

基于 SHAP 的消费信贷借款人信用风险影响因素研究¹

孙超，姚小义，刘轶

(湖南大学，湖南、长沙，410006)

摘要：本文基于捷信消费金融在 Kaggle 的公开数据集，比较了树模型机器学习方法的分类预测效果，选取 LightGBM 模型作为分类基准模型，并运用 SHAP 模型对分类结果归因分析，可视化消费信贷借款人信用风险的影响。实证结果表明，借款人个人基本特征的 SHAP 绝对值普遍高于其他类型特征，引起的借款人预期信用风险变化更大；借款人的标准化信用评分、贷款年金收入比率等特征的 SHAP 值走势呈现明显的阶段性；教育水平、性别、年龄等特征对借款人信用风险的交互影响较为明显。

关键词：消费信贷；信用风险；SHAP；LightGBM 模型

中图分类号： F8

文献标识码： A

一、引言

在全面促进消费、加快构建以国内大循环为主体、国内国际双循环相互促进的新发展格局背景下，消费信贷对转变我国经济发展方式、优化经济结构、改善人民生活水平发挥着重要作用。受益于消费升级、普惠金融政策的推动和大数据、人工智能等先进技术在各领域的广泛应用，我国的消费信贷行业保持良好的发展势头。2020 年底，我国消费金融公司总体的资产规模第一次超过 5000 亿元，贷款余额为 4927.8 亿元，同比增长 4.34%。

我国消费信贷行业在发展中一直存在个人征信短板、贷前数据采集不足、数据维度较少等多种问题，虽然目前发展平稳良好，但其背后的信用风险、欺诈风险等各种风险并没有消失。随着消费信贷行业的市场竞争日益激烈，消费金融公司的生存压力增加，部分公司为了增加业务量放松了对借款人信用风险的把控，对公司经营的合规性以及发展的稳健性都造成了不利影响。随着监管机构对于消费信贷的监管逐渐深入和细化，消金公司亟待提升对自有借款人的信用风险把控能力。因此，提高消费信贷信息的集中、整合与加工能力，准确识别高信用风险借款人群体特征进而控制风险具有重要的现实意义。

本文主要应用基于 GBDT 的树类机器学习模型预测消费信贷借款人违约情况，并运用 SHAP 模型进行归因分析，分析单因素和多因素之间交互作用对消费信贷借款人信用风险的影响。本文的研究丰富了 SHAP 模型在消费信贷领域的应用，增强了机器学习模型信用评估结果的可解释性，并可视化各因素对消费信贷借款人信用风险的非线性影响。本文的可视化结果兼具样本整体和个体两个维度，丰富了借款人信用风险研究的维度。此外，本文从固定时间窗和固定次数两个角度构造时序特征因素来呈现信息时效性分类的作用，发现此类特征影响较大，具有一定现实借鉴意义。

二、文献综述

借款人信用风险影响因素相关研究中，国内外学者分别研究了学历^①、工作年限^①、借款目的^①、贷款期限^①、账户信息^①、消费行为^①等硬信息特征和借款描述^①、借款人照片^①等软信息特征对借款人信用风险的影响。为深入探究影响机制，学者还研究了特征的交互影响。一方面，通过设置因素相乘交互项，孙武军，樊小莹(2016)发现借款人工作经历越丰富或学历越高，其借贷成功率越高。并且二者的叠加效应会削弱其在单特征时对借贷成功率的正向影响。^①另一方面，通过对样本分类，分组运用模型实证，彭红枫，林川(2018)根据性别、年龄段、学历、收入划分人人贷借款样本，研究借款陈述中六种类型词语对借款成功率的影响。发现词语信号对不同年龄和收入阶段的借款人影响差异较大，而性别和学历差异对词语信号作用影响较小。^①王正位等(2020)根据消费信息含量不同，以消费金额高低、消费次数多少、消费历史时间长短划分花呗和信用卡样本，发现消费金额较低、次数较少、时间较短的样本中，消费行为信息有限，花呗授信预测借款人违约效果较差。^①

¹ 基金项目：湖南省社会科学基金项目“基于大数据的网络借贷信用风险评估研究（16YBA078）”。

近年来，机器学习逐渐成为借款人信用风险评估的主流方法。学者运用随机森林、支持向量机和卷积神经网络等模型进行信用风险评估。其中，基于 GBDT 改进的多种机器学习算法在借款人违约分类上有较好性能并得到广泛使用。Ma 等(2018)用 XGBoost 和 LightGBM 对 Lending Club 数据建模比较，发现机器学习算法可使贷款违约减少约 1.17 亿美元。⁰Xia 等(2018)运用 XGBoost 测试五个消费信贷数据集借款人违约风险，发现对 XGBoost 而言，贝叶斯超参数优化的搜索效果优于随机搜索、网格搜索和手工搜索。⁰ Taha, Malebary(2020) 运用优化的 LightGBM 检测信用卡交易欺诈行为，运用贝叶斯超参数优化训练参数，发现该算法优于其他分类器。⁰Al Daoud(2019)，马晓君等(2020)对比了 XGBoost、LightGBM 和 CatBoost 的分类预测效果，发现 LightGBM 的稳健性最好，运行速率最快，分类准确度和精度表现较好。^{0,0}

