International Business & Economics Studies
ISSN 2640-9852 (Print) ISSN 2640-9860 (Online)
Vol. 8, No. 3, 2026
www.scholink.org/ojs/index.php/ibes

Original Paper

Data-Empowered Multi-level Weighted Rule Evaluation for Customer Classification in the Tobacco Industry—A Foundation for Market Status Identification and Predictive Regulation

SHEN Xiaoxia[1], XUE Jianbo[1], WANG Ting[1], LI Rongxiang[1], & WANG Kaixuan[1]

[1] Liangshan Prefecture Branch of Sichuan Provincial Tobacco Company, Liangshan, Sichuan, 615000, China

Fund Project

Science and Technology Project of Sichuan Provincial Tobacco Company, "Research and Application of Data-Empowered Precise Identification and Predictive Regulation Mechanism for Cigarette Market Status" (Project No. SCYC202428).

Received: April 26, 2025 Accepted: June 15, 2026 Online Published: July 8, 2026

doi:10.22158/ibes.v8n3p221 URL: http://dx.doi.org/10.22158/ibes.v8n3p221

Abstract

Changes in consumer spending patterns have a significant impact on the market, and grasping consumption trends is crucial to the development of enterprises. Taking Liangshan tobacco marketing as the research object, this paper analyzes retail business data to classify retailers into different types, capturing changes in cigarette consumption to support tobacco marketing decisions. By establishing a retailer quality scoring model, retailers are evaluated from four dimensions: basic compliance, data accuracy, sales behavior normativity, and initiative, ultimately forming a comprehensive evaluation of retailers. Based on the evaluation results, retailers are classified into four categories: Benchmark, Stable, Follower, and Lagging. This classification provides an effective evaluation standard for subsequent marketing decisions, enables efficient resource allocation, and encourages retailers to actively respond to market changes, thereby enhancing customer satisfaction.

Keywords

Retailer Classification, Quality Scoring, Marketing Decision-making, Customer Satisfaction

1. Introduction

The trend of consumption diversification drives enterprises to create diverse consumption scenarios.

On the one hand, these scenarios satisfy different consumer needs, on the other hand, the creation of such scenarios enhances the core competitiveness of enterprises. The two reinforce each other, enabling enterprises to gain unique market advantages amidst the continuously evolving consumption trends. Digital technology is currently one of the key technologies for meeting these needs and strengthening competitiveness. As a special industry, the tobacco sector, under the growing trend toward health consciousness, urgently needs to grasp the changing trends in cigarette consumption. Guided by market demand, it should optimize the allocation and utilization of resources to cope with future uncertainties.

2. Literature Review

Existing research has made significant progress in customer segmentation and evaluation, primarily focusing on clustering analysis of customer behaviors and characteristics using machine learning algorithms, as well as constructing multi-dimensional evaluation systems. Zhao et al. (2024) employed the K-means algorithm to segment customers into different groups. Barrera et al. (2023) developed a multi-criteria decision support system for customer segmentation based on the sorting transcendence method. Wang et al. (2022) constructed a customer profile dimension recognition model. Zhou et al. (2021) obtained web content data and employed data mining for customer classification. Hashemian et al. (2024) studied e-book platforms, achieving increased subscription sales through customer segmentation. Cao (2024) and Mokashi (2024) highlighted the value of machine learning in customer segmentation and compared the similarities and differences among various machine learning algorithms.

Although corresponding algorithms have achieved notable progress in customer segmentation, there remains a gap in their direct applicability to scenarios with mandatory requirements for data authenticity and compliance. Therefore, based on the specific customer classification needs of the tobacco industry, a multi-level classification approach that combines quantitative data (such as scan days, sales-to-purchase ratio, deviation rate, inventory accuracy) with qualitative evaluations (such as willingness to cooperate, self-calibration behavior) is essential, integrating objective indicators with "soft judgments" to form a highly operable customer classification framework.

3. Model Construction

With the core objective of "optimizing resource allocation through precise classification to better serve consumers," this study constructs a management principle system for customer classification in the tobacco industry, forming a "Goal—Principle—Indicator" transmission chain that provides the policy foundation and logical starting point for establishing a multi-level weighted evaluation model.

3.1 Indicator Definition

Based on the actual business conditions of tobacco retailers, the classification indicators are developed as shown in Table 1.

