

直播带货下的青年群体购买意愿影响因素数据挖掘研究

——以K市为例

王锴文

云南经济管理学院, 高职学院, 云南昆明

收稿日期: 2024年3月26日; 录用日期: 2024年4月7日; 发布日期: 2024年5月13日

摘要

背景和目的: 以Y省K市青年群体为研究对象, 利用统计问卷和数据挖掘技术, 剖析直播带货下的青年群体购买意愿影响因素。方法: 通过问卷设计致力于将关键因素转化为具体问题, 在问卷发放和数据收集后, 通过聚类分析和机器学习中的特征工程对数据集进行预处理, 通过L2正则化项的逻辑回归模型对不同处理策略下的数据展开模型分析并验证其有效性。结论: 本文通过模型分析得出最优数据预处理策略, 并探究K市网络带货环境下青年群体购买意愿的影响因素。

关键词

青年群体, 网络直播购买意愿, 数据挖掘

Data Mining Research on the Influencing Factors of Young People's Purchase Intention under Live Streaming

—Taking K City as an Example

Kaiwen Wang

Yunnan College of Business Management, Higher Vocational College, Kunming Yunnan

Received: Mar. 26th, 2024; accepted: Apr. 7th, 2024; published: May 13th, 2024

Abstract

Background and Purpose: Taking the youth group in K city of Y province as the research object,

this paper analyzes the influencing factors of the purchase intention of the youth group under the live broadcast with goods by using the statistical questionnaire and data mining technology. Method: Through the design of questionnaires, we are dedicated to transforming key factors into specific questions. After distributing the questionnaires and collecting data, we preprocess the dataset by cluster analysis and feature engineering in machine learning. We then use a logistic regression model with L2 regularization terms to analyze the data under different processing strategies and verify its effectiveness. Conclusion: This paper obtains the optimal data preprocessing strategy through model analysis, and explores the influencing factors of young people's purchase intention under the environment of online shopping in K city.

Keywords

Youth Group, Webcast Purchase Intention, Data Mining

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着数字技术的飞速发展，网络购物方式不断创新。短视频和社交平台的崛起，为消费者提供了全新的在线购物体验。中国产业研究院的数据显示，2022年通过这些渠道进行网购的比例高达72%，同比增长了6%。尤其值得注意的是，直播购物作为一种新兴的网购方式，其用户比例已达66%，同比增长19%，仅次于传统电商平台，成为第二大网络购物方式[1]。针对这一趋势，本研究聚焦于Y省K市的青年群体，旨在探究影响他们参与网络直播购物意愿的因素。

在现有的研究中，张宝生等[2] [3] [4]发现，网络直播可视性、互动性、真实性、娱乐性对消费者的购买意愿产生推动作用，而优质内容、主播个人魅力及信任等因素也对购买决策有正向影响。胡晓丽等[5]利用电商购物历史数据，提出了基于CNN-LSTM的用户购买行为预测模型，实现用户属性、商品属性及用户行为特征的自动抽取与选择；学者李伊林、祝歆等[6] [7]基于历史交易数据，通过融合模型取得了比单一模型更好的预测效果。权李之[8]则基于二元逻辑回归模型研究了“直播+电商”模式下农产品购买意愿的影响因素，发现质量品质、互动交流、价格等因素影响明显。

尽管以上研究为本文提供了宝贵参考，但在直播购物的影响因素研究和分析方法上仍有改进空间：

1) 用户购买预测通常基于历史购物数据，而结合主观影响因素数据的研究不足；2) 数据集预处理方法上，多是采用机器学习中的特征选择；3) 预测模型主要利用算法的融合模型展开，侧重于模型预测效果的对比，而非借助传统模型对直播购物影响因素的分析；4) 少数研究仅从消费者基础信息及主观购买意愿态度调查入手，缺乏其他客观影响因素的调查分析。对此，本文将从以下两方面进行创新与补充：1) 通过问卷设计细化并分解主观和客观影响因素，利用聚类分析和特征选择相结合的技术进行数据预处理，最大化挖掘数据中的潜在信息和特征。2) 采用L2正则化的逻辑回归模型对处理后数据进行分析，既可保证模型稳定性、降低复杂度，还能通过对比不同数据预处理效果来确定最优预处理策略；同时，模型系数将用于探讨青年群体网络直播购买意愿的影响因素，为相关领域研究提供新的视角与实证参考。