然而，机器学习模型输出的重要特征存在不一致性，且其可解释性存在较大缺陷。针对这个问题，Lundberg 和 Lee(2017)提出了具有稳定性、一致性和合理性的 SHAP 解释模型⁰，Lundberg, Erion 和 Lee(2018)进一步提出 Tree SHAP，用于树模型机器学习方法，如 XGBoost 和 LightGBM 等。此外，还将 SHAP 值扩展到交互效果，以图形的方式呈现潜在隐藏的成对交互关系。⁰目前，SHAP 模型已被广泛应用于建筑⁰、新闻与传媒⁰、交通^{0,0}等领域。

综上，目前研究涵盖的借款人信用影响因素相对有限，对于因素影响的交互作用，已有文献大多通过因素相乘交互项或对样本分组实证，用特征系数呈现其对信用风险的影响，较为笼统。在各种机器学习方法中，基于 GBDT 的机器学习方法在信贷风险评估方面普遍性能更佳。Tree SHAP 模型很好的对接了 LightGBM 等基于 GBDT 的树类机器学习模型，并且其在消费信贷领域的研究还有待补充。因此，本文主要应用基于 GBDT 的树类机器学习模型，结合 Tree SHAP，分析单因素和多因素之间交互作用对消费信贷借款人信用风险的影响，以期为消费信贷的信用风险控制提供经验借鉴。

三、研究设计

(一) SHAP 模型介绍

SHAP (SHapley Additive exPlanations)，基于经典博弈论 Shapley 值，是一种加性特征归因方法 (Additive feature attribution methods)。这类方法存在唯一解，且具有局部准确性、一致性和缺失性三个理想性质。SHAP 值将条件期望与博弈论中的经典 Shapley 值结合，赋予每个特征 ϕ_i ：

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (1)$$

其中，N 是所有输入特征的集合， $f_x(S) = E[f(x)|x_S]$ 是函数在输入特征子集 S 的条件下的期望值。 ϕ_i 为正值且越大表示对解释函数输出的正向影响越大， ϕ_i 为负值且绝对值越小表示对解释函数输出的负向影响越小。

本文运用机器学习模型实现借款人违约的二分类预测，预测结果为借款人违约概率。SHAP 值与违约概率的关系则通过 sigmoid 函数体现，即

$$\hat{y} = \frac{1}{e^{-SHAP_y} + 1} \quad (2)$$

其中， \hat{y} 为借款人违约预测概率， $SHAP_y$ 为借款人 y 的 SHAP 值。

SHAP 值可以进行特征级可视化和样本级可视化。SHAP 值计算采用样本级穷举式，计算样本中各特征对预测结果的边际影响，因此 SHAP 值可表示单个样本中各特征的贡献。此外，将一个特征的所有样本 SHAP 值进行统计即可得到特征级可视化。

Lundberg 等人在 2018 年提出的 Tree SHAP，其改进方面除了降低 SHAP 值计算的复杂度，能基于局部解释进而解释全局模型之外，还提出 SHAP 交互值分析特征交互影响。Tree SHAP 将 SHAP 值中的交互效应与主效应分离，产生特征属性矩阵，表示所有特征对给定模型预测的影响。其使用博弈论中的“Shapley 交互作用指数”来捕捉局部交互效应：

$$\Phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!} V_{ij}(S) \quad (3)$$

当 $i \neq j$, 且

$$\nabla_{ij}(S) = f_x(S \cup \{i, j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S) \quad (4)$$

$$= f_x(S \cup \{i, j\}) - f_x(S \cup \{j\}) - [f_x(S \cup \{i\}) - f_x(S)] \quad (5)$$

在公式(3)中, 特征 i 和特征 j 之间的 SHAP 交互值在每个特征之间平均分配, 因此 $\Phi_{i,j} = \Phi_{j,i}$, 总的交互效应为 $\Phi_{i,j} + \Phi_{j,i}$ 。可视化方面, SHAP 依赖图 (SHAP dependence plot) 捕捉了模型中相互作用的垂直弥散效果, 这些效果可以通过每个样本点的交互值呈现。

