

基于Firth逻辑回归与随机森林的 在线购买意向预测

朱宇露

南京邮电大学经济学院, 江苏 南京

收稿日期: 2026年2月26日; 录用日期: 2026年3月9日; 发布日期: 2026年4月1日

摘要

在数字经济由规模扩张向质量驱动转型的背景下, 预测用户购物意图对提升电商转化率至关重要。本研究利用UCI在线会话数据集作为基准测试样本, 旨在验证Firth逻辑回归与随机森林模型在处理类别不平衡及数据稀疏性问题上的预测效能, 为实时意图预测提供算法框架参考。研究发现: 第一, 实时行为是电商转化的核心指标, 其中页面价值对购买概率有极强的正向驱动作用, 尽管其存在统计内生性, 但在动态识别逻辑下是捕捉访客从“浏览”转向“决策”阶段的关键先导指标, 而退出率则产生显著的负向抑制效应; 第二, 外部情境具有显著调节作用, 大促月份通过激活消费者的价格预期与限时心理, 明显提升购买概率; 第三, 新老访客存在决策异质性, 回访者对页面流畅度、周末窗口更为敏感, 而新访客则更依赖高价值页面的即时引导。本文为电商企业动态识别高价值会话、进行智能化干预及实施差异化运营提供了指导性的算法支撑。

关键词

电子商务转化, 在线购物意图, Firth逻辑回归, 随机森林模型, 算法验证

Predicting Online Purchase Intention via Firth Logistic Regression and Random Forest

Yulu Zhu

School of Economics, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: February 26, 2026; accepted: March 9, 2026; published: April 1, 2026

Abstract

As the digital economy transforms from scale expansion to quality-driven growth, it is significant to

predict users' shopping intention to enhance the conversion rates of e-commerce. This study uses the UCI online session dataset as a benchmark sample to verify the prediction efficiency of Firth Logistic Regression and Random Forest models in dealing with the problems of category imbalance and data sparsity, and provides an algorithm framework for real-time intention prediction. The findings reveal that: (1) Real-time behavior is the primary indicator of e-commerce transformation, in which page values have a strong positive driving effect on purchase probability. Despite its inherent statistical endogeneity, it serves as a critical precursor signal for capturing the transition from the "browsing" to the "decision-making" stage within dynamic recognition logic. Conversely, exit rates have a significant negative inhibitory effect; (2) External contexts play a significant moderating role. In big promotion months, the purchase probability is significantly improved by activating consumers' price expectation and limited-time psychological triggers; (3) Significant decision heterogeneity exists between new and returning visitors. Old visitors are more sensitive to page fluency and weekend windows, while new visitors rely more on the real-time guidance of high-value pages. This study provides guiding algorithm decision support for e-commerce enterprises to dynamically identify high-value sessions, carry out intelligent intervention, and implement differentiated operational strategies.

Keywords

E-Commerce Conversion, Online Purchase Intention, Firth Logistic Regression, Random Forest Model, Algorithm Validation

Copyright © 2026 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

在数字经济高度发展的背景下，全球电子商务市场正经历从规模扩张向质量驱动的范式转型，根据相关数据显示，全球电子商务市场规模预计将从 2023 年的 16.29 万亿美元增至 2033 年的约 67.05 万亿美元[1]。然而，随着市场竞争的加剧和技术的快速迭代，跨境电商企业遭遇了前所未有的挑战与机遇，尤其是在用户行为这一领域，借助大数据技术去挖掘用户行为特征，已然成为企业构建核心竞争力、达成精准营销以及实现个性化推荐的关键途径。

电子商务用户在线转化数据常呈现出稀疏性以及类别不平衡的特性，即仅有少数最终达成交易。数据分布的偏态性会让传统的 Logistic 回归模型在最大似然估计时出现系统性偏误，甚至还会引发模型无法收敛的“完全分离”问题。因此，本文选取 UCI 在线购物数据集作为算法验证的基准测试样本，高度稀疏的典型特征使其成为验证 Firth 逻辑回归与随机森林模型稳健性的理想“玩具数据”，为跨境电商企业的精细化运营以及转化率优化提供算法支撑和实践指导。

