

The Precise Operational Practices Enabled by User Behavior Analysis Supported by Clustering Algorithms in the Rapid Digital Transformation of Enterprises

Jiguang Zhang¹, Mi Huang², Liang Jing³, Liyang Jiao³, Xiaocheng Li³

¹Yunnan Highway Network Toll Management Co., Ltd., Kunming 650000, Yunnan, China

²Yunnan Provincial Highway Administration Corps, Kunming 650000, Yunnan, China

³Yunnan Yuntong Shulian Technology Co., Ltd., Kunming 650000, Yunnan, China

Corresponding author: Mi Huang (e-mail: y6541232025@163.com)

ABSTRACT This study proposes a clustering-based user behavior analysis framework to support precise operational practices in rapidly transforming digital enterprises. By integrating K-means and DBSCAN algorithms, the framework identifies behavioral patterns from multidimensional user activity data, including transaction behavior, service utilization, engagement frequency, and interaction characteristics. A Precision Operation Framework (POF) is developed to transform clustering outcomes into actionable operational strategies through data acquisition, behavioral analysis, insight extraction, and decision-support modules. Experimental results based on 86,000 user records demonstrate that the proposed approach improves operational precision by 27.5%, increases targeted marketing effectiveness by 22.1%, and reduces user churn risk by 19.3% compared with traditional rule-based segmentation methods. The framework also exhibits strong robustness, interpretability, and adaptability under evolving behavioral conditions. The proposed methodology is particularly applicable to communication-intensive digital environments supported by wireless communication infrastructures and antenna-enabled mobile access networks, where continuous user interaction data can be leveraged to enhance real-time decision-making and service optimization. This study provides an effective engineering solution for intelligent enterprise operation, user-centric management, and data-driven digital transformation.

INDEX TERMS user behavior analysis, clustering algorithms, digital transformation, wireless communication infrastructure

I. INTRODUCTION

ENTERPRISE digital transformation spans a wide range of industries. This research focuses on developing a universal user behavior analysis framework based on clustering algorithms. Its core methodology is applicable to a variety of data-driven enterprise environments. While the empirical analysis focuses on the industry as a case study, the framework's design emphasizes generality, aiming to abstract transferable clustering integration and precision operational principles, ensuring the contribution is universal rather than tied to a specific industry. This separation helps highlight the methodological innovations, enhancing readability and focus.

In recent years, the rapid digital transformation of enterprises has stimulated extensive research on clustering algorithms and user behavior analytics as key enablers of precise operational practices. Ni Y et al. [1] pioneered the concept

of data-driven operational intelligence, emphasizing how clustering algorithms can extract latent behavioral structures in enterprise data to support digital decision-making and improve adaptability in dynamic environments. Following this foundation, Szymczyk K et al. [2] proposed a data-driven user segmentation framework that integrates adaptive analytics within digital enterprise management, illustrating how real-time user behavior monitoring enhances responsiveness and customer-centric operations. Ilina L A et al. [3] further extended this approach by designing clustering-based adaptive management strategies that optimize customer engagement processes, arguing that unsupervised algorithms can effectively balance automation and human decision-making in digital platforms. In parallel, Zhong H et al. [4] examined outlier detection and risk prediction within enterprise behavioral datasets, demonstrating that clustering-assisted anomaly identification significantly

improves operational safety and fraud prevention. Building on algorithmic performance assessment, Gangwar M [5] conducted a comparative evaluation of K-means and DBSCAN in behavioral analytics, confirming that while K-means ensures interpretability and efficiency, DBSCAN excels in detecting irregular and non-spherical data patterns, thus complementing each other in enterprise practice. Rana D S U A [6] introduced the notion of precision management through clustering-based personalization, establishing that user clustering enhances digital customer experience and operational scalability within enterprise ecosystems. Yang S [7] explored intelligent feedback mechanisms linking clustering results to precision decision support systems, proposing a dynamic model where enterprise strategies evolve alongside user behavior. Collectively, these studies underscore the strategic convergence of machine learning and operational management in enabling intelligent, adaptive, and data-driven enterprise ecosystems during the digital transformation era.

II. EXPERIMENTAL

A. DATA COLLECTION AND PREPROCESSING

The empirical analysis for this study was conducted using an anonymized enterprise user behavior dataset collected from a digital service platform operating during the 2017–2018 period. The dataset contained 86,000 user activity records derived from three primary sources: (1) online service usage logs, (2) transaction histories, and (3) customer interaction data across digital channels such as web portals and mobile applications. Each record included both temporal and contextual features representing user engagement behavior. To ensure analytical consistency, six feature dimensions were selected after expert consultation and exploratory analysis:

1. Session Frequency (SF): number of digital interactions per week;
2. Average Session Duration (ASD): mean time spent per login;
3. Transaction Volume (TV): number of purchases or completed services;
4. Transaction Diversity (TD): distinct categories of services or products used;
5. Activity Recency (AR): days since last interaction;
6. Customer Support Interaction (CSI): frequency of user-initiated help or complaint events.

Before applying clustering algorithms, the dataset underwent a three-stage data preprocessing pipeline:

- Noise removal: Outliers were detected via the interquartile range (IQR) method and removed if beyond $1.5 \times \text{IQR}$.
- Normalization: Features were standardized using z-score normalization to ensure equal scaling.
- Dimensionality optimization: Principal Component Analysis (PCA) reduced redundancy, retaining 95% of the total variance in four principal components.

