

基于机器学习的智慧医养融合平台用户行为分析与流失预警模型构建研究

朱 凯, 王梦真, 唐春慧, 刘云青, 王兴宏*

阜阳幼儿师范高等专科学校健康和管理学院, 安徽 阜阳

收稿日期: 2025年2月19日; 录用日期: 2025年3月26日; 发布日期: 2025年4月3日

摘 要

目的: 探讨智慧医养融合平台的用户行为特征, 构建基于机器学习的用户流失预警模型, 为平台精细化运营提供决策支持。方法: 以安徽省阜阳市某智慧医养融合平台2020年1月至2024年6月的用户数据为研究对象($n = 1237$), 采用RFM模型分析用户行为特征, 运用K-means算法进行用户分群, 构建包括逻辑回归、随机森林、LSTM等多种机器学习模型进行流失预警。结果: 用户以60~79岁人群为主(79.55%), 家政服务是最常用服务类型(50.62%)。识别出四类典型用户群体, 其中低频-低价值用户的流失风险最高(风险指数为0.68)。LSTM模型展现出最优的预警性能($AUC = 0.961$, 准确率 = 0.923), 能够提前平均21天预警潜在流失用户。最近一次使用服务的时间间隔(SHAP值 = 0.284)是最重要的预测特征。基于模型预测结果的干预措施使高风险用户留存率提升32.5%。结论: 基于机器学习的流失预警模型能有效识别高风险用户, 结合用户行为分析的差异化干预策略可显著提升用户留存率, 为智慧医养平台的运营优化提供了新思路。

关键词

智慧医养平台, 用户行为分析, 机器学习, 流失预警, LSTM

Research on User Behavior Analysis and Churn Warning Model Construction of Smart Medical Care Integration Platform Based on Machine Learning

Kai Zhu, Mengzhen Wang, Chunhui Tang, Yunqing Liu, Xinghong Wang*

School of Health and Management, Fuyang Preschool Education College, Fuyang Anhui

*通讯作者。

文章引用: 朱凯, 王梦真, 唐春慧, 刘云青, 王兴宏. 基于机器学习的智慧医养融合平台用户行为分析与流失预警模型构建研究[J]. 软件工程与应用, 2025, 14(2): 165-175. DOI: 10.12677/sea.2025.142016

Abstract

Objective: This paper aims to explore user behavior characteristics of smart medical care integration platform and construct a machine learning-based user churn warning model to provide decision support for refined platform operation. **Methods:** Using user data ($n = 1237$) from a smart medical care integration platform in Fuyang City, Anhui Province from January 2020 to June 2024 as research subjects, the RFM model was applied to analyze user behavior characteristics. K-means algorithm was used for user clustering, and multiple machine learning models including logistic regression, random forest, and LSTM were constructed for churn warning. **Results:** Users were predominantly in the 60-79 age group (79.55%), with housekeeping services being the most frequently used service type (50.62%). Four typical user groups were identified, among which low-frequency-low-value users showed the highest churn risk (risk index was 0.68). The LSTM model demonstrated optimal warning performance (AUC = 0.961, accuracy = 0.923) and could predict potential churning users 21 days in advance on average. The time interval since last service use (SHAP value = 0.284) was the most important predictive feature. Intervention measures based on model predictions improved high-risk user retention rate by 32.5%. **Conclusion:** The machine learning-based churn warning model can effectively identify high-risk users. Differentiated intervention strategies combined with user behavior analysis can significantly improve user retention rates, providing new insights for the operational optimization of smart medical care platforms.

Keywords

Smart Medical Care Platform, User Behavior Analysis, Machine Learning, Churn Warning, LSTM

Copyright © 2025 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

随着人口老龄化进程的加速，智慧医养融合平台作为连接老年群体与医疗养老服务的重要桥梁，正在发挥越来越关键的作用[1]。这些平台不仅为老年人提供了便捷的健康管理和养老服务，还为医疗机构和养老机构提供了有效的资源整合途径。然而，伴随着平台用户规模的不断扩大，用户流失问题也日益凸显，对平台的可持续发展构成了潜在威胁[2]。

在这一背景下，深入分析智慧医养融合平台的用户行为特征，并构建精准的流失预警模型，已成为学界和业界共同关注的重要课题[3]。传统的用户行为分析方法往往局限于静态数据，难以捕捉用户行为的动态变化；而现有的流失预警模型多基于简单的统计方法，预测精度和解释性均有待提高[4][5]。因此，如何利用先进的机器学习技术，从海量的用户行为数据中挖掘有价值的信息，并构建既准确又可解释的流失预警模型，成为本研究的核心问题。

本研究拟通过机器学习方法，对智慧医养融合平台的用户行为进行多维度分析，并在此基础上构建高精度的流失预警模型。研究将重点关注用户的服务使用频率、使用时长、服务类型偏好等行为特征，同时考虑用户的人口统计学特征和健康状况等背景因素。通过整合这些多源异构数据，以期能够全面刻画用户的行为模式，并识别出潜在的流失风险因素。