(二) 模型构建

1. 数据来源

捷信集团主要布局消费信贷业务, 近年来中国发展为其最重要的市场。截至 2019 年 3 月底, 中国市场信贷金额占集团总额比例达 63.9%。2010 年, 捷信集团在中国成立全资子公司——捷信消费金融有限公司。之后, 捷信消费金融迅速成长为中国持牌消费金融公司中第一梯队的企业。虽然近两年受到疫情影响, 捷信消费金融的业务量和资产规模有所下降, 但总体来说仍属于消费金融公司中的头部公司, 其线下的获客模式目前仍是兴业消金、中银消金等银行系消费金融公司以及传统商业银行开展零售消费金融业务的主要模式, 具有较好代表性, 其信贷数据集在数据完整性、指标设计、数据挖掘等方面也能够为其他消费金融公司提供经验。

本文采用捷信集团 2018 年 8 月公开的信贷数据集²来分析消费信贷借款人信用风险。数据集共包含 7 个数据表, 主要记录了借款人贷款申请基本信息、第三方机构征信数据和平台历史贷款数据共三方面的数据信息, 共 221 维特征。其中, 主表包含 307511 笔贷款信息, 其中正常还款 282686 笔, 违约 24825 笔, 违约率为 8.07%, 数据存在严重不平衡。

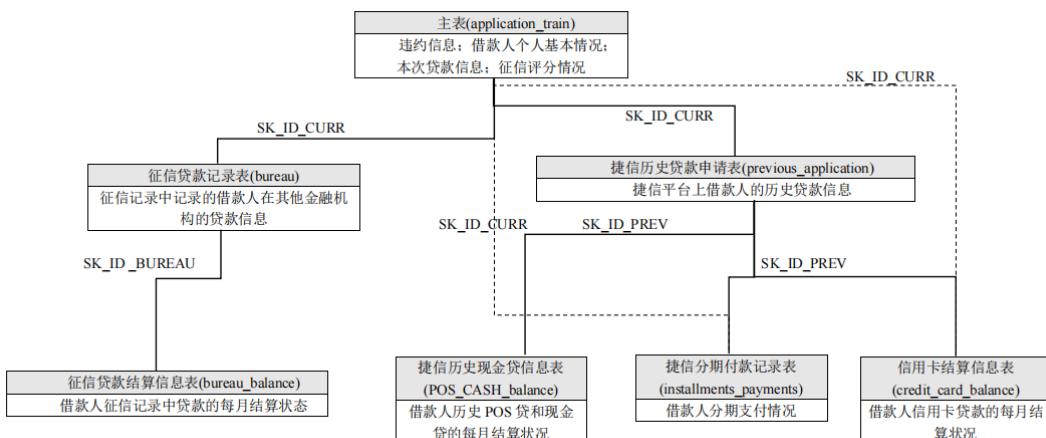


图 1 数据集结构

图 1 介绍了各表连接关系, 其中, “SK_ID_CURR”为主表贷款唯一编号, “SK_ID_BUREAU”为征信记录贷款唯一编号, “SK_ID_PREV”为捷信平台历史贷款唯一编号。各数据表有三种关联方式, 一是主表通过“SK_ID_CURR”关联征信信息和捷信历史贷款信息; 二是征信信息之间通过“SK_ID_BUREAU”关联; 三是捷信不同类型历史贷款信息之间通过“SK_ID_PREV”关联。

2. 数据清洗

本文对于异常值清理, 主要从横向和纵向两个角度。纵向是对特征变量序列中的极端值以及不符合现实情况的数据进行空值转换、零值转换或剔除。横向是对特征间关系不符合现实应有数量关系的单笔贷款数据剔除。此外, 本文剔除了缺失值占比超过 60% 的变量, 剩余变量中, 整数型变量用众数填充, 非整数型数值变量用均值填充, 类别变量用“XNA”填充, 并对客户逾期 30 天的次数、最近一次修改手机号码时间等特殊意义变量以零值填充。

² 捷信集团官方网站 <https://www.kaggle.com/c/home-credit-default-risk/data>

(三) 特征构造与特征选择

本文运用 One-Hot 编码处理多类别特征，并根据主表与其他数据表贷款的一对多关系，将其它数据表信息通过不同的统计方式汇总到主表中。

首先，本文在单数据表中构建了 11 个衍生特征，如表 1 所示。其次，本文在将特征信息汇总到主表时，构造了简单统计特征、时序特征和条件特征三类特征。简单统计特征是对数值型特征用最大值、最小值等简单统计量聚合汇总，对类别特征按特征包含的类别计数聚合汇总。时序特征有固定时间窗和固定次数两个角度。本文设定了两种单位的时间窗，对于贷款特征的时间以天数计量的，本文设定的时间窗为 30 天、90 天、180 天、365 天和 730 天；对于贷款特征的时间以月数计量的，本文设定的时间窗为 1 个月、3 个月、6 个月、12 个月和 24 个月，统计不同时间窗内的各特征简单统计量。本文设定的固定次数为 1 次、3 次、5 次、10 次，统计最近几次的各特征简单统计量。条件特征是在某一固定特征类别下其它特征的简单统计量，如贷款状态为进行中的征信记录贷款剩余天数等。