Table 1. Classification Evaluation Indicators

Indicator	Symbol	Unit/Range	Ideal Direction	Description
Price Collection Retailer	x_1	{0,1}	—	Used for marking only, not directly participating in classification
Basic Sample Retailer	x_2	{0,1}	1	Included in sample management
System Maintained	x_3	{0,1}	1	System status
Customer Segment (Monthly)	x_4	{L,M,S}	—	Segment classification, does not affect individual quality
Inventory Accuracy	x_5	[0,100]%	High	Single-specification accuracy ≥95% qualifies
Scan Days	x_6	Days	≥27 (non-Chinese New Year) / ≥15 (Chinese New Year)	Elimination if <27 for two consecutive months
Sales-to-Purchase Ratio (140 days)	x_7	[0,∞)	Within standard interval	Elimination if outside interval for two consecutive months
System Sales Deviation Rate	x_8	[0,100]%	Low (<5%)	Reflects inventory tampering behavior
Negative Inventory Spec Count	x_9	Non-negative integer	0	Negative inventory indicates non-compliance
Cigarette Scan Transaction Count	x_{10}	Positive integer	Reasonable	Judged in conjunction with sales volume
Total Cigarette Sales Volume	x_{11}	Positive integer	Reasonable	—
Average Sales per Transaction	x_{12}	Cartons/transaction	1–2 cartons	Too high or too low raises suspicion
Concentrated Scan Transaction Count	x_{13}	Non-negative integer	Small	Large number of scans within a short period

Concentrated Scan Ratio	x_{14}	[0,100]%	Low (<10%)	Concentrated transactions/total transactions
Carton Sales Ratio	x_{15}	[0,100]%	Normal (30%~70%)	Abnormal if too high or too low
Monthly Inventory-to-Sales Ratio	x_{16}	[0,∞)	Reasonable (1~3)	Abnormal indicates inaccurate inventory

3.2 Customer Quality Scoring Model

A comprehensive score is constructed through weighted summation across four dimensions:

$$S=w_1 \cdot Q_{\text{Compliance}}+w_2 \cdot Q_{\text{Accuracy}}+w_3 \cdot Q_{\text{Behavior}}+w_4 \cdot Q_{\text{Initiative}}$$

Based on classification management requirements, the comprehensive score is divided into evaluation intervals as shown in Table 2.

Table 2. Comprehensive Score Evaluation Intervals

Category	Score Range	Additional Conditions	Label
Benchmark	$S\geq 90$	Inventory accuracy ≥99%, no negative inventory, deviation rate <5%, scan days full attendance	Fully autonomous, scan-on-sale, real-price modification, completely consistent data
Stable	75 \le S < 90	Inventory accuracy ≥95%, deviation rate <10%, scan days ≥27	Proactive cooperation, occasional omissions, minimal deviation
Follower	$60\leq S<75$	Does not meet Stable criteria but no hard elimination triggers	Requires supervision, significant data deviation
Lagging	$S<60$ or triggers R1–R5	—	Resistant and perfunctory, data has no value

3.2.1 Basic Compliance Score $Q_{\text{Compliance}}$

The Basic Compliance Score reflects the extent to which a retailer adheres to the relevant management regulations for tobacco retail. It is primarily composed of the scan days score and the sales-to-purchase ratio score, and the smaller of these two scores is taken as the Basic Compliance Score. The corresponding models are specified as follows.

(1) Scan Days Score:

$$S_{\text{days}}=min\left(100,\frac{x_6}{27}\times 100\right)$$

Where

S_{days}: The score derived from the number of days on which consumers scan the marketing-specific QR code during cigarette purchases in the retail process.

x_6: It refers to the scan days for months other than the Spring Festival month. If the statistical month is a Spring Festival month (e.g., January or February), the denominator 27 is replaced by 15.

(2) Sales-to-Purchase Ratio Score:

First define the standard interval for the sales-to-purchase ratio:

$$L=\frac{0.8}{max(1,r)},U=\frac{1.2}{min(1,r)},r=\frac{\text{Average daily wholesale sales over 140 days}}{\text{Average daily wholesale sales over 365 days}}$$

The sales-to-purchase ratio score is defined as a piecewise function:

$$S_{\text{ratio}}=\begin{cases}100, & x_7\in[L,U]\\ max(0,100\text{-}200\times min(|x_7\text{-}L|,|x_7\text{-}U|)), & x_7\notin[L,U]\end{cases}$$

(3) Basic Compliance Comprehensive Score:

$$Q_{\text{Compliance}}=min(S_{\text{days}},S_{\text{ratio}})$$

3.2.2 Data Accuracy Score Q_{Accuracy}

The Data Accuracy Score reflects the authenticity of the data. It is composed of three components: inventory accuracy, system sales deviation rate, and a penalty for negative inventory. The final Data Accuracy Score is then calculated by integrating these three items.

