

基于Stacking方法的银行客户产品认购预测

陈奕然, 魏正元*, 张亚雯

重庆理工大学数学科学学院, 重庆

收稿日期: 2025年12月9日; 录用日期: 2025年12月31日; 发布日期: 2026年1月9日

摘要

本文通过对银行客户数据的挖掘与建模, 旨在预测客户是否会购买银行产品。采用融合随机森林、LightGBM、XGBoost及多层感知机的Stacking集成学习方法, 先以四种模型作为基学习器挖掘数据中线性、非线性及复杂特征模式, 再通过逻辑回归元学习器整合优化预测结果。实验结果显示, 该集成模型预测准确性显著优于单一模型, 在客户产品认购行为预测任务中表现出色。在应用中, 基于高认购概率模型输出的重要特征与客户行为标签完成6个客群的划分, 形成多维度用户画像体系, 为精准营销与客户关系管理提供支持。

关键词

Stacking集成学习, 数据挖掘, 机器学习, 树模型, 神经网络

Bank Customer Product Subscription Prediction Based on the Stacking Method

Yiran Chen, Zhengyuan Wei*, Yawen Zhang

School of Mathematical Sciences, Chongqing University of Technology, Chongqing

Received: December 9, 2025; accepted: December 31, 2025; published: January 9, 2026

Abstract

This study aims to predict whether customers will purchase bank products through data mining and modeling of bank customer data. A Stacking ensemble learning method integrating Random Forest, LightGBM, XGBoost, and Multilayer Perceptron (MLP) is adopted: firstly, these four algorithms serve as base learners to explore linear, non-linear, and complex feature patterns in the data, and then a Logistic Regression meta-learner is employed to integrate and optimize the prediction results. Experimental results demonstrate that the proposed ensemble model significantly outperforms

*通讯作者。

single models in prediction accuracy and performs excellently in the task of forecasting customers' product subscription behaviors. In the application phase, six customer groups are divided based on the key features output by the high subscription probability model and customer behavior labels, and a multi-dimensional user portrait system is established, thereby providing robust support for precision marketing and customer relationship management.

Keywords

Stacking Ensemble Learning, Data Mining, Machine Learning, Tree-Based Model, Neural Network

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

理财产品销售已成为银行零售业务的核心收入与利润来源,如何精准把握客户购买意愿、优化营销效率,成为银行提升竞争力的关键命题。居民对资产保值的需求持续提升,而传统营销模式多依赖过往营销结果与人工经验判断,主观性强且效率低下,不仅导致营销资源浪费,还难以满足客户多样化的金融需求[1]。银行积累的海量客户交易记录与行为数据,为通过技术手段破解这一痛点提供了基础,精准的购买意愿预测模型已成为银行优化资源配置、实现个性化营销的核心支撑[2]。

国内外学者已围绕客户购买意愿预测开展诸多探索。国际研究中, HJ chang 等[3]结合聚类分析与关联规则,通过挖掘客户个人信息与购买行为特征构建预测模型,为潜在客户识别提供了新思路; YS Cho [4]基于消费频次、金额等指标设计加权算法,有效提升了消费行为预测的准确性; Kabir 等[5]则验证了随机森林算法在捕捉客户购买模式、提升预测精度中的优势。国内研究方面,陈良凯等[6]率先将研究焦点从理财产品本身转向客户购买意愿预测;金铮[7]通过对比实验证实,经梯度下降优化的逻辑回归模型在金融产品认购预测中表现优异;方鼎等人[8]将 XGboost 与 LightGBM 模型应用于银行产品认购预测,其 MAP 评测指标优于传统算法;祝歆等[9]与付长凤[10]的研究均表明,逻辑回归与支持向量机的融合模型,预测效果显著优于单一模型;余堃等[11]通过优化抽样方法改进随机森林算法,提升了大样本数据下的分析精度;宋伟等[12]融合集成学习思想构建的分类器,在时序数据处理中展现出高准确率。

尽管现有研究已验证多种机器学习算法的有效性,但仍存在明显不足[13]:多数研究对客户购买意愿的影响特征缺乏全面系统的分析,未能充分借助数据可视化手段挖掘特征间的内在关联;同时,单一算法在处理银行客户数据特征复杂、部分客群样本稀缺的场景时,易出现性能瓶颈,而集成算法的深度应用仍有待拓展。

基于此,以银行客户交易与行为数据为对象,通过数据可视化分析特征分布及关联关系,明确关键影响因素;以机器学习算法为基础,采用 Stacking 集成学习方法,构建银行客户产品购买意愿预测模型。同时结合特征重要性分析开展客户分群形成用户画像,精准呈现不同客群的购买偏好与潜力。本文旨在通过预测模型与用户画像的双重支撑,帮助银行精准识别潜在客户,为制定差异化营销策略提供决策依据,进而降低运营成本,提升营销效率。