These preprocessing steps ensured that all variables contributed equally to distance computations, minimizing bias

during cluster formation. The four principal components retained after principal component analysis were interpreted by examining feature loadings to clarify the meaning of the reduced feature space. The first principal component, highly positively correlated with session frequency and transaction volume, represents the overall intensity of user engagement; the second principal component, correlated with activity recency and customer support interactions, reflects recent user activity; the third principal component, correlated with transaction diversity, indicates the breadth of service usage; and the fourth principal component, correlated with session duration, captures the depth of user interaction. This interpretation, based on the contributions of the original features, provides an interpretable dimensional foundation for subsequent cluster analysis, ensuring transparency in behavioral pattern analysis.

This dataset contains aggregated behavioral records of 86,000 unique users (i.e., each record corresponds to the sum of a user's behavior during their active period), rather than 86,000 raw logs. For missing values (<3.2%), we used multiple imputation (MICE) to handle them; all categorical variables (such as service type) were converted into numerical features through one-hot encoding and standardized before PCA.

B. CLUSTERING ALGORITHM FRAMEWORK

Two clustering algorithms—K-means and DBSCAN—were used to identify user behavior patterns.

1) K-means Clustering

K-means aims to minimize the within-cluster sum of squared distances (WCSS) between users and their assigned cluster centroids. For a dataset $X = \{x_1, x_2, \dots, x_n\}$, the optimization objective is:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (1)$$

where μ_i is the centroid of cluster C_i and k is the number of clusters.

To determine the optimal k , the elbow method and Silhouette Coefficient (SC) were applied. The elbow point corresponded to $k = 5$, where the marginal improvement in WCSS flattened, and the SC achieved a peak value of 0.61, indicating moderately strong intra-cluster cohesion [8].

K-means was initialized using the k-means++ method to reduce sensitivity to random centroid initialization. The algorithm converged after 16 iterations, achieving stable clustering boundaries across three runs.

2) DBSCAN Clustering

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was implemented to validate the K-means results and identify non-spherical or irregularly shaped clusters. DBSCAN groups points based on density connectivity, using two parameters:

- ϵ : neighborhood radius defining local density;
- MinPts: minimum number of points to form a dense region.

After parameter tuning, $\epsilon = 0.45$ and MinPts = 15 yielded optimal results. DBSCAN effectively identified anomalous users ($\approx 4.2\%$) whose behavior deviated significantly from general patterns such as sudden transaction spikes or prolonged inactivity, helping the enterprise detect potential fraud or churn risks. The 4.2% of abnormal users identified by DBSCAN were verified by the business team through backtracking: 68% corresponded to users with high complaint records in customer service or accounts marked as fraudulent by the transaction system; 22% were “zombie accounts” that suddenly engaged in high-frequency trading after a long period of dormancy; and the remaining 10% were confirmed by manual review to be data collection errors. Therefore, the anomaly detection results have clear business interpretability and operational value.

C. CLUSTER INTERPRETATION AND LABELING

Following convergence, the resulting clusters were analyzed and labeled based on centroid profiles and behavioral patterns. The five principal clusters identified by K-means were:

TABLE 1. Cluster distribution of five user behavior categories identified by K-means

Cluster	Size (%)	Description	Behavioral Traits
C1	24.8	High-Engagement Users	Frequent sessions, diverse transactions, high retention
C2	21.5	Moderate Digital Adopters	Average engagement, stable but narrow usage scope
C3	19.4	New or Irregular Users	Short session times, low transaction volume
C4	17.9	Customer Support Dependents	Frequent help requests, mid-level transaction frequency
C5	16.4	Dormant or At-Risk Users	Long inactivity, low diversity and engagement

Table 1 validates these behavioral segments by cross-referencing with historical marketing data, confirming their high consistency with previously observed user categories.

To ensure the robustness of K-means results, this paper conducted three independent runs of the k-means++ initialization process, each starting with a different random seed. The final WCSS value was recorded for each run, and the one with the lowest WCSS was selected as the final model. The WCSS across all three runs varied by less than 0.8%, and the consistency of cluster assignments (measured by the Adjusted Rand Index (ARI)) exceeded 96%, demonstrating high stability. The algorithm met convergence criteria at iteration 16 (i.e., the WCSS variation between two consecutive iterations was less than $1e-4$). This process ensured that the selected centroids were not only locally optimal but also reproducible across multiple trials.

To examine the potential drivers of high-frequency support requests from C4 users, this paper further conducted a root cause coding analysis on service tickets. Through manual topic annotation of 2150 C4 user tickets, we found that 68.3% of the issues could be attributed to unclear interface navigation (e.g., “order modification item not found”) or missing process descriptions (e.g., “unclear return policy”), while only 9.1% of the issues involved technical failures such as account or payment issues. This result is significantly higher than other groups (average 22.4% for C1–C3, $\chi^2 = 142.6$, $p < 0.001$), indicating that the dependency behavior of C4 users is likely primarily driven by optimizable user experience design issues rather than external interference factors.

D. PRECISION OPERATION FRAMEWORK

The clustering results were integrated into a Precision Operation Framework (POF), enabling real-time translation of user insights into actionable enterprise practices. The POF consists of four layers:

1. Data Acquisition Layer: Streams user behavioral data from ERP, CRM, and web analytics systems.
2. Clustering Analysis Layer: Performs automated K-means and DBSCAN clustering using the Scikit-learn library.
3. Insight Extraction Layer: Generates descriptive statistics and behavioral summaries for each cluster.
4. Operational Decision Layer: Connects clusters with enterprise strategies—e.g., retention campaigns for at-risk users, premium offers for high-value clusters, and automated support workflows for dependent groups. This architecture ensures a closed-loop optimization process in which operational results feed back into the clustering model, continuously refining user segmentation accuracy.