2. 理论框架

本研究在综合现有文献的基础上，构建了一个理论框架，用以分析Y省K市青年群体参与网络直播

购物意愿的影响因素。研究的核心在于将理论与实证数据相结合，通过以下步骤实现研究目标：

1) 问卷设计：在问卷设计阶段，将关键分析要素转化为具体问题，尽可能获取影响青年群体购买意愿的变量。2) 数据收集与处理：在问卷发放和数据收集完成后，采用聚类分析和特征工程对数据集进行预处理，旨在筛选出对购买意愿有显著影响的特征。3) 模型建立与验证：利用处理后数据，构建逻辑回归模型，并通过 L2 正则化来避免过拟合，确保模型的泛化能力，并验证模型的有效性。4) 核心问题解决：i) 运用聚类分析及结合特征选择的方法对数据进行预处理，建立影响购买意愿的决策模型，并对模型进行分析与评估，以识别最有效的数据预处理策略。ii) 结合问卷调查数据和模型分析结果，探讨直播带货环境下青年群体购买意愿的影响因素。通过上述步骤，本文形成了一个完整的研究思路，具体见图 1 所示。

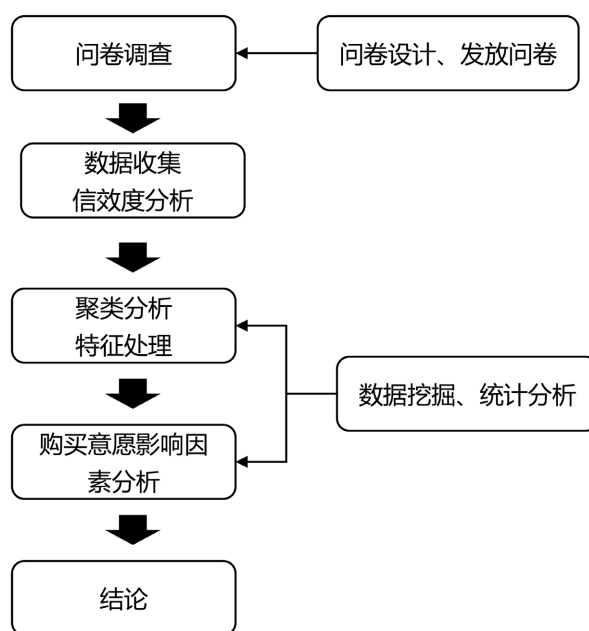


Figure 1. Theoretical framework
图 1. 理论框架

3. 研究方法

3.1. 购买意愿的影响因素

网络直播购物作为一种新兴的电子商务模式，其购买意愿的形成是一个涉及多方面因素的综合过程。本文通过对相关文献的梳理与分析，将影响网络直播购买意愿的因素归纳为以下几个关键点：1) 消费者特征，包括消费者的基本信息，如性别、年龄、收入水平等。2) 接触网络直播程度，如观看网络直播频率、时长等，反映了消费者对直播内容的熟悉与参与度。3) 信任与风险，消费者在直播体验中感知到的信任程度和潜在风险，如对主播的可信度、商品质量的担忧等。4) 网络直播体验，直播的便利性、互动性、氛围体验等都直接影响消费者的观看和购买体验，从而影响购买意愿。5) 使用黏性，包括消费者在直播平台上的购物频率、对低价优惠活动的响应等，这些因素能够增强消费者对直播平台的忠诚度，进而影响其购买意愿。

基于上述要点，本文设置了以下自变量(解释变量)与因变量(被解释变量)。其中 X5~X10 为态度得分变量，分值为 1~5 分，分别表示非常不同意、不同意、不确定、同意、非常同意。具体如表 1 所示。

Table 1. Variable description

表 1. 变量说明

变量	变量说明
X1	网络购物经历, X1 = 0 表示无, X1 = 1 表示有。
X2	网购年限, 网购兴起于 2007 年左右, 规定其范围在 1~14。
X3	每周观看网络直播的频率, 设置为 0 次、1~5 次、6~10 次、10 次以上。
X4	每次观看网络直播的平均时长, 设置为 15 分钟以内、15 分钟至 30 分钟、30 分钟以上。
X5	网络直播购物比传统网购更方便
X6	网络直播购物中的优惠活动、质量和服对我购物影响很大
X7	观看我喜欢、感兴趣的主播对我购物影响很大
X8	直播中主播业务能力的好坏对我购物影响很大
X9	网络直播购物更利于我与他人交流商品信息
X10	国家出台的关于网络直播购物监管标准会让直播购物更加安全透明
X11	近三个月内网络直播购物的次数, 包含 3 次及以下、4~7 次、8~10 次、10 次以上。
X12	近三个月内网络直播购物的总消费, 包含 300 元以下、300~500 元、500 元~700 元、700 元~1000 元、1000 元以上。
SEX	性别, SEX = 0、1 表示男性、女性。
X13	年龄段, X13 = 0、1 表示 21~30 岁、31~40 岁。
X14	学历情况(包含在读), 设置为专科以下、专科(高职高专)与本科、硕士及以上。
X15	平均每月可支配收入, 设置为 2000 元以下、2000~4000 元、4000~6000 元、6000 元以上。
Y2	今后继续通过网络直播进行购物的意愿情况, 因变量。Y2 = 0、1 表示不会或不确定、会。

