Prediction of Installment Default Risk of Financial Products Based on Consumers' Buy-Now-Pay-Later Behavioral Characteristics

Ruiyang Hong

Acabridge College, Shanghai 200000, China
hongruiyang913@gmail.com

Abstract. With the widespread adoption of buy-now-pay-later (BNPL) in the consumption of high-priced goods, installment payments have not only enhanced the accessibility of consumption but also increased the exposure to default risks. Moreover, a single traditional indicator is no longer sufficient to accurately depict the heterogeneity of risks. This article focuses on the installment payment scenarios for high-priced goods, using 7,380 consumer transaction data as a sample. By combining K-means clustering and Logistic regression, it conducts a study on default risk prediction from the perspective of behavior. The research classifies consumers into three categories based on their consumption behavior characteristics. These are: the high-income, high-brand preference, high-consumption group (Cluster 1), the low-income, low-brand preference, low-consumption group (Cluster 0), and the group with balanced characteristics in the middle (Cluster 2). The results show that there are significant differences in risk among the three groups. The loss rate of Cluster 1 is the highest (0.22), followed by Cluster 0 (0.17), and Cluster 2 has the lowest rate (0.10). High-spending customer segments do not necessarily represent the lowest-risk categories. This study reveals the correlation between consumption behavior patterns and installment default risk, thereby refining the risk identification framework for installment financing. It provides empirical support for financial institutions to implement tiered risk management and differentiated risk control strategies.

Keywords: BNPL, Consumer behavior, Credit risk prediction, K-means clustering, Logistic regression.

1 Introduction

Over the last few years, the accelerated growth of installment payment solutions, including BNPL, within high purchase programs involving consumers, has precipitated a distinguished shift in the consumer payment decision-branching processes. In comparison with the conventional one-time paying scheme, installment purchasing does not only disperse the pay weight but also reduces the incumbency against the present cash spending. In some ways, it increases the availability of goods of high prices to consumers and the fast growth of online retail and consumer finance markets. Nevertheless, although this form of payment lessens temporary liquidity pressure, it

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_128

can also raise the effect of huge, multi-period debt on the stability of cash flow of consumers and their ability to repay it to (or continue to) pay their financial obligations, which enhances the likelihood that economic risks may be experienced.

If this were under installment payment, consumers would no longer make decisions based on their level of income and already have credit or not. Rather, in most cases, they respond to their consumption behavior as having been affected by a combination of various factors including sensitivity to price, preference towards a given brand, product positioning, and subjective risk perception at the same time. They imply that a variety of consumer behaviour, having been presented to the same installments products, will vary largely regarding the repayment trajectories and risk-related results. Here, one demographic trait or fixed credit parameter has not been effective in creating a complete risk profile of consumers in installment financial product. Consequently, this has complicated the question of risk identification and made it more heterogenous [1].

It has been demonstrated in previous research that the use of traditional statistical models including logistic regression is more stable and interpretable in credit risk prediction, having been utilized considerably in personal credit segments and consumer finance sectors [2, 3]. As the data collection ability has been enhanced, researchers have progressively introduced the variables of transaction behavior, consumption history, and machine learning algorithms into the discipline of risk modeling, so as to offset the flaws of the classical models in terms of nonlinear associations, and interpersonal differences. Related research has established that real behavior information like the frequency of purchase, the distribution of consumption amount, choice of prices and brand loyalty can be used to explain default risk incrementally on top of demographic factors.

This aspect stands out well in the case of BNPL installment payment. Installing tools on the one hand affects the perception consumers have on price and debt, they postpone form of payments, which causes more aggressive consumption in decisions to group behaviourally homogeneous, that exposes greater risk of agglomerating and cumulative effect on household financial strains. Hence, determining the risk variations of consumers under installment conditions in behavioral terms has become a significant trend in credit risk study [4,5].