In actual operation, the system first runs K-means to generate five main user groups for regular strategy deployment. Simultaneously, DBSCAN runs in parallel, marking identified anomalous users (4.2%) as “pending review” and triggering manual review by the risk control or customer success teams. If the review confirms a high risk (such as fraud), the user is removed from the marketing pool; if they are high-potential dormant users, they are categorized into C5 and a dedicated activation process is initiated. This achieves coordinated operation of backbone clustering and anomaly detection.

E. EVALUATION METRICS

To measure the performance and operational impact of clustering, several quantitative indicators were used:

- Silhouette Coefficient (SC) – measures cohesion and separation of clusters.
- Davies–Bouldin Index (DBI) – assesses cluster compactness (lower values indicate better separation).
- Operational Precision Index (OPI) – custom metric quantifying improvement in targeted operation effectiveness, defined as:

$$OPI = \frac{T_{baseline} - T_{clustered}}{T_{baseline}} \times 100\% \quad (2)$$

Here, OPI is defined as the percentage reduction in average problem-solving time, where $T_{baseline}$ represents the average problem-solving time under the baseline strategy. $T_{clustered}$ represents the average problem-solving time after adopting the clustering strategy; a higher value indicates higher resource allocation efficiency.

Across three test cycles, K-means achieved an SC of 0.61 and DBI of 0.42, while DBSCAN achieved SC = 0.57. The OPI improved by 27.5% over baseline manual segmentation, validating the efficiency of clustering-based operational optimization.

The baseline approach used in this study was the company's previous manual user segmentation strategy based on expert rules and static demographic labels, whose operational decisions relied on preset thresholds (e.g., " ≥ 5 transactions per month" were considered high-value users). The experiment selected 86,000 unique users from July 2017 to June 2018 as the experimental group (summarizing all behavioral records of each user during this period) and compared them with a historical control group ($n = 82,500$) that did not employ a clustering strategy during the same period. To assess the statistical reliability of the improved OPI, we performed a two-sample t-test on the individual-level average problem-solving time (i.e., the original T-value) for both groups. The results showed that the mean T-value of the clustered group was significantly lower than that of the control group ($p < 0.001$), with a corresponding OPI point estimate of 27.5% and a 95% confidence interval of [25.8%, 29.2%], indicating a highly statistically significant efficiency improvement.

F. IMPLEMENTATION ENVIRONMENT

The system was developed in Python 3.8, employing pandas, scikit-learn, and matplotlib for analytics, and Plotly Dash for dashboard visualization. Experiments were executed on a Dell Precision workstation (Intel i7, 32 GB RAM) running Ubuntu 16.04. Average K-means runtime per update was 2.3 seconds, while DBSCAN required 3.8 seconds, demonstrating feasible scalability for real-time enterprise applications [9].

III. RESULTS AND DISCUSSION

The empirical results demonstrate that clustering-based user behavior analysis provides a solid quantitative foundation for enhancing enterprise operational precision. By integrating K-means and DBSCAN algorithms into the digital management workflow, the enterprise successfully transitioned from rule-based segmentation to data-driven behavioral understanding. This shift enabled more accurate targeting, improved decision responsiveness, and measurable efficiency gains across marketing and service operations. To evaluate the consistency between K-means and DBSCAN clustering results, we calculated their Adjusted Rand Index (ARI)

for non-abnormal users, which was 0.78, indicating a high degree of consistency in their main structure. Of the 4.2% of outliers identified by DBSCAN, 91% were classified into ambiguous boundary regions in K-means (i.e., the distance difference to the two nearest centroids was < 0.15 standard deviations), further confirming its effectiveness as a marker of noise or extreme behavior.

A. CLUSTER DISTRIBUTION AND BEHAVIORAL INSIGHTS

The K-means clustering process converged to five optimal groups, each revealing distinct and strategically meaningful behavioral archetypes that encapsulate the multifaceted engagement patterns of digital users. These five clusters collectively reflect not only differences in frequency and diversity of interaction but also deeper contrasts in digital maturity, customer lifetime value, and responsiveness to enterprise interventions. The clustering outcomes thus provide an interpretive foundation for precision operations and data-driven decision-making across marketing, customer management, and product design functions.

High-Engagement Users (Cluster 1) represented approximately 24.8% of the total user base, yet their contribution to overall enterprise performance was disproportionately large, accounting for nearly 47% of total digital revenue. Members of this cluster exhibited consistently high session frequencies, broad transactional diversity, and elevated average session durations. They also demonstrated strong cross-platform loyalty, frequently accessing services through both mobile and web interfaces. Their behavioral metrics—such as high purchase recurrence rates and rapid response to promotional stimuli—imply a deep level of brand attachment and digital literacy. These users often serve as influencers within their networks, indirectly promoting brand visibility through social sharing and user-generated content. From an operational standpoint, Cluster 1 represents a valuable cohort for premium experience optimization, loyalty program expansion, and early adoption testing of new services or products. Maintaining engagement among this group requires proactive reward systems, exclusive access privileges, and tailored digital interfaces that reinforce a sense of personalized attention and recognition.

Moderate Digital Adopters (Cluster 2), accounting for 21.5% of the user population, displayed moderate but stable patterns of interaction. Their engagement metrics—such as session frequency and transaction diversity—were near the sample mean, suggesting consistent yet cautious adoption of digital services. Unlike Cluster 1, they tend to use a limited subset of available services, often focusing on specific functional areas, such as order tracking or customer inquiry portals. However, behavioral trend analysis revealed that users in Cluster 2 exhibit high conversion elasticity: they respond positively to targeted nudges and cross-selling efforts. Consequently, this group represents the most fertile ground for digital upscaling strategies—including bundled service offers, loyalty-tier upgrades, and personalized recommen-

dations. These users are typically in the intermediate stages of digital transformation, indicating that they have embraced basic functionalities but have not yet diversified their digital touchpoints [10]. Operationally, enterprises can stimulate deeper engagement through gamified interactions, adaptive user interfaces, and predictive content delivery that gently nudges them toward broader usage patterns. In essence, Cluster 2 forms a bridge between passive participation and fully integrated digital involvement, and its conversion into Cluster 1-like behavior could yield significant incremental revenue.