3.2. 特征工程

3.2.1. 数据预处理

数据预处理主要是对变量数据进行异常值处理和标准化, 数据标准化处理公式为:

$$x^* = \frac{x - \mu}{\sigma} \quad (1)$$

式中, μ 为变量 X 的均值, σ 为其标准差。

3.2.2. 聚类分析

聚类分析旨在将数据分类到不同的类或簇中, 使得同一簇内的对象具有很高的相似性, 而不同簇中的对象具有很大的相异性。本文采用 R 型聚类分析的步骤如下:

- 1) 数据预处理: 采用标准化处理, 消除量纲影响;
- 2) 距离度量方法选择: 采用计算相关系数得出相关距离, 公式如下。

$$D(x, y) = 1 - r(X_k, X_l) = 1 - \frac{\sum_{i=1}^n (X_{ki} - \bar{X}_k)(X_{li} - \bar{X}_l)}{\sqrt{\sum_{i=1}^n (X_{ki} - \bar{X}_k)^2 (X_{li} - \bar{X}_l)^2}} \quad (2)$$

其中 $X_k = (X_{1k}, \dots, X_{nk})'$ 是第 k 个变量的观测值向量, \bar{X}_k 为第 k 个变量的均值, 计算公式为 $\bar{X}_k = \frac{\sum_{i=1}^n X_{ik}}{n}$ 。

3) 聚类方法: 采用系统聚类, 通过相关距离选择最接近的一对变量合并成新类。使用聚类分析避免了人为主观的分类影响, 有助于构造新变量特征。

3.2.3. 特征选择

特征选择的目的是为了在优化模型的同时尽量不影响拟合效果，具体如下：

1) 嵌入法

嵌入法是一种让模型“自行选择”哪些特征的方法，即特征选择和模型拟合同时进行。利用机器学习的算法模型进行拟合，得到各变量特征的权值系数，根据权值系数从大到小选择特征，主要是通过人为规定的阈值进行变量选择，若变量对应的属性值小于规定阈值，则视作不重要且可被移除的变量。规定的阈值越小则可选择变量个数越多，反之可选择变量个数则越少。

2) 包装法

与嵌入法类似，包装法也是一个特征选择和模型拟合同时进行的方法，不同的是，包装法使用模型本身来选取特征，而非人为规定阈值来进行变量选取。包装法主要以 `coef_` 属性获得变量重要性，主要通过递归消除特征法对模型进行多次拟合。

3.3. 模型分析及评估

3.3.1. 逻辑回归模型

Logistic 回归模型定义如下：

已知样本 $T = \{(X_1, Y_1), \dots, (X_N, Y_N)\}$ ，其中 $X_k = [x_1, \dots, x_n]$ 为第 K 个样本对应变量的数据， $Y_k \in \{0, 1\}$ 且 $Y_k = 0$ 表示反例， $Y_k = 1$ 则为正例。设 $\theta = [\theta_1, \dots, \theta_n]$ 即为模型内自变量 X 所对应系数， b 表示模型截距项，二分类逻辑回归模型具体形式为：

$$P(Y = 1 | X_i) = Y(X_i) = \frac{1}{1 + e^{-(\theta^T X_i + b)}} \quad (3)$$

式中， $Y(X_k) \in [0, 1]$ ，表示样本 X_k 对应事件发生的概率，通过判断其发生概率是否大于设定阈值来判断样本对应事件是否会发生。通常通过最大化似然函数来求解模型参数 θ ，但本文使用机器学习中的损失函数来衡量模型预测错误程度，最小化损失函数等价于最大化似然函数。

3.3.2. L2 正则化

L2 正则化(即 Ridge 回归)方法为模型添加一个先验知识：即模型参数 θ 服从零均值正态分布，该方法的目标函数为：

$$-\ln L(\theta) = -\sum_i \left[y_i \ln p(x_i) + (1 - y_i) \ln p(1 - p(x_i)) + \frac{1}{2\sigma^2} \theta^T \theta \right] \quad (4)$$

L2 正则化有助于使拟合误差最小化，提高预测精度，并降低特征权重以简化模型，避免过拟合。它确保参数 θ 接近于 0 但不为 0，平滑参数空间，降低模型复杂度，并消除多重共线性。

3.3.3. 交叉验证

交叉验证是一种评估模型有效性的统计方法，比单纯的模型预测精度更稳定全面。 k 折交叉验证将数据分为 k 个部分，每次用一部分作为测试集，其余作为训练集，迭代 k 次后计算平均精度作为评估分数。在本文中，首先使用 5 折交叉验证进行模型参数选择，其次使用 10 折交叉验证进行模型有效性评估。

