Research on Campus Beauty Consumption Characteristics and Precision Recommendation Based on Multi-source Data and Machine Learning

Jingyuan Yang*, Xiang Li, Xiuyun Zhang, Yawen Zhao, Xinru Guo, Zhen Hao
School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China
**Corresponding Author*

Abstract: Driven by the growing appearance economy and self-pleasing consumption, China's beauty sector continues expanding, despite major challenges, including slowing traffic growth, intensifying product homogenization and weak consumer brand loyalty. University students are significant potential consumers, but many of them are unaware of their skin condition and face the high trial expense and unsuitable products. This study combines the sales record data of e-commerce with online reviews and questionnaire feedback from Zhengzhou university students. It aims at investigating the campus beauty consumption, student segmentation and product-campus demand matching by utilizing text mining, sentiment analysis, descriptive statistics, K-Means clustering and association rule mining. A mobile app for cosmetic recommendations is also created. The results indicate that cost-effective items outsell other items on campus. There are three groups of students, namely those who use basic skincare; those who are cautious about buying skincare as they have concerns about their skin's reaction; and those who use products often and have preferences. There is a strong correlation between product reputation and sales. Main grievances involve irritation, allergies, greasiness and effectiveness issues, with demand also fluctuating according to seasons and campus activities. These findings shed light on beauty consumer groups. This application facilitates rational purchases for students, targeted marketing to campus and personalized services. The approach and results can also be applied to other groups of consumers.

Keywords: Beauty Consumption; Text Mining; Sentiment Analysis; K-Means Clustering; Precision Recommendation Application; User Profiling

1. Introduction

1.1 Research Background

The consumer segment has more and more spending power with its rising household disposable income. The appearance economy and self-reward consumption have gained widespread popularity. Covering skincare, color cosmetics and fragrance categories, China's beauty market maintains sustained expansion. Figure 1 presents statistics released by iiMedia Research. The market scale hit 516.9 billion yuan in 2023, and is projected to exceed 640 billion yuan by 2027, with a stable annual growth rate ranging from 5% to 8%. Nevertheless, such expansion is accompanied by a host of industry challenges. Traffic dividends are gradually disappearing, accompanied by surging customer acquisition costs. Product homogenization has become increasingly severe. Most brands allocate more resources to marketing campaigns rather than research and development, resulting in limited technical barriers within the sector. Consumer preferences evolve rapidly, leading to weak brand loyalty. Viral cosmetic items can hardly retain long-term popularity, which creates intense competition across the whole industry. Therefore, brands have strong demand for data-driven insights to identify consumers' genuine demands.

College students are a core potential group in the beauty market. They are open-minded about spending and willing to spend on themselves, and they are easily influenced by social media recommendations and livestream sales. At the same time, they face clear pain points. First, most students do not really know their skin type, so they follow trends and buy viral products blindly, which can easily lead to allergies or breakouts, and the trial-and-error cost is high. Second, they are weak at judging information:

 Copyright @ STEMM Institute Press

false advertising and paid soft articles are everywhere, so it is hard to tell whether a product actually works. Third, their spending mindset is not mature yet, and impulse or over-budget purchases are common. Fourth, most products on the market are designed for working professionals and do not fit campus life well, where students want something affordable, gentle, and convenient. [1,2]

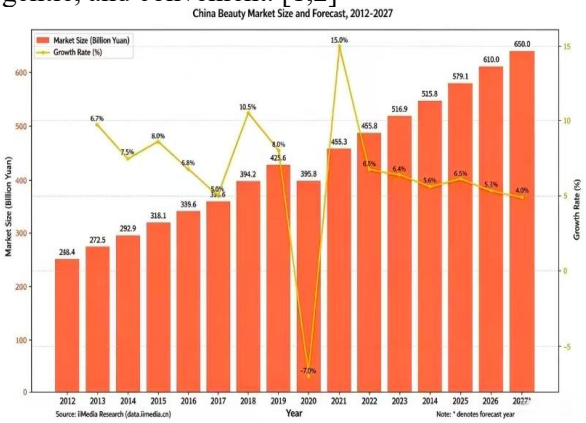


Figure 1. Market Size and Forecast of China's Beauty Market, 2012-2027

Zhengzhou is an education hub in central China. It has nearly 100 universities and over a million enrolled students, making it an important growth area for the regional beauty market. Annual campus beauty spending in China has already passed 30 billion yuan. Products priced between 50 and 200 yuan are the mainstream choice for students, and domestic brands are increasingly accepted. But right now, brands don't have a clear campus strategy. Campuses also lack professional beauty services and precise recommendation channels. So supply and demand don't match up well. This study focuses on students at Zhengzhou universities, examines their beauty consumption characteristics, and builds a recommendation service, which is both typical and necessary. [2,3]

1.2 Research Significance

This study has both theoretical and practical value. On the theoretical side, we combine e-commerce big data with questionnaire data from a specific group of people, and use natural language processing and machine learning methods to build a "skin type - consumption - product" analysis framework for campus users. This enriches research on campus beauty consumption and precision recommendation, and offers a reusable analysis paradigm.