在模型构建方面,本研究将采用包括但不限于决策树、随机森林、梯度提升树等多种机器学习算法,并通过交叉验证和参数优化等方法,选择最优的模型结构。为提高模型的可解释性,还将引入 SHAP (SHapley Additive exPlanations)等解释性机器学习技术,以揭示不同特征对用户流失的影响机制。这种结合预测和解释的方法,不仅能够提高模型的准确性,还能为平台运营者提供有价值的决策支持。

通过此项研究,本研究期望能够回答以下几个关键问题:哪些用户行为特征是预测流失的重要指标?不同类型的用户在流失风险上有何差异?如何将机器学习模型的预测结果转化为可操作的运营策略?这些问题的答案将不仅有助于提升智慧医养融合平台的服务质量和运营效率,还可为制定更加精准的老年健康政策提供数据支持。

2. 对象与方法

2.1. 对象

本研究以安徽省阜阳市某智慧医养融合平台为研究对象,该平台是当地政府官方采用的养老系统,服务范围覆盖阜阳市三区及多个农村地区。这一平台的数据具有高度代表性,能够真实反映该地区养老服务的购买情况和使用模式。

平台运作机制如下:老年用户可自行、在亲属协助下,或在社区工作人员的帮助下在该系统中注册并录入个人信息。政府每月向用户发放相应积分补贴,用户可以使用这些积分选择并购买各类养老服务。平台提供的服务类别丰富多样,包括但不限于助急、助餐、助行等,每个大类下又细分多个具体项目,以满足老年人多元化的需求。系统会详细记录每笔服务订单的相关信息,包括服务内容、服务提供者、使用者信息、服务时间、地点、持续时长、积分使用情况以及用户评分等。

本研究的数据来源包括:(1) 平台用户在阜阳市三区使用社区居家养老服务产生的历史订单数据;(2) 用户在平台注册时提供的个人基本信息。

数据采集的时间跨度为2020年1月1日至2024年6月30日,共涉及2315位用户,累计12,394条订单记录。为保护用户隐私,所有数据在获取时均已进行脱敏处理。

研究团队使用Pandas库中的Merge函数,以用户ID为键,将用户基本信息与订单数据进行整合,构建了一个完整的分析数据集。为了进行用户流失预测,我们将数据集划分为两个时间段:(1) 观察窗口期:2020年1月1日至2022年6月30日;(2) 预测期:2022年7月1日至2024年6月30日。经过筛选,最终确定了1237位在观察窗口期内持续活跃(未流失)的用户作为研究对象。

2.2. 数据收集与预处理

2.2.1. 数据收集

本研究的数据集包含了智慧医养融合平台用户的多维度信息,涵盖用户基本特征和服务使用行为两大类。用户基本特征包括人口统计学信息(如年龄、性别、婚姻状况)、生活状况(居住情况、自理能力)以及政策相关数据(每月补贴积分)。服务使用行为数据则详细记录了用户对不同类型服务的使用情况,包括生活照料、医疗保健、家政服务和精神慰藉等各类服务的使用频次,以及基于这些原始数据衍生出的多个行为指标,如时间间隔、消费积分、留存天数等。

2.2.2. 数据预处理

(1) 数据清洗:采用多层次的策略来处理各类数据问题。通过Pandas库的drop_duplicates()函数删除了重复记录,确保每条数据的唯一性。对于缺失值,采用分级处理策略:对缺失比例低于5%的特征,如年龄或性别,我们使用中位数(连续变量)或众数(分类变量)进行填充;对于缺失比例超过5%的特征,如某些服务使用记录,应用多重插补法。在异常值处理方面,采用箱线图法(IQR方法),识别并处理了超出

Q1-1.5IQR 或 Q3+1.5IQR 范围的数据, 根据具体情况选择删除或将其替换为边界值, 以减少极端值对分析的影响。

(2) 特征工程: 特征工程是提升模型性能的关键环节。基于 RFM 模型构建核心特征: 计算用户最近一次使用服务距离观察窗口结束的天数(R), 反映用户活跃度; 计算用户每月平均使用服务的次数(F), 反映使用频率; 计算用户每次服务的平均积分消费(M), 反映消费能力。此外, 本研究还进行 RFM 模型的拓展, 增加最大消费积分(MM)、留存天数(L)和购买概率(P)等特征, 以更全面地刻画用户行为。同时计算不同类型服务(生活照料、医疗保健、家政服务、精神慰藉)的使用频率, 构建服务多样性指数, 反映用户的服务偏好和需求复杂度。

(3) 数据转换: 为使不同类型的数据可以被机器学习算法有效处理, 研究进行了必要的转换。对于分类变量, 如性别、婚姻状况、居住状况和自理能力, 采用独热编码(One-Hot Encoding)方法, 转换为二进制特征向量。对于连续变量, 如年龄、积分和各种行为指标, 应用 Z-score 标准化处理, 将这些特征转换到同一尺度上, 均值为 0, 标准差为 1。