2. 理论基础与研究假设

2.1 研究现状

目前，学术界对电子商务的研究主要集中在用户行为分析、个性化推荐与精准营销三大维度。Novak [2]等(2000)较早通过构建定量模型，界定了 Web 环境下“心流”体验的构成要素，并证实了这种沉浸式在线体验对消费者购物行为及网络使用习惯的显著影响；杨峰[3]等提出融合显式反馈与语义信息的图卷积推荐模型 TDGCN-L，为大规模电商平台的多信号融合推荐系统提供参考；同时，成保梅[4]等强调应融

合情境因素，通过计算融合情境相似度来获取电子商务用户当前情境的近似情境集，对电子商务用户-兴趣项-情境构建三维模型，以达到用户兴趣挖掘准确率较高、综合性能较好、召回率较高的目的；在精准营销与用户画像方面，毛明扬[5]等提出通过大数据挖掘技术手段可以从海量数据中自动学习，自动提取多样化的复杂特征，全面了解用户的行为习惯、偏好属性、潜在需求等深层次的用户特征，构建准确的用户模型，进而为用户提供更加精准的服务。综上所述，虽然已有研究识别了用户行为变量的价值，但在类别不平衡数据环境下，结合情境因素对转化机制进行深度推演的研究仍需进一步加强。

2.2. 研究假设推演

(1) 实时行为特征对购买意向的影响研究

根据精细加工可能性模型(ELM)，消费者在电商环境下的决策从信息处理与响应方式出发对态度行为改变进行理解[6]。“刺激-机体-反应”(SOR)模型则解释和预测环境刺激对个体认知结构、情感体验及行为决策的作用机制[7]，网页环境特征作为外部刺激，通过影响消费者的认知与情感状态，最终驱动行为响应。页面价值代表了该页面在用户最终完成购买交易中所起到的平均价值贡献，虽然其因包含成交金额而具有统计内生性，但它实质上是用户从“信息浏览”转向“决策转化”阶段的强表征信号，是中心路径的核心变量，退出率则反映了购物路径的摩擦力。基于此，本文提出如下假设：

H1: 实时行为特征显著影响在线购买意向，其中页面价值具有显著的正向预测作用，而退出率具有显著的负向抑制作用。

(2) 外部情境与访客类型的交互影响研究

根据理性行为理论(TRA)及情境强度理论，个体决策并非孤立存在，而是受到外部环境及个人过往经验的共同调节。理性行为理论主张，行为意愿受到行为的态度与主观规范共同作用[8]。一方面，时间情境如大促月份、周末等，通过时间的紧迫感和改变消费者的心理来影响决策。凭借“低价”带来的实惠感，以及国内高速物流体系的支撑，“双十一”曾一度让消费者快速地享受到以高折扣购买商品带来的满足感[9]，显著提升了消费意愿。另一方面，回访者和新访客在信息摄取习惯方面存在着较为突出的差异，回访者由于具有购买产品经验，对于网站结构的敏感度以及对时间窗口的把控能力一般要比新访客更强，呈现出更明显的异质性特征。基于此，本文提出如下假设：

H2: 外部情境因素显著调节购买意向，且该影响在不同访客类型之间呈现出显著的异质性。

3. 研究设计与数据准备

3.1. 数据来源与样本说明

本文基于 UCI Machine Learning Repository 网站的在线购物者购买意向数据集进行分析，共计 12,330 条电商会话数据。通过数据预处理，剔除缺失值及异常值，进行数据标准化、类别不平衡处理，在预处理后的数据的基础上进行模型构建与分析。虽然 UCI 数据集在反映直播电商等最新业态方面存在时效性局限，但其高度稀疏和类别不平衡的特征，使其成为验证 Firth 逻辑回归解决“完全分离”问题及小样本偏差的理想基准数据集。

3.2. 变量定义与测度

本文选取了 Revenue (是否最终产生了购买行为)为被解释变量，解释变量主要包含用户实时访问行为和网页质量指标，前者为访客在单次会话中访问的不同类型页面的数量以及在各类别页面上花费的总时长，后者包括 Bounce Rate、Exit Rate 和 Page Values，这些是由电子商务网站中每个页面所衡量的 Google Analytics 指标。其中 Page Values = (交易收入 + 总目标价值)/唯一页面访问量，由于其包含交