2. Stacking 集成模型

2.1. 预测框架

本研究框架由 4 个部分组成(见图 1)。

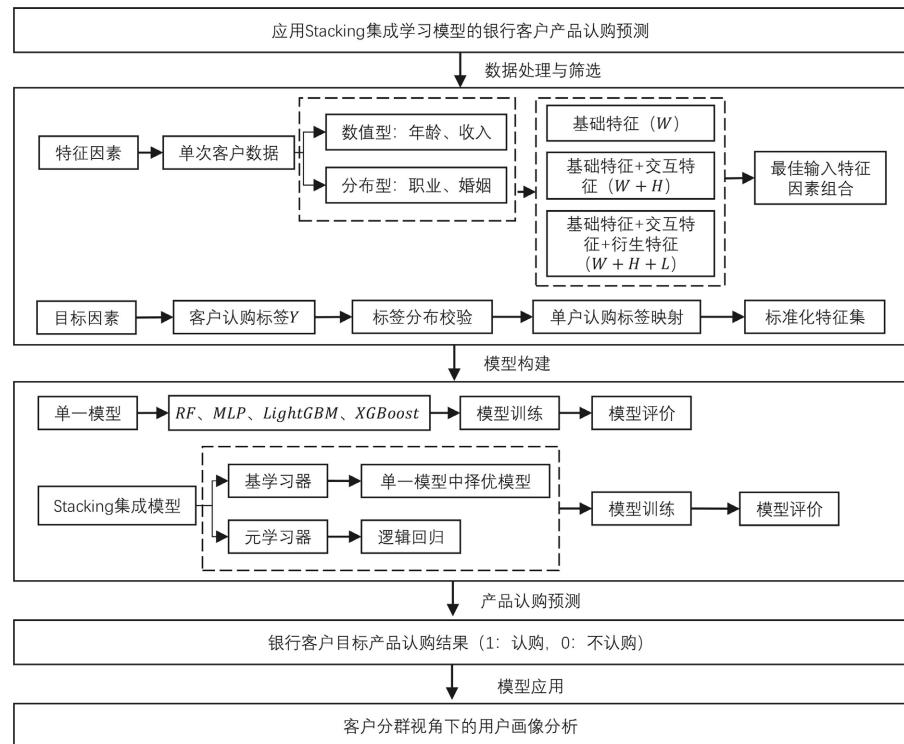


Figure 1. Framework for predicting bank customers' product subscription
图 1. 银行客户认购产品预测模型框架

- 1) 数据处理与筛选。将单次客户数据作为特征因素，字段划分为数值型 X_{num} 与分类型变量 X_{cat} ，分别采用 KS、卡方检验验证分布一致性，缺失、异常值处理后，特征工程阶段由基础特征(W)扩展到基础特征 + 交互特征($W+H$)，进一步叠加衍生特征，形成基础特征 + 交互特征 + 衍生特征($W+H+L$)，筛选最佳入模特征因素组合。将客户认购标签(subscribe)作为目标因素 Y ，先对标签分布做校验，再将其映射为单户维度的 0/1 标识，最终完成标准化特征集的构建。
- 2) 集成模型构建。采用 10 种覆盖经典分类算法、树模型以及神经网络的不同机器学习模型开展单模型预测，根据模型评价择优 4 种模型作为 Stacking 集成框架的基学习器，通过组合 4 个基学习器与 1 个元学习器来优化模型预测性能，提升银行客户产品认购行为预测精度。
- 3) 产品认购预测。基于构建的标准化特征集与 Stacking 集成模型，结合 5 折交叉验证与量化评估体系，通过多轮模型迭代优化，最终输出银行客户对目标产品的认购结果(1 代表“认购”，0 代表“不认购”)，完成产品认购预测任务。
- 4) 模型应用。结合重要特征开展客户分群视角下的用户画像分析。

2.2. 数据说明与处理

本文数据来自天池大数据众智平台(<https://tianchi.aliyun.com>)，数据集包含 22,500 条样本的训练集(train.csv)与 7500 条样本的测试集(test.csv)，数据总量共 30,000 条，各字段具体含义见表 1。

为验证训练集与测试集的变量分布一致性，对数值型变量 X_{num} 采用 KS 检验，计算样本分布 $F_{\text{sample}}(x)$ 与理论分布 $F_{\text{ref}}(x)$ 的最大偏差统计量，如公式(1)所示：

$$D = \max_x |F_{\text{sample}}(x) - F_{\text{ref}}(x)| \quad (1)$$

对分类型变量 X_{cat} 采用卡方检验, 计算实际频数 O_i 与期望频数 E_i 的拟合统计量, 如公式(2)所示:

$$\chi^2 = \sum_{i=1}^j \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

式中, j 为 X_{cat} 个数, i 为 X_{cat} 索引。表 2、表 3 检验得出, 两类变量在训练集与测试集分布整体一致。

Table 1. Bank customer data field description
表 1. 银行客户数据字段说明

字段	说明
age	年龄
job	职业: admin, unknown, unemployed, management...
marital	婚姻: married, divorced, single
default	信用卡是否有违约: yes or no
housing	是否有房贷: yes or no
contact	联系方式: unknown, telephone, cellular
month	上一次联系的月份: jan, feb, mar, ...
day_of_week	上一次联系的星期几: mon, tue, wed, thu, fri
duration	上一次联系的时长(秒)
campaign	活动期间联系客户的次数
pdays	上一次与客户联系后的间隔天数
previous	在本次营销活动前, 与客户联系的次数
poutcome	之前营销活动的结果: unknown, other, failure, success
emp_var_rate	就业变动率(季度指标)
cons_price_index	消费者价格指数(月度指标)
cons_conf_index	消费者信心指数(月度指标)
lending_rate3m	银行同业拆借率 3 个月利率(每日指标)
nr_employed	雇员人数(季度指标)
subscribe	客户是否进行购买: yes 或 no

Table 2. KS test for numerical variables
表 2. KS 检验表

数值型变量	D 统计量	p 值	显著性水平	分布是否一致
age	0.0085	0.8090	0.05	是
duration	0.0080	0.8658	0.05	是
campaign	0.0131	0.2858	0.05	是
pdays	0.0126	0.3289	0.05	是
previous	0.0052	0.9982	0.05	是
emp_var_rate	0.0065	0.9687	0.05	是
cons_price_index	0.0162	0.1041	0.05	是

续表

cons_conf_index	0.0097	0.6628	0.05	是
lending_rate3m	0.0062	0.9819	0.05	是
nr_employed	0.0169	0.0785	0.05	是

Table 3. Distributional variable chi-square test**表 3. 卡方检验表**

分类型变量	χ^2 统计量	自由度	p 值	显著性水平	分布是否一致
job	19.9164	11	0.0465	0.05	否
marital	4.9645	3	0.1744	0.05	是
education	3.7780	7	0.8050	0.05	是
default	3.1650	2	0.2055	0.05	是
housing	0.3740	2	0.8294	0.05	是
loan	2.7069	2	0.2583	0.05	是
contact	0.4208	1	0.5165	0.05	是
month	9.5339	9	0.3895	0.05	是
day_of_week	3.4266	4	0.4891	0.05	是
poutcome	0.4119	2	0.8139	0.05	是

数据无显性缺失, 但 job、marital 等 6 个离散特征存在 “unknown” 形式的隐性缺失(default 字段缺失占比最高, 约 20%), 见图 2。鉴于样本整体质量较高, 不剔除样本, 将 “unknown” 作为独立类别编码。针对异常值检测结果见图 3, 结合业务逻辑与统计方法差异化处理, age 特征的极端值对应真实高净值客户群体, 予以保留; 对 duration 字段采用 Winsorizing 方法, 取 1% 分位数 $Q_{0.01}$ 和 99% 分位数 $Q_{0.99}$ 为阈值截断, 如公式(3)所示:

$$X_{\text{duration}} = \begin{cases} Q_{0.01}, & \text{if } X_{\text{duration}} < Q_{0.01} \\ X_{\text{duration}}, & \text{if } Q_{0.01} \leq X_{\text{duration}} \leq Q_{0.99} \\ Q_{0.99}, & \text{if } X_{\text{duration}} > Q_{0.99} \end{cases} \quad (3)$$

对 campaign 字段, 依据 IQR 法则计算第一四分位数 Q_1 与第三四分位数 Q_3 , 得四分位距 $IQR = Q_3 - Q_1$, 以 Upper Bound = $Q_3 + 1.5 \times IQR$ (过度营销对应阈值 10 次) 为上限, 对超出值进行截断。处理后的数据更贴合业务场景, 有效信息得以保留。

2.3. 银行客户产品认购的 Stacking 集成模型

为提升银行客户产品认购行为预测的准确性, 本文采用多算法融合策略, 筛选性能优异的单一模型作为 Stacking 集成框架的基学习器, 通过 4 个基学习器与 1 个元学习器的层级架构优化预测性能。基学习器从原始训练数据中学习特征表示, 元学习器则基于基学习器的输出进一步学习集成决策规则, 最终作出预测, 以此降低单一模型的偏差。本文构建的 Stacking 集成模型见图 4, 具体实施步骤如下:

1) 单一模型择优。将数据集按 8:2 的比例随机划分为训练集 \mathcal{D}_{tr} 与验证集 \mathcal{D}_{val} 。表 4 展示了各类模型在未调参条件下于验证集 \mathcal{D}_{val} 上的性能表现。