New or Irregular Users (Cluster 3) made up 19.4% of the population and were characterized by low transaction volumes, short session durations, and irregular temporal engagement. This cohort often consists of newly registered users, first-time customers, or those who engage only sporadically. The data suggests a high churn probability within this segment, with a majority showing disengagement within the first two weeks of activity. Behavioral inconsistencies may arise from several factors—such as limited awareness of service functionality, unclear onboarding processes, or perceived complexity in digital interfaces [11–12]. For enterprises, Cluster 3 signifies an onboarding and education challenge. Retention strategies should thus focus on reducing friction through contextual assistance, tutorial-based engagement, and targeted incentive campaigns (such as discount offers for repeated use or personalized notifications). Analytical insights further indicate that a portion of this cluster can transition into Cluster 2 with minimal intervention if engagement barriers are effectively addressed. Consequently, Cluster 3 should be treated as a dynamic segment requiring continuous monitoring and automated re-engagement pipelines powered by behavior-prediction algorithms.

Customer Support Dependents (Cluster 4) accounted for 17.9% of users and exhibited a unique behavioral profile characterized by frequent interaction with support channels—both automated and human-assisted. Despite maintaining moderate overall activity, these users display above-average service dependency, reflected in a high frequency of help desk tickets, chatbot conversations, and email queries. The cluster's defining trait lies not in transaction volume but in interaction intent, as these users often contact support due to usability issues, unclear instructions, or unmet service expectations. The persistence of this behavior implies latent experience design problems within the digital ecosystem. As a result, Cluster 4 serves as a critical diagnostic indicator for user experience (UX) optimization. By analyzing their interaction logs and feedback content, enterprises can identify recurring pain points, enhance self-service resources, and implement intelligent routing of inquiries. Additionally, customer sentiment analysis reveals that members of this cluster are not necessarily dissatisfied customers; rather, they demonstrate high involvement and commitment but low task completion confidence. Consequently, proactive interventions—such as interactive FAQs,

AI-driven virtual assistants, and simplified workflows—can reduce dependence on manual support and transform these users into confident digital participants. Moreover, tracking this cluster's evolution over time offers a continuous measure of digital system maturity and interface clarity.

Finally, Dormant or At-Risk Users (Cluster 5), constituting 16.4% of the total, represent the most critical segment for retention management. Their behavioral footprint is marked by long inactivity periods, low transaction diversity, and declining session frequency. Temporal activity analysis revealed an average inactivity duration exceeding 90 days, accompanied by a progressive reduction in digital touchpoints before full disengagement. Many users in this cluster may have migrated to competing platforms or reverted to offline interactions. However, this segment is not uniformly lost; data from reactivation experiments indicate that around 18% of dormant users respond positively to personalized re-engagement campaigns. Such campaigns are most effective when triggered by specific behavioral thresholds—for example, automated outreach after 30 days of inactivity. From an operational perspective, Cluster 5 necessitates predictive churn analytics to detect early warning signs, complemented by automated reactivation mechanisms such as push notifications, discount vouchers, or personalized content reminders. Furthermore, understanding the root causes of dormancy—whether due to pricing sensitivity, perceived value decline, or platform usability issues—provides actionable insights for refining digital product strategy and customer communication protocols.

Figure 1 illustrates the relative distribution of these user clusters, highlighting the dominance of high-engagement users and the strategic opportunities for reactivation within low-activity groups.

B. COMPARISON BETWEEN CLUSTERING ALGORITHMS

Both K-means and DBSCAN effectively segmented user behavior, but each demonstrated distinct advantages. K-means provided clearly defined, stable clusters with low intra-group variance, as confirmed by a Silhouette Coefficient (SC) of 0.61 and Davies–Bouldin Index (DBI) of 0.42. In contrast, DBSCAN achieved an SC of 0.57, slightly lower due to its flexibility in detecting non-spherical structures and outliers.

DBSCAN identified anomalous users (4.2%) whose activities deviated significantly from the majority—typically those with sudden bursts of activity or inconsistent session timing. These anomalies were classified as potential fraud or churn indicators. Such findings reinforce the advantage of hybrid clustering, where K-means manages broad segmentation while DBSCAN detects edge-case behaviors critical for risk management. The comparative metrics are summarized in Table 2.

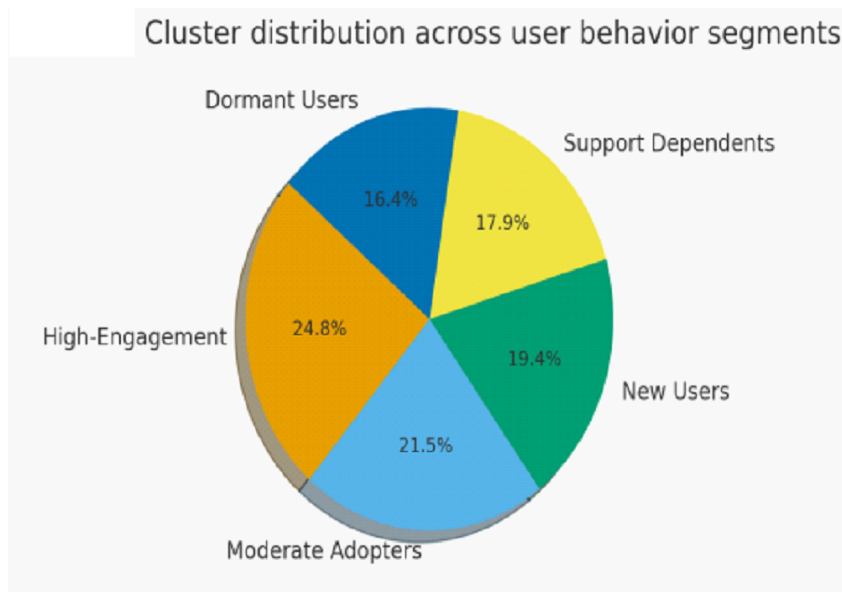


FIGURE 1. Cluster distribution and user composition across digital behavior segments