总体来说，通过以上方式，本文共构造 797 个特征。结合主表原有的 110 个特征，共有 907 个特征。删除缺失率大于 60% 的特征后，共剩余 459 个特征。

表 1 单数据表衍生特征

特征名称	计算公式	公式含义
社会经历程度(EMPLOYED_BIRTH_RTO)	DAYS_EMPLOYED/DAYS_BIRTH	工作时长/年龄
还款收入压力(INCOME_CREDIT_RTO)	AMT_INCOME_TOTAL/AMT_CREDIT	收入/贷款金额
家庭人均收入(INCOME_MEMBERS_RTO)	AMT_INCOME_TOTAL/AMT_FAM_MEMBERS	收入/家庭人数
年度还款收入压力 (ANNUITY_INCOME_RTO)	AMT_ANNUITY/AMT_INCOME_TOTAL	贷款年金/收入
还款时间压力(ANNUITY_CREDIT_RTO)	AMT_ANNUITY/AMT_CREDIT_EDIT	贷款年金/贷款金额
贷款申请的缩减金额(AMT_minus)	AMT_CREDIT_AMT_APPLICATION	捷信历史贷款的审批金额-申请金额
消费贷款意愿 (AMT_GOODS_APPLICATION_minus)	AMT_GOODS_PRICE_AMT_APPLICATION	捷信历史贷款的标的商品价格-申请金额
消费贷款审批风险 (AMT_GOODS_CREDIT_minus)	AMT_GOODS_PRICE_AMT_CREDIT	捷信历史贷款的标的商品价格-审批金额
还款提前结束情况 (DAYS_LAST_DUE_change)	DAYS_LAST_DUE-DAYS_LAST_DUE_1ST_VERSION	捷信历史贷款的实际到期日-初始到期日
当月提前还款天数 (DAYS_PAYMENT_INSTALMENT_minus)	DAYS_ENTRY_PAYMENT-DAYS_INSTALMENT	捷信历史贷款的分期付款实际支付日期-固定还款日期
当月超前还款金额 (AMT_PAYMENT_INSTALMENT_minus)	AMT_PAYMENT_AMT_INSTALMENT	捷信历史贷款的实际支付分期还款金额-固定分期还款金额

本文通过计算 WOE (Weight of Evidence, 证据权重) 和 IV (Information Value, 信息值) 进行特征选择，剔除 IV<0.02 的特征变量。此外，考虑特征经济意义，将少数信贷评估研究中重要但相关性较弱的特征考虑进来。特征选择后，本文共剩余 165 个特征变量，其中，数值型特征 146 个，类别型特征 19 个。

(四) 机器学习模型建立与选择

本文运用 SMOTE Tomek 采样来改善数据不平衡下分类模型性能。SMOTE-Tomek 是一种基于数据清洗的过滤采样算法，利用 SMOTE 对原始数据过采样来扩大样本集，通过移除采样后数据集中的 TomeLink 对，删除类间重叠样本。本文首先运用 Python 中 sklearn 模块的 train_test_split() 函数按训练集 80%、测试集 20% 的比例随机划分数据集，得到训练集数据

245975 条，测试集数据 61494 条。之后对训练集数据运用 SMOTE Tomek 采样，处理数据不平衡问题，重采样后的训练集数据共 450582 条，违约样本和非违约样本比例持平。

本文采用贝叶斯优化调节机器学习模型的超参数。贝叶斯优化利用先验知识调节参数，并通过减少计算任务来加速寻找最优参数的进程，不易陷入局部最优。由于本文主要应用基于 GBDT 的机器学习模型，因此贝叶斯优化的概率代理模型选择非参数模型中的 Tree Parzen Estimator (TPE)。本文将经过 SMOTE Tomek 采样的训练集数据应用到机器学习模型的贝叶斯调参过程，应用 Python 的 Hyperopt 模块优化超参数，采用五折交叉验证，将二分类的逻辑回归函数作为目标函数，以 AUC 指标评价筛选出最优参数结果。各模型最优参数结果如表 2 所示：