4. 实证检验和结果分析——以 K 市为例

4.1. 调查数据简析

由于实际条件所限，本研究通过在线问卷平台发放并回收了 636 份问卷，其中异常问卷共计 105 份，实际回收问卷 531 份，排除作为预测集测试所预留的 50 份问卷，实际收到有效问卷为 481 份，问卷有效

率为 75.63%。

同时，由于直播购买意愿(Y2)中愿意：不愿意或不确定的两种分类占比约为 9:2，即存在样本分类不均衡问题。针对此情况，本文使用 Python 中的 sklearn 库时在模型参数中对参数 class_weight 设置为 ‘balance’。此设置会根据各类别的样本数量自动调整其权重，使得少数类别得到更多的关注，从而解决分类样本不平衡问题。

4.1.1. 信度检验

Table 2. Reliability analysis

表 2. 信度分析

统计量	Cronbach’s Alpha	基于标准化项的 Cronbach’s Alpha	项数
信度检验结果	0.903	0.942	8

对量表中的 6 个态度问题进行信度检验，结果如表 2。Cronbach’s Alpha 系数大于 0.9，说明问卷量表的信度高，结果可信。

4.1.2. 效度检验

Table 3. Construct validity test result

表 3. 结构效度检验结果

统计量	KMO 值	Bartlett 的球形度检验		
		近似卡方	df	sig
效度检验结果	0.906	3582.254	28	0.000

对量表中的 6 个态度问题进行效度检验如表 3，问卷中量表题对应的 KMO 值为 0.906，且通过 Bartlett 的球形度检验，说明问卷的结构效度非常好，量表问题设置合理且能有效反映研究变量。

4.2. 聚类分析

4.2.1. 聚类个数选择

本文的 R 型聚类分析使用 R 软件进行，利用 Gap 统计值、mclust 包、Calinski-Harabasz 准则三种方法对比分析确定聚类个数 k。

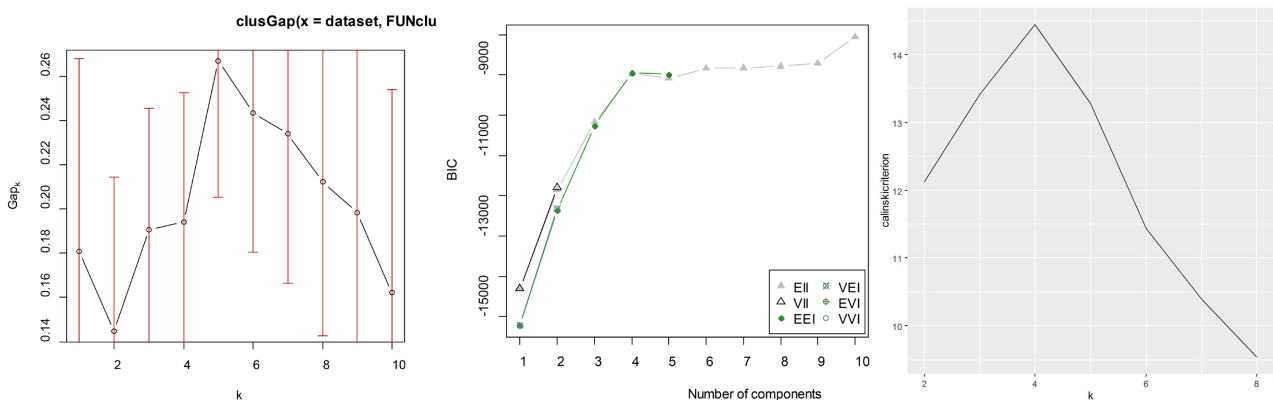


Figure 2. Gap statistics (left), mclust package clustering analysis (middle), Calinski-Harabasz criterion analysis (right)
图 2. Gap 统计值(左)、mclust 包聚类分析(中)、Calinski-Harabasz 准则分析(右)

通过图2可以看出：1) $k = 5$ 时对应 Gap 统计值最大；2) $k = 4$ 后对应的 BIC 值变动趋势逐渐平滑是不错的选择；3) $k = 4$ 时对应的 C-H 值最大，即变量聚类数目为 4 时聚类效果最佳。对比三种方法对应结果，本文选定聚类数目 $k = 4$ 来进行聚类分析处理。

