E-Commerce Letters 电子商务评论, 2025, 14(7), 2525-2534
Published Online July 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.1472463

Hans汉斯

基于集成学习的电商消费者复购行为预测及可解释性分析

李　盼，杨继宇

贵州大学数学与统计学院，贵州　贵阳

收稿日期：2025年6月13日；录用日期：2025年6月25日；发布日期：2025年7月29日

摘　要

随着大数据时代的到来，电商行业的发展迎来了全新的机遇和挑战，基于电商平台的海量用户行为数据，如何深入挖掘并精准预测消费者的复购行为，对电商平台提高客户忠诚度和运营效果有着重要意义。本文基于阿里巴巴天池大赛提供的电商消费者行为数据，旨在构建一个基于集成学习与SHAP解释机制的复购行为预测模型，从而对消费者复购行为进行预测与探究。在数据预处理阶段，本研究首先对原始数据集进行缺失值处理；针对数据不平衡问题，本文采用随机下采样方法对非重复购买样本进行下采样，从而提高模型对重复购买样本的识别能力。接着，通过特征提取和特征选择操作，最终选择了18个特征作为后续电商消费者复购行为预测与模型优化分析，以全面刻画用户在平台上的行为广度、多样性、活跃度与平台黏性。随后，基于梯度提升决策树(GBDT)、极致梯度提升(XGBoost)、随机森林算法(RF)、K近邻(KNN)和决策树算法(DT)这五种基分类器，采用贝叶斯优化对各模型的超参数进行自动调优，并运用Stacking集成学习策略构建最终预测模型。实验结果表明，Stacking集成模型在Accuracy (0.7750)、Recall (0.8677)、Precision (0.7824)及F1值(0.8228)指标上均优于单一模型，具有更好的泛化能力。最后，通过SHAP解释性分析识别出影响复购的关键因素：用户行为多样性特征(cat_unique、action_2_freq、item_unique)对复购预测具有显著正向影响，反映消费行为广度与深度对用户忠诚度的促进作用；用户画像特征(age_range、gender)及部分计数特征(time_count、cat_count)对预测贡献不显著。本研究为电子商务平台通过数据驱动方法优化个性化营销策略、提高客户留存率提供了具有实践指导意义的决策支持。

关键词

集成学习，SHAP解释性方法，电子商务，复购行为

Prediction and Interpretability Analysis of E-Commerce Consumers' Repurchase Behavior Based on Ensemble Learning

Pan Li, Jiyu Yang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

文章引用: 李盼, 杨继宇. 基于集成学习的电商消费者复购行为预测及可解释性分析[J]. 电子商务评论, 2025, 14(7): 2525-2534. DOI: 10.12677/ecl.2025.1472463

Received: Jun. 13th, 2025; accepted: Jun. 25th, 2025; published: Jul. 29th, 2025

Abstract

With the advent of the big data era, the e-commerce industry is facing unprecedented opportunities and challenges. Leveraging massive user behavior data from e-commerce platforms, in-depth exploration and accurate prediction of consumer repurchase behavior have become crucial for enhancing customer loyalty and operational effectiveness. This study utilizes e-commerce consumer behavior data from the Alibaba Tianchi Competition to construct a repurchase prediction model based on ensemble learning and SHAP interpretability mechanisms, aiming to investigate and forecast consumer repurchase behavior. During the data preprocessing phase, the study first addresses missing values in the original dataset. To tackle the class imbalance issue, random under-sampling is applied to non-repurchase samples, thereby improving the model's ability to identify repurchase cases. Through feature extraction and selection, 18 features are ultimately selected to comprehensively characterize users' behavioral breadth, diversity, activity level, and platform engagement for subsequent repurchase prediction and model optimization analysis. Subsequently, based on five base classifiers—Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGBoost), Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree (DT)—we employed Bayesian optimization for automated hyperparameter tuning of each model and adopted a Stacking ensemble learning strategy to construct the final predictive model. Experimental results demonstrate that the Stacking ensemble model outperforms individual models across multiple evaluation metrics, including Accuracy (0.7750), Recall (0.8677), Precision (0.7824), and F_1-score (0.8228), exhibiting superior generalization capability. Finally, SHAP interpretability analysis identifies key factors influencing repurchase: user behavior diversity features (cat_unique, action_2_freq, item_unique) show significant positive effects on repurchase prediction, reflecting how behavioral breadth and depth enhance customer loyalty. In contrast, demographic features (age_range, gender) and certain count-based features (time_count, cat_count) contribute minimally to prediction accuracy. This research provides e-commerce platforms with data-driven decision support for refining personalized marketing strategies and improving customer retention, offering practical guidance for business applications.