(1) Indicator Score Definitions:

Inventory Accuracy Score:

$$S_{\text{stock}}=x_5$$

System Sales Deviation Rate Score:

$$S_{\text{bias}}=max(0,100\text{-}2\times x_8)$$

Negative Inventory Penalty:

$$P_{\text{neg}}=min(10\times x_9,100)$$

Each negative inventory specification incurs a 10-point deduction, up to a maximum of 100 points.

If x_9=0, then P_{neg}=0, and the score is simply the average of the first two items; if the deduction due to excessive negative inventory results in a negative value, the score is set to 0.

(2) Comprehensive Score:

$$Q_{\text{Accuracy}}=max\left(0,\frac{S_{\text{stock}}+S_{\text{bias}}}{2}\text{-}P_{\text{neg}}\right)$$

3.2.3 Sales Behavior Normativity Score Q_{Behavior}

(1) Concentrated Scan Ratio Score:

$$S_{\text{concent}}=max(0,100\text{-}2\times x_{14})$$

(2) Average Sales per Transaction Deviation Score:

Let the reasonable interval be cartons per transaction. If $x_{12}<0.2$, the score is directly set to 0; otherwise:

$$S_{\text{unit}}=\begin{cases}0, & x_{12}<0.2\\ 100, & 0.5\le x_{12}\le 5\\ max(0,100-20\times(x_{12}-5)), & x_{12}>5\\ max(0,100-20\times(0.5-x_{12})), & 0.2\le x_{12}<0.5\end{cases}$$

(3) Carton Sales Ratio Score:

Let the reasonable interval be $[20\%,80\%]$. Each 10 percentage points of deviation from the interval boundary incurs a 20-point deduction:

$$S_{\text{strip}}=\begin{cases}100, & 0.2\le x_{15}\le 0.8\\ max(0,100-200\times(0.2-x_{15})), & x_{15}<0.2\\ max(0,100-200\times(x_{15}-0.8)), & x_{15}>0.8\end{cases}$$

(4) Comprehensive Score:

Take the arithmetic mean of the above three scores:

$$Q_{\text{Behavior}}=\frac{S_{\text{concent}}+S_{\text{unit}}+S_{\text{strip}}}{3}$$

3.2.4 Initiative Score $Q_{\text{Initiative}}$

(1) Base Score and Bonus Items:

$$Q_{\text{Initiative}}=20\cdot x_2+10\cdot x_3+S_{\text{inv}}+20\cdot x_1$$

Parameter descriptions:

- x_2=1: Basic sample retailer, +20 points
- x_3=1: System maintained, +10 points
- x_1=1: Price collection retailer (price modification rate > 50%), +20 points
- S_{inv}: Monthly inventory-to-sales ratio reasonableness score

(2) Monthly Inventory-to-Sales Ratio Reasonableness Score:

The reasonable interval is $[0.5,4]$. Scores within the interval receive 50 points; otherwise, linear deduction applies (each unit deviation from the interval incurs a deduction, ensuring total deduction does not exceed 50):

$$S_{\text{inv}}=\begin{cases}50, & 0.5\le x_{16}\le 4\\ max(0,50-20\times(0.5-x_{16})), & x_{16}<0.5\\ max(0,50-20\times(x_{16}-4)), & x_{16}>4\end{cases}$$

Note: The coefficient 20 is derived from "a deviation of 2.5 units results in a 50-point deduction" (e.g., deviation from 4 to 6.5 incurs a 50-point deduction, equivalent to 20 points per unit).

(3) Total Score Truncation:

The initiative score has a maximum of 100 points:

$Q_{\text{Initiative}}=min(100,20x_2+10x_3+S_{\text{inv}}+20x_1)$

3.3 Hard Elimination Rules

Triggering any of the following rules results in the data collection retailer being classified as "Lagging"

and excluded from comprehensive score calculation.

The elimination rules are defined as logical conditions:

$$\text{Eliminated}=R_1 \vee R_2 \vee R_3 \vee R_4 \vee R_5$$

(1) Definitions of Each Elimination Condition:

R1: Let be the Chinese New Year month indicator (true for January–February, false otherwise):

$$R_1=((\neg F) \wedge (x_6<27)) \vee (F \wedge (x_6<15))$$

and this state persists for two consecutive months.