On the practical side, the study pinpoints the pain points and needs of student beauty consumers. It can help college students spend more rationally and cut trial-and-error costs, and it gives brands quantitative evidence for product optimization and precision marketing on campus. The recommendation app built in this study can also serve as a direct example for building campus beauty service platforms, benefiting all parties. [4]

1.3 Research Approach and Innovations

Figure 2 shows the full technical route of this study. It follows the logic of "data collection and preprocessing-analysis and modeling- pattern extraction-app implementation- conclusion". First, three categories of data were gathered: e-commerce product information, online user reviews, and questionnaire data collected from universities in Zhengzhou. Multiple analytical techniques were then adopted for subsequent investigation. Reputation-related complaints were identified using sentiment analysis, consumption patterns were examined using descriptive statistics, the segmentation of users and products was done using clustering and the matching principles using association rule mining. Lastly, a precision recommendation application was designed and developed on the basis of the above outcomes, forming a closed-loop research framework.

This work presents two major innovations. To begin with, it targets a specific demographic: university students in Central China. The localized analysis incorporates Zhengzhou's climatic conditions and campus consumption scenarios, overcoming the limitations of generalized research. Secondly, multi-source data are integrated for model construction. A functional vertical recommendation application is developed to translate research results into an interactive service with both scientific rigor and practical value.



Figure 2. Mind Map of the Research Approach

2. Data Sources and Research Methods

2.1 Data Sources

Four types of data are used in this study:

(1) Campus survey data from Zhengzhou universities. We ran online questionnaires and offline interviews to collect information on students' beauty consumption behavior, purchase preferences, usage scenarios, budgets, and skin types. The samples covered students from different types of universities in Zhengzhou.

(2) Product data from e-commerce platforms. Seven popular campus products were picked from three mainstream affordable brands - Innisfree, Maybelline, and Inoherb. For each one, we recorded price, sales volume, ingredients, and suitable skin types.

(3) Our dataset was sourced from user-generated review content on mainstream e-commerce platforms. Specifically, we gathered public customer reviews and their corresponding follow-up comments for seven target products, and subsequently applied these textual materials to two core research tasks: reputation mining and sentiment analysis.

(4) Public industry data. Beauty industry reports and market size data published by authoritative third-party organizations were gathered for background analysis and trend comparison.

2.2 Data Preprocessing

The multi-source data went through cleaning and standardization before analysis. For questionnaire responses, we dropped invalid, duplicate, and logically inconsistent samples, and filled in missing values where reasonable. For reviews, emojis, ads, and meaningless repeated text were removed, followed by word segmentation and stop-word removal. Product fields were normalized, duplicates were stripped and outliers managed. The newly added variables included spending power, product attributes, sentiment tendency and scenario preference. The numeric data was scaled and the text labels were encoded for further analysis. At this point, the dataset was ready for later analysis.

2.3 Research Methods

The current study integrated a step-by-step analytical process with machine learning techniques. They are reviewed below in detail, from review processing to descriptive analysis, consumer grouping, purchasing associations and consumer demand trends.

(1) Collected e-commerce reviews were analyzed using text mining techniques as well as using sentiment analysis. Using word frequency, recurring words were identified from raw text data which were tokenized and cleaned. Using a sentiment dictionary and manual validation, the reviews were classified into positive and negative reviews, and an overall sentiment score per product was then calculated. This approach enabled us to gauge user satisfaction and pinpoint core user complaints, and we further generated positive and negative word clouds to visualize key reputation patterns. [5,6]

(2) Descriptive data analysis. Frequency counts, proportions and mean values were calculated first. Charts report the demographics, budget, channels of purchase, type of products and how often students use makeup, to give an overview of campus consumption.

(3) K-Means clustering analysis. Clustering took into account five aspects: budget, frequency of makeup application, primary needs, skin type and usage scenarios. Three was found to be the optimum number of clusters from the elbow method. Then using K-Means, students were segmented and products were classified in accordance with their fit with the campus, and differences between the groups were found. The resulting profiles provide links between students' buying potential, skin issues and general situations of use and product features.[7]

(4) Association rule analysis was performed using the Apriori algorithm, where the strength of the rules was determined by their support and confidence.

Strong rules connected budgets, purchasing channels, how the product was used, the function of the product, and product category preference. They are good for describing consumer behavior and recommending products according to scenarios. Instead of recommending products merely because they are popular, these relationships enable the system to take into account when and why students may require a particular product [8].

(5) We used scatter plots and correlation coefficients to investigate the relationship between sales and sentiment scores (correlation and trend analysis). This analysis indicates the effect of reputation on market performance. The monthly demand index computed with the addition of seasons and campus activities

 Copyright @ STEMM Institute Press

explained the annual consumption pattern at the level of each product category [9, 10].

3. Empirical Analysis Results

3.1 Text Analysis of Product Reputation and Pain Points

Table 1. Keywords in User Reviews of Cosmetics

Brand	Sentiment	High-Frequency Words	Medium-Frequency Words	Low-Frequency Words
Innisfree	Positive	decent, effective, oil control, loose powder	cheap, like it, long-lasting, easy to use, powder texture, fine, cost-effective	genuine, voluminous hair, setting makeup, recommend, good deal, shipping, apply makeup, skin type, satisfied, quality, convenient, review, appearance, brand, purchase
Innisfree	Negative	useless, irritating, oily skin	fake smooth feel, damaged packaging, breakouts, greasy, oily residue	bad attitude, trash, got worse, return, not oily, gap, stinging, poor review, cheap feeling, irritant, poorly made, low star, waste, allergy, awful, rough, bad review
Maybelline	Positive	makeup removal, clean	easy to use, decent, gentle, eyes, eye and lip, repurchase	friend, good deal, makeup remover, recommend, packaging, not stinging, logistics, like it, ability, effective, unlimited, purchase, removes well, cost-effective, eye cream, brand, feel, worth it, refreshing, product
Maybelline	Negative	irritating, useless	greasy, false eyelashes, oily skin, stinging, damaged packaging, cleansing oil	not oily, oily, water-oil
Inoherb	Positive	color, like it	moisturizing, decent, effective, good-looking	good review, pictures, review, texture, brand, mom, moisturizing, lipstick, suitable, satisfied, quality, brightens skin tone, feel, lip balm, received, on lips, overall, packaging, actual product, repurchase, purchase, logistics
Inoherb	Negative	color difference, disappointed	ugly, return, useless, price difference	greasy, cleansing oil, oily shine, no good, waste, sticky, fake product, damaged, extremely bad, fake, inferior, terribly bad, irritating, allergy, gap, poor review, disgusting, itchy, cheap, pungent, trash, mismatch, regret