Even though the related research has been continuously enriched, some of the limitations on existing literature of the segmented scenario of installment purchase of high-priced goods can be identified to possess high leverage properties and high budget constraints. The vast majority of studies are aimed at risk prediction on an overall sample level, as well as very few consumer behavior heterogeneity are characterized. Behavioral variables, on the other hand, may be factored as such directly into regression equations as independent explainers. and systematic analysis of the risk structure in terms of various combinations of behavior is less systematically considered in the light of consumer segmentation. The interaction of the behavioral factors in determining different user groups and in addition associating installments default risks is yet to be further examined.

On this basis, the installation purchase of high-priced goods scenario is the object of the research presented in this article providing a background of consumer transaction

data and installment behavior data. The first one is that user groups sharing common characteristics of consumption behavior are defined by clustering techniques. On the basis of this, the logistic regression model is conducted to examine the differences in risks between various behavior groups. The article provides empirical reference to risk-stratification management as well as the differentiated risk control strategies of installment financial products based on the view of behavioral heterogeneity in complementing the research framework to identify the risk of installment financial products.

2 Method

2.1 Data Source and Explanation

This paper uses a set of online retail and installment payment transaction data as a research sample, which includes a total of 7,380 consumer transaction records. Each record corresponds to a single installment purchase behavior of a consumer within a specific time window, and the sample unit is at the "consumer - transaction" level.

The variables included in the sample can be classified into three categories. The first category consists of consumer basic characteristic variables, which are mainly used to describe the economic background of consumers, such as indicators like average income level. The second category consists of consumption and behavioral characteristic variables, including brand affinity, price selection level, price sensitivity, and aesthetic orientation, etc. These are used to reflect the decision preferences and behavioral patterns of consumers in the scenario of purchasing high-priced goods on an installment basis. The third category consists of risk outcome variables. In this paper, the churn rate is adopted as the proxy indicator for default risk, which is used to measure the potential risk levels of different consumers in the installment scenario. The setting of these variables provides a basis for subsequent consumer clustering analysis and the comparison of default risk differences.

2.2 Variable Selection

Based on the research objective, this paper constructs a variable system from the perspective of consumer behavior, dividing the research variables into three categories: dependent variables, core behavioral independent variables, and control variables, in order to systematically depict the risk differences in the scenario of purchasing high-priced goods through installment payments.

This article uses the churn rate as a proxy variable for default risk, to measure the potential risk level of consumers in the installment payment scenario. This indicator is widely used in the research of consumer finance and customer management to describe users' tendencies towards default, withdrawal or risk exposure. Especially in cases where it is impossible to directly observe the long-term actual default results, it is regarded as a reasonable and operational risk proxy indicator [6,7]. Therefore, the churn rate is applicable to the analysis of consumer risk heterogeneity in this article as well as the comparative study of risk levels among different market segments.

The core independent variables mainly describe the behavioral characteristics of consumers in the context of installment purchase of high-priced goods, including indicators such as average income level, brand affinity, price selection level (such as product grade), price sensitivity, quality orientation and aesthetic orientation. These variables respectively reflect the installment payment behavior patterns of consumers from dimensions such as economic capacity, consumption preferences and decision-making orientations. Existing research has shown that behavioral variables such as purchase intensity, price selection, and brand preference can provide incremental explanatory information beyond traditional demographic characteristics in credit risk analysis, helping to reveal risk differences among different consumer groups [3,4,8].

In the process of variable screening, this paper combines correlation analysis (Pearson and Spearman coefficients) and one-way analysis of variance (ANOVA) to conduct a preliminary test on the statistical correlation between behavioral variables and risk outcomes. This is done to eliminate variables that are not significantly associated with default risk, thereby ensuring explanatory power while avoiding model complexity.

To control the interference of individual background differences on risk outcomes, this paper introduces necessary control variables in the model analysis, including basic characteristics such as the consumer's region. Regional variables are categorized based on the level of urban economic development, dividing the sample into first-tier, new first-tier, second tier, and other cities, in order to depict the systematic differences in staging risks under different regional economic environments. Such demographic and background variables are commonly regarded as basic control variables in credit risk research [2,9].