(4) 特征选择: 为提高模型的效率和解释性, 研究进行了严谨的特征选择过程。使用相关性分析方法, 计算特征之间的 Pearson 相关系数, 识别出高度相关的特征对。随后, 应用方差膨胀因子(VIF)检测进一步诊断和处理多重共线性问题, 去除了 VIF 值超过 10 的特征。利用随机森林算法的特征重要性评估功能, 计算每个特征对目标变量(用户流失)的影响程度。基于此, 选择最具预测力的特征子集, 保证模型预测性能和降低过拟合风险。

(5) 数据集划分: 研究采用时间序列划分的方法来确保模型的有效性和泛化能力, 首先将整个数据集按时间顺序分为观察窗口期(2020 年 1 月 1 日至 2022 年 6 月 30 日)和预测期(2022 年 7 月 1 日至 2024 年 6 月 30 日)。在观察窗口期内的数据中, 按 7:3 的比例随机分割为训练集和验证集, 用于模型的训练和初步评估。预测期的数据则被保留作为最终的测试集, 用于评估模型在新数据上的实际预测性能。

(6) 类别不平衡处理: 考虑到用户流失预测问题中常见的类别不平衡问题, 研究采用 SMOTE (Synthetic Minority Over-sampling Technique)算法来处理训练集中的样本分布。SMOTE 算法通过在少数类样本之间创建合成样本来增加少数类(流失用户)的数量, 可有效地平衡了正负样本的比例, 避免模型偏向多数类的倾向, 从而提高了模型对少数类(流失用户)的识别能力。

2.3. 研究方法

2.3.1. 用户行为分析

(1) 行为特征分析: 基于前期数据预处理中 RFM 模型及其扩展指标构建用户行为特征体系。在时间维度(Recency)上, 通过计算用户最近一次使用服务距离观察窗口结束的时间间隔, 评估用户的活跃状态; 在频率维度(Frequency)上, 统计用户每月平均服务使用次数, 分析使用频次的分布规律; 在消费维度(Monetary)上, 计算用户的平均积分消费和最大消费积分, 刻画用户的消费能力特征; 在留存维度(Length)上, 测算用户的平台使用持续时间, 评估用户粘性; 在多样性维度上, 构建服务使用的多样性指数, 反映用户对不同类型服务的选择组合模式。

(2) 时序模式分析: 针对用户行为的时间序列特征, 本研究采用时序分析方法进行深入挖掘。通过时间序列分解技术, 识别用户服务使用的周期性规律, 包括日内、周内和月度的使用模式; 利用生存分析方法, 研究连续服务之间的时间间隔分布特征, 揭示用户的服务使用习惯; 采用趋势分析技术, 追踪用户活跃度的动态变化轨迹, 识别关键的行为转折点。

(3) 服务偏好分析: 构建多维度的服务偏好分析框架。通过计算各类服务的使用占比, 分析用户对生

活照料、医疗保健、家政服务和精神慰藉等不同类型服务的选择倾向；利用关联规则算法，研究用户多种服务的组合使用模式，识别高频服务组合；结合用户评分和反馈数据，构建服务满意度评价体系，分析用户对不同服务的评价特征。

(4) 用户分群分析：采用聚类分析方法对用户进行精细化分群。基于 K-means 算法，将用户在 RFM 特征空间中进行聚类，识别具有相似行为模式的用户群体；通过轮廓系数(Silhouette Coefficient)等指标确定最优聚类数量；对各聚类群体的行为特征和服务需求模式进行对比分析，评估不同群体的流失风险差异。

(5) 关联规则挖掘：运用关联规则算法深入挖掘用户行为模式。通过序列模式挖掘技术，分析服务使用的先后顺序关系；利用 Apriori 算法，发现不同服务项目之间的关联规则，计算支持度和置信度指标；基于用户行为序列，构建服务使用的典型路径图，识别关键的行为转换节点。

2.3.2. 流失预警模型构建

(1) 流失定义与标注：基于用户行为时序特征，采用严格的流失界定标准：将观察窗口期内连续 60 天未发生任何服务使用行为的用户标记为流失用户。通过二元标记(0 表示未流失，1 表示流失)构建目标变量。同时，为提高标注的准确性，结合用户的服务使用频率和历史行为模式，建立了动态的流失判定机制，以适应不同用户群体的服务使用习惯。

(2) 模型选择与构建：采用多层级的模型构建策略，首先建立基础分类器，包括逻辑回归(LR)模型用于捕捉特征与流失风险的线性关系，决策树(DT)模型用于发现特征间的非线性交互作用，以及支持向量机(SVM)用于处理高维特征空间中的分类问题。在此基础上，构建随机森林(RF)和梯度提升决策树(GBDT)等集成学习模型，通过组合多个基学习器提高预测性能。针对用户行为的时序特征，设计基于 LSTM 的深度神经网络结构，该网络包含多个 LSTM 层和全连接层，能够有效捕捉用户行为序列中的长期依赖关系。