4.2.2. 距离度量及系统聚类

在系统聚类的方法选择上，本文基于学者胡雷芳的研究[9]，选定最长距离法与类平均法进行系统聚类分析，以 $k = 4$ 为聚类数目，图3展示了两种方法的聚类结果。

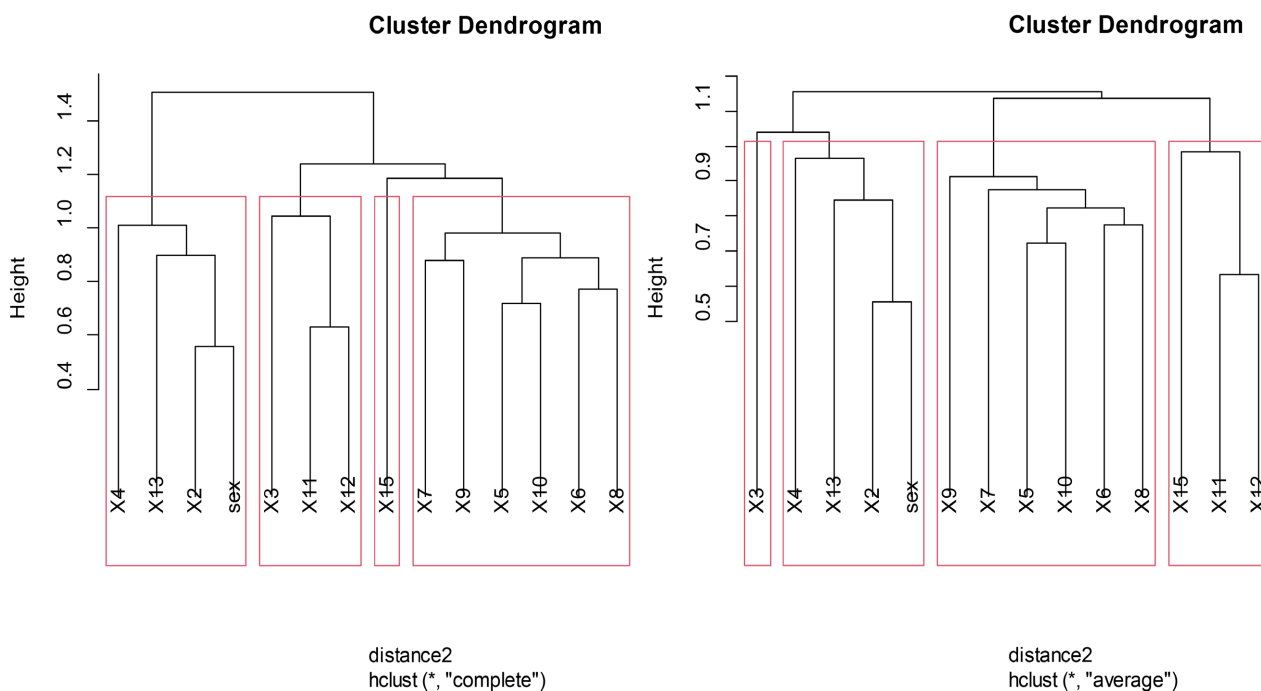


Figure 3. Systematic cluster diagram (left: longest cluster method, right: class average method)

图3. 系统聚类图(左：最长聚类法，右：类平均法)

观察上图后发现两种不同的聚类方法得到了相似的聚类结果：1) X5 至 X10 的变量为一个大类，反映了受访者态度得分变量。2) X2、X4、X13、SEX 为一个类，体现受访者基本信息。3) 变量 X3、X11、X12、X15，尽管两种聚类结果存在细微差异，但总体反映了受访者的收入水平和参与直播购物的情况。

进一步对态度得分变量(X5 至 X10)的聚类进行分析：1) 两种聚类方法所得系统聚类结果一致；2) 其中，X5 与 X10 可归为类 I，代表直播购物环境体验，由直播购物的便利性和政策保障体现。X6、X8 可归为类 II，代表直播购物服务体验态度，通过直播购物所提供的优惠和服务、主播的业务能力来体现。X7、X9 可归为类 III，体现直播互动体验态度，以个人喜爱的主播和交流体验为主。

基于以上分析，文章对三个态度得分变量进行细化和降维处理： $XA_1 = X5 + X10$ 表示直播购物环境体验态度， $XA_2 = X6 + X8$ 表示直播购物服务体验态度， $XA_3 = X7 + X9$ 表示直播购物互动体验态度。三类态度变量评分标准为：变量中至少含有一个“不同意”或“非常不同意”选项的，视作否定态度，其对应的态度得分变量小于 5 分；至少含有一个“一般”或“不确定”选项的，视作中立态度，其对应的态度得分变量在 5~7 分之间；均为“同意”及以上选项的，视作肯定态度，其对应的态度得分变量在 8~10 分之间。具体变量聚类情况如表 4 所示。

Table 4. Variable clustering table
表 4. 变量聚类表

变量	变量说明
XA_1	客观环境因素态度得分, $XA_1 = X5 + X10$
XA_2	服务体验态度态度得分, $XA_2 = X6 + X8$ 。
XA_3	互动体验态度态度得分, $XA_3 = X7 + X9$
XA_1L	环境体验态度; 设置为否定、中立、肯定。
XA_2L	服务体验态度; 设置为否定、中立、肯定。
XA_3L	互动体验态度; 设置为否定、中立、肯定。