Before modeling, this paper carried out unified numeralization and coding processing for the variables. Numerical variables were directly used for analysis and were standardized in subsequent steps to eliminate the influence of dimensional differences on model estimation. For categorical behavior variables, depending on whether they have an inherent sequential relationship, either label encoding (Label Encoding) or one-hot encoding (One-hot Encoding) is used for transformation to meet the requirements of clustering analysis and Logistic regression models for numerical input. The above-mentioned coding method has a wide range of applications in credit risk modeling and consumer behavior research, which helps to enhance the operability and interpretability of the model while ensuring the preservation of information [5,10].

3 Results

3.1 Consumer Segmentation Using K-Means and PCA

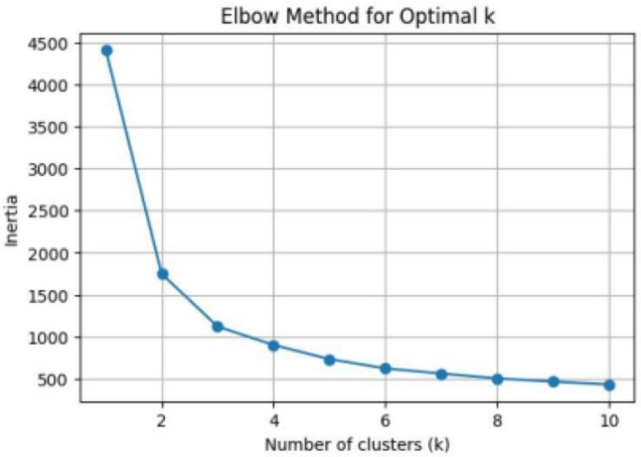


Fig. 1. Elbow method for optimal k

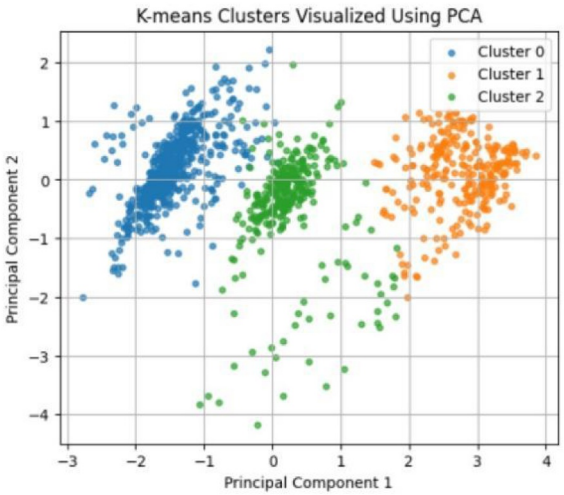


Fig. 2. K-means clusters visualized using PCA

Based on the behavioral characteristics of consumers (including income level, brand affinity, average product price and evaluation score), this paper uses the K-means clustering method to group users by their behaviors. Using the Elbow Method (see Figure 1), it can be observed that when the number of clusters k = 3, the decrease in within-group sum of squares begins to slow down significantly, thus three clusters are selected as the optimal clustering result. Based on this, the standardized behavioral

variables were clustered, and principal component analysis (PCA) was utilized to reduce the high-dimensional features to a two-dimensional space for visualization (see Figure 2). The results show that different consumer groups exhibit relatively distinct distribution ranges in the principal component space, indicating structural differences in consumer behavior dimensions such as consumption intensity, price selection, and brand preference. This result indicates that the clustering method based on behavioral variables can effectively identify user groups with different consumption patterns, providing a stratified foundation for subsequent risk difference analysis.