After text cleaning, word segmentation, and word frequency statistics on user reviews of the seven products from the three brands, we extracted positive and negative keywords separately, as shown in Table 1, which gives a direct view of each brand's core strengths and pain points. [5,6]

The table leads to several conclusions. Innisfree's strengths lie in basic moisturizing and gentle oil control, while its pain points are irritation from some products, weak cleansing power, and plain packaging. Maybelline is strong in makeup removal and cleansing as well as moisturizing, well-pigmented lip products, but some of its removers leave a greasy residue and its lip products have color differences and a drying feel. Inoherb offers basic skincare effects at a high cost-performance ratio, but some products feel heavy and sticky, and their effects are hard to perceive.

The three brands share the same highlights: low prices, acceptable basic functions, and ease of use. They also share common pain points: irritation and allergy, greasy skin feel, effects that do not match the advertising, and color differences in makeup. Users with sensitive skin react most strongly to irritation, while users with oily skin complain most about greasiness and makeup wearing off. Each brand has its own direction for improvement.

3.2 Basic Characteristics of Campus Beauty Consumption

Based on descriptive statistics of the questionnaire samples, the budget, skin type, and makeup frequency distributions of students at Zhengzhou universities are shown in Table 2. [2,4]

The statistics show that 83% of students spend no more than 200 yuan per month on beauty products, so affordable consumption is absolutely dominant. The most common skin

Copyright @ STEMM Institute Press

types are combination-oily and dry, and 14% of the students have sensitive skin; affected by Zhengzhou's climate, redness and oiliness are the core skin problems.Just 11 percent of students apply makeup on a daily basis. The prevailing trend centers on subtle cosmetic looks, with primary attention devoted to fundamental skin care. Across the student population, core purchasing considerations lie in product mildness and low irritation, demonstrable efficacy, as well as cost-effectiveness.

Table 2. Questionnaire Sample of Students at Zhengzhou Universities

Dimension	Category	Count	Percentage
Beauty Budget	50-100 yuan	35	33.98%
	100-200 yuan	26	25.24%
	Under 50 yuan	25	24.27%
	Over 200 yuan	17	16.51%
Skin Type	Combination-oily	33	32.04%
	Dry	33	32.04%
	Sensitive	14	13.59%
	Oily	11	10.68%
	Not sure	12	11.65%
Makeup Frequency	Only for dates/parties	42	40.78%
	Hardly ever	34	33.01%
	Only for class/going out	16	15.53%
	Daily	11	10.68%

3.3 Consumer Segmentation Based on K-Means

Taking budget, makeup application frequency, core demands and scenario preferences as clustering dimensions, we adopted the elbow method to determine the optimal cluster number K=3. The K-Means algorithm was then employed to categorize university students in Zhengzhou into three consumer segments with differentiated characteristics. The clustering outcomes are presented in Figure 3, while detailed information of each group is summarized in Table 3.

Table 3. Characteristics of Consumer Groups at Zhengzhou Universities

Group Label	Share	Key Features	Key Needs
Basic-Demand Users	40%	Budget ≤ 50 yuan, low makeup frequency, focuses on basic cleansing and moisturizing, highly price-sensitive	Low price, basic functions, high value for money
Cautious Sensitive-Skin Users	30%	Budget 50-100 yuan, medium makeup frequency, high share of sensitive skin, cares about ingredient fit	Non-irritating, repairing, fits dorm skincare
Sophisticated High-Frequency Users	30%	Budget ≥ 101 yuan, high makeup frequency, mostly oily or combination-oily skin, values long wear and portability	Portable makeup, oil control and long wear, fits touch-ups

features	featuresNormalized	prediction
[150.0,2250.0,0.0]	[0.5134233898080166,1.4483728278132 11,-1.0829655685741755]	1
[75.0,1250.0,0.0]	[-0.46588418704801554,-0.5145535046178515,-1.0829655685741755]	1
[150.0,1250.0,0.0]	[0.5134233898080166,-0.5145535046178515,-1.0829655685741755]	1
[25.0,2250.0,2.0]	[-1.1187559049520 37,1.4483728278132 11,0.0729458673132864]	1
[25.0,1750.0,5.0]	[-1.1187559049520 37,0.4669096615976 7977,1.806813021144479]	0
[150.0,1250.0,0.0]	[0.5134233898080166,-0.5145535046178515,-1.0829655685741755]	1
[25.0,1250.0,0.0]	[-1.1187559049520 37,-0.5145535046178515,-1.0829655685741755]	1
[75.0,1250.0,2.0]	[-0.46588418704801554,-0.5145535046178515,0.0729458673132864]	2
[250.0,1750.0,2.0]	[1.8191668256160594,0.4669096615976 7977,0.0729458673132864]	0

Figure 3. K-Means Clustering Results of Student Groups

The clustering findings reveal that university students' beauty consumption in Zhengzhou exhibits distinct budget stratification, where skin type serves as the driving force and consumption scenarios remain the focal point. The three clusters have substantial differences in requirements, which creates a foundation for precise product positioning.