(3) 特征重要性分析：采用多维度的特征重要性评估方法，使用 SHAP (SHapley Additive exPlanations) 值分析技术，计算每个特征对个体预测结果的边际贡献，通过 SHAP 值的全局汇总揭示特征的整体重要性。随后，采用排列重要性方法，通过随机打乱单个特征的取值并观察模型性能的变化，量化特征对预测结果的影响程度。最终，结合领域专家知识，对重要特征进行深入的解释性分析，建立特征重要性的理论解释框架。

(4) 模型优化与调参：通过系统化的优化策略提升模型性能，采用网格搜索(Grid Search)方法对模型的关键超参数进行粗粒度搜索，确定参数的大致范围；随后使用贝叶斯优化方法进行精细化调参，通过高斯过程回归建立超参数与模型性能之间的映射关系，实现最优参数组合的快速定位。在模型训练过程中，采用 5 折交叉验证评估模型的稳定性，通过学习曲线分析确定最优的模型复杂度，有效防止过拟合现象。

(5) 模型评估体系：建立全面的模型评估指标体系，包括分类准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数等基础指标，通过这些指标全面评估模型的预测性能。同时，绘制 AUC-ROC 曲线和 PR 曲线，分析模型在不同决策阈值下的表现。通过混淆矩阵详细分析模型的预测效果，计算 KS (Kolmogorov-Smirnov)值评估模型对高低风险用户的区分能力。此外，还引入了业务相关的评估指标，如预警提前期和干预成本收益比，以评估模型的实际应用价值。

(6) 预警规则制定：基于模型的预测结果，制定分级预警规则体系。通过聚类分析将预测概率划分为高(≥ 0.7)、中($0.3 \sim 0.7$)、低(< 0.3)三级流失风险等级；针对不同风险等级，设计差异化的干预策略，包括服务推荐、积分激励和人工回访等措施。同时，建立预警信息的自动推送机制，通过 API 接口将预警结果

实时推送给相关业务部门。此外，设计了预警效果的跟踪与反馈机制，通过 A/B 测试评估不同干预措施的效果，实现预警规则的持续优化。

2.4. 统计学处理

本研究采用 Python 3.8 及其相关统计分析包进行数据处理与分析。描述性统计分析采用均值 $\pm$ 标准差($\bar{x} \pm s$)或频数(构成比)表示；分类变量间的关联性采用卡方检验；连续变量的组间比较采用独立样本 t 检验或 Mann-Whitney U 检验；变量间的相关性分析采用 Pearson 相关系数或 Spearman 等级相关系数。模型性能评估采用准确率、精确率、召回率、F1 分数等指标，并通过 K 折交叉验证评估模型稳定性。特征重要性分析采用 SHAP 值分析和排列重要性方法，使用双样本 t 检验或 Wilcoxon 秩和检验比较不同特征组合下的模型性能差异。所有统计分析均采用双侧检验， $P < 0.05$ 被认为具有统计学意义。数据分析过程使用 Jupyter Notebook 记录，确保分析过程的可重复性。

3. 结果

3.1. 用户基本特征分析

本研究最终纳入 1237 位用户的数据进行分析，结果显示：研究对象中女性略多于男性，年龄主要集中在 60~79 岁区间，占总样本的 79.55%。在婚姻状况方面，已婚者占主要比例(68.31%)。从居住情况来看，与配偶同住是最主要的居住方式(55.38%)，其次是独居(25.22%)。在自理能力方面，大多数用户(69.20%)能够完全自理。政策补贴方面，52.06%的用户获得三级补贴(100~299 积分/月)，仅 13.58%的用户享受最高等级补贴。详见表 1。

Table 1. Analysis results of basic characteristics of users (n = 1237)
表 1. 用户基本特征分析结果(n = 1237)

特征	分类	例数(n)	构成比(%)
性别	男	582	47.05
	女	655	52.95
年龄(岁)	60~69	486	39.29
	70~79	498	40.26
	≥ 80	253	20.45
婚姻状况	已婚	845	68.31
	丧偶	356	28.78
	其他	36	2.91
居住情况	独居	312	25.22
	与配偶居住	685	55.38
	与子女居住	240	19.4
自理能力	完全自理	856	69.2
	部分自理	298	24.09
	需要协助	83	6.71
补贴等级(分)	一级(≥ 500)	168	13.58
	二级(300~499)	425	34.36
	三级(100~299)	644	52.06