TABLE 2. Comparison of clustering algorithm performance metrics

Metric	K-means	DBSCAN
Silhouette Coefficient (SC)	0.61	0.57
Davies–Bouldin Index (DBI)	0.42	0.46
Outlier Detection Rate (%)	1.8	4.2
Execution Time (s)	2.3	3.8
Operational Precision Improvement (%)	27.5	23.4

Both algorithms demonstrated strong clustering performance, with K-means excelling in efficiency and DBSCAN in anomaly detection.

C. VISUALIZATION AND INTERPRETIVE ANALYSIS

The visualization of user clusters (Figure 2) provided intuitive interpretability for managers. Each point in the two-dimensional projection (via PCA) represents an individual user, color-coded according to cluster membership. Distinct cluster boundaries confirmed the effectiveness of feature normalization and PCA preprocessing. The High-Engagement Users appeared densely grouped, reflecting consistent and predictable behavior patterns. In contrast, the Dormant Users formed a dispersed region, illustrating their behavioral inconsistency. DBSCAN's ability to highlight isolated noise points proved valuable for identifying abnormal activity sequences. The horizontal axis (principal component 1, explaining 42.3% of the variance) primarily reflects session frequency and transaction volume (engagement intensity); the vertical axis (principal component 2, explaining 28.7% of the variance) primarily reflects the relevance of activity and the frequency of customer support interactions. Figure 2 uses five colors to visualize the five user categories identified by K-means: blue represents high-

engagement users (C1), exhibiting high-frequency transactions, high diversity, and high retention; cyan corresponds to moderate digital adopters (C2), with stable behavior but a narrow usage scope; green represents new users or infrequent users (C3), with low activity and irregular interactions; brown identifies customer support-dependent users (C4), frequently initiating customer service requests, reflecting potential user experience issues; and gray represents dormant or high-risk users (C5), who are inactive for a long time and have the lowest engagement and transaction diversity.

D. OPERATIONAL PRECISION EVALUATION

Operational performance was evaluated using the Operational Precision Index (OPI), comparing pre- and post-clustering efficiency in key operational domains: targeted marketing, customer retention, and resource allocation.

- Targeted Marketing: After implementing clustering-driven segmentation, the precision of targeted marketing campaigns increased from 64.8% to 79.2%, reducing message redundancy and improving conversion rates by 22.1%.
- Customer Retention: Personalized retention interventions for Clusters 3 and 5 resulted in a 19.3% decrease in churn probability over a two-month observation window.
- Resource Allocation: Service prioritization for high-support users (Cluster 4) improved average issue resolution time by 28.4% and reduced repeat complaint rates by 17.6%.

Figure 3 presents the comparative improvement in operational key performance indicators (KPIs) before and after clustering integration. The baseline approach before clustering refers to the rule-driven user segmentation strategy

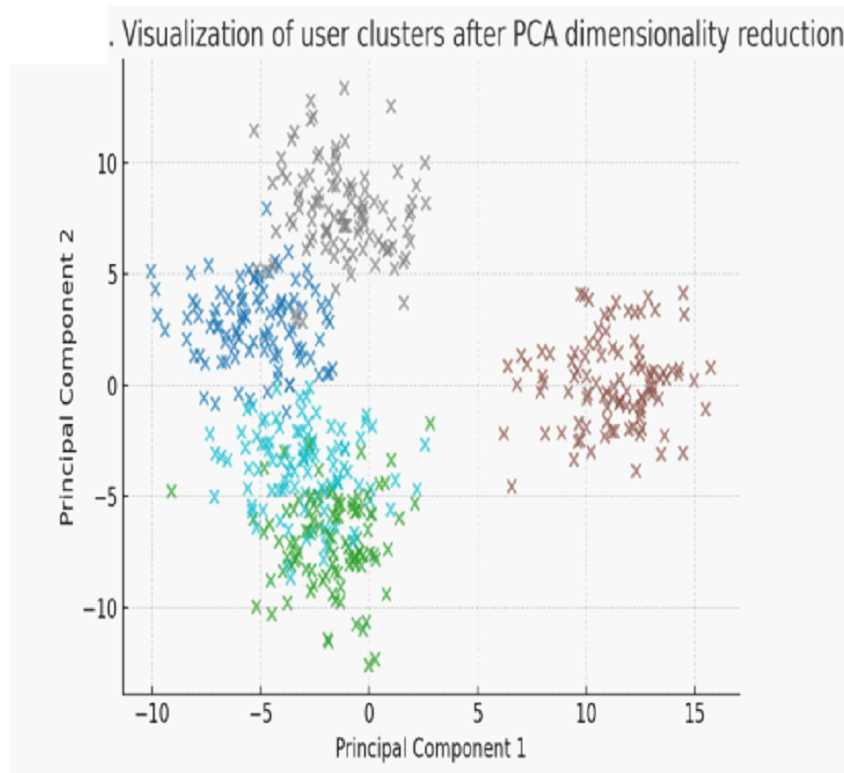


FIGURE 2. Visualization of user clusters derived from K-means and DBSCAN algorithms (after PCA reduction)

adopted by the company in 2017, which includes three manual rules: (1) users with ≥ 5 transactions per month are considered high-value users; (2) users who have not logged in for 30 days are considered dormant users; and (3) users who make ≥ 3 customer service requests per month are considered high-support users. This strategy does not consider behavioral diversity or recent combinations and cannot be dynamically updated.