Product Name	Price Fit	Function Fit	Scenario Fit	features	featuresNormalized	prediction
Innisfree Exfoliating Gel	5.0	4.0	4.0	[5.0,4.0,4.0]	[1.1870186660251187,-0.6454972243879 04,-0.6454972243879 04]	2
Maybelline Eye & Lip Makeup Remover	5.0	4.0	5.0	[5.0,4.0,5.0]	[-1.5485269498851367,-0.6454972243879 04,1.2909944487358847]	1
Innisfree Green Tea Moisturizing Facial Cleanser	4.0	4.0	4.0	[4.0,4.0,4.0]	[-0.2214037231835242,-0.6454972243879 04,-0.6454972243879 04]	2
Innisfree Red Pomegranate Skincare Set	4.0	5.0	4.0	[4.0,5.0,4.0]	[-0.2214037231835242,1.2909944487358847,-0.6454972243879 04]	0
Maybelline Color Sensational Light Lipstick	4.0	4.0	5.0	[4.0,4.0,5.0]	[-0.2214037231835242,-0.6454972243879 04,1.2909944487358847]	1
Innisfree Fresh Moisturizing Mask Set	5.0	5.0	4.0	[5.0,5.0,4.0]	[1.1870186660251187,1.2909944487358847,-0.6454972243879 04]	0

Figure 4. K-Means Clustering Results of Product Fit

3.4 Product-Campus Fit Layering

We constructed a three-dimensional evaluation system encompassing price compatibility, functional compatibility and scenario compatibility (rated on a 1–5 scale). This system was calibrated against findings from reputation text analysis. The K-Means clustering approach was adopted to classify the seven products into three tiers: high compatibility, moderate compatibility and low compatibility. The clustering outcomes are presented in Figure 4, while compatibility details are summarized in

 Copyright @ STEMM Institute Press

Table 4.

Table 4. Product-Campus Fit Layering Results

Fit Level	Products	Main Fit Reasons	Optimization Direction
High Fit	Innisfree moisturizing set, Inoherb repair mask	Price matches mainstream budgets, meets frequent moisturizing and repair needs, fits dorm scenarios	Faster absorption, better fit
Medium Fit	Inoherb facial cleanser, Maybelline light lipstick	Basic functions up to standard, moderate price, fits cleansing and commuting scenarios, with minor fit issues	Improve freshness, more moisturizing
Low Fit	Innisfree alcohol-based products, Maybelline makeup remover, Inoherb functional skincare	Overpriced or irritating ingredients, high bad-review rate among sensitive-skin users, poor scenario fit	Replace irritating ingredients, not promoted for now

Compatibility stratification outcomes indicate that high-compatibility products constitute the core options for campus recommendation. Moderate-compatibility items can achieve stronger competitiveness following targeted optimization, whereas low-compatibility products require adjustments prior to entering the campus market. Such conclusions establish a priority foundation for product recommendation and brand improvement.

3.5 Scenario-Based Precision Recommendation Logic

Drawing on user clustering outcomes and product compatibility stratification, together with consideration of prevalent consumption scenarios among Zhengzhou university students, we constructed a three-dimensional "user–scenario–product" matching system. The core matching principles are summarized in Table 5

Table 5. Scenario-Based Precision Recommendation Rules

User Group	Frequent Scenarios	Matched Products	Matching Logic
Basic-Demand Users	Daily basic care in dorms	Inoherb facial cleanser	Unit price 30-50 yuan, focuses on cleansing and moisturizing, high value fits low budgets
Cautious Sensitive-Skin Users	Spring repair, bedtime skincare	Innisfree moisturizing set + Inoherb repair mask	Alcohol- and fragrance-free, gentle repair, fits redness and sensitive skin care
Sophisticated High-Frequency Users	8 a.m. classes, club activities, summer touch-ups	Maybelline light lipstick + Innisfree oil-control loose powder	Small and portable, long wear and oil control, fits long days out and touch-ups

This matching system is constructed based on authentic survey data and closely linked to consumption scenarios of Zhengzhou university campuses. Boasting transparent and interpretable logic, it can deliver relatively satisfactory recommendation accuracy without reliance on sophisticated black-box algorithms, thereby securing favorable user acceptance. It underpins the core logic of the application's recommendation function elaborated subsequently.

3.6 Correlation Analysis between Reputation and Sales

We generated a scatter plot taking cumulative product sales as the horizontal axis and review sentiment scores as the vertical axis, illustrated in Figure 5. The findings reveal an overall significant positive correlation between sales volume and sentiment scores, indicating that product reputation serves as the core driver of campus product popularity.