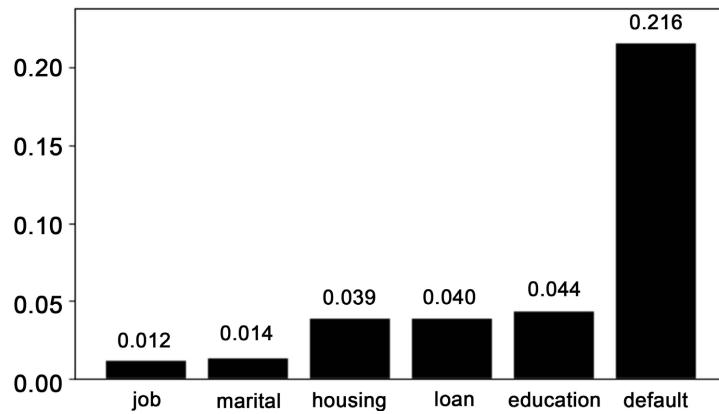


Figure 2. Implicit missing variables and the corresponding proportions
图 2. 隐性缺失变量及所占比例

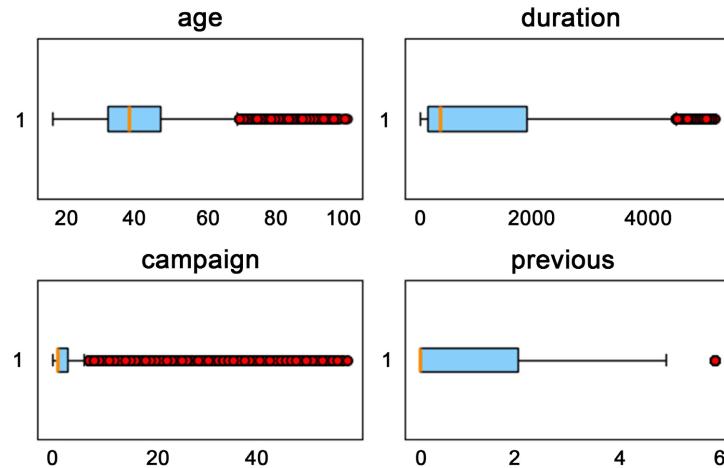


Figure 3. Box plot for outlier detection
图 3. 异常值检测箱型图

2) 基学习器确定。选取准确率(Acc)最优的4个模型作为Stacking框架第一层的基学习器, 分别为随机森林(RF)、多层感知机(MLP)、LightGBM、XGBoost, 记为 \mathcal{M}_m ($m=1,2,3,4$), 四类模型分别代表不同的算法范式(集成树、神经网络、梯度提升), 计算模型在验证集上预测结果的皮尔逊相关系数 $\rho(\mathcal{M}_i, \mathcal{M}_j)$ ($i, j \in \{1, 2, 3, 4\}, i \neq j$), 验证其输出的低相关性即具备多样性。通过Optuna超参数优化方法对每个基学习器独立进行超参数寻优, 各模型在数据集上的最优参数配置见表5。

3) 元特征生成。对每个基学习器 \mathcal{M}_m , 采用 $K=5$ 折交叉验证训练: 将 \mathcal{D}_{tr} 随机划分为 K 个子集 $\{\mathcal{D}_{\text{tr}}^k\}_{k=1}^K$, 每次以 $\mathcal{D}_{\text{tr}}^k$ 为验证集、 $\bigcup_{t \neq k} \mathcal{D}_{\text{tr}}^t$ 为训练集训练模型, 得到第 k 折的预测值 $\hat{y}_{m,k}$ 。将 K 折预测值组合, 形成基学习器对 \mathcal{D}_{tr} 的元特征 $Y_{\text{tr}}^m = \{\hat{y}_{m,1}, \hat{y}_{m,2}, \dots, \hat{y}_{m,|\mathcal{D}|}\}$; 同时, 用全量训练的 \mathcal{M}_m 对原始测试集 \mathcal{D}_{te} 预测, 得到元特征 $Y_{\text{te}}^m = \{\hat{y}_{m,|\mathcal{D}|+1}, \dots, \hat{y}_{m,|\mathcal{D}|+|\mathcal{D}|}\}$ 。最终, 4个基学习器的元特征组合为新训练集 $X_{\text{meta-tr}} = [Y_{\text{tr}}^1, Y_{\text{tr}}^2, Y_{\text{tr}}^3, Y_{\text{tr}}^4]^T$ (维度为 7500×4), 新测试集 $X_{\text{meta-te}} = [Y_{\text{te}}^1, Y_{\text{te}}^2, Y_{\text{te}}^3, Y_{\text{te}}^4]^T$ (维度为 4235×4)。

4) 最终预测。选取逻辑回归(LR)模型 $\mathcal{M}_{\text{meta}}$ 作为第二层元学习器, 以 $X_{\text{meta-tr}}$ 与原始标签 y_{tr} 训练 $\mathcal{M}_{\text{meta}}$; 将 $X_{\text{meta-te}}$ 输入训练好的 $\mathcal{M}_{\text{meta}}$, 得到最终预测结果 \hat{y}_{te} 。