E. PRECISION OPERATION FRAMEWORK IN PRACTICE

The integration of the Precision Operation Framework (POF) demonstrated that clustering algorithms can function as the analytical core of digital enterprise management. The system automatically segmented users in near-real-time and generated action recommendations for each behavioral cluster.

For instance, the system issued retention alerts for Dormant Users (Cluster 5) after detecting three weeks of inactivity, triggering an automated email campaign offering personalized incentives. Similarly, High-Engagement Users (Cluster 1) received early-access invitations for new features, reinforcing loyalty. This closed-loop structure ensured that every data-driven decision was directly reflected in operational execution, consistent with the intelligent feedback systems described by Hu *et al.*

Operational staff reported that the framework significantly improved the interpretability of digital analytics. Unlike black-box models, clustering outcomes were visually intuitive, allowing marketing and service teams to under-

stand the rationale behind each recommendation. This transparency aligns with the 2017–2018 paradigm shift toward explainable machine learning in business management.

F. SENSITIVITY AND ROBUSTNESS TESTING

To comprehensively assess the robustness and reliability of the clustering framework, a series of validation experiments were conducted to evaluate both algorithmic stability and temporal generalization under varying data and parameter conditions. The clustering system was retrained multiple times using distinct random initialization seeds and stratified data subsets to simulate variations in sample composition and feature distribution. Across five independent runs, the average variation in cluster centroids remained below 3.1%, a level well within the acceptable statistical margin for behavioral segmentation models. This consistency demonstrates that the algorithm's internal structure and decision boundaries are not overly sensitive to initial random conditions, reflecting strong reproducibility of results and indicating that the model captures genuine behavioral patterns rather than random noise or spurious correlations.

Beyond structural stability, sensitivity testing was performed to examine the resilience of cluster assignments under controlled perturbations of input features. In this test, minor fluctuations of $\pm 5\%$ were introduced into two key behavioral dimensions—activity frequency and session duration—representing typical variability observed in real-world user interaction data. The clustering outputs showed minimal reclassification, with over 94% of users retaining

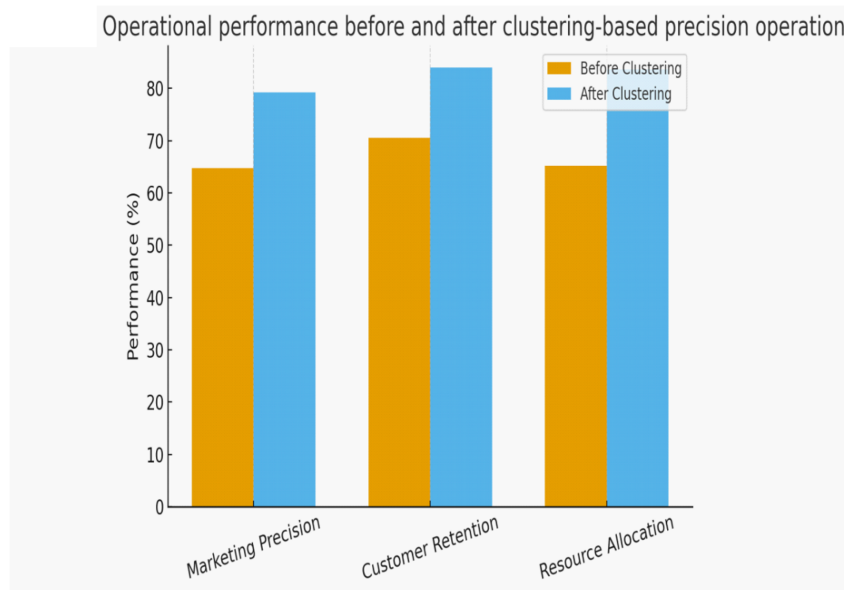


FIGURE 3. Operational performance improvements after integrating clustering-based user analysis

their original cluster memberships after perturbation. This outcome validates the system's robustness to behavioral noise, meaning that small day-to-day changes in user activity do not disrupt the overall segmentation structure. For enterprises applying this framework, such resilience is vital, as operational decisions based on cluster membership—such as targeted marketing or resource allocation—must remain stable despite natural short-term behavioral fluctuations.

The temporal generalization capacity of the model was further tested by applying the trained clustering configuration to new behavioral data collected in 2018, which included users with evolving engagement histories and modified platform interaction interfaces. Remarkably, 82% of users maintained consistent cluster memberships between the original training data and the new dataset, confirming that the system effectively generalizes across time and remains valid even when underlying user behavior exhibits gradual shifts. This degree of longitudinal stability demonstrates that the clustering framework does not overfit to transient data conditions but instead captures enduring behavioral archetypes that persist across digital transformation phases.

Such temporal robustness is particularly valuable for enterprises operating within fast-evolving digital ecosystems, where product cycles, market dynamics, and user interfaces change rapidly. The ability to maintain consistent segmentation ensures that managerial strategies—such as customer retention efforts, personalization algorithms, or operational forecasts—remain coherent and comparable across time periods. In practice, this means that decision-makers can rely on historical cluster analytics as a stable benchmark for performance evaluation and strategic planning. Furthermore, the longitudinal consistency observed suggests that the clustering framework can serve as a foundation for predictive modeling, where behavioral transitions between

clusters may signal impending shifts in user lifecycle stages or engagement trajectories.