表 2 各模型贝叶斯参数调优结果

参数	随机森林	GBDT	XGBoost	LightGBM	CatBoost
学习率(learning_rate)	-	0.0903	0.0327	0.0596	0.1454
树的最大深度(max_depth)	9	5	7	8	7
树的样本采样比例(subsample/bagging_fraction)	-	0.5387	0.8547	0.8161	-
树的特征采样比例 (colsample_bytree/feature_fraction)	-	-	0.6635	0.9163	-
树的最大特征使用数量(max_features)	65	-	-	-	-
迭代次数(n_estimators/num_iterations)	249	75	243	261	191
树的叶子最大数量 (max_leaf_nodes/num_leaves)	-	309	-	125	-
叶子节点最小样本数量(min_child_weight)	-	-	8	-	-
节点分裂所需的最小损失函数下降值(gamma)	-	-	1.304	-	-
L1 正则化项(reg_alpha)	-	-	0.3536	0.6168	-
L2 正则化项(reg_lambda/l2_leaf_reg)	-	-	0.2433	0.5739	4.6579
one-hot 编码最大规模(one_hot_max_size)	-	-	-	-	4

本文将各树模型的参数设置为调参后的最优参数，运用 AUC、F1Score、Recall、Precision、Brier 分数和 KS 值评价模型分类性能。同时，也与解释性较强的逻辑回归模型表现对比。各模型评价指标结果如表 3 所示：

表 3 各模型评价指标对比

	AUC	Recall	F1_score	Precision	Brier	KS 值
逻辑回归	0.6337	0.6423	0.1936	0.1139	0.4349	0.2000
随机森林	0.7028	0.6601	0.2338	0.1420	0.3516	0.3080
GBDT	0.7361	0.7065	0.2485	0.1508	0.3472	0.3540
XGBoost	0.7618	0.6897	0.2755	0.1721	0.2974	0.3940
LightGBM	0.7729	0.6844	0.2892	0.1833	0.2758	0.4120
CatBoost	0.7631	0.6679	0.2794	0.1766	0.2825	0.3920

根据表 3，总体来看，树模型的分类表现相比逻辑回归模型有明显提升。XGBoost、LightGBM 和 CatBoost 三个模型表现最佳且分类性能相似，AUC 均在 0.77 左右，Brier 分数在 0.28 左右，LightGBM 在这两个指标上的表现最好。此外，根据模型召回率 Recall 和 KS 值，LightGBM 和 XGBoost 召回率比 CatBoost 更高，高风险样本被预测到的概率更高，且在模型区分度 KS 值方面，LightGBM 表现更好。考虑到模型总体表现和运行时间效率，本文选择 LightGBM 为分类预测模型进行 SHAP 解释性分析。

四、实证结果与分析

(一) 特征重要性分析

本文梳理了 SHAP 输出的 LightGBM 前 30 个重要特征，并结合分类性能相似的 XGBoost 和 CatBoost 的特征重要性排名以及特征现实意义，最终呈现 29 个重要特征的 LightGBM 排

名情况，如表 4 所示。

表 4 重要特征情况

特征类别	特征内容	特征重要性排名	特征名称
个人基本信息	婚姻状况	1	婚姻状况：已婚
		12	婚姻状况：单身或未婚
		20	婚姻状况：公证结婚
		28	婚姻状况：离异
	征信情况	3	征信机构 2 的标准化评分
		10	征信机构 3 的标准化评分
		22	征信机构 1 的标准化评分
		85	前 1 个季度内被查询征信的次数
	教育水平	5	教育水平：高等教育（大学及以上）
		9	教育水平：中等或中等特殊教育
	收入类型	7	收入类型：普通工作
		11	收入类型：商业助理
		13	收入类型：退休人员
		31	收入类型：国家公务员
	性别	8	性别
	车产信息	19	是否有车产
	社会经历程度	30	工龄年龄比率
本次贷款信息	还款压力	26	贷款年金与金额比率
		75	贷款年金收入比率
历史贷款信息	分期付款准时性	2	最近 10 笔分期付款中延迟还款次数
		14	最近 5 笔分期付款中提前还款次数
	利率	6	捷信贷款利率水平为高值组次数
		17	最近 1 次贷款状态为已批准
	贷款状态众数（固定次数）	21	2 年以上捷信贷款状态众数为已批准
		23	2 年以内捷信贷款状态众数为已批准
	当月逾期天数（固定时间窗）	18	2 年以上消费贷和现金贷当月平均逾期天数(放宽条件)
		25	2 年以内消费贷和现金贷当月的平均逾期天数
征信记录贷款信息	贷款笔数（固定时间窗）	34	1 年内征信记录贷款次数
	剩余应还数额	55	征信记录贷款剩余应还金额总和