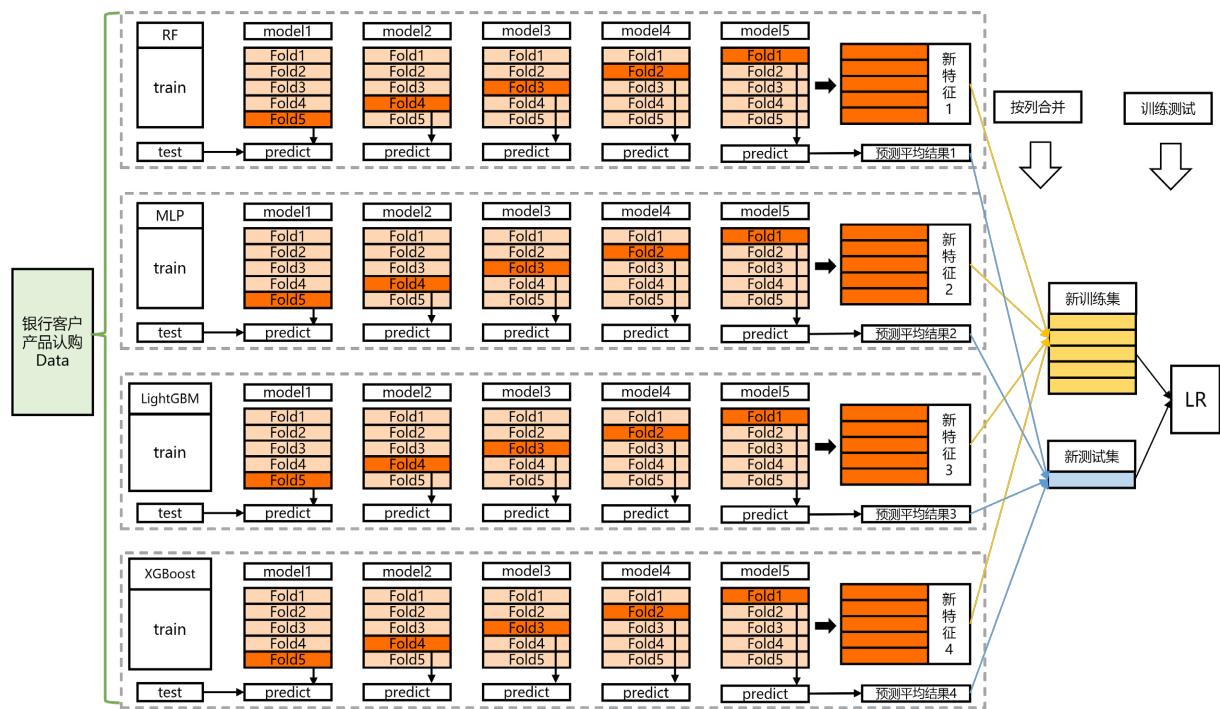


Figure 4. Stacking ensemble learning framework diagram
图 4. Stacking 集成学习框架图

Table 4. Comparison of several models on the validation set
表 4. 若干模型关于验证集的对比

模型名称	Acc	AUC
SVM	0.883	0.757
MLP	0.889	0.880
XGBoost	0.902	0.892
LightGBM	0.900	0.887
Decision Tree	0.853	0.668
Random Forest	0.899	0.882
Gradient Boosting	0.888	0.874
Logistic Regression	0.879	0.732
Linear Discriminant	0.879	0.746
K-Nearest Neighbors	0.881	0.724

Table 5. Optimal parameter settings of various models
表 5. 各模型最优参数设置

模型	参数名称	参数取值
RF	max_depth	80
	n_estimators	1500

续表

MLP	alpha	0.004
	hidden_layer_sizes	(128, 64)
	n_estimators	2500
	learning_rate	0.003
	num_leaves	20
	reg_alpha	3.368
LightGBM	reg_lambda	4.753
	max_depth	6
	min_child_samples	71
	subsample	0.984
	colsample_bytree	0.814
	max_depth	3
	min_child_weight	6
	gamma	4.233
XGBoost	subsample	0.811
	colsample_bytree	0.810
	reg_alpha	1.318
	reg_lambda	2.732

2.4. 模型评估

鉴于银行客户认购产品问题属于二分类任务，模型输出仅有两种结果：认购(yes)与不认购(no)。根据预测结果与真实类别，可将样本划分为四类：真正例(TP)：真实值为“认购”且模型预测为“认购”的样本数；假负例(FN)：真实值为“认购”但模型预测为“不认购”的样本数；假正例(FP)：真实值为“不认购”但模型预测为“认购”的样本数；真负例(TN)：真实值为“不认购”且模型预测为“不认购”的样本数。基于上述混淆矩阵元素，本研究采用以下核心指标对模型性能进行量化评估。

Acc 如下公式所示，衡量模型在所有样本中预测正确的比例，是分类任务中最常用的基础评价指标。受试者工作特征曲线(ROC)以假正例率为横轴、真正例率为纵轴，能够直观且全面地刻画模型在不同阈值下的性能权衡。ROC 曲线下的面积(AUC)则将 ROC 曲线的信息浓缩为一个数值，其取值范围为[0, 1]，可作为衡量模型整体区分能力与综合预测性能的重要指标。