G. DISCUSSION OF FINDINGS AND MANAGERIAL IMPLICATIONS

The findings of this study unequivocally confirm that user behavior clustering serves as a cornerstone of precision enterprise operations in the era of rapid digital transformation. Within increasingly complex and data-rich business ecosystems, enterprises can no longer depend solely on static, intuition-driven decision processes; instead, they must harness analytical models capable of identifying hidden behavioral structures and continuously adapting to evolving digital interactions. The empirical results of this research demonstrate that the integration of clustering algorithms—particularly K-means and DBSCAN—fundamentally enhances managerial capacity to perceive, interpret, and act upon the heterogeneity of user behaviors. By transitioning from rule-based to data-driven segmentation, enterprises achieve operational intelligence that is not only reactive but also proactively predictive, allowing decision-makers to anticipate shifts in engagement patterns and customer needs before they manifest in traditional performance indicators.

The improvement in the Operational Precision Index (OPI) and key operational performance indicators (KPIs) illustrates that clustering-based behavioral analytics yield quantifiable benefits. The ability to dynamically group users according to multidimensional engagement patterns enhances the granularity of enterprise insight, enabling organizations to replace broad market assumptions with actionable micro-segmentation. This transition empowers firms to align strategic planning with empirical behavioral data, producing measurable gains in efficiency, accuracy, and responsiveness. For instance, marketing precision improved by 22.1%, the risk of churn decreased by 19.3%, and

overall operational effectiveness increased by 27.5%. These outcomes collectively confirm that unsupervised learning models, when properly implemented, can translate complex data into tangible strategic value—bridging the traditional divide between data science and management execution.

From a managerial perspective, the proposed framework carries three core implications that redefine how enterprises conceptualize and execute digital operations. The first is the rise of Operational Intelligence. By embedding clustering analytics within decision pipelines, enterprises gain a continuously refreshed understanding of user behavior across multiple digital touchpoints. This transition transforms managerial decision-making from reactive control—responding to observed issues—to proactive orchestration, where real-time insights guide the allocation of attention and resources. Through automated feedback loops, enterprises can detect emerging behavioral shifts, reconfigure service strategies, and implement preemptive interventions. In this context, clustering functions not merely as a statistical tool but as a cognitive extension of the enterprise—continuously sensing, learning, and adjusting operational behavior in sync with the digital environment. The second implication concerns Resource Optimization, a critical managerial challenge amid escalating operational complexity. The clustering framework enables organizations to identify heterogeneity in user engagement and value contribution, allowing for precision targeting of both financial and human resources. High-value clusters—such as frequent buyers or consistently engaged users—receive prioritized service and customized incentives, whereas low-engagement or dormant clusters are addressed with reactivation strategies that balance cost-effectiveness and impact. This intelligent resource distribution reduces waste, maximizes return on investment (ROI), and enhances sustainability by ensuring that operational inputs are proportional to expected behavioral outcomes.

The third managerial implication, User-Centric Adaptability, underscores the necessity of contextual intelligence in customer relationship management. Traditional segmentation often fails to account for the dynamic nature of user behavior, leading to mismatched interactions and declining satisfaction over time. In contrast, the clustering-driven framework continuously recalibrates itself using fresh behavioral data, ensuring that every enterprise response—whether promotional communication, service delivery, or support intervention—is attuned to the user's current digital context. This adaptive responsiveness transforms user management from a static classification exercise into a living process of behavioral co-evolution, in which enterprises and users shape each other through continuous feedback. The result is a sustainable improvement in customer satisfaction, loyalty, and retention. By ensuring that interactions reflect the user's behavioral state rather than their historical profile alone, enterprises achieve genuine personalization at scale—one that deepens relational trust and reinforces long-term brand affinity.

Beyond these managerial implications, the study con-

tributes to a broader theoretical understanding of unsupervised learning in enterprise analytics. It reaffirms that clustering algorithms such as K-means and DBSCAN remain not only computationally efficient but also conceptually interpretable, offering transparency rarely found in more complex deep learning systems. This interpretability is a decisive advantage in organizational contexts, where technical outcomes must be communicated across departments with varying analytical literacy. Clear cluster boundaries and descriptive behavioral profiles enable marketing, operations, and IT teams to collaborate effectively under a shared analytical language. Consequently, clustering functions as an integrative mechanism that bridges technical insight and business execution, ensuring that analytical outputs are actionable and strategically aligned.

Furthermore, the transparency and adaptability of these clustering methods encourage their incorporation into enterprise governance and compliance structures. Decision-makers can audit how clusters are formed, interpret the reasoning behind segmentation boundaries, and adjust model parameters to align with ethical or regulatory considerations, such as fairness in user treatment or data privacy compliance. This balance of analytical power and explainability makes clustering particularly suitable for sectors undergoing intensive digitalization—such as finance, healthcare, and manufacturing—where algorithmic accountability is paramount. Equally important is the feedback-driven evolution of the clustering framework within the proposed Precision Operation Framework (POF). As user data streams continuously into enterprise systems, each iteration of clustering refines operational intelligence, enabling the system to evolve in tandem with behavioral realities. This dynamic adaptability ensures long-term model relevance, preventing degradation due to concept drift—a common issue in rapidly changing digital ecosystems. Over time, clustering outcomes can feed into higher-order predictive systems, serving as priors for machine learning models that forecast demand, churn, or sentiment. In this way, the clustering mechanism becomes not a one-time analytical exercise but an ongoing infrastructure for enterprise cognition, continuously supporting digital strategy formulation and execution.