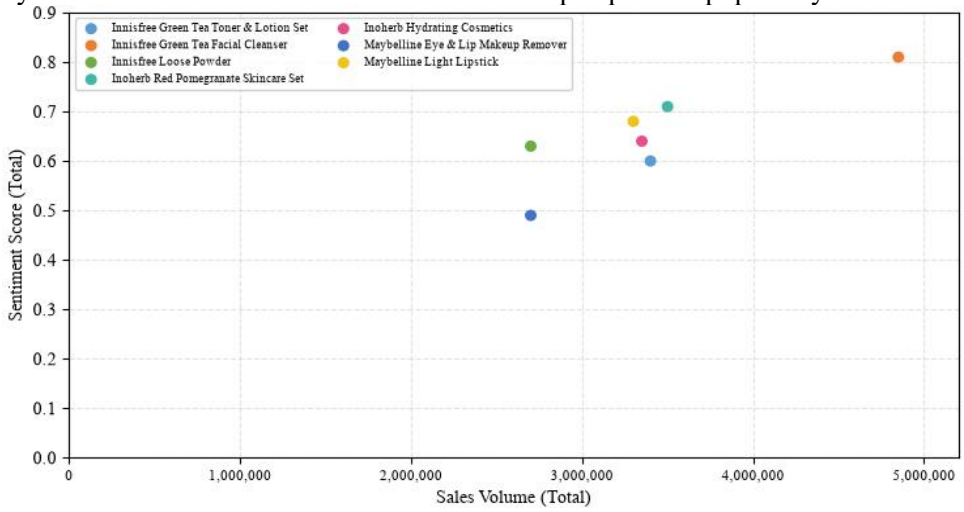


Figure 5. Scatter Plot of Product Reputation and Sales

Among the sampled products, the Innisfree Green Tea Facial Cleanser and Inoherb Red Pomegranate Set achieve both robust sales performance and favorable sentiment scores, qualifying them as campus bestsellers. The Maybelline Light Lipstick records upper-middle sales accompanied by relatively positive sentiment metrics, implying promising growth potential. While the Innisfree Green Tea Toner & Lotion Set enjoys strong sales, its sentiment rating remains low; issues concerning greasiness and skin irritation constrain further market expansion. The Maybelline Eye & Lip Makeup Remover exhibits poor performance in both sales volume and sentiment evaluation, calling for targeted optimization.

3.7 Temporal Trends of Campus Beauty Consumption

By integrating Zhengzhou's climate data and campus schedule data, a monthly demand index of the various categories was obtained. Seasonal patterns and campus events were found to impact the trend, as seen in Figure 6. The index thus takes account of the skincare demands caused by the weather and also of different time periods in which students' daily plans or shopping opportunities shift.

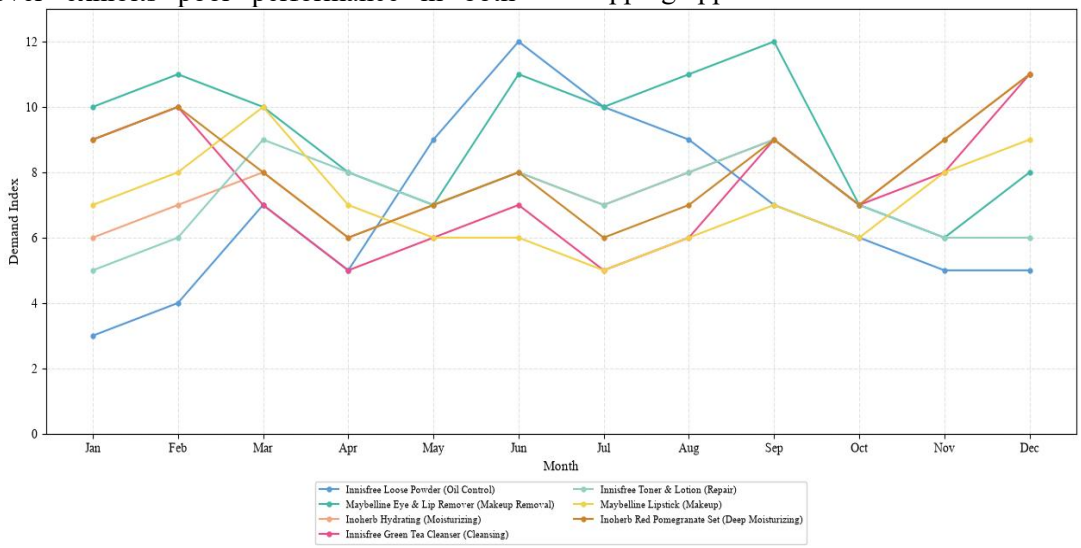


Figure 6. Monthly Consumption Trends of Different Categories

Oil-control and setting products are in strong demand in summer and at the start of the school year, peaking in June, which fits graduation season and military training. Makeup removal and cleansing products see their highest demand in September when new students arrive, and stay stable all year as a rigid need. Moisturizing and repair products rise clearly in winter, peaking in December, driven by the dry climate and holiday demand. Sensitive-skin repair products peak in spring, matching Zhengzhou's windy, dry spring when allergies are common.

Overall, all categories see sales peaks around graduation season, the start of the school year, and year-end, with lows in April and July. Different categories fluctuate periodically with seasons and scenarios, and this pattern also provides data support for the seasonal recommendation function of the app.

4. Design and Implementation of the Precision Beauty Recommendation Application

Building upon user segmentation findings, product-campus compatibility assessment and seasonal consumption patterns obtained from empirical analysis, this study develops a mobile-end beauty precision recommendation application oriented to university student groups. Different from mainstream commercial beauty platforms that pursue comprehensive functions, this application sets three core design goals: data-driven product matching, full consideration of real-world campus scenarios, and guiding students to form rational purchasing habits. In its design the application provides easy interactions with simple learning, which reflects students' habits, as they just want to browse and make decisions without learning much. The whole system sits on four layers, namely user layer, application layer, algorithm layer and data layer, as shown in Figure 7, which completes the whole process from multi-source data input to end-user interactive feedback.

