E., Messina, E., & Sormani, G. (2025, April). Dynamic User Profiling for Personalized Tourism Recommendations Using Reinforcement Learning Models. In *International Conference on Advanced Information Networking and Applications (pp. 206-215)*. Cham: Springer Nature Switzerland.
- [27] Wang, L., Zhang, D., Yang, F., Zhao, P., Liu, J., Zhan, Y., ... & Zhang, Q. (2025, August). LettinGo: Explore User Profile Generation for Recommendation System. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2 (pp. 2985-2995)*.
- [28] Hui, G., LiQing, Z., Mang, C., & ShiKun, X. (2024). Research on Personalized Recommendation Algorithms Based on User Profile. *International Journal of Advanced Computer Science & Applications*, 15(3).
- [29] Yu, J., & Li, B. (2024, January). Research and Design of A Personalized Recommendation System for Ceramic Products Based on User Profiling and Cluster Analysis. In *2024 Asia-Pacific Conference on Software Engineering, Social Network Analysis and Intelligent Computing (SSAIC) (pp. 581-586)*. IEEE.
- [30] Gu, Y., Ding, Z., Wang, S., & Yin, D. (2020, January). Hierarchical user profiling for e-commerce recommender systems. In *Proceedings of the 13th international conference on web search and data mining (pp. 223-231)*.
- [31] Amoretti, M., Belli, L., & Zanichelli, F. (2017). UTravel: Smart mobility with a novel user profiling and recommendation approach. *Pervasive and mobile computing*, 38, 474-489.
- [32] Huang, J. (2024). Personalized Recommendation Method for Cultural Creative Products in Tourism Cities Based on User Profiles. *Procedia Computer Science*, 243, 1133-1142.
- [33] Bonomo, M., Ciaccio, G., De Salve, A., & Rombo, S. E. (2019, August). Customer recommendation based on profile matching and customized campaigns in on-line social networks. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 1155-1159)*.
- [34] Lin, Q., Duan, L., & Huang, J. (2024). Personalized Pricing Through Strategic User

Profiling in Social Networks. *IEEE/ACM Transactions on Networking*, 32(5), 3977-3992.

- [35] Mohamed, S., Sethom, K., Namoun, A., Tufail, A., Kim, K. H., & Almoamari, H. (2022). Customer Profiling Using Internet of Things Based Recommendations. *Sustainability*, 14(18), 11200.