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

3. 结果与分析

为验证本文所构建模型在银行客户产品认购预测任务中的有效性，本实验构建多种分类模型，结合 5 折交叉验证量化评估性能，并将本文 Stacking 集成模型与主流机器学习、神经网络模型展开对比，以验证集成框架的优势。

3.1. 对比实验

表 6 呈现了所有模型在 5 折交叉验证下的核心评价指标。由表中数据可知，本文构建的 Stacking 集

成模型在两项指标上均表现最优，其准确率达 0.971，充分证明集成框架通过融合多基模型的优势信息，有效降低了单一模型的预测偏差与方差，提升了模型的泛化能力。

Table 6. Prediction results of different models
表 6. 不同模型预测结果

模型名称	Acc	AUC
Stacking	0.971	0.887
XGBoost	0.963	0.880
LightGBM	0.964	0.885
MLP	0.927	0.802
RF	0.956	0.886
LR	0.923	0.855

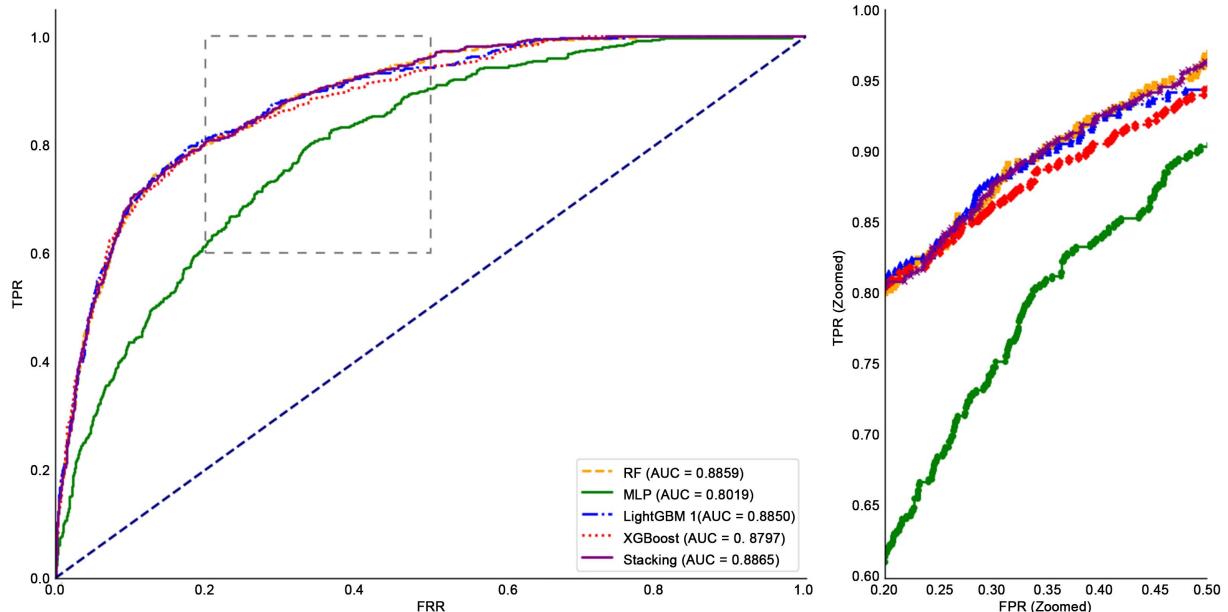


Figure 5. Comparison of multi-model results on ROC
图 5. ROC 多模型结果对比

为更直观对比各模型的区分能力，借助图 5 绘制所有模型的 ROC 曲线及对应 AUC 值；同时结合图 6 的多指标预测结果对比，进一步验证模型性能：

从 ROC 曲线的形态来看，本文所构建的 Stacking 模型的 ROC 曲线始终处于其他模型之上，且曲线下的 AUC 值最大，这表明该模型在正确识别认购客户与避免误判非认购客户这两者的权衡能力上更为出色；而多指标结果(AUC、验证集 Acc、测试集 Acc)则清晰显示，Stacking 模型在三项核心指标上均优于单一模型，说明其泛化能力与预测稳定性更突出。