易收入，存在内生性偏误风险。该指标并非购买行为的诱因，而是通过观察用户走过的路径价值，来反推其成交概率；此外，数据集还包括地区、访客类型、衡量访问日期是否为周末、月份等。变量名称及含义表如表 1 所示：

Table 1. Variable names and meanings
表 1. 变量名称及含义表

变量名	变量名称	变量定义
Administrative	管理类页面访问量	用户访问的账户管理、结算跳转等管理性质页面的数量。
Administrative_Duration	管理类页面停留时间	用户在上述管理性质页面上花费的总时间(秒)。
Informational	信息类页面访问量	用户访问的关于公司、联系信息、配送政策等信息页面的数量。
Informational_Duration	信息类页面停留时间	用户在上述信息类页面上花费的总时间(秒)。
ProductRelated	产品相关页面访问量	用户访问的商品详情页、分类页等产品相关页面的数量。
Product Related_Duration	产品相关页面停留时间	用户在产品相关页面上花费的总时间(秒)。
BounceRates	跳出率	从该页面进入网站后未进行任何后续操作即离开的访问占比。
Exit Rates	退出率	用户在该页面结束会话并离开网站的次数占该页面总浏览量的比例。
Page Values	页面价值	衡量该页面在用户最终完成购买交易中所起到的平均价值贡献。
Special Day	日期临近度(节日敏感度)	衡量访问日期与特定节日(如母亲节、情人节)的接近程度(0 到 1 之间)。
Month	月份	访问发生的月份(1 月至 12 月)。
Operating Systems	操作系统	用户访问网站时使用的操作系统代号。
Browser	浏览器	用户使用的浏览器代号。
Region	地区	用户所在的地理区域代号。
Traffic Type	流量类型	用户进入网站的渠道来源(如搜索、直接输入、广告等)。
Visitor Type	访客类型	区分新访客(New Visitor)、回访者(Returning Visitor)或其它(Other)。
Weekend	是否为周末	标记该访问会话是否发生在周六或周日。
Revenue	是否购买	记录该次会话是否最终产生了购买行为。

表 2 为部分变量的描述性统计分析表：

Table 2. Descriptive statistics of selected variables**表 2.** 部分变量描述性统计表

Variable	N	Mean	St. Dev.	Min	Max
Administrative	12,330	2.32	3.32	0	27
Administrative_Duration	12,330	80.82	176.78	0	3,398.75
Informational	12,330	0.5	1.27	0	24
Informational_Duration	12,330	34.47	140.75	0	2,549.38
ProductRelated	12,330	31.73	44.48	0	705
ProductRelated_Duration	12,330	1,194.75	1,913.67	0	63,973.52
Bounce Rates	12,330	0.02	0.05	0	0.2
Exit Rates	12,330	0.04	0.05	0	0.2
Page Values	12,330	5.89	18.57	0	361.76
Special Day	12,330	0.06	0.2	0	1

表 3 为多重共线性检验表，VIF (方差膨胀因子)均小于 10，因此变量间无严重的多重共线性，可以适用于后续的回归分析。

Table 3. Multicollinearity test table**表 3.** 多重共线性检验表

Variable	GVIF	Df	GVIF ^{1/(2*Df)}
Administrative	1.816348	1	1.347719
Administrative_Duration	1.663146	1	1.28963
Informational	1.831389	1	1.353288
Informational_Duration	1.692937	1	1.301129
Product Related	4.907133	1	2.215205
ProductRelated_Duration	4.871247	1	2.20709
Bounce Rates	2.001281	1	1.414667
Exit Rates	2.201454	1	1.48373
Page Values	1.073652	1	1.036172
Special Day	1.259902	1	1.122454
Month	1.413063	9	1.019395
Visitor Type	1.142772	2	1.033927
Weekend	1.011995	1	1.00598

同时，进一步比较转化率分布及购买/非购买组的均值差异(见表 4 和表 5)，观察得到显著的稀疏性与类别不平衡特征，即仅有少数用户最终在电商平台购买产品，且两组均值差异明显。