综上分析表明, 所选变量可有效地区分不同维度的受访者态度, 并且通过加和的方式对态度得分变量进行了降维处理, 使得后续分析可以更加集中和简化。

4.3. 基于不同数据预处理情况下的模型分析

本文对聚类以及聚类 + 特征选择处理后数据以 7:3 的比例将数据分为训练集与测试集, 利用 python 中的 sklearn 机器学习库展开分析。主要步骤为: 1) 设定模型相关参数取值范围, 拟合训练集数据后得到交叉验证得分图; 2) 使用步骤 1 中设定的参数取值范围, 将参数迭代地代入模型中, 拟合训练集数据, 并使用交叉验证来评估模型的性能。3) 对比得出最优变量系数, 通过拟合训练集数据得出模型具体结果。4) 模型拟合效果评估, 主要依据模型预测精度和交叉验证得分。

4.3.1. 聚类分析下的模型分析

对 Y2 的 L2 正则化的逻辑回归拟合中, 最优算法为 saga, C 值(L2 正则化强度倒数)的分析如图 4 所示。

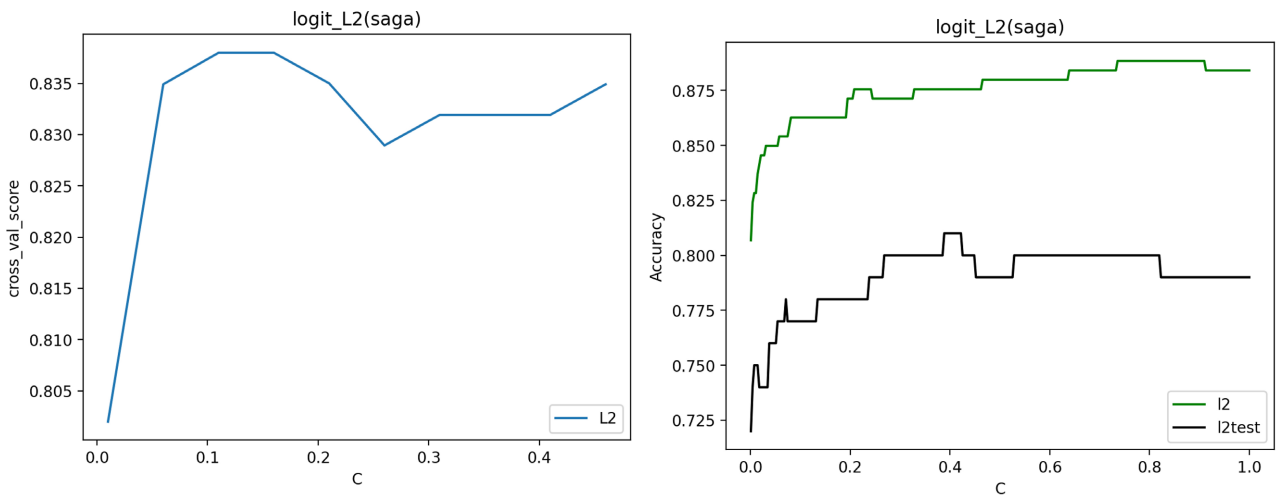


Figure 4. (Left) C-value—cross-validation score plot; (right) C-value—model prediction accuracy plot