与此同时，这也进一步体现出小样本场景下神经网络模型存在的局限性，从侧面凸显出 Stacking 集成框架借助多基模型互补来提升预测可靠性的价值。

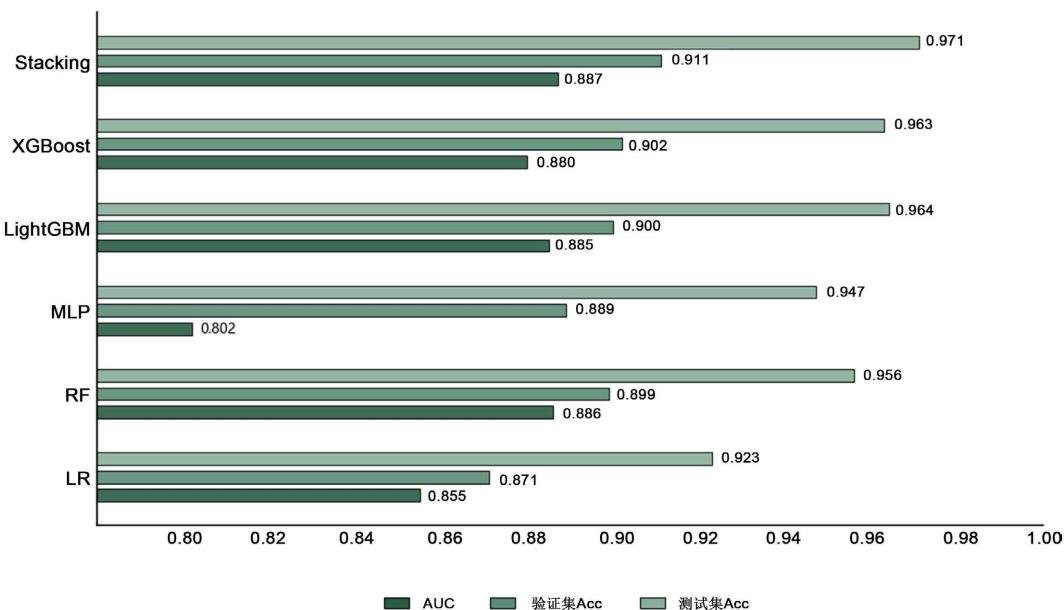


Figure 6. Comparison of prediction results of different models
图 6. 不同模型预测结果对比

3.2. 重要特征分析

结合图 7 分析结果, 前 8 个高权重特征可归为三大维度, 共同构成银行客户决策的核心驱动因素: 在客户互动维度, “联系时长”决定客户沟通充分程度, “联系间隔天数”影响客户对营销的关注热度, “联系月份”对应消费周期与空闲节奏, 三者共同锁定客户沟通时的认购意愿; 宏观经济维度里, “同业拆借利率”关联产品成本预期, “消费者信心指数”“雇员人数”反映经济景气度下客户的风险偏好, “消费价格指数”影响实际购买力, 这些信号直接决定客户理财决策优先级; 客户基础属性维度的“年龄”, 则以差异化的理财需求与风险承受能力, 成为客户群体划分、产品匹配的基础依据。

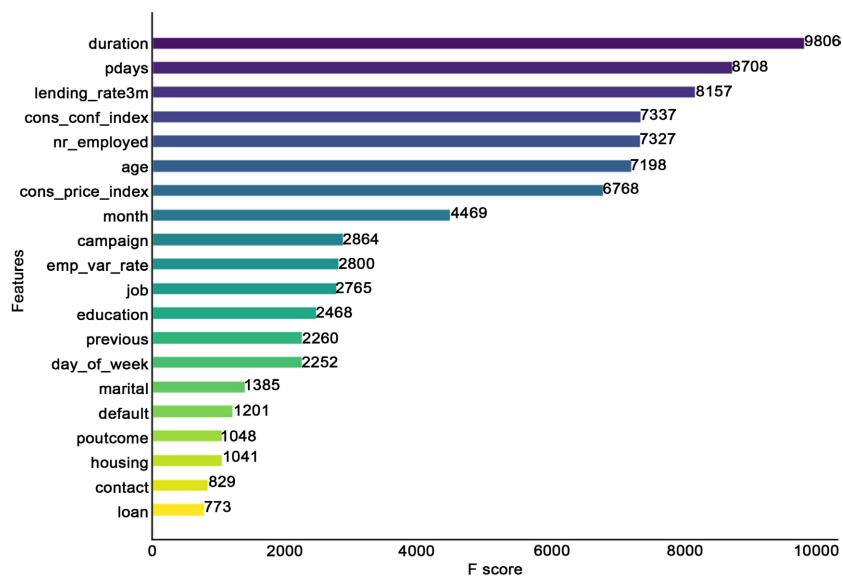


Figure 7. Feature importance ranking
图 7. 重要特征排序

3.3. 客户分群及用户画像

本文基于所构建的高认购概率模型,以模型输出的核心特征为依据,通过 duration 特征的分位数(50%、75%, 100%)将客户分为低、中、高三档活跃客户,同时结合 poutcome 特征将客户界定为老用户(上一次营销成功客户)和新用户(上一次没成功的客户),两两组合形成了 6 个客群见表 7。图 8 展示了基于该模型对 6 个客群开展的重要特征分析,本文进一步依托模型挖掘的客群差异化特征,系统总结了各客群的核心属性与认购潜力特点,最终形成个性化营销策略见表 8。