根据表 4，本文的重要特征主要分为四个方面：个人基本信息、捷信本次贷款信息、捷信历史贷款信息和征信记录贷款信息。其中，个人基本信息包括 17 个特征，捷信本次贷款基本信息包括 2 个特征，捷信历史贷款信息包括 8 个特征，征信记录贷款信息包括 2 个特征。总体而言，借款人的个人基本信息特征和在消费金融平台的历史贷款信息特征的重要性排名相对靠前，借款人自身的素养和经历是评估借款人信用风险更关键的因素。此外，本文构造的时序特征和借款人个人基本信息的比率特征在评估借款人信用风险时重要性排名较高，说明消费信贷信用评估中信息的时效性分类和借款人情况的相对性较为重要。

（二）单特征影响分析

本文选取部分代表性特征进行 SHAP 图分析。根据公式（2）中 SHAP 值和违约概率的转换关系，借款人的违约预测概率随借款人 SHAP 值增加而增加。当特征的 SHAP 值为正值时，会促使借款人个体的 SHAP 值增加，从而对借款人违约预测概率产生正向影响。

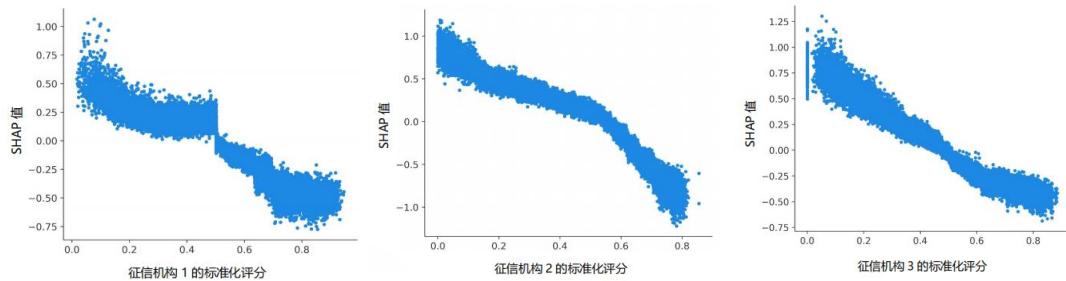


图 2 征信信息 SHAP 分析图

根据图 2, 借款人信用分和 SHAP 值总体呈负相关, 即借款人信用分越高时违约预测概率越低, 说明了模型预测结果的有效性。三家征信机构的评分标准稍有不同, 机构 2 对借款人的信用评分偏低, 评分标准相对严格, 其信用评分前 55% 的借款人 SHAP 值均为正值, 而其他机构 SHAP 值为正值的借款人在前 50%; 机构 1 的信用评分标准相对宽松, 其 SHAP 值分布范围集中在 -0.75 到 0.75 左右, 小于机构 2 的 SHAP 值 -1.2 到 1.1 的范围, 对借款人信用的区分度稍有欠缺。此外, 机构 2 前 10% 借款人的 SHAP 值高于其他两个机构, 对于高违约风险人群的识别度更优。总体来说, 三家征信机构评分均是预测借款人违约概率的重要特征, 但各征信机构评分标准以及对不同风险类别客户的识别能力不同。

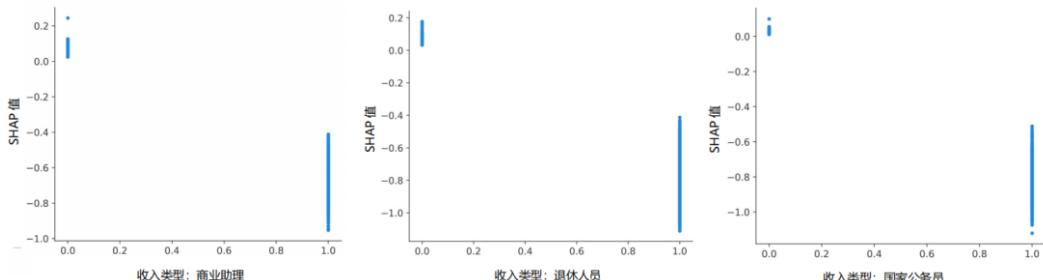


图 3 职业性质 SHAP 分析图

根据图 3, 公务员和退休人员两类借款人 SHAP 值相对较低, 信用风险降低程度最明显。这两类借款人的特点是稳定, 具有稳定的收入来源、工作变动程度较小, 经济实力较为固定, 信用风险水平更容易评估, 预期信用风险水平较低。商业助理类型的借款人 SHAP 值稍高, 其职业性质要求借款人掌握更多金融知识, 对金融产品和相关行业的了解更为深入, 金融素养较高, 其在自身的资产管理和日常消费中能够形成风险可控的规则, 预期信用风险水平较低。总体来看, 职业性质背后的借款人收入稳定性、工作可持续性和金融素养是借款人信用风险的重要影响因素。收入越稳定, 工作变动程度小、不易失业以及金融素养较高的职业借款人预期信用风险水平较低。