3.2 Behavioral Differences Across Consumer Segments

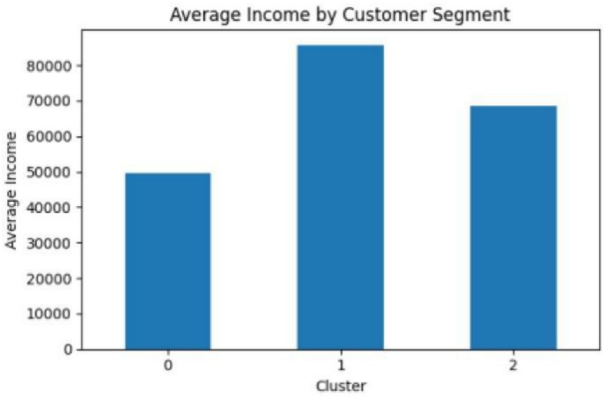


Fig. 3. Average income by customer segment

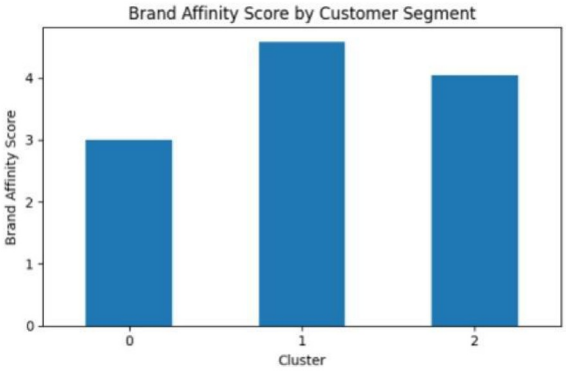


Fig. 4. Brand affinity score by customer segment

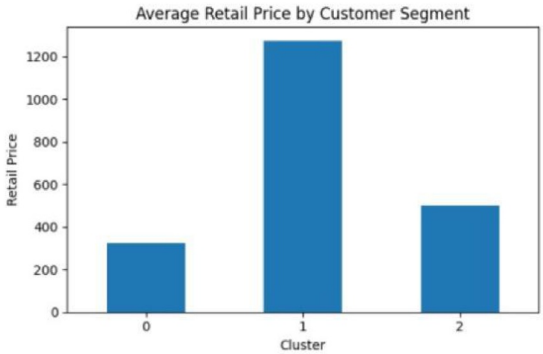


Fig. 5. Average retail price by customer segment

In order to further portray the differences in key behavioral characteristics of different consumer subdivisions, this paper makes a comparative analysis of the three core dimensions of each cluster in terms of income level, brand affinity and commodity price selection (see Figure 3 - Figure 5). The results show that the three types of consumers show a highly consistent hierarchical structure in the above behavioral characteristics. Cluster 1 is significantly higher than other groups in terms of average income, brand affinity score and average purchase price, showing obvious high-value consumption characteristics; Cluster 0 is at the lowest level in all three indicators, reflecting stronger price sensitivity and lower consumption levels; Cluster 2 falls between the two in terms of various indicators and exhibits a relatively balanced consumption behavior pattern. The consistency stratification results indicate that the clustering method not only distinguishes different consumer groups but also forms stable and interpretable differences in multiple behavioral dimensions.