Table 7. Customer group classification and distribution by activity level and user type
表 7. 基于活跃度与用户类型的客群分类及数量分布

poutcome/duration	低活跃度	中活跃度	高活跃度
新用户	10720	4563	3777
老用户	1847	1062	531

Figure 8. Customer segmentation feature importance ranking

图 8. 客户分群重要特征排序一览

Table 8. Customer segmentation portraits and suggestions
表 8. 客户分群画像及建议

客群	特点	策略建议
新用户&活跃度低	年长 + 多营销→易认购	增老年客群营销次数
新用户&活跃度中	低利率 + 长沟通→易认购	利率下行期延长沟通
新用户&活跃度高	就业稳→易认购	就业向好时加大营销
老用户&活跃度低	敏感宏观经济	景气时重点触达
老用户&活跃度中	近联系 + 长时长→易认购	提联系密度及时长
老用户&活跃度高	就业多→易认购	就业活跃期强营销

4. 结论与展望

本研究围绕客户认购预测任务展开, 经实验验证与分析, 得出具体结论如下:

1) 针对性的银行客户数据处理策略有效。一致性检验消除训练集与测试集的分布偏差, 对隐性缺失值实施独立类别编码、对异常值结合营销场景差异化处理, 保障数据质量与业务贴合度。

2) 预测模型在客户认购预测中表现优异。该框架经优质基学习器筛选、参数优化以及元学习器融合输出, 高精度预测客户认购行为, 有效识别潜力客户特征, 且具备稳定泛化性能。

总之, 本研究构建的 Stacking 集成预测模型, 可结合客户分群画像的精准定位能力, 辅助银行优化差异化营销策略, 既实现对高认购潜力客群的精准触达, 降低无效成本, 又能针对不同客群的需求特征提供个性化服务, 提升客户体验与满意度。未来研究将进一步拓展样本规模、补充客户多维度行为特征, 或优化模型层级结构以增强其复杂场景适应能力; 同时, 本研究的技术思路与客群分析框架, 亦可迁移至银行客户流失预警、信贷风险评估等其他客户行为预测场景, 为银行业务数字化转型提供决策支持。

基金项目

- 1) 重庆市教委科学技术研究项目, 编号: KJQN202502003。
- 2) 重庆理工大学本科教育教学改革研究重点项目, 编号: 2024ZD09。
- 3) 重庆理工大学研究生课程思政示范课程《应用数理统计》, 编号: gzlsz202501。
- 4) 重庆理工大学研究生创新项目, 编号: gzlcx20243281。

参考文献

- [1] 叶梦源. 关于银行客户关系管理的优化策略[J]. 现代营销, 2023, 796(2): 112-114.
- [2] 罗常伟, 王双双, 尹峻松, 等. 集成学习研究现状及展望[J]. 指挥与控制学报, 2023, 9(1): 1-8.
- [3] Chang, H., Hung, L. and Ho, C. (2007) An Anticipation Model of Potential Customers' Purchasing Behavior Based on Clustering Analysis and Association Rules Analysis. *Expert Systems with Applications*, **32**, 753-764. <https://doi.org/10.1016/j.eswa.2006.01.049>
- [4] Cho, Y.S., Moon, S.C., Oh, I., Shin, J. and Ryu, K.H. (2013) Incremental Weighted Mining Based on RFM Analysis for Recommending Prediction in U-Commerce. *International Journal of Smart Home*, **7**, 133-144. <https://doi.org/10.14257/ijsh.2013.7.6.13>
- [5] Kabir, M.R., Ashraf, F.B. and Ajwad, R. (2019) Analysis of Different Predicting Model for Online Shoppers' Purchase Intention from Empirical Data. 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, 18-20 December 2019, 1-6. <https://doi.org/10.1109/iccit48885.2019.9038521>
- [6] 陈良凯, 黄登仕. 商业银行理财产品消费者购买行为的影响因素分析[J]. 财经科学, 2014(4): 53-60.
- [7] 金铮. 基于机器学习算法的商业银行精准营销模型研究[J]. 现代营销(经营版), 2020(10): 120-121.
- [8] 方鼎. 梯度提升模型在银行产品预测上的应用[D]: [硕士学位论文]. 武汉: 华中科技大学, 2017.
- [9] 祝歆, 刘潇蔓, 陈树广, 李静, 张天宇. 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.
- [10] 付长凤. 基于机器学习融合算法的网络购买行为预测研究[J]. 北京印刷学院学报, 2020, 28(S1): 262-264.
- [11] 李慧, 李正, 余堃. 一种基于综合不放回抽样的随机森林算法改进[J]. 计算机工程与科学, 2015, 37(7): 1233-1238.
- [12] 宋伟, 张帆, 叶阳东, 等. 基于 SAX 方法的时间序列分类问题的多阶段改进研究[J]. 计算机工程与科学, 2016, 38(5): 988-996.
- [13] Tékomabou, S.C.K., Gherghina, S.C., Toulni, H., Neves Mata, P., Mata, M.N. and Martins, J.M. (2022) A Machine Learning Framework towards Bank Telemarketing Prediction. *Journal of Risk and Financial Management*, **15**, Article 269. <https://doi.org/10.3390/jrfm15060269>