Keywords

Ensemble Learning, SHAP Interpretability Method, E-Commerce, Repurchase Behavior

1. 引言

随着大数据时代的到来，电商行业的发展迎来了全新的机遇和挑战。1 月 17 日，中国互联网络信息中心发布的第 55 次《中国互联网络发展状况统计报告》显示，截至 2024 年 12 月，我国网民规模达 11.08 亿人，互联网普及率达 78.6%；网络支付用户规模达 10.29 亿人，网络购物用户规模达 9.74 亿人，网上零售额、移动支付普及率稳居全球第一。这一现象表明，电商行业正处于高速发展的阶段，消费者的行为模式和需求也变得日益复杂。为此，基于电商平台的海量用户行为数据，如何深入挖掘并精准预测消费者的复购行为，对电商平台提高客户忠诚度和运营效果有着重要意义。

复购率作为衡量客户忠诚度的重要指标，用其去预测消费者复购行为已经成为国内外学者的热点研究。张李义[1]等研究表明，相比于其他模型，用 XGBoost 模型去预测消费者重复购买意愿，能够更好地

捕捉消费者的购买行为特征，进而显著提升复购行为的预测效果。巫月娥[2]从网络品牌的角度研究了复购行为的影响因素，强调了通过理解消费者复购决策的驱动因素来进行精准营销的重要性。Tingting L 等[3]提出了一种深度行为与情感感知的个性化推荐模型，通过结合动态用户行为建模和情感分析，构建了基于协同过滤与深度学习的混合推荐框架。该模型能自适应变化的用户偏好和情感状态，有效提升推荐准确性和消费预测效果。但在现有的研究中，对复购行为进行预测仍然存在一些局限性，如数据隐私、模型解释性不足等相关问题。

为了克服模型“黑箱”问题并提高模型的透明度与可信度，引入模型可解释性分析显得尤为重要。通过对模型进行可解释性分析，不仅能够揭示模型的决策过程，还能帮助电商平台理解哪些因素在消费者行为预测中起到关键作用，从而为业务决策提供更有依据的支持。纪守领[4]等总结了现有的机器学习模型可解释性研究，并探讨了多种可解释性技术的应用，特别是局部可解释模型(LIME)和 SHAP 值等方法，这些技术为电商平台提供更透明、更可信的模型结果。

为此，本研究在 GBDT、XGBoost、随机森林算法(RF)、KNN 和决策树算法(DT)共 5 种基分类器的基础上，基于融合思想构建集成分类器，旨在对电商消费者的复购行为进行预测，并结合 SHAP 解释性方法对模型进行可解释性分析，深入探讨影响消费者复购行为的关键因素，在提升模型性能的同时，为业务人员提供具备解释力的智能决策支持。

2. 数据来源与特征选择

2.1. 数据来源与预处理

本研究采用的数据集来源于阿里巴巴天池大赛，包含匿名用户在“双十一”购物节前 6 个月及当天的完整购物记录，其中是否重复购买为分类标签。数据集主要由天猫平台的用户画像和用户行为日志两大部分组成，对数据进行预处理后，按照用户唯一 ID 对数据进行整合，最终获得 260,864 条样本，数据集的主要字段及其详细说明见表 1 所示。其中，非重复购买者样本 244,912 例，重复购买者样本 15,952 例，仅占 6.12%，存在显著的类别不平衡问题，具体分布如图 1 所示。