Table 4. Frequency of purchasing behavior
表 4. 购买行为频率表

Revenue	Count	Percentage
FALSE	10,422	84.52555
TRUE	1908	15.47445

Table 5. Analysis of mean difference of purchasing behavior
表 5. 购买行为均值差异分析

Variable	Revenue_FALSE Mean	Revenue_FALSE SD	Revenue_TRUE Mean	Revenue_TRUE SD
Administrative	2.117732	3.202481	3.393606	3.730527
Administrative_Duration	73.740111	171.017970	119.483244	201.114732
Informational	0.451833	1.211582	0.786164	1.521176
Informational_Duration	30.236237	133.909306	57.611427	171.619131
Product Related	28.714642	40.744717	48.210168	58.267365
Product Related Duration	1069.987809	1803.797757	1876.209615	2312.214392
Bounce Rates	0.025317	0.051877	0.005117	0.012185
Exit Rates	0.047378	0.051231	0.019555	0.016463
Page Values	1.975998	9.072424	27.264518	35.191954
Special Day	0.068432	0.209054	0.023166	0.123521

4. 实证分析

4.1. Firth 惩罚似然 Logistic 回归模型

(1) 模型原理

在拟合 Logistic 回归模型过程中, 当样本量较小, 尤其是解释变量较多, 数据分布不均衡, 不满足渐近统计推断前提时, 会出现分离问题, 此时模型的最大似然估计不存在或者不可靠[10]。因此, 本文构建 Firth 惩罚似然 Logistic 回归模型, 该模型的原理是在小样本情况下, 由于得分函数的无偏性及曲率导致参数估计值的偏差, 该方法的基本思想就是在得分函数中加入惩罚项, 从而将参数最大似然估计值的偏差减少[10]。

(2) 模型结果

根据 Firth 回归模型结果(见表 6)显示, Page Values 是驱动消费者决策的核心指标, 其 OR 值为 4.5765 ($P < 0.05$), 意味着在其他条件不变的情况, Page Values 每提升一个单位, 用户最终下单的概率大约增加 3.5765 倍, 该指标具有内生性, 量化了页面对用户决策的支撑强度, 是识别高意向会话逻辑中权重最高的预测因子; Product Related_Duration 亦推动消费者购买产品, 其 OR 值为 1.1198 ($P < 0.05$), 意味着在其他条件不变的情况, Product Related_Duration 每提升一个单位, 用户最终下单的概率大约增加 0.1198 倍。相反, Exit Rates 对购买转化产生显著的负向抑制作用, 其 OR 值为 0.4782 ($P < 0.05$), 表明该因素的增加显著降低了用户的转化意向。

在时间维度的回归分析中, 11月(Month Nov)的OR值在所有月份中最大(8.5426), 且在1%水平下显著, 对购买决策展现出最强的正向预测效应。这意味着在控制其他变量的情况下, 11月访问的用户其购买概率显著提升, 充分体现了年度大规模促销活动对消费者决策逻辑的重塑作用。

Table 6. Logistic regression results
表 6. Logistic 回归结果表

Variable	Estimate	OR	Lower_CI	Upper_CI	P_Value
(Intercept)	-3.5038	0.0301	0.0077	0.0831	0.0000
Administrative	0.0049	1.0049	0.9835	1.0267	0.6530
Administrative_Duration	-0.0213	0.9789	0.9141	1.0455	0.5306
Informational	0.0319	1.0324	0.9789	1.0879	0.2383
Informational_Duration	0.0122	1.0122	0.9511	1.0742	0.6963
Product Related	0.0018	1.0018	0.9995	1.0040	0.1300
ProductRelated_Duration	0.1131	1.1198	1.0160	1.2439	0.0221
Bounce Rates	-0.2012	0.8178	0.5969	1.1068	0.1953
Exit Rates	-0.7378	0.4782	0.3793	0.5977	0.0000
Page Values	1.5209	4.5765	4.1975	5.0017	0.0000
Special Day	-0.1118	0.8943	0.5566	1.4016	0.6325
Month Mar	1.0992	3.0017	1.0888	11.6473	0.0316
Month May	1.0522	2.8639	1.0539	11.0171	0.0377
Month June	1.3030	3.6802	1.2179	15.0101	0.0188
Month Jul	1.6937	5.4394	1.9077	21.4977	0.0006
Month Aug	1.6104	5.0049	1.7558	19.7766	0.0013
Month Sep	1.6024	4.9649	1.7503	19.5664	0.0013
Month Oct	1.6021	4.9637	1.7662	19.4602	0.0011
Month Nov	2.1451	8.5426	3.1354	32.9364	0.0000
Month Dec	1.0065	2.7360	0.9903	10.6258	0.0526
Visitor Type Other	-0.4549	0.6345	0.2113	1.5964	0.3548
Visitor Type Returning_Visitor	-0.3320	0.7175	0.6074	0.8492	0.0001
Weekend	0.0962	1.1010	0.9576	1.2640	0.1759