The application forms a closed-loop service workflow covering skin condition assessment, personalized recommendation, reference for

 Copyright @ STEMM Institute Press

purchasing decisions and post-use feedback collection. All recommendation outputs of functional modules are derived from questionnaire statistics, review text mining and clustering results in this research, instead of commercial advertising data. This design is intended to address the pain point that existing beauty software is flooded with sponsored content, which reduces the credibility of recommendation results. Supported by local campus survey data, the application can provide relatively objective reference suggestions for student users.

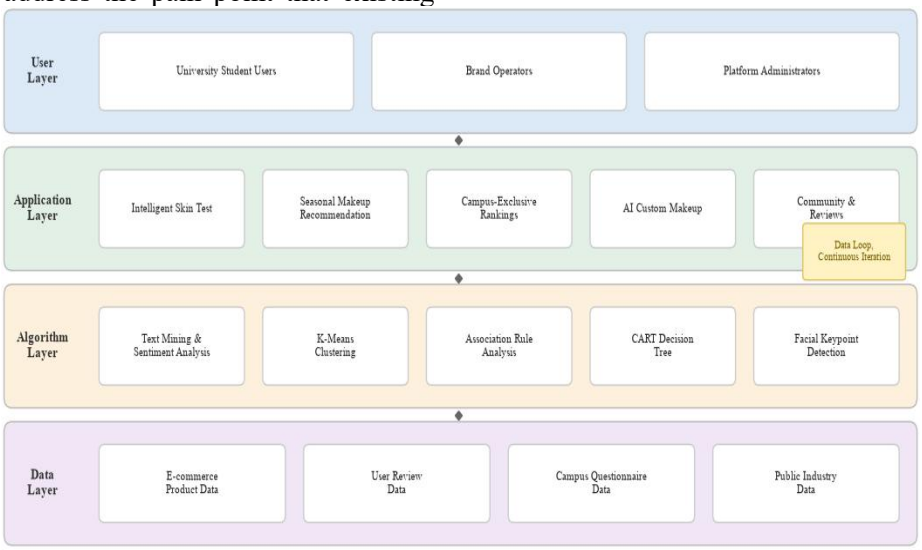


Figure 7. Architecture of the Precision Beauty Recommendation System

4.1 Core Functional Module Design

Each functional module is developed targeting the actual pain points summarized from questionnaires and e-commerce comment texts.

(1) Intelligent skin-type assessment. Survey statistics indicate that 11.65% of college students cannot accurately identify their own skin type. To avoid wrong judgment caused by users' self-evaluation, this module sets five simple questions about skin oil distribution, dryness performance, sensitivity reaction and pore status. The CART decision tree model classifies users into four skin categories: dry skin, oily skin, combination skin and sensitive skin, and the cross-validation accuracy on the sample set reaches more than 80%. After determining the skin type, the system automatically marks high-risk irritating ingredients such as alcohol, essence and high-concentration acid components that may trigger skin problems for corresponding groups. The whole skin-testing process can be finished within one minute, and the process is optimized according to user dropout rate in internal tests to ensure high completion rate of assessment tasks.

(2) Scenario-oriented seasonal recommendation. The monthly demand index constructed in Section 3.7 serves as the basis for this module. Combined with Zhengzhou's seasonal climate characteristics and typical campus time nodes, the recommendation content is dynamically adjusted: barrier-repair skincare is prioritized in spring with high incidence of skin allergy; oil-control and long-lasting makeup products are pushed during summer high-temperature period and military training; moisturizing care products become the main recommendation in dry winter. All recommended products are screened from items with high campus matching degree, and users can directly check product parameters and aggregated real user comments. The matching logic integrates three dimensions including seasonal features, campus usage scenarios and individual skin attributes, which is summarized from association rule mining results.

(3) Campus-exclusive product ranking list. Differing from general e-commerce best-seller lists heavily affected by marketing investment, the ranking synthesizes three dimensions of data: actual sales volume of sample brands, sentiment score calculated from comment corpus and student preference weight obtained from questionnaires. The list is refreshed once a month. Paid advertisements and dubious bulk reviews are removed. Its three sub-lists cover overall reputation, new product recommendations and best value. Skin types and price tags are available on every entry to help students narrow their product search quickly.

Copyright @ STEMM Institute Press

These rankings also differentiate the most popular items, the best-value alternatives and the newly available products.

(4) AI-assisted cosmetic color and style matching. Users take a front-on photo for the system to detect and analyze skin colors, face shape and facial proportions. The system suggests shades of lipstick and tones of foundation according to the type of skin and budget. It is not just a color theory based tool, but includes color deviation feedback from reviews, which will be useful for students that cannot test offline and can get different colors from what they see online. This gives a practical approach to color matching based on the review information, especially when there is no true representation of the final result in photos of products on the Web.

(5) Campus user review community. This user-generated community, for students only, allows them to post their real product experiences, makeup tips and questions. They have been sorted by skin type, product type and use. Soft ads and fake marketing posts are blocked by rules, addressing concerns raised in interviews. Community reviews, too, contribute to sentiment evaluation, boosting the review pool and future suggestions. By arranging the postings by the relevant attributes, the users may be able to find posts from other users who have similar skin concerns or needs.