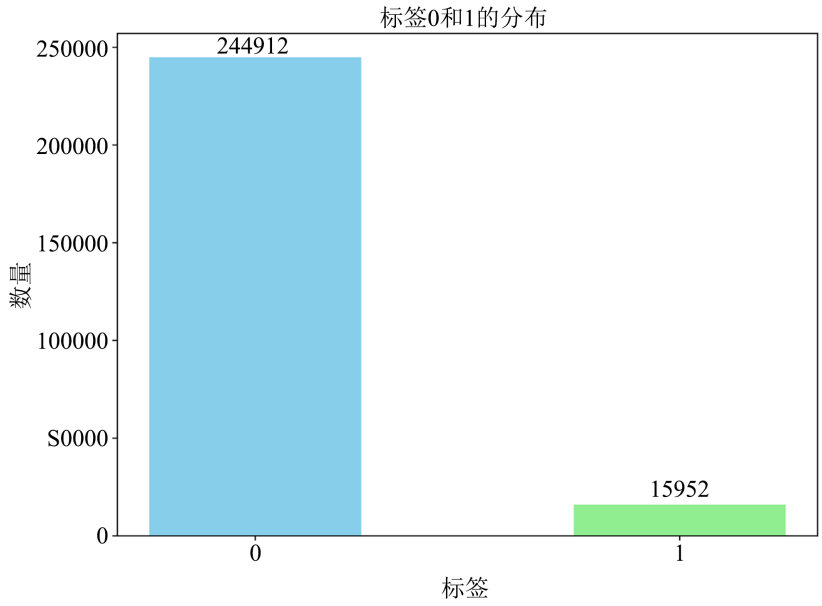

Figure 1. The class distribution of samples in the dataset
图 1. 数据集中样本类别分布

Table 1. A detailed description of the key variables in the dataset
表 1. 数据集主要字段描述

数据类型	字段名称	描述
用户行为日志	user_id	购物者的唯一 ID 编码
	item_id	商品的唯一编码
	cat_id	商品所属品类的唯一编码
	merchant_id	商家的唯一 ID 编码
	brand_id	商品品牌的唯一编码
	time_tamp	购买时间(格式：mmdd)
	action_type	包含{0, 1, 2, 3}，0 表示单击，1 表示添加到购物车，2 表示购买，3 表示添加到收藏夹
用户画像	user_id	购物者的唯一 ID 编码
	age_range	用户年龄范围。1：<18 岁；2：[18, 24]；3：[25, 29]；4：[30, 34]；5：[35, 39]；6：[40, 49]；7：≥50 岁；8：≥50 岁；0 和 Null 表示未知
	gender	用户性别。0 表示女性，1 表示男性，2 和 Null 表示未知
目标变量	label	包含{0, 1}，1 表示重复买家，0 表示非重复买家

在数据预处理阶段，针对用户画像特征年龄范围 gender，0 和 Null 均表示未知，统计特征值为 0 和 Null 的样本共有 95,131 条，见图 2，占总样本数据的 22.43%，不能将其直接删除，为此我们用 0 填充 Null。对于用户画像特征 gender，2 和 Null 均表示未知，表 2 结果显示性别未知的数据有 16,862 条，占总样本数据的 3.98%，同理用 2 填充 Null。对于用户行为日志数据，只有特征 brand_id 含有缺失值，共 91,015 条，占总样本数据的 0.17%，占比较小，故直接删除缺失值。按照用户唯一 ID 对数据进行整合后，将其按照 8:2 的比例随机划分为训练集和测试集。针对训练集中的类别不平衡问题，我们保留了全部重复购买样本，并对非重复购买样本采用随机下采样方法进行处理，最终使两类样本比例达到 1:1 的平衡状态。

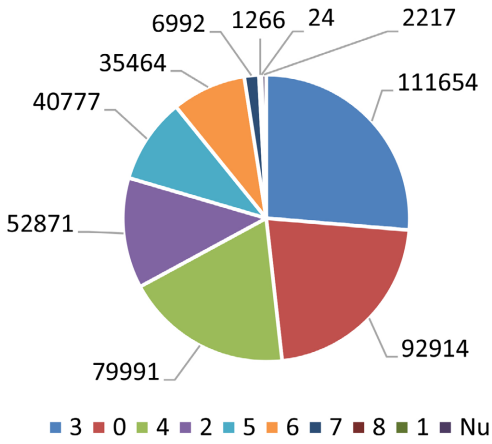

Figure 2. Distribution of age ranges within user profile data
图 2. 用户画像年龄范围分布

Table 2. Distribution of gender within the user profiling features
表 2. 用户画像特征性别分布

gender 取值	数量
0	285,638
1	121,670
2	10,426
Null	6436

2.2. 特征提取与特征选择

特征工程是机器学习中至关重要的一部分，其主要目的是通过对原始数据进行预处理、转换和提取有用的特征，以便为模型提供有效的信息，提升模型性能。在处理电商消费者行为数据时，合理的特征提取尤为关键，它能够帮助模型更好地理解消费者的偏好与行为模式，进而提高复购率预测的准确性。本研究针对电商平台的消费者行为数据，构建了一系列多维度的特征，以支撑复购率预测模型的有效性。