(3) 平均边际效应

为了更加直观地衡量各个变量的实际贡献情况, 本文计算了平均边际效应以量化各个变量对购买概率的实际贡献。结果显示(见表 7), PageValues 的平均边际效应是 0.1255, 这表明在样本平均水平上, 该指标每提高一个单位, 用户下单的概率将增加 12.55 个百分点。这一结果有力证实了该页面在用户最终

完成购买交易中所起到的平均价值贡献实质上刻画了页面在诱导用户从“随机浏览”向“确定性决策”转化过程中的关键信息效用。Exit Rates 的平均边际效应是-0.061，这表明 Exit Rates 每增加一个单位，电商用户购买产品转化概率会降低 6.1%，由此可知用户在该页面结束会话并离开网站的次数占该页面总浏览量的比例对在线购买意向起负向抑制作用。

不仅如此，本文对月份(Month)的平均效应进行计算，11 月的边际效应在所有月份中最大，为 0.3064，且在 1%水平下显著。这意味着在控制其他变量的情况下，11 月电商平台访问的用户相比其他月份，其购买产品概率显著提升 30.64%，充分体现了年度促销活动对用户购买产品的激励作用。

Table 7. Average marginal effect
表 7. 平均边际效应

Variable	dF/dx	Std.Err.	z	P > z	
Administrative	0.0004	0.0009	0.4314	0.6661	
Administrative_Duration	-0.0018	0.0028	-0.6307	0.5282	
Informational	0.0026	0.0022	1.1689	0.2425	
Informational_Duration	0.0010	0.0026	0.3712	0.7105	
Product Related	0.0001	0.0001	1.4626	0.1436	
Product Related_Duration	0.0097	0.0043	2.2774	0.0228	*
Bounce Rates	-0.0172	0.0127	-1.3544	0.1756	
Exit Rates	-0.0610	0.0094	-6.4590	0.0000	***
Page Values	0.1255	0.0062	20.1433	0.0000	***
Special Day	-0.0098	0.0194	-0.5052	0.6134	
Month Mar	0.1454	0.0976	1.4892	0.1364	
Month May	0.1248	0.0802	1.5563	0.1196	
Month June	0.1996	0.1328	1.5033	0.1328	
Month Jul	0.2832	0.1445	1.9604	0.0500	*
Month Aug	0.2645	0.1413	1.8720	0.0612	.
Month Sep	0.2624	0.1404	1.8696	0.0615	.
Month Oct	0.2607	0.1383	1.8853	0.0594	.
Month Nov	0.3064	0.1154	2.6556	0.0079	**
Month Dec	0.1328	0.0957	1.3878	0.1652	
Visitor TypeOther	-0.0350	0.0275	-1.2757	0.2021	
Visitor Type Returning_Visitor	-0.0300	0.0087	-3.4592	0.0005	***
Weekend	0.0080	0.0061	1.3200	0.1868	

(4) 评价指标

本文绘制 ROC 曲线并计算 AUC 值以便全面评估 Firth 逻辑回归模型对在线购物意向的预测效果。

如图 1 所示, 模型的 AUC 值为 0.874。根据统计学评价标准, AUC 处于(0.8, 0.9)区间代表模型具有优异的预测效能。红色曲线迅速向左上角偏移, 表明模型在处理类别不平衡数据时, 能够有效平衡敏感度与特异性, 准确识别出具有真实购买意愿的用户, 模型拟合效果较好。