R2: Let:

$$\begin{cases} L=\dfrac{0.8}{max(1,r)} \\ U=\dfrac{1.2}{min(1,r)} \\ r=\dfrac{\text{Average daily wholesale sales over 140 days}}{\text{Average daily wholesale sales over 365 days}} \end{cases}$$

Then:

$$R_2=(x_7<L \vee x_7>U)\text{for two consecutive months}$$

$$R_3=(x_5<95\%)\text{for two consecutive on-site verifications}$$

$$R_4=(x_8>40\%) \wedge (x_8>30\%\text{for two consecutive months})$$

$$R_5=(x_9>3)\text{persisting for two or more consecutive months}$$

Note: If elimination is triggered, the data collection retailer is directly classified as “Lagging” and excluded from comprehensive score calculation.

4. Validation Framework

4.1 Data Preparation

This study included monitoring data from 168 cigarette retail sample households, covering indicators such as price collection retailer identification, monthly customer segment, inventory accuracy, scan days, sales-to-purchase ratio, negative inventory specification count, cigarette scan transaction count, total cigarette sales volume, carton sales ratio, and monthly inventory-to-sales ratio. The data spans four market types (urban network, town, rural, and agricultural network), with broad customer tier coverage and comprehensive operational indicators, ensuring robust representativeness.

4.2 Validation Indicator Outputs

Table 3. Sample Distribution Overview

Category	Count
Scan days < 27	87
Inventory accuracy < 95%	41

Sales-to-purchase ratio out of range (<0.8 or >1.2)	33
System sales deviation rate > 40%	12
Negative inventory specifications > 3	9

4.2.1 Customer Quality Scoring Model Results

Substituting each retailer's corresponding evaluation indicator values, the dimensional comprehensive scores are computed as shown in Table 4.

Table 4. Sample Dimensional Scores

Dimension	Mean	Standard Deviation	Minimum
Basic Compliance	72.4	28.3	0
Data Accuracy	83.7	18.2	0
Sales Behavior Normativity	91.2	12.4	0
Initiative	34.6	18.9	0

4.2.2 Customer Quality Classification Results

With weights set as w_1=0.3,w_2=0.3,w_3=0.2,w_4=0.2, the calculation results are presented in Table 5.

Table 5. Customer Quality Classification Results

Category	Number of Shops	Percentage	Description
Benchmark	32	11.4%	Comprehensive score ≥90, inventory ≥99%, full scan attendance or only 1–2 days missed, no negative inventory, deviation rate <5%
Stable	78	27.9%	75≤S<90, inventory ≥95%, scan days ≥27, deviation rate <10%
Follower	54	19.3%	60≤S<75, or scan days insufficient or moderate deviation rate without elimination triggers
Lagging	116	41.4%	Triggers any R1–R5 rule (e.g., scan days <27, sales-to-purchase ratio out of range, inventory accuracy <95% consecutive, negative inventory >3, deviation rate >40%)

Through the customer quality scoring model, retailers are classified into four categories: Benchmark, Stable (accounting for 27.9%), Follower, and Lagging. The evaluation results differ only marginally from the empirical judgments of marketing management personnel, demonstrating the practical value of the model. Based on these evaluation results, marketing managers need to adopt effective measures to optimize resource allocation, reduce the proportion of Lagging retailers, and facilitate the transition of retailers toward higher-quality tiers, thereby continuously enhancing consumer satisfaction.

5. Conclusion

This paper, grounded in the five principles of voluntary participation and fairness, merit-based selection with bottom-line constraints, dynamic adjustment with orderly entry and exit, data quality orientation, and classification serving monitoring purposes, constructs a customer classification evaluation model encompassing four dimensions: basic compliance.

References

Barrera, F., Segura, M., & Maroto, C. (2023). *Multiple criteria decision support system for customer segmentation using a sorting outranking method*. Expert Systems with Applications.

Cao, L. (2024). Comparison of Machine Learning Methods in Customer Segment. *Highlights in Science, Engineering and Technology*.

Hashemian, F., Maleki, N., & Zeinali, Y. (2024). From User Behavior to Subscription Sales: An Insight Into E-Book Platform Leveraging Customer Segmentation and A/B Testing. *Services Marketing Quarterly*.

Mokashi, R. (2024). *Enhancing Pattern Classification Accuracy Through Customer Segmentation-Using Machine Learning Algorithms*. SN Computer Science.

Wang, C. (2022). Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach. *Information Processing & Management*.

Zhao, J. (2024). Customer segmentation application based on K-Means. *Applied and Computational Engineering*.

Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services*.