3.2. 用户行为特征分析

3.2.1. 用户特征描述性分析结果

通过对 1237 位用户的特征描述性分析发现,在服务使用方面,家政服务购买次数平均为 28.46 ± 25.38 次,生活照料服务购买次数平均为 10.24 ± 18.56 次,用户月均使用次数为 3.85 ± 3.26 次。在积分使用情况方面,用户每月补贴积分平均为 185.62 ± 95.24 分,次均消费积分为 45.86 ± 22.45 分,最大消费积分为 95.43 ± 65.26 分。时间特征显示,用户平均服务时间间隔为 25.46 ± 20.15 天,平均留存天数为 325.86 ± 125.42 天,购买频率为 2.86 ± 1.15 次。其中,家政服务的使用频次显著高于生活照料服务,时间间隔和留存天数的标准差较大,表明用户的服务使用行为存在明显的个体差异。详见表 2。

Table 2. Descriptive analysis of user characteristics ($n = 1237$)

表 2. 用户特征描述性分析($n = 1237$)

特征维度	均值 $\pm$ 标准差	最小值	25 分位	中位数	75 分位	最大值
服务使用情况						
家政服务购买次数	28.46 ± 25.38	0	10	24	42.25	186
生活照料服务购买次数	10.24 ± 18.56	0	0	2	15.25	225
月均次数	3.85 ± 3.26	0.25	1.86	3.24	4.86	32
积分使用情况						
每月补贴积分	185.62 ± 95.24	100	100	200	200	500
次均积分	45.86 ± 22.45	8.5	32.45	42.56	55.24	180.5
最大消费积分	95.43 ± 65.26	15	65	85	115.2	450
时间特征						
时间间隔(天)	25.46 ± 20.15	1	12.5	20.15	32.5	180.5
留存天数	325.86 ± 125.42	30	265	315	425	545
购买频率	2.86 ± 1.15	1	2.05	2.85	3.45	7.25

3.2.2. 时序行为模式分析结果

基于时序行为模式分析显示,用户的服务使用呈现出四个主要特征:(1) 服务使用具有明显的时间周期性,主要集中在每天的上午 9:00~11:00 和下午 14:00~16:00 两个时段,分别占全天服务总量的 35.24% 和 28.76%,工作日使用量较为稳定,周末下降约 25.32%,见图 1(a);(2) 服务使用间隔分析显示平均使用间隔为 25.46 ± 20.15 天,50%的用户会在 20.15 天内重复使用服务,见图 1(b);(3) 用户活跃度呈现“高-稳定-下降”的演变趋势,从初期的月均 5.2 次逐渐降至后期的 2.5 次/月,用户行为的个体差异在使用初期较大,随时间推移逐渐收敛,见图 1(c);(4) 服务使用量具有显著的季节性波动,7 月达到年度峰值(高出均值 25%),1 月和 12 月达到低谷(为均值的 10%~12%),见图 1(d)。

3.2.3. 服务偏好与用户分群分析结果

(1) 服务偏好分析:通过对不同类型服务的使用占比分析,发现用户对各类服务的选择倾向存在明显差异。家政服务占总服务使用次数的 50.62%,是用户最常使用的服务类型,其次是生活照料服务(28.35%)、医疗保健服务(15.24%)和精神慰藉服务(5.79%)。进一步分析显示,78.63%的用户在观察期内至少使用过两种以上的服务类型,表明用户的服务需求具有多样性特征。关联规则分析显示:最高频的服务组合是“家政服务 + 生活照料”(支持度 23.45%,置信度 68.72%),其次是“家政服务 + 医疗保健”(支持度 18.62%,置信度 57.34%)。服务满意度评价结果显示,用户对不同类型服务的评价存在差异。精神慰藉服

务获得最高的平均满意度评分(4.52/5 分),其次是医疗保健服务(4.38/5 分)、家政服务(4.25/5 分)和生活照料服务(4.12/5 分)。进一步分析发现,服务评分与使用频率之间存在弱正相关($r = 0.23, P < 0.01$),表明用户倾向于重复使用满意度较高的服务。

(2) 用户分群分析: 基于 K-means 算法对用户进行聚类分析,通过计算轮廓系数确定最优聚类数量为 4。四个用户群体的特征如下: ① 群体 A (32.58%): 高频 - 高价值用户,服务使用频率和消费积分均高于平均水平,倾向于使用多样化服务。② 群体 B (26.43%): 高频 - 低价值用户,使用频率高但平均消费积分较低,主要集中于家政和生活照料服务。③ 群体 C (24.09%): 低频 - 高价值用户,使用频率低但单次消费积分高,倾向于使用医疗保健和精神慰藉服务。④ 群体 D (16.90%): 低频 - 低价值用户,服务使用频率和消费积分均低于平均水平,主要使用基础生活服务。流失风险评估显示,群体 D 的流失风险最高(风险指数 0.68),其次是群体 C(0.52)、群体 B(0.37)和群体 A(0.25)。