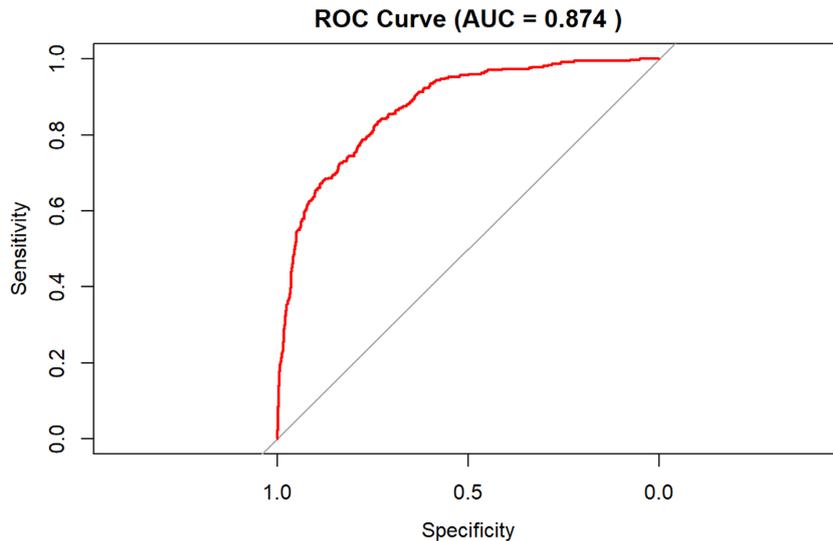


Figure 1. ROC curve
图 1. ROC 曲线

4.2. 随机森林稳健性检验模型

(1) 模型原理

随机森林是一种集成学习算法, 它通过构建多棵决策树来提高预测准确性。随机森林借助 Bagging 算法对样本信息随机抽样产生多个训练集, 然后对每个训练集采用决策树作为基分类器, 根据多棵树多数投票结果作为最终的预测值。根据决策树中选择的最佳变量作为分类节点, 从而对变量进行重要性排序 [11], 最终通过投票(分类)或平均(回归)方式综合所有树的结果, 从而提高模型的稳定性和泛化能力。

(2) 模型结果

基于本次分析的随机森林特征重要性数据(共 17 个输入变量), 随机森林模型展现了更优越的预测性能并提供了可靠的特征重要性排序。输入变量的重要性测度可视化如下图 2 所示。在这一集成学习中, 从对输出变量预测精度影响的角度看, Page Values、Exit Rates、Bounce Rates、Product Related 被列为预测用户购买产品行为的前四大重要特征; 从对输出变量异质性下降程度影响角度看, Page Values、Product Related_Duration、Exit Rates、Product Related、Month 的重要性依次位于前五。基于此, 随机森林模型的结果与前文 Firth 惩罚似然 Logistic 回归模型结果互相印证, 研究结论更具说服力。

4.3. 异质性分析

为了进一步探究不同特征群体在购买决策机制上的差异, 本文根据访客的历史交互记录, 将总样本划分为新访客与回访者两类子样本。结果如表 8 所示。

页面价值(Page Values)在两组中均在 1%的水平下显著为正, 新访客的系数为 1.9854, 略高于回访者的系数。这表明高价值的页面信息对消除新用户的认知摩擦、促成首次下单具有更强的判别价值和预测效力。退出率(Exit Rates)仅对回访者具有显著的负向影响, 而对新访客并不显著。这说明回访者对购物流程的顺畅度更为敏感, 任何导致流失风险的负面指标都会显著抑制其复购意愿。

同时,回访者的购买产品行为还受到访问日期与特定节日(如母亲节、情人节)的接近程度(Special Day)和是否为周末(Weekend)的显著影响。其中,特殊日期对回访者有显著负向影响,系数为-0.9459,反映出回访者在大促前期的“持币观望”心理;而周末效应则表明回访者更倾向于在非工作日进行决策。