图 4. (左) C 值——交叉验证得分图; (右) C 值——模型预测准确率图

基于上图结果对比, 模型参数 C 值 = 0.4 时交叉验证得分与准确率的表现均较为良好, 故选定 C 值 = 0.4 作为模型参数。

4.3.2. 聚类 + 嵌入法下的模型分析

针对带 L2 正则化的逻辑回归模型而言, 规定区间[0, 模型系数最大的值]作为阈值取值范围, 得到

聚类降维 + 嵌入法——10 折交叉验证得分情况如图 5 所示。

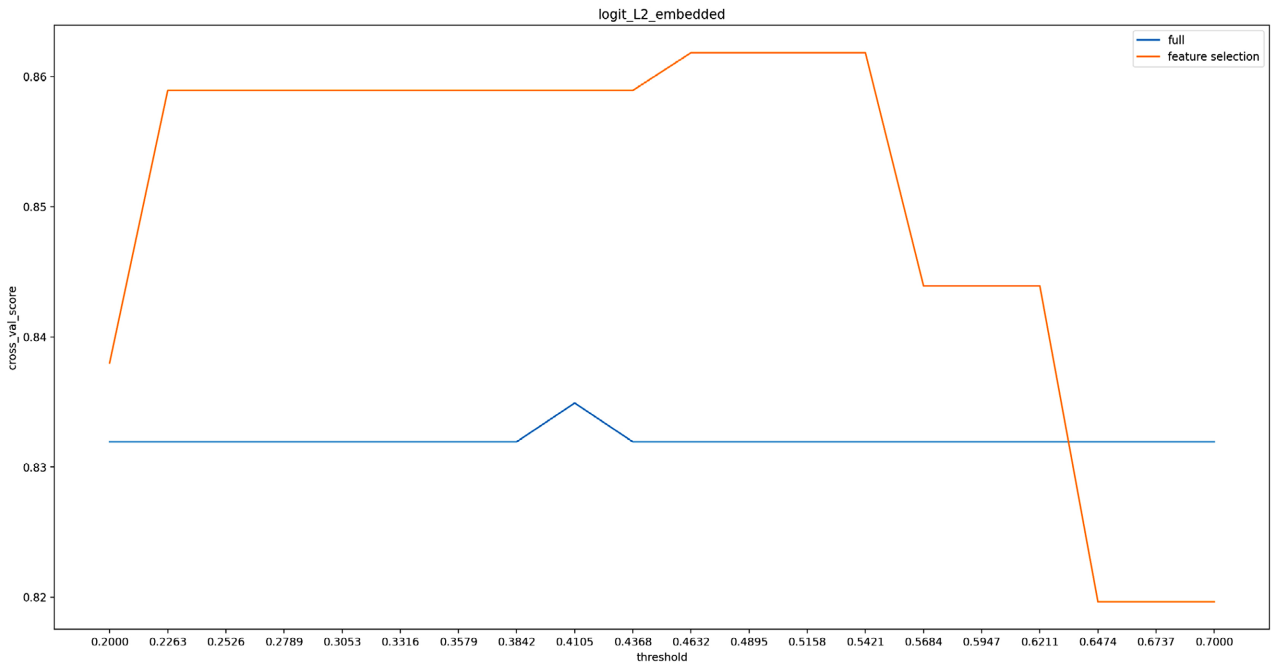


Figure 5. Clustering Dimensionality Reduction + Embedding Method—10-fold Cross-validation Score Graph

图 5. 聚类降维 + 嵌入法——10 折交叉验证得分图

在聚类 + 嵌入法下的模型分析中，最优阈值为 0.47，通过最优阈值得到特征选择的变量个数为 7，这些变量是：X3、XA_1L、XA_2L、X12、SEX、X13、X15。

4.3.3. 聚类 + 包装法下的模型分析

针对带 L2 正则化的逻辑回归模型而言，通过包装法得到变量选择个数——10 折交叉验证得分曲线如图 6 所示。

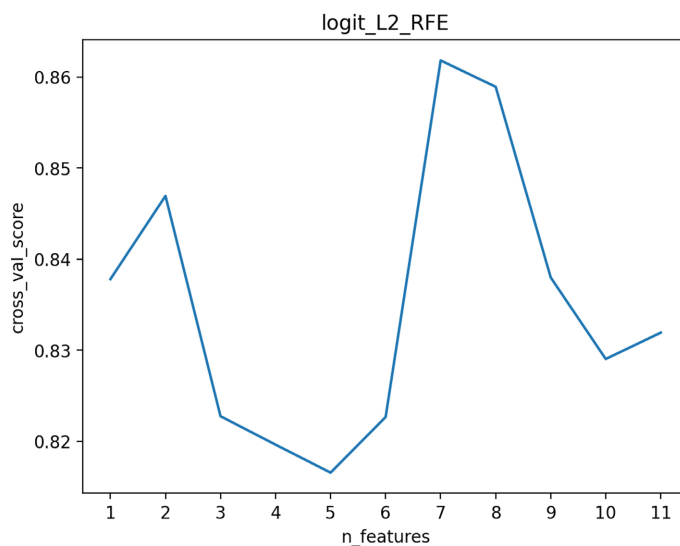


Figure 6. Number of selected variables—10-fold cross-validation score plot

图 6. 变量选择个数——10 折交叉验证得分图

通过图像可得 7 个变量对应的交叉验证得分最高，各变量综合重要性排名具体如表 5 所示。

Table 5. Ranking of comprehensive importance of variables

表 5. 变量综合重要性排名

变量	X2	X3	X4	XA_1L	XA_2L	XA_3L	X11	X12	SEX	X13	X15
重要性名次	4	1	5	1	1	3	2	1	1	1	1

通过表内排名情况可以发现：X3、XA_1L、XA_2L、X12、SEX、X13、X15 共 7 个变量重要性排名最高，即为受访者基础信息、态度变量得分(受访者客观环境、服务因素的态度评价)、网络直播及直播购物基本信息。

4.3.4 模型有效性分析

基于三种不同数据预处理情况下的模型构建后，具体有效性评估指标如表 6 所示。

Table 6. Model cross-validation score

表 6. 模型交叉验证得分

模型	10 折交叉验证得分
逻辑回归(基于聚类降维数据)	0.84
逻辑回归(基于聚类降维 + 嵌入法)	0.85
逻辑回归(基于聚类降维 + 包装法)	0.85