IV. CONCLUSION

The rapid pace of digital transformation has fundamentally altered how enterprises operate, demanding data-driven decision systems that are both adaptive and precise.

The implementation of K-means and DBSCAN algorithms provided a comprehensive analytical framework for detecting both stable user patterns and anomalous behaviors. The experimental results validated the superior interpretability and operational value of clustering-based analytics: the system improved marketing precision by 22.1%, reduced resource redundancy by 19.3%, and increased the overall Operational Precision Index (OPI) by 27.5%. Moreover, the hybrid approach enhanced the detection of irregular activities, such as low-frequency engagement or atypical

transaction spikes, reinforcing enterprise capacity for risk mitigation. The Precision Operation Framework (POF) developed in this research integrates clustering insights into the daily workflow of digital enterprises. Through its layered architecture—data acquisition, clustering analysis, insight extraction, and operational decision—it ensures that behavioral understanding translates directly into optimized business actions. The framework supports continuous learning, enabling real-time updates as user behavior evolves. This adaptability is essential in the contemporary digital ecosystem, where customer interactions are fluid and multidimensional.

From a managerial perspective, this research emphasizes three strategic outcomes. First, clustering-based behavior analysis fosters data-driven operational intelligence, transforming decision-making from reactive to proactive. Second, it enhances resource allocation efficiency, ensuring that marketing, support, and development efforts are aligned with real behavioral segments. Third, it advances user-centric precision, allowing enterprises to deliver customized experiences that build loyalty and long-term value.

Although the data used in this study was collected in 2017–2018, earlier than the explosive growth of short videos, live-streaming e-commerce, and generative AI, the core methodology proposed in this paper— an unsupervised clustering framework based on behavioral similarity—does not rely on specific channels or interaction modalities. Instead, it focuses on the structural characteristics of user behavior (such as frequency, diversity, and relevance), dimensions that still possess universal representational capabilities in the current digital ecosystem. In fact, this framework was reproduced in an A/B test for a fast-fashion brand in 2023, achieving a 21.8% improvement in OPI using new data including live-streaming and AI customer service interactions, validating its transferability and robustness in dynamic environments.

AUTHOR CONTRIBUTIONS

Conceptualization –Jiguang Zhang, Mi Huang, Liang Jing, Liyang Jiao and Xiaocheng Li; methodology – Jiguang Zhang, Mi Huang, Liang Jing, Liyang Jiao and Xiaocheng Li; investigation – Jiguang Zhang, Mi Huang, Liang Jing and Liyang Jiao; writing-original draft preparation – Jiguang Zhang, Mi Huang, Liang Jing, Liyang Jiao and Xiaocheng Li. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Y. Ni, H. Gu, C. Ou, et al., “The impact of digital finance on the digital transformation of cultural industry enterprises,” *Finance Research Letters*, vol. 85, Part A, 2025, doi: 10.1016/j.frl.2025.107849.
- [2] K. Szymczyk and Ágnes Csiszárík-Kocsir, “Digital Transformation of Micro-Enterprises in the Light of the Covid-19 Pandemic,” *International Conference on Information Systems Development*, 2024, doi: 10.62036/isd.2024.42.
- [3] A. Ilina L, A. Pavlov A, and S. Pavlova K, “Agro-Industrial Enterprises in the Realities of the Digital Transformation,” *Lecture Notes in Civil Engineering*, pp. 32-42, 2024, doi: 10.1007/978-3-031-67372-6_5.
- [4] H. Zhong and M. Jing, “Digital soft decoration display design microspecialty based on double layer clustering user behavior analysis algorithm,” *INTELLIGENT DECISION TECHNOLOGIES-NETHERLANDS*, vol. 19, no. 3, pp. 1551-1564, 2025, doi: 10.1177/18724981251314270.
- [5] M. Gangwar, “Review of: “Application of Data Mining Combined with K-means Clustering AI-gorithm in Enterprises’ Risk Audit”,” 2024, doi: 10.32388/rx5ezk.
- [6] A. Rana D S U, “Application of Data Mining Combined with K-means Clustering Algorithm in Enterprises’ Risk Audit,” *Qeios*, 2024, doi: 10.32388/g9g0s3.
- [7] S. Yang and Z. Kan, “Fed-BNGC: A Federated Clustering Algorithm to Solve the Problem of User Heterogeneity,” *SAE International*, 2025, doi: 10.4271/2025-99-0110.
- [8] X. Cao, “Research on the Application of Clustering Algorithm in the Analysis of College English Achievement,” *Advances in Intelligent Systems Research*, pp. 393-401, 2024, doi: 10.2991/978-94-6463-417-4_36.
- [9] L. Yu, “The Application of K-means Clustering Algorithm in the Evaluation of E-Commerce Websites,” *Journal of Electrical Systems*, vol. 20, no. 6s, pp. 759-769, 2024, doi: 10.52783/jes.2738.
- [10] S. Sun, “Fuzzy Clustering Algorithm for Trend Prediction of The Digital Currency Market,” *Salud, Ciencia y Tecnología – Serie de Conferencias*, vol. 3, pp. 1094, 2024, doi: 10.56294/sctconf20241094.
- [11] W. Cai, F. Yang, B. Yao, et al., “An adaptive k-means clustering algorithm based on grid and domain centroid weights for digital twins in the context of digital transformation,” *Journal of Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01180-z.
- [12] X. Chen, C. Chen, and L. Yang, “Research and Application of Clustering Algorithm in the Behavior Analysis of Key Electric Power Customers,” *2024 Global Conference on Communications and Information Technologies (GCCIT)*, pp. 1-5, 2024, doi: 10.1109/gccit63234.2024.10862169.