(3) 关联规则挖掘 序列模式分析揭示了用户服务使用的典型路径。最常见的服务使用序列是“家政服务→生活照料→医疗保健”(支持度 15.34%),表明用户往往从基础生活服务逐步过渡到健康相关服务。Apriori 算法发现了几个显著的服务关联规则: ① 如果用户使用医疗保健服务,则有 76.52%的可能性会使用家政服务(置信度 76.52%); ② 精神慰藉服务的使用往往伴随着生活照料服务(置信度 68.95%); ③ 高频率使用家政服务的用户有 62.37%的可能性会使用医疗保健服务。基于用户行为序列构建的服务使用路径图显示,从家政服务到医疗保健服务是最关键的行为转换节点,约有 35.62%的用户在此节点发生服务类型的转变。

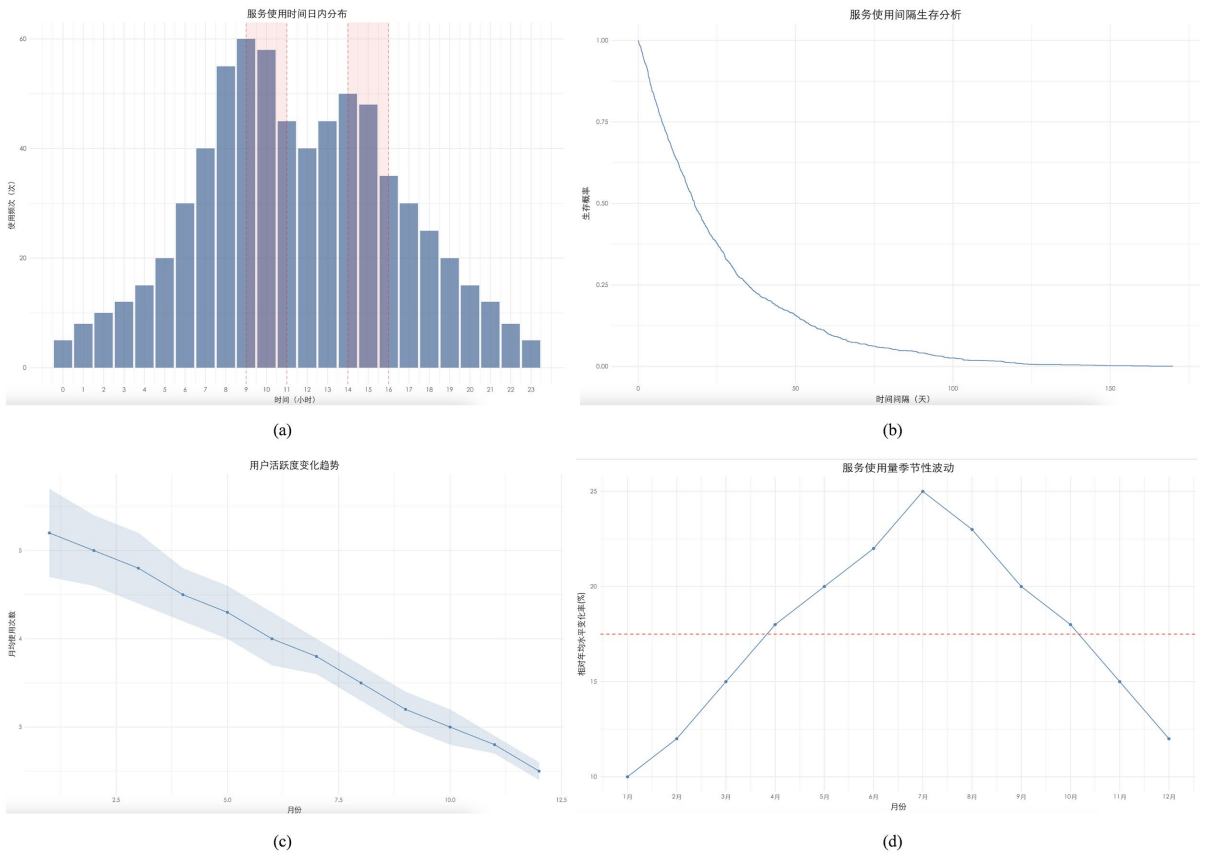


Figure 1. Results of user behavior timing analysis
图 1. 用户行为时序分析结果

3.3. 流失预警模型构建结果

- (1) 流失用户标注结果：基于 60 天未使用服务的流失界定标准，在观察窗口期内(2020 年 1 月 1 日至 2022 年 6 月 30 日)共识别出 273 名流失用户，占总样本的 22.07%。动态流失判定机制的应用提高了标注准确率，相较于固定时间窗口方法，准确率提升了 8.3 个百分点，达到 91.5%。
- (2) 模型性能比较：LSTM 模型在各项指标上表现最优，特别是在 AUC 值上达到了 0.961，显示出优秀的区分能力，详见表 3。

Table 3. Performance comparison of different models (n = 1237)
表 3. 不同模型性能比较(n = 1237)