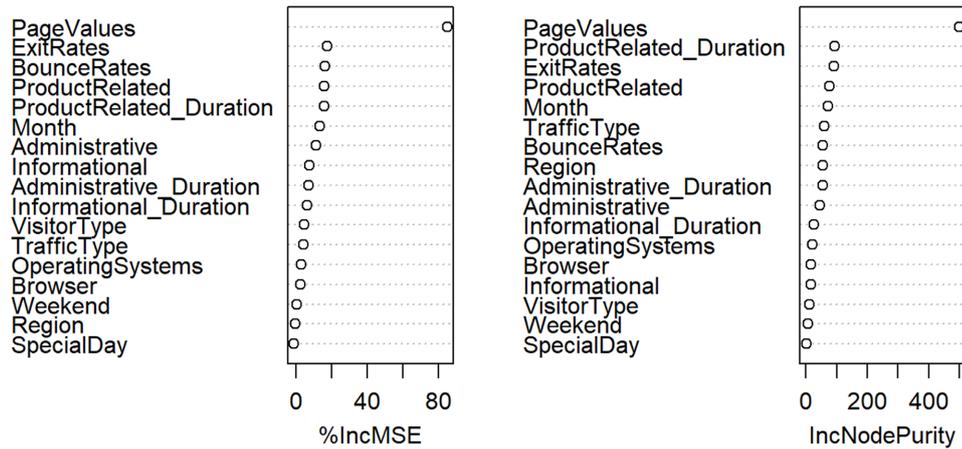


Figure 2. Visualization of importance measure of input variables in random forest model
图 2. 随机森林模型输入变量重要性测度可视化

Table 8. Heterogeneity analysis
表 8. 异质性分析

Variable	Coeff_New	P_Val_New	Group_New	Sig_New
(Intercept)	-1.4016	0.0000	New_Visitor	***
Page Values	1.9854	0.0000	New_Visitor	***
ExitRates	-0.1457	1.0000	New_Visitor	
Product Related_Duration	0.4631	1.0000	New_Visitor	
Special Day	0.0353	0.9559	New_Visitor	
Weekend	-0.2222	0.2053	New_Visitor	
Variable	Coeff_Ret	P_Val_Ret	Group_Ret	Sig_Ret
(Intercept)	-2.2550	0.0000	Returning_Visitor	***
Page Values	1.3689	0.0000	Returning_Visitor	***
Exit Rates	-0.9472	0.0000	Returning_Visitor	***
Product Related_Duration	0.2587	0.0000	Returning_Visitor	***
Special Day	-0.9459	0.0000	Returning_Visitor	***
Weekend	0.2135	0.0055	Returning_Visitor	**

4.4 稳定性检验

(1) 安慰剂检验

为了进一步排除不可观测因素对研究结论的潜在干扰,本文采用了安慰剂检验。通过对处理组进行

500 次随机置换并重新建立 Firth 逻辑回归模型，得到解释变量 Page Values 的 P 值和密度分布(见图 3)。核密度分布曲线接近于[0, 1]上的均匀分布，符合安慰剂检验的理论预期。结果显示，随机模拟产生的 P 值主要分布在非显著区域($P > 0.05$)，而本文基准模型得到的实际 P 值(0.0000)远低于该分布。这一结果有力地证明了页面价值对在线购买意向的正向驱动作用的研究结论具有极强的稳健性。

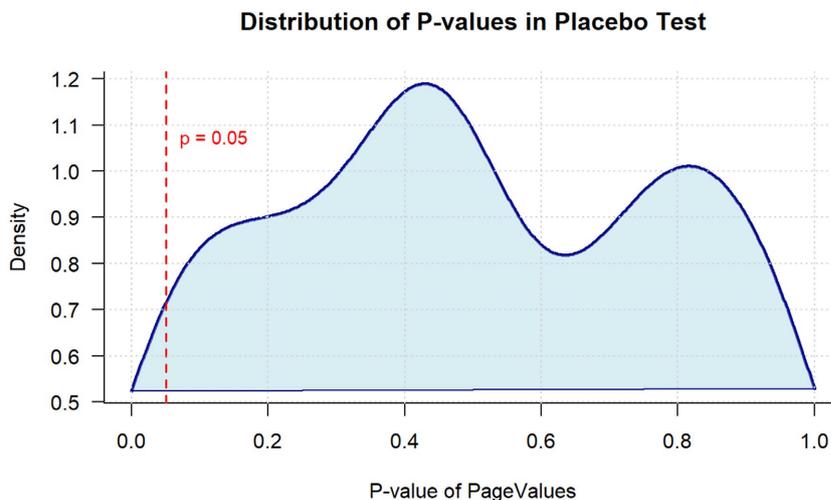


Figure 3. Placebo test
图 3. 安慰剂检验

(2) H_L 检验

Hosmer-Lemeshow 检验结果显著($P = 1.183e-05 < 0.05$)，这主要归因于本研究样本量较大($N > 10000$)，使得该统计量对细微偏差极度敏感。然而，观察校准曲线可以发现，模型预测值与实际观察值在总体趋势上保持了高度一致，本文认为该模型在捕捉消费者购买意图方面具备足够的稳健性与实践参考价值。