针对处理过的数据集，本文从商家、用户、用户与商家的交互信息以及用户与商品的交互信息四个方面来构建新特征。在特征提取之前我们将每个用户所访问的所有商品、商品类别、商家、品牌、用户行为类别以及购买的时间分别加入空格符进行拼接，使之成为针对每一个单一维度的用户特征，构成用户对各个维度的访问路径。接下来，从访问路径中分别提取商家、用户、用户与商家的交互信息以及用户与商品的交互信息四个方面的特征。提取的特征描述如表 3 所示：

Table 3. The extracted features along with their corresponding descriptions
表 3. 提取出的特征及描述

类别	特征	描述
商家维度	cat_count	商品类别总数
	mercgant_count	商家总数
	brand_count	品牌总数
用户维度	itme_count	浏览或购买的商品总数
	time_span	时间戳
	time_count	时间戳总数
	action_count	行为类别总数
用户与商家的交互信息	cat_unique	不同商品类别(去重后)
	mercgant_unique	不同的商家数(去重后)
	brand_unique	不同的品牌数(去重后)
	item_unique	每个用户浏览或购买的不同商品数
	user_merchant_op_count	用户与不同商家的互动次数
用户与商品的交互信息	action_0_freq	点击商品的频率
	action_1_freq	将商品添加到购物车的频率
	action_2_freq	购买商品的频率
	action_3_freq	将商品添加到收藏夹的频率

最终，删除无关特征后我们选取了 18 个特征用于后续的模型训练，具体特征如表 4 所示：

Table 4. The features derived from the feature engineering process
表 4. 特征工程提取出的特征

特征数	特征
18	age_range, gender, time_count, action_count, item_count, cat_count,merchant_count, brand_count, item_unique, cat_unique, merchant_unique, brand_unique, action_0_freq, action_1_freq, action_2_freq, action_3_freq, time_span, user_merchant_op_count

3. 模型的构建与实验结果分析

3.1. Stacking 集成学习算法

Stacking 是一种集成学习方法，通过结合多个模型的预测结果来构建一个更强的预测模型。其核心思想是将多个基模型的输出进行组合，并利用一个“元模型”进行最终的预测。在构建预测模型时，单一模型往往受到其算法局限性的制约，而通过集成多个模型，Stacking 能够有效减少过拟合，提高模型的泛化能力，从而在处理复杂问题时取得更优的性能表现。

在本研究中，我们基于随机森林(RF)、支持向量机(SVM)、K 最近邻(KNN)、决策树(DT)和梯度提升机(GBDT)这五种常用的基分类器构建 Stacking 集成学习模型。

3.2. 模型评估

本文，我们采用了多种评价指标来评估模型性能，包括准确度(Accuracy)、召回率(Recall)、精确度(Precision)、F1 值(F_1-score)、ROC 曲线和 AUC 值。这些指标能够从不同角度全面评估模型的性能，确保我们能够充分了解模型在不同场景下的表现。

下面我们利用混淆矩阵中的四个基本统计指标对上面提到的评估指标进行介绍，混淆矩阵如表 5 所示。其中：TN 表示当样本真实标签为负类且预测为负类，代表真反例数；FP 表示当样本真实标签为负类且预测为正类，代表假正例数；FN 表示当样本真实标签为正类且预测为负类，代表假反例数；TP 表示当样本真实标签为正类且预测为正类，代表真正例数。

Table 5. Confusion matrix
表 5. 混淆矩阵

混淆矩阵		模型的预测标签	
		0	1
真实标签	0	TN	FP
	1	FN	TP

1) 准确度(Accuracy)

准确度(Accuracy)用来衡量所有被预测正确的样本，计算公式如下：

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}. \tag{1}$$

2) 召回率(Recall)

召回率(Recall)用来衡量所有正类样本中被预测为真正正类的比例，计算公式如下所示：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

3) 精确度(Precision)

精确度(Precision)用来衡量被预测为正类的样本中真正正类所占的比例，计算公式如下：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3}$$