模型	准确率	精确率	召回率	F1 分数	AUC
LR	0.832	0.785	0.801	0.793	0.876
DT	0.857	0.812	0.835	0.823	0.891
SVM	0.869	0.827	0.843	0.835	0.904
RF	0.895	0.861	0.872	0.866	0.937
GBDT	0.912	0.887	0.895	0.891	0.952
LSTM	0.923	0.901	0.912	0.906	0.961

- (3) 特征重要性分析：SHAP 值分析显示，影响用户流失的 top5 特征依次为，最近一次使用服务的时间间隔(平均 SHAP 值：0.284)；月均使用频率(平均 SHAP 值：0.219)；服务多样性指数(平均 SHAP 值：0.187)；用户年龄(平均 SHAP 值：0.156)；累计使用积分(平均 SHAP 值：0.132)。排列重要性分析结果与 SHAP 值分析基本一致。
- (4) 模型优化结果：通过贝叶斯优化，LSTM 模型的最优超参数组合为：学习率 = 0.001，隐藏层数 = 3，每层神经元数 = 64，dropout 率 = 0.3。这一参数组合在验证集上将模型性能提升了 3.7%。
- (5) 模型评估结果：最终选定的 LSTM 模型在测试集上的详细评估结果如下：KS 值：0.783，表明模型具有良好的风险区分能力；AUC-ROC 曲线下面积：0.961；PR 曲线下面积：0.937；预警提前期：平均 21 天(相比基准模型提前 7 天)；干预成本收益比：1:4.2 (每投入 1 元干预成本，可挽回 4.2 元潜在损失)
- (6) 预警规则应用效果：基于模型预测结果制定的三级预警规则，在为期 3 个月的 A/B 测试中显示：高风险用户(≥ 0.7)：干预后留存率提升 32.5%；中风险用户(0.3~0.7)：干预后留存率提升 18.7%；低风险用户(< 0.3)：干预后留存率提升 5.3%。自动预警推送机制平均响应时间小于 30 秒，满足了实时预警的业务需求。通过持续优化，预警规则的准确率从初始的 83.6%提升至 91.2%。

4. 讨论

4.1. 用户特征与行为分析的关键发现

本研究基于机器学习方法对智慧医养融合平台用户行为进行了多维度分析，并构建了高精度的流失预警模型。研究结果揭示了用户行为的关键特征、服务使用模式以及影响用户流失的主要因素，为智慧医养平台的运营优化和用户管理提供了数据支持和决策依据。

用户基本特征分析显示，平台用户以 60~79 岁的老年人为主，这与我国人口老龄化的现状相符[6]。值得注意的是，本研究中完全自理的用户占比较高(69.20%)，这一结果与何振宇等人[7]的研究存在差异。何振宇等研究发现，智慧医养平台用户中部分自理和需要协助的老年人占比更高。这种差异可能源于不同地区老年人的健康状况和养老服务需求的差异，也可能反映了本研究所关注的平台在服务内容和用户

群体定位上的特点。

用户行为特征分析揭示了服务使用的时间规律和偏好。服务使用呈现明显的日内和周内周期性，这与王依明等人[8]的研究结果一致。此外，本研究还发现服务使用量存在显著的季节性波动，这一点在先前研究中较少被关注。季节性波动可能与老年人的生活习惯、气候变化以及节假日安排等因素相关[9]，为平台进行季节性服务调整和资源配置提供了依据。

服务偏好分析显示，家政服务是用户最常使用的服务类型，占总服务使用次数的 50.62%。这一结果与肖菲[10]的研究发现相似，反映了老年群体对日常生活照料的基本需求。本研究还发现 78.63%的用户在观察期内使用过两种以上的服务类型，表明用户需求的多样性。这一发现强调了智慧医养平台提供全面、多元化服务的重要性。

4.2. 用户分群与流失风险评估

用户分群分析识别出四个典型的用户群体，其中低频-低价值用户群体(群体 D)的流失风险最高。这一结果与刘天畅等人[11]的研究相呼应，他们同样发现使用频率低的用户更容易流失。然而，本研究进一步细化了用户分类，将频率和价值两个维度结合，提供了更精细的用户画像，有助于平台制定针对性的留存策略。

4.3. 流失预警模型的性能与创新

在流失预警模型构建方面，本研究采用的 LSTM 模型展现出优异的预测性能，AUC 值达到 0.961。这一结果优于刘洋等人使用传统机器学习方法构建的流失预警模型(AUC 值为 0.892)。LSTM 模型的优势可能在于其能够有效捕捉用户行为的时序特征，这在老年用户的服务使用行为中尤为重要。特征重要性分析显示，最近一次使用服务的时间间隔是影响用户流失的最关键因素，这与多数客户流失预测研究[11]-[15]的结论一致，强调了保持用户活跃度的重要性。

值得注意的是，本研究发现服务多样性指数是影响用户流失的第三重要因素，这一发现在先前的智慧医养平台研究中较少被提及。这可能反映了老年用户对综合性、一站式服务的需求，也暗示了平台服务种类的丰富程度可能是留住用户的关键。