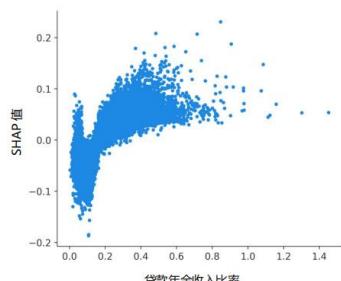


图 4 贷款年金收入比率 SHAP 分析图

根据图 4, 总体来说, 贷款年金和收入比率与 SHAP 值总体呈正相关, 还款的资金压力越大, 预期借款人发生信用风险的可能性提高越多, 与现实规律较为一致。其中, 借款人的贷款年金占年收入比率超过 80% 的样本数量非常少, 且预期信用风险增加水平不稳定, 考虑到借款人收入中一部分会用来支付各类生活成本, 80% 可认为是一个风控极限。贷款年金与收入比率为 20% 时是转折点, 贷款年金占年收入小于 20% 时, SHAP 值基本为负值, 违约预测概率降低, 此时借款人的还款的资金压力相对较小, 贷款回收情况相对良好。贷款年金占年收入大于 20% 时, SHAP 值基本为正值, 借款人违约预测概率增加, 且在占比小于 40%

时, SHAP 最大值直线上升, 预期信用风险增长较快。此时, 借款人还款压力上升, 但会由于借款人具体收入水平、自有资金水平、风险承受能力等多方面不同导致年金占比相同的借款人违约预测概率不同, 其 SHAP 值呈现逐渐分散趋势。

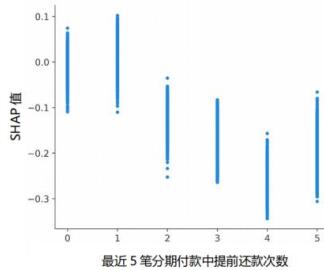


图 5 分期付款信息 SHAP 分析图

根据图 5, 最近 5 笔还款中, 每两笔为一个阶段, 随着贷款笔数增加和阶段变化, 借款人信用风险预测水平稍有降低。无提前还款行为和仅有一次提前还款行为的借款人 SHAP 值在 0 附近摆动, 其预期信用风险变动方向存在不明确性。可以看出, 偶尔一次的提前还款或没有提前还款并不能给借款人的信用风险准确定性。因为此类借款人中既存在自身信用水平良好但暂未生成有效的多次分期付款记录的, 又存在分期付款有延迟还款行为的。前者的预期信用风险降低, SHAP 值为负值; 后者的预期信用风险增加, SHAP 值为正值。因而此类借款人 SHAP 值在 0 附近摆动, 借款人信用风险变动方向不确定。

随着分期付款的提前还款次数增加, 借款人信用风险降低的定性更准确。最近 5 笔分期近乎每次都提前还款的借款人预期信用风险可降低程度较大, SHAP 值最低可达到-0.3 左右。有提前还款行为的借款人相对更加厌恶风险, 此类借款人对资金持有的成本比较敏感, 当筹集到足够资金时, 将尽早归还应付贷款以避免进一步的利息支出和不良征信记录生成。此外, 该借款群体的借款策略更趋于保守, 不易发生非理性借贷行为, 进而降低了借款违约行为发生的可能性。^[2]

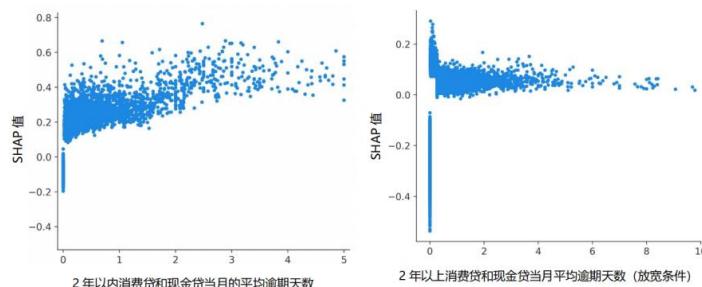


图 6 逾期信息 SHAP 分析图

根据图 6, 总体来说, 不同时间段下, 借款人没有逾期行为时 SHAP 值均为负值, 借款人预期信用风险相对较小; 借款人的平均逾期天数不为零时, SHAP 值均为正值, 借款人预期信用风险会有不同程度增加, 符合现实认知。具体来说, 2 年内即短期贷款的平均逾期天数 SHAP 值随逾期天数增加而缓慢增加, SHAP 值分布范围从最初的 0.1 至 0.4, 到后来的 0.35 至 0.6, 借款人预期信用风险会有所提高。2 年以上即长期贷款平均逾期天数的 SHAP 值基本无明显趋势变化, 取值在 0 至 0.1 之间, 这个指标无论是在影响程度还是在逾期界定的严格性方面都较前一个指标更弱, 因此其对于借款人预期信用风险的影响较有限。