4) F_1-score

F_1-score 是综合考虑精确度和召回率的调和平均值，尤其适用于数据不平衡的情况，计算公式如下：

$$\text{F}_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

5) ROC 和 AUC

AUC 的含义是 ROC 曲线下的面积，用来衡量模型对不同类别的判别能力。AUC 的值在 0 到 1 之间，越接近于 1，代表分类器的预测能力越好。

3.3. 实验对比与结果分析

本文采用贝叶斯优化方法对五种基分类器的超参数进行自动搜索，实验经过 50 次迭代评估不同的超参数组合性能，并根据每次评估结果自动获取表现最优的超参数组合，结果见表 6。接下来在贝叶斯优化后的五个基分类器的基础上，根据 Stacking 集成学习思想构建模型，在同一测试集上进行预测，输出多个模型性能评估指标，包括准确度(Accuracy)、召回率(Recall)、精确度(Precision)、F_1-score 和 AUC。对评估指标开展横向对比分析，以揭示集成策略对个体分类器优势特征的融合能力。各指标结果如表 7 所示，实验结果显示，Stacking 集成模型在整体性能上优于单一模型，其中 Accuracy (0.7750)、Recall (0.8677)、Precision (0.7824)及 F_1-score (0.8228)达到了所有方法中的最高值。准确率(Accuracy)提升最为显著，相比次优模型 RF 和 KNN 提高了 5%，F_1-score (0.8228)达到最优，表明其在平衡召回率和精确率方面表现突出。尽管其 AUC (0.6616)略低于 KNN (0.6993)，但 Stacking 通过整合多模型的优势，显著提升了分类的稳定性和综合预测能力。

Table 6. Space of optimal hyperparameters obtained through Bayesian optimization
表 6. 贝叶斯优化的最优超参数空间

模型	超参数名	最优值
GBDT	learning_rate	0.0274
	max_depth	6
	n_estimators	150
XGBoost	learning_rate	0.0132
	max_depth	9
	n_estimators	492
RF	max_depth	11
	min_samples_leaf	3
	min_samples_split	3
	n_estimators	341

续表

KNN	n_neighbors	4
	weights	1
DT	max_depth	11
	min_samples_leaf	3
	min_samples_split	6

Table 7. Comparative analysis of the ensemble learning approach and five individual base classifiers
表 7. 集成学习方法与五种基分类器的性能对比

模型	Accuracy	Recall	Precision	F1-Score	AUC
GBDT	0.7	0.771	0.771	0.771	0.6168
XGBoost	0.7	0.771	0.771	0.771	0.6455
RF	7250	0.8355	0.7529	0.7921	0.6025
KNN	0.725	**0.8677**	0.7333	0.7949	**0.6993**
DT	0.65	0.7387	0.7387	0.7387	0.5416
Stacking 集成模型	**0.775**	**0.8677**	**0.7824**	**0.8228**	0.6616

图 3 更直观地展示了各个模型的性能，图中结果显示，Stacking 集成模型在 Accuracy、Precision、Recall 和 F_1-score (0.8228)四个核心指标上均表现最优，这与上面的分析是一致的。