3.3 Risk Differences Across Segments (Churn as Proxy for Default Risk)

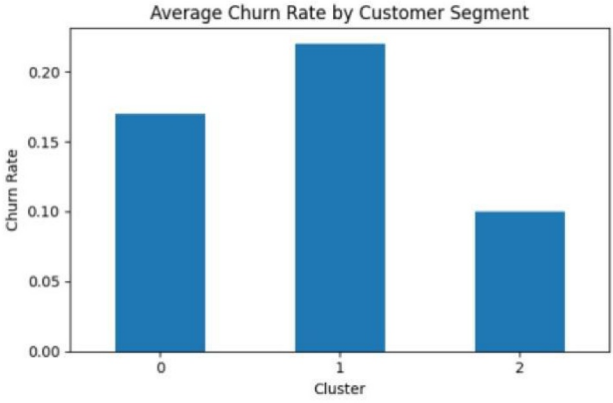


Fig. 6. Average churn rate by customer segment

To compare the differences in risk outcomes among different consumer segments, this article further calculated and compared the average churn rate of each cluster and used it as a proxy indicator for installment default risk (see Figure 6). The results show that there are significant differences in risk performance among different consumer groups. Among them, Cluster 1 has the highest average churn rate (about 0.22), followed by Cluster 0 (about 0.17), while Cluster 2 has the lowest churn rate (about 0.10). This result indicates that there is significant heterogeneity in the level of installment default risk among different consumer segments. Even among user groups with higher overall consumption ability or brand preference, the risk level is not necessarily lower. Consumer segmentation helps identify risk differences that are difficult to distinguish by traditional indicators.

3.4 Summary of Segment Profiles

Table 1. Summary of segment profiles

cluster_label	avg_income	brand_affinity_score	retail_price	review_star	churn_rate	num_customers	share_%
0	49701.78	2.99	325.19	3.97	0.17	542	49.1
1	85711.23	4.58	1275.05	4.68	0.22	261	23.7
2	68384.50	4.04	501.74	4.10	0.10	300	27.2

Table 1 summarizes the overall characteristics of three types of consumers in terms of income level, brand preference, price choice, and risk indicators, clearly presenting significant differences in consumption behavior and risk performance among different consumer groups. Among them, Cluster 0 has the lowest average income level (49701), and its brand affinity and consumer price are both in a low range, indicating that this group is more sensitive to prices and their consumption decisions are more cost constrained. Its corresponding churn rate is 0.17, reflecting a certain degree of repayment pressure and moderate risk level.

In contrast, Cluster 1 has the highest average income and brand affinity (85,711; 4.58), and its average product price is significantly higher than that of other groups (1,275), indicating strong purchasing power and brand preference. However, the attrition rate of this group was also the highest (0.22), indicating that in the context of high-priced installment purchases, higher consumption intensity may be accompanied by greater exposure to installment risks. Cluster 2 is in the middle range in terms of income level, consumption prices and brand preferences. Its churn rate is the lowest (0.10), indicating relatively stable consumption behavior and a lower risk level. Overall, the differences in behavioral characteristics among different consumer groups are clearly correlated with their risk profiles.

4 Discussion

The consumer segmentation results show that in the scenario of installment purchase of high-priced goods, consumers exhibit a clear hierarchical structure in behavioral characteristics such as income level, brand affinity, price selection, and preference orientation. This result indicates that installment consumption behavior is not solely determined by a single factor of economic ability but is the result of the combined action of multidimensional behavioral characteristics. Introducing behavioral variables such as brand preference and price orientation can help to more accurately depict consumers' differentiated behavior patterns in installment decisions. This finding is consistent with regulatory observations on the complexity of BNPL usage behavior, where installment risk is often nested within consumers' actual behavioral combinations, rather than being fully reflected by a single income or credit indicator [9].

Combining the risk results, it can be found that there are significant differences in risk levels among different consumer segments, and these differences are not completely linearly related to income levels. Although some high-value consumer groups have strong economic capabilities, they may face greater short-term cash flow pressure due to higher installment amounts or frequency of installment usage, resulting in a higher level of risk. In contrast, groups whose consumption structure is more in line with their economic ability have relatively more stable risk performance. This phenomenon indicates that in the BNPL scenario, risk exposure is more likely to come from installment usage and liability structure, rather than simply differences in payment ability. This conclusion is consistent with regulatory research on the risk characteristics of the BNPL market [9].

Furthermore, the results of this study indicate that behavioral variables have important supplementary value in identifying installment financial risks. Compared to traditional demographic characteristics, behavioral variables can more directly reflect consumers' decision-making styles and risk exposure paths in staged scenarios. Combining behavioral variables with consumer segmentation methods can help identify risk differences at the level of behavioral combinations, thereby enhancing the refinement and interpretability of risk identification. Previous studies have also pointed out in the systematic review of BNPL regulation and policy evolution that in the future, risk management and regulatory focus should be more on consumer behavior patterns and the degree of risk concentration among different groups, rather than relying solely on single transactions or static characteristic indicators [10].