模型交叉验证得分结果表明：1) 聚类 + 特征选择构建的模型有效性优于仅通过聚类处理构建的模型；2) 基于聚类 + 嵌入法与包装法构建的模型有效性差异不大，但考虑到嵌入法通常需要人为设定阈值，可能会引入主观性并影响结果稳定。而包装法则非人为选取特征、更加自动化且易于操作。

综上，本文选定基于聚类降维数据所构建模型与基于聚类降维 + 包装法构建模型对预留的预测数据集进行预测分析，结果如表 7 所示：

Table 7. Summary of model prediction analysis results

表 7. 模型预测分析结果汇总

模型	正类预测准确率	负类预测准确率
逻辑回归(基于聚类降维数据)	92%	54%
逻辑回归(基于聚类降维 + 包装法)	92%	67%

在正类预测、负类预测的准确率上，基于聚类降维 + 包装法构建的模型效果更优，故其为最优预测模型，模型的具体系数项情况如表 8 所示。

Table 8. Model coefficient terms

表 8. 模型系数项

变量	coef
X3	0.6061
XA_1L	1.6727
XA_2L	1.1256
X12	0.8582

续表

sex	0.0399
X13	-0.8064
X15	0.5889
Intercept (截距项)	-11.8707

通过模型系数可发现变量影响程度情况为：X3 (每周观看直播频率)、XA_1L (直播购物环境评价)、XA_2L (直播购物服务评价)、X12 (近三个月内直播购物总额)、X13 (年龄段)、X15 (平均每月可支配收入) 共计 6 个变量的影响程度较高,除 X13 与 Y2 呈负相关关系外,其余变量对 Y2 均呈正相关关系。XA_1L、XA_2L、X12、X13、X15 共 5 个变量对应系数绝对值大于 0.5, 反映了这些变量影响程度较大。

5. 结论

本文通过问卷调查、数据预处理、模型拟合最终选定基于聚类降维 + 包装法构建的逻辑回归为最优预测模型, 并针对模型系数项分析探究了变量影响程度情况。综合分析结果可得: 1) 基于聚类 + 特征选择的预处理数据在模型表现上优于仅基于聚类分析后的数据模型表现, 反映了特征选择可去除无关或冗余的特征, 从而提高模型的预测能力和解释性; 2) 直播购物的客观环境体验和服务体验对购买意愿有显著的正向影响。这意味着提升用户体验是增加销售的关键; 3) 性别对于购买意愿的影响较小, 而年龄段呈负相关关系。这可能意味着年轻人更倾向于通过网络直播平台进行购物, 或者年轻人对于直播购物的接受度更高; 4) 平均每月可支配收入作为个人隐私数据, 在直接调查中可能难以获取, 可考虑通过问题变换的方式侧面收集。以上分析不仅为理解网络直播购物的购买意愿提供了实证分析, 也为直播平台和商家提供了策略制定的参考依据。

参考文献

- [1] 赵梓煜, 刘孜旺. “网红经济 + 直播电商”对消费者购买行为影响分析[J]. 现代商贸工业, 2024, 45(4): 35-37. <https://doi.org/10.19311/j.cnki.1672-3198.2024.04.012>
- [2] 张宝生, 张庆普, 赵辰光. 电商直播模式下网络直播特征对消费者购买意愿的影响——消费者感知的中介作用[J]. 中国流通经济, 2021, 35(6): 52-61. <https://doi.org/10.14089/j.cnki.cn11-3664/f.2021.06.005>
- [3] 刘平胜, 石永东. 直播带货营销模式对消费者购买决策的影响机制[J]. 中国流通经济, 2020, 34(10): 38-47. <https://doi.org/10.14089/j.cnki.cn11-3664/f.2020.10.004>
- [4] 赵冬玲. 网络直播时代的品牌曝光和销售转化探究——以购物类直播平台为例[J]. 商业经济研究, 2018(1): 62-64.
- [5] 胡晓丽, 张会兵, 董俊超, 等. 基于 CNN-LSTM 的用户购买行为预测模型[J]. 计算机应用与软件, 2020, 37(6): 59-64.
- [6] 李伊林, 段海龙, 林振荣. 数据平衡与模型融合的用户购买行为预测[J]. 计算机应用与软件, 2022, 39(9): 50-55+86.
- [7] 祝歆, 刘潇蔓, 陈树广, 等. 基于机器学习融合算法的网络购买行为预测研究[J]. 统计与信息论坛, 2017, 32(12): 94-100.
- [8] 权李之. “直播 + 电商”模式下农产品购买意愿影响因素的模型构建[J]. 中国市场, 2023(23): 125-128. <https://doi.org/10.13939/j.cnki.zgsc.2023.23.125>
- [9] 胡雷芳. 五种常用系统聚类分析方法及其比较[J]. 浙江统计, 2007(4): 3.