4.2 Technical Architecture and User Experience Optimization

The project is divided into front and back end parts with the use of a lightweight mobile interface, modular services and secure storage. Mainstream frameworks, not experimental tools, are also mature and provide reliability and easy maintenance for a team of students. With five skin-test questions, one tap and three clicks, users get recommendations and product details. The contents of the homepage are adjusted based on the user's skin information and browsing history. No technical jargon related to skincare is used so new students can follow it, and support for visual or motor impairments is planned for future releases.

From the beginning of the design, privacy is built in. Information gathered at the time of skin assessment is used only for giving personalized recommendations and will not be sold to third parties or used for other marketing purposes. All data from the user is encrypted. The in-app feedback feature is designed to gather bug reports and suggestions, and certain features were modified based on user feedback during internal testing.

4.3 Application Practical Value

The application takes statistical results and turns them into actual service tools. Skin identification, risk ingredient reminders and budget-aware matching help students make safe and economical purchases to limit the occurrence of skin reactions. Behavioral data can help brands recognize preferences across skin-type groups and better design formulations, specifications and packaging. At the industry level, there are not many vertical recommendation platforms that can be found on campus currently. The application's closed loop between data and service can also guide service research on other consumer groups.

5. Discussion and Practical Implications

Based on multi-source data analysis of Zhengzhou college students' beauty consumption, this research puts forward targeted suggestions for student consumers, cosmetic brands and the whole beauty industry respectively.

5.1 Enlightenment for College Students: Realize Rational Beauty Consumption

Survey data reflects that a considerable number of college students have cognitive deficiencies in beauty consumption. Firstly, students should confirm their own skin type before purchasing cosmetics. 13.59% of respondents belong to sensitive skin and another 11.65% cannot judge their skin condition, and such groups are more likely to buy irritating products blindly and suffer economic losses. Consumers should check repeated complaints in reviews before buying. Common complaints about irritation, greasiness and exaggerated claims of efficacy from brand to brand give more helpful advice than advertising. If many users report bad experiences with the same product, it might be a real problem that consumers cannot see from the promotion. Secondly, beauty spending should be put to good use.

Sample data indicated that it is advisable to spend around 60% of the beauty budget on basic skincare, 30% on makeup and 10% on tools to limit impulse buying and needless stockpiling. Cheap fake cosmetics can be damaging to the

 Copyright @ STEMM Institute Press

skin, so official stores and authorized retailers are better than just opting for the cheapest choice. The allocation given is suggested as a practical guide to budgeting, and not a rule which every student must follow in their spending.

5.2 Enlightenment for Brands: Optimization Strategies for Campus-Oriented Products

K-Means resulted in the identification of three student segments, all with a need for differentiated products and marketing, which brands should provide. Cautious sensitive-skin users in total account for 30% of users, but few products are especially tailored for them. Focusing on the non-irritating repair of the barrier function may help to resolve irritation issues, as may removal or limiting of alcohol and strong fragrance. Another 30% are experienced users who apply cosmetics often and are looking for portability and long-lasting performance. Premium full-size product packaging is not ideal for long class days when compact packaging and durable oil-control formulas are sought. The other 40% want basic products at a lower price, so they would find excellent products priced at 50-100 yuan to suit their requirements. These disparities indicate that differential emphasis should be given to the issues of affordability, ingredient safety and performance among the three groups.

Brands need to address four common grievances when developing products: irritation, greasy textures, and ineffectiveness as well as a mismatch of shades. Product improvements are more important than more promotion. Limited edition sets to align with freshman enrollment, graduation, and military training may perform better than general seasonal campaigns. This means promotions will be timed to student needs in addition to the general market-wide advertising time.

5.3 Enlightenment for the Industry: Construct Multi-party Cooperative Campus Service Ecosystem

Campus beauty services could be developed in a joint effort by the universities, cosmetic companies and e-commerce platforms.

Spaces for skincare and makeup suggestions can be established within the student affairs offices of the university. The “user–scenario–product” model could be applied in e-commerce platforms where a section for students could be defined to match a product with user needs and user purposes. The service would turn the segmentation results from the study into information which students can make use of before buying. For the whole industry, when analyzing consumer demands, it is advisable to combine sales data, text comment mining and offline questionnaire survey, instead of relying on single data source. Relevant departments need to strengthen supervision over false reviews and hidden soft advertisements.

The analytical framework of this research is not limited to beauty consumption field. The complete workflow from comment text mining, user clustering to scenario-based recommendation can be migrated to other student consumption tracks such as dormitory supplies and sports goods. The core idea that segmented consumer groups deserve scenario-customized services rather than one-size-fits-all platforms has general reference significance for consumer-oriented industry research.

6. Conclusions and Outlook

6.1 Research Conclusions

Taking students at Zhengzhou universities as the research subject, we combined e-commerce big data with questionnaire responses, and used text mining, clustering, and association analysis to study campus beauty consumption; on top of that, we built a recommendation app. The main conclusions are as follows.

First, campus beauty consumption is dominated by affordable products, with light makeup the norm and basic skincare the focus; being gentle and cost-effective is what students care about most. 83% of students keep their monthly budget within 200 yuan, and the 50-100 yuan range has the largest group. Combination-oily and dry skin dominate, sensitive skin accounts for 14%, and redness plus oiliness are the core skin problems. Only 11% of respondents wear makeup every day. Three clearly different profiles are identified by clustering. Low-demand users spend the least and are less likely to put on cosmetics and like simple, basic items. Cautious sensitive-skin users, by contrast, are willing to spend more on mild formulas and are extra diligent in checking the ingredients. Heavy users spend and use the most and focus on the need for longevity and appropriateness to situations. Therefore, it is important that product positioning takes into account each of the segments rather than uniform solutions.