5 Conclusion

This article takes the scenario of installment purchase of high-priced goods as the research background, and based on consumer transaction and behavioral characteristic data, constructs a risk identification and analysis framework from a behavioral perspective. By combining cluster analysis and logistic regression methods, this article identified risk differences under different consumer behavior patterns. The research results indicate that consumers exhibit a clear hierarchical structure in terms of their

behavioral characteristics such as income level, brand preference, and price selection. Different combinations of behaviors correspond to significantly different levels of risk, suggesting that behavioral variables have important supplementary value in the identification of installment financial risks.

Although this article has certain exploratory significance in terms of its analytical perspective and method design, it still has several research limitations. Firstly, the research was conducted based on single-period cross-sectional transaction data, failing to depict the dynamic process of consumers' installment risks over time. Therefore, it was difficult to reflect the long-term repayment behavior and the mechanism of risk accumulation. Secondly, in this paper, the churn rate is used as an indicator to represent the risk of default. It is mainly employed to identify the relative risk differences among different consumer groups, but it cannot directly observe actual default events or individual default probabilities. Therefore, these conclusions are more applicable in the context of risk stratification and heterogeneity analysis. However, there are still applicable boundaries in the context of credit scoring applications that are oriented towards prediction. Furthermore, the behavioral variables employed in this study mainly come from observable characteristics at the transaction level. They have not yet incorporated aspects such as consumers' financial status or macroeconomic factors, which may to some extent limit the comprehensiveness of the risk explanation.

Future research can further expand upon the analytical framework presented in this paper. On the one hand, with the availability of multi-period installment transactions and actual default data, subsequent research can incorporate longitudinal data to depict the dynamic evolution path of consumer risks. On the other hand, while retaining the explanatory advantage of behavioral variables, a more complex predictive model can be combined to expand the application value of behavioral characteristics in phased financial risk management. At the same time, it is necessary to further combine the issues of BNPL multi-platform stacking and information disclosure emphasized in regulatory reports, and supplement longer-term and cross-platform liability data to test the extrapolation of the behavioral clustering results in the real market environment.

References

1. Di Maggio, M., Kermani, A., & Majlesi, K.: Buy now, pay later and household finance. National Bureau of Economic Research Working Paper No. 30010 (2022)
2. Silva, W. B., Meurer, A. C., & Cunha, D. P.: Credit risk evaluation using logistic regression and artificial intelligence techniques. Journal of Applied Statistics **47**(13–15), 2879–2894 (2020)
3. Hand, D. J., & Henley, W. E.: Statistical classification methods in consumer credit scoring: A review. Journal of the Royal Statistical Society: Series A (Statistics in Society) **160**(3), 523–541 (1997)
4. Vasuki, M., Celestin, M., Kumar, A. D., & Asamoah, P. J.: Evaluating credit scoring models with statistical regression techniques in consumer finance. Brainae Journal of Business, Sciences and Technology **9**(2), 180–190 (2025)
5. Bellotti, T., & Crook, J.: Support vector machines for credit scoring and discovery of significant features. Expert Systems with Applications **36**(2), 3302–3308 (2009)

6. Donou-Adonsou, F., & Leslie-Piper, N.: BNPL and financial fragility in U.S. households. Finance Research Letters **86**, 108423 (2025)
7. Khandani, A. E., Kim, A. J., & Lo, A. W.: Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance **34**(11), 2767–2787 (2010)
8. Wu, Y., Zhang, Y., Lu, T., Xu, Y., & Sun, Y.: Consumer behavior and credit default: An empirical analysis based on self-regulation theory. Journal of Economic Management **2**(2), 97–123 (2023)
9. Bureau, C. F. P.: Buy now, pay later: market trends and consumer impacts (2022)
10. Krause, D.: Buy Now, Pay Later: An Analysis of Consumer Financial Protection Bureau Research and Regulatory Actions (2025)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