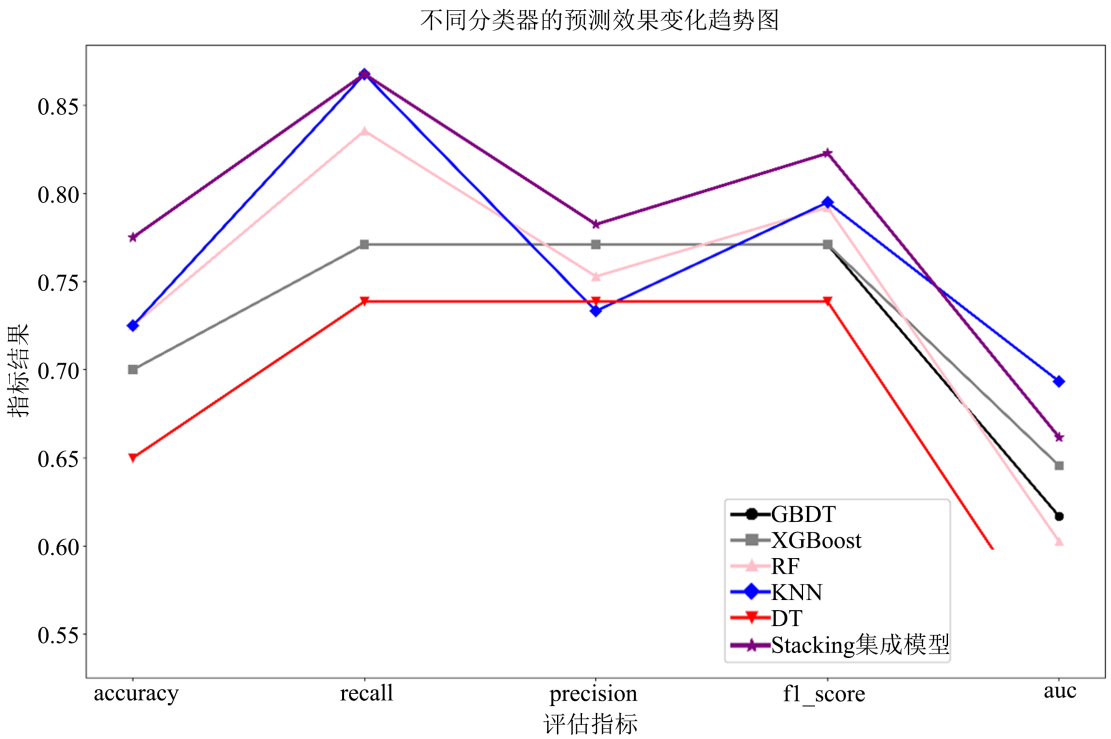

Figure 3. Comparative trends of evaluation metrics among various models
图 3. 不同模型的评估指标趋势

4. 模型可解释性分析

4.1. SHAP 解释性方法

SHAP (Shapley Additive Explanations)是一种基于博弈论、与模型无关的解释性方法，旨在提供模型的透明性和可解释性，特别是对于复杂的机器学习模型，如随机森林(RF)、梯度提升树(GBDT)和深度学习模型等。SHAP 方法的核心思想源于 Shapley 值[5]，最早由经济学家 Lloyd Shapley [6]提出，Shapley 值是一个带有正负号的数值，正负号代表该特征对于预测的输出结果的贡献是积极的还是消极的，值的大小代表该特征对模型预测结果的贡献大小。

4.2. 电商消费者复购行为预测模型的可解释性分析

为了确定各特征对电商消费者复购行为预测是积极影响还是消极影响，本研究绘制了各个特征的 SHAP 摘要图。图中的每个点代表一个样本，点的颜色代表特征的取值，颜色从蓝色到红色表示特征的取值逐渐增大。特征对应的样本点分布越靠近 0，表明该特征的重要性就越低，反之越高。

图 4 SHAP 摘要图结果显示，cat_unique、action_2_freq、item_unique、time_span 以及 merchant_unique 五个特征对模型的输出影响最大，其中 action_2_freq、cat_unique 以及 item_unique 对复购预测具有显著正向，即随着特征取值的增加，特征对模型预测样本为重复买家的正向贡献就增加，也就是说特征的取值越大将该样本预测为重复买家的概率就越大，这反映了消费行为的广度和深度对用户忠诚度的促进作用。time_span 的红点分布集中在正区间，说明长期活跃用户更可能复购，而短期用户(蓝点)对预测贡献接近零或负向。merchant_unique 红色的点大多分布在负向区域和 0 附近，可以认为随着特征取值的增加，特征对模型输出的负向贡献增加，也就是说，随着用户接触的商家数量增加会降低其复购率。