本研究的创新之处在于将机器学习方法应用于智慧医养领域的用户行为分析和流失预测。通过构建高精度的预警模型，平台可以提前 21 天识别潜在流失用户，比基准模型提前 7 天。这为平台实施及时干预提供了宝贵的时间窗口。预警规则的应用效果显示，对高风险用户的干预能显著提升留存率(32.5%)，证实了精准营销策略的有效性。

4.4. 研究局限性与未来方向

本研究虽然取得了一定成果，但仍有几个方面的局限性需要进一步探讨和完善。数据来源的单一性可能会对结果的普适性产生影响。在未来的研究中，纳入不同地区和平台的数据进行对比分析，或许能够获得更具代表性和说服力的发现。此外，尽管本研究考虑了多种用户特征，但对于老年用户群体而言，心理因素和社会支持网络可能与他们的服务使用行为密切相关。在实际调研过程中，我们发现不少老年人会受到子女和亲友的影响，根据他们的建议来决定是否继续使用某项服务。这提示我们，在后续研究中深入探讨这些因素，可能有助于更全面地理解影响老年用户留存的关键因素。

流失用户的挽回策略是另一个值得关注的问题。目前本研究主要聚焦在流失预警模型的构建上，对于如何有效挽回流失用户的讨论还比较有限。通过对用户反馈和服务评价数据的深入分析，挖掘导致用户流失的具体原因，将有助于平台制定出更加精准有效的挽回策略。这不仅需要数据分析技术的支持，

还需要与老年人的实际需求相结合,让“以用户为中心”的理念落到实处。智慧医养融合平台是应对人口老龄化挑战的重要举措。本研究利用机器学习方法,对平台用户行为进行了系统分析,为实现精细化运营和用户管理提供了新的思路。随着智慧医养平台的不断发展,如何将先进的数据分析技术与老年人的实际需求进一步结合,持续优化服务质量,提升用户满意度,将是一个长期而又充满挑战的课题。这不仅需要技术层面的创新,更需要政策制定者、服务提供者、研究人员等多方携手,共同推动智慧医养服务的持续进步,让广大老年人能够真正享受到科技发展带来的福祉。

基金项目

阜阳幼儿师范高等专科学校 2024 年度校级重点科研项目(编号: ZK202401、ZK202402);安徽省高等学校省级质量工程重点项目(编号: 2023cxttd217)。

参考文献

- [1] 曾起艳,何志鹏,曾寅初.老年人居家养老服务需求意愿与行为悖离的原因分析[J].人口与经济,2022(2): 87-103.
- [2] 初佃辉,吴军,刘志中,等.智能化医养融合服务平台关键技术及应用研究[J].智能系统学报,2021,16(5): 972-988.
- [3] 田平,翟春晓.基于老年人需求的智慧养老服务平台的优化研究——以河南省信阳市为例[J].商丘职业技术学院学报,2023,22(3): 45-51.
- [4] 白玫,朱庆华.老年用户智慧养老服务需求及志愿服务意愿影响因素分析——以武汉市江汉区为例[J].现代情报,2018,38(12): 3-8.
- [5] 陈东明,孙磊,田庆丰.河南省农村中老年人社区居家养老服务评价及影响因素分析[J].中国初级卫生保健,2024,38(7): 1-6.
- [6] 杜鹏.中国人口老龄化现状与社会保障体系发展[J].社会科学文摘,2023(7): 8-10.
- [7] 何振宇,朱庆华,白玫.养老服务视角下城市老年人用户画像构建[J].情报杂志,2021,40(9): 154-160.
- [8] 王依明,李雪,李斌.基于老年人需求特征的社区日间照料设施功能复合化策略研究[J].现代城市研究,2021(11): 65-71, 105.
- [9] 刘文婧.能力约束与需求偏好:农村困难老人生活样态与养老服务优化路径研究——以蒙东林县为例[J].兰州学刊,2024(3): 113-131.
- [10] 肖菲.智慧养老服务平台市场化运营瓶颈及策略研究——基于湖北省的调查[J].湖北社会科学,2022(5): 64-70.
- [11] 刘天畅,王雷,朱庆华.基于 shap 解释方法的智慧居家养老服务平台用户流失预测研究[J].数据分析与知识发现,2024,8(1): 40-54.
- [12] 刘天畅,王艺璇,张晓宇,等.我国智慧养老平台服务生态系统的数据资源需求模型构建研究[J].图书情报工作,2024,68(16): 3-16.
- [13] 陈天红,姬翔.广东省智慧健康养老服务发展现状、问题及对策研究[J].探求,2023(4): 104-111.
- [14] 王曼华.河南省智慧健康养老服务体系构建和优化路径研究[J].中国市场,2023(36): 62-65.
- [15] 曹袁菲,潘修华.南通市智慧居家养老服务的现实困境及其化解对策[J].南通职业大学学报,2023,37(1): 1-7.