5. 结论与建议

本研究通过对 12,330 条电子商务用户会话数据构建 Firth 逻辑回归模型分析发现，实时行为特征是决定用户转化意向的关键逻辑所在，其中页面价值(PageValues)以 4.5765 的优势比(OR 值)成为显著的正向驱动力，而退出率(ExitRates)则呈现出显著的负向抑制效应。外部情境分析证实，针对具有显著“大促心理预期”的特定时间窗口，系统可利用模型识别的高价值会话进行动态干预，致使转化概率出现跨越式增长。

基于此，本文提出以下建议：首先，构建基于“意图信号”的实时监测与智能化干预机制。电商平台应将页面价值作为识别高意向会话的实时信号，利用数字人等 AI 技术开展精准内容介入，通过提供即时咨询或者个性化优惠来促成临门一脚的购买产品转化；其次，企业需重视访客类型存在的非对称性影响，对于表现出“持币观望”心理且对路径顺畅度极为敏感的回访者，应优化购物链路的完整性，并在促销前期借助预热策略精准释放其消费潜力；最后，建立动态的流失预警与差异化留存体系。针对模型识别出的高退出率路径，平台需区分用户属性实施差异化补救。本研究虽基于 UCI 基准数据集验证了算法的稳健性，但仍存在研究不足之处，仍需进一步接入近三年的实时脱敏数据进行模型重校，以捕捉内容电商等新业态下的行为特征，从而在动态的市场竞争中实现用户价值的持续增长。

参考文献

- [1] 曹志斌, 郑森. 电子商务平台的用户行为分析与个性化推荐策略研究[J]. 商场现代化, 2025(23): 45-47.

-
- [2] Novak, T.P., Hoffman, D.L. and Yung, Y. (2000) Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Marketing Science*, **19**, 22-42. <https://doi.org/10.1287/mksc.19.1.22.15184>
- [3] 杨峰, 耿秀丽. 利用 TDGCN-L 优化电商推荐: 整合显式反馈以提高用户满意度[J/OL]. 重庆工商大学学报(自然科学版), 1-10. <https://link.cnki.net/urlid/50.1155.N.20251231.0933.002>, 2026-02-11.
- [4] 成保梅, 韩景灵. 融合情境因素的电子商务用户兴趣挖掘仿真[J]. 计算机仿真, 2020, 37(4): 326-329.
- [5] 毛明扬, 马焕坚. 基于大模型的用户画像分析与数字人个性化电商营销策略[J]. 数字技术与应用, 2025, 43(12): 144-146.
- [6] Cheng, P., Wang, W. and Yang, S. (2024) Doing the Right Thing: How to Persuade Travelers to Adopt Pro-Environmental Behaviors? An Elaboration Likelihood Model Perspective. *Journal of Hospitality and Tourism Management*, **59**, 191-209. <https://doi.org/10.1016/j.jhtm.2024.04.009>
- [7] Mehrabian, A. and Russell, J.A. (1974) *An Approach to Environmental Psychology*. The MIT Press.
- [8] 李新宇. 优衣库国际化品牌形象对中国消费者购买意愿的影响研究[D]: [硕士学位论文]. 南宁: 广西民族大学, 2023.
- [9] 孟佳惠. 诚信归位电商生态方能行稳致远——2025 年度电子商务领域“双十一”信用预警[J]. 中国信用, 2025(11): 20-29.
- [10] 韩婷. Firth 惩罚最大似然估计在 logistic 回归中解决分离问题时的应用[D]: [硕士学位论文]. 晋中: 山西医科大学, 2013.
- [11] 易莹莹, 宋锡文. 我国流动人口健康影响因素重要性的研究——基于随机森林模型实证分析[J]. 西北人口, 2020, 41(4): 15-26.