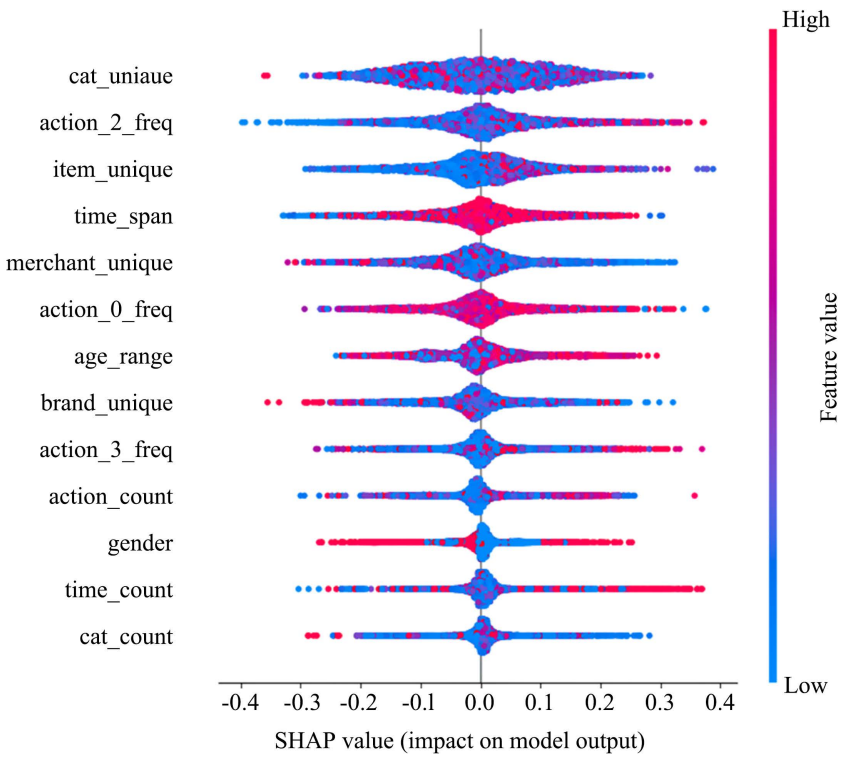

Figure 4. Summary plot of SHAP values for consumer characteristics
图 4. 消费者特征 SHAP 摘要

brand_unique、age_range、gender、action_count 等其他特征对模型预测贡献较低，其中用户画像信息 age_range、gender 样本点集中在 SHAP 值靠近 0 的区域，且蓝红混杂，说明它们对复购预测的区分度较弱；time_count 和 cat_coun 两个特征的分布较对称，对模型结果输出影响不显著。

5. 结论

本研究构建了基于集成学习和可解释性分析相结合的电商消费者复购行为预测模型框架。实验结果表明，Stacking 集成模型相较于 GBDT、XGBoost、随机森林算法(RF)、KNN 和决策树算法(DT)这五个基分类器显著提升了模型性能，其中准确率(Accuracy)提升最为显著，相比次优模型 RF 和 KNN 提高了 5%，F_1-score 相对于次优模型 KNN 提升了 2.79%。Stacking 集成模型的多项指标优势表明，其能够有效整合基分类器的互补性优势，突破单一模型的性能瓶颈，增强模型对复购样本的识别能力，从而优化了商家对具有复购意愿的用户制定营销策略的途径。对于模型可解释性分析来说，本文采用 SHAP 方法对 Stacking 集成模型从全局维度进行了解释性探讨。研究结果表示，用户行为多样性特征(cat_unique, action_2_freq, item_unique)对复购预测具有显著正向影响，反映消费行为广度与深度对用户忠诚度的促进作用；time_span 特征与复购行为大致呈正相关，即长期活跃用户复购概率更高；merchant_unique 特征值的增加可能会降低消费者的复购行为，即随着用户接触的商家数量增加会降低其对商品的复购意愿；用户画像特征(age_range, gender)及部分计数特征(time_count, cat_count)对模型输出贡献不显著。基于此，平台商家可以重点针对高多样性、高活跃度和高购买频次的用户群体，制定差异化的营销和留存策略。

参考文献

[1] 张李义, 李一然, 文璇. 新消费者重复购买意向预测研究[J]. 数据分析与知识发现, 2018, 2(11): 10-18.

[2] 巫月娥. 网络品牌视角下网络消费者重复购买的营销策略[J]. 企业经济, 2013, 32(1): 105-108.

[3] Li, T.T., Wu, Y.L., *et al*. (2025) Deep Learning-Based Analysis of E-Commerce Enterprises: User Behavior and Consumption Prediction. *Journal of Organizational and End User Computing* (*JOEUC*), **37**, 1-36. https://doi.org/10.4018/JOEUC.379722

[4] 纪守领, 李进锋, 杜天宇, 等. 机器学习模型可解释性方法、应用与安全研究综述[J]. 计算机研究与发展, 2019, 56(10): 2071-2096.

[5] Liu, Y., Liu, Z., Luo, X. and Zhao, H. (2022) Diagnosis of Parkinson's Disease Based on SHAP Value Feature Selection. *Biocybernetics and Biomedical Engineering*, **42**, 856-869. https://doi.org/10.1016/j.bbe.2022.06.007

[6] Levine, D.K. (2018) Introduction to the Special Issue in Honor of Lloyd Shapley: Eight Topics in Game Theory. *Games and Economic Behavior*, **108**, 1-12. https://doi.org/10.1016/j.geb.2018.05.001