Copyright @ STEMM Institute Press

Second, complaints about inexpensive cosmetics are irritation and allergy, greasiness, and poor effectiveness and color mismatch. Negative-review text mining validates the most frequently raised concerns, namely irritation, allergy and greasiness. Sentiment scores are strongly correlated with sales, with generally more positively reviewed products being more successful. This confirms that reputation affects sales. The Innisfree green tea facial cleanser and the Inoherb red pomegranate skincare set are some of the bestsellers. There are also seasonal and campus demand trends with oil-control and setting products in the summer and early in semesters, makeup removers and cleansers when freshmen register, moisturizing and repair products in the winter, and sensitive-skin products in the spring. Together with the trends, these connections support the "user - scenario - product" framework that associates segments with daily skincare for the dorm, recovery in springtime, eight o'clock classes, club activities and summer touch-up. This descriptive approach achieved fairly high predictive accuracy with relatively simple algorithms. The advantage of this is that it makes the recommendations clear to the consumer, and addresses variations in the seasonal environment and in campus operation.

Third, the vertical beauty app built in this research serves three needs at once: students choosing products, brand operators reaching the campus market, and campuses filling a service gap. The platform has five functional modules - skin assessment, seasonal recommendation, campus product ranking, AI-assisted makeup design, and community communication - which together form a closed loop of "assessment - recommendation - consumption - transaction". The skin assessment module runs on a CART decision tree and reaches about 80% accuracy, and it flags risky ingredients such as fragrance and ethanol to warn users. For brand merchants, the data the app accumulates supports product optimization by skin type and scenario-based marketing. On an industry level, it is a solution to the lack of beauty services on campus and sets up a virtuous cycle in which data leads to services and services generate still more data. The service can gather more data about evolving preferences as the students interact with the recommendations and enter their own reviews.

6.2 Limitations and Future Work

However, clear limitations remain. This research was conducted only in Zhengzhou universities, limiting the regional representativeness of the samples. The product database includes limited brands and items and should be extended to local emerging brands. The application itself can still improve. Real-time integration of user behavior might be beneficial to recommendation responsiveness. The next phase of the research will expand the areas surveyed and the sample surveyed, and incorporate fresh feedback to enhance the application. Further data gathering is required to confirm whether the Zhengzhou purchasing trends are similar to those of students at universities across China.

Currently, the recommendation mechanism is based on rules for most of its reasoning. These rules are transparent and interpretable but not adapted to individuals. Dynamic recommendations could be provided with the approach of collaborative filtering and deep learning with long-term behavior and immediate feedback. Continuous engagement is expected through improvements which are based on real user input instead of assumptions, with planned pilot trials with university associations and cosmetic brands. It would also be useful to test whether the proposed features still make sense beyond the scope of research and in-house trials.

There are three directions for future work. First, representative universities from eastern, central and western China will allow cross-regional comparisons of the segmentation and matching patterns. Second, the use of convolutional neural networks that process facial images and skin indicators could help to enhance skin-type identification, and real-time behavioral data could make recommendations more dynamic. Third, each update of the app should be rigorously tested using A/B testing and feedback cycles. Possible partnerships include university institutions and e-commerce sites. Finally, accessibility and privacy safeguards should be given more attention as the number of users grows. These advances will facilitate testing for generalizability and advancing technical performance, as well as assessing whether personalized services provide measurable benefits.

References

[1] Zheng Y, Li C. Body socialization: the practical logic of female college students' cosmetics consumption. China Youth Study, 2025(5): 110-119. (in Chinese)

 Copyright @ STEMM Institute Press

[2] Yu Q. Investigation and analysis of female college students' makeup consumption behavior. China Business (Theory Research), 2019(2): 11-12. (in Chinese)
[3] Xia B, Han M, Zhu Y. Experiential marketing of L'Oréal in China: a case study based on consumer behavior of Chinese female university students//Proceedings of the International Conference on Economic Management and Cultural Industry. Paris: Atlantis Press, 2021.
[4] Li B. Research on marketing strategy of Chinese domestic cosmetics based on female college students' consumption characteristics//Proceedings of the 2021 6th International Conference on Social Sciences and Economic Development (ICSSSED 2021). Paris: Atlantis Press, 2021.
[5] Kaur G, Sharma A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. Journal of Big Data, 2023, 10: 61.
[6] Mutinda J, Mwangi W, Okeyo G. Sentiment analysis of text reviews using lexicon-enhanced BERT embedding (LeBERT) model with convolutional neural network. Applied Sciences, 2023, 13(3): 1445.
[7] Khan R H, Dofadar D F, Alam M G R, et al. LRFS: online shoppers' behavior-based efficient customer segmentation model. IEEE Access, 2024, 12: 96462-96480.
[8] Song H, Zhang H, Xing Z. Research on personalized recommendation system based on association rules. Journal of Physics: Conference Series, 2021, 1961(1): 012027.
[9] Zhang M. Consumer behavior analysis based on Internet of Things platform and the development of precision marketing strategy for fresh food e-commerce. PeerJ Computer Science, 2023, 9: e1531.
[10] Latha Y M, Rao B S. Amazon product recommendation system based on a modified convolutional neural network. ETRI Journal, 2024, 46(4): 633-647.

Copyright @ STEMM Institute Press