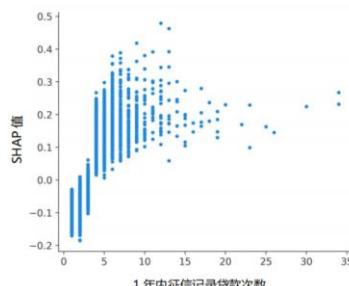


图 7 征信记录历史贷款笔数 SHAP 分析图

根据图 7, 随着征信记录的其他机构贷款笔数增加, SHAP 值逐渐增大, 预期借款人信用风险增长加快。一年以内的贷款距离时间较短, 借款人的还款行为和信贷习惯等可参考性较强, 在短时间内贷款次数过多可明显反映出借款人的资金紧缺程度, SHAP 正值较大, 预期信用风险增加较明显。贷款次数为 3 次时是一个分界线。3 次以内时, SHAP 值基本为负值且在 -0.2 之内, 预期借款人信用风险小幅降低。4 次及以上时, SHAP 值快速转变为正值并在贷款约 13 次时达到峰值 0.5 左右, 此时借款人违约预测概率有相对较大程度增加。因此, 借款人近期贷款次数过多时, 预期借款人存在较大的违约风险, 判断界限为 3 次较合适。

(三) 多特征交互影响分析

本文根据两特征交互影响程度大小, 选取在特征取值范围内, SHAP 交互绝对值基本达到 0.05 及以上的部分特征进行分析, 呈现交互作用较为显著的影响。受限于运行环境的内存, 本文随机抽取 SMOTE Tomek 采样过后训练集数据 15770 条, 测试集数据 9224 条作为新的训练集和测试集来输出特征交互图。

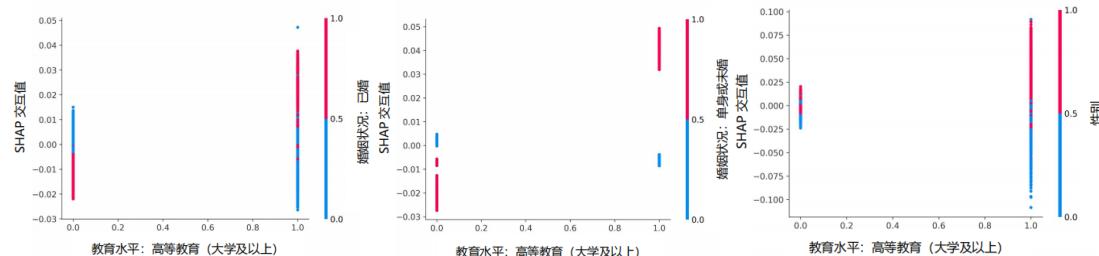


图 8 教育水平与婚姻状况、性别的 SHAP 交互影响图

根据图 8, 在学历同为高等教育的条件下, 已婚借款人的 SHAP 交互最大值小于未婚借款人, SHAP 交互值变动区间较大, 说明在借款人学历高时, 已婚借款人相比未婚借款人被认为信用风险增加程度更小, 信用风险集中度较低。在高学历人群中, 单身或未婚的借款人相比已婚借款人有相对更高的风险偏好, 风险容忍度更高, 受到约束相对较少, 信用风险发生可能性较大; 同时, 高学历在一定程度上反映了借款人金融知识储备较多, 更可能运用多种金融工具达到消费目的, 风险关联性和累积程度可能较强, 助长了其预期信用风险增加。

对于教育水平和性别的交互作用影响, 高学历女性的 SHAP 值为正值, 借款人预期信用风险增加; 高学历男性的 SHAP 值为负值, 预期信用风险降低。出现这种情况可能是因为, 一方面, 高学历女性从事的职业性质和社交圈可能会使借款人的日常以及必要的消费水平高于其他类型借款人; 另一方面, 现实生活中, 高学历女性在薪资收入和职业发展前景等方面相对弱于高学历男性借款人。如此一来, 高学历女性收入水平相对更低但消费水平相对更高, 可留存资金被压缩, 自身资产相对较少, 可承担的风险能力相对较差, 从而使其预期信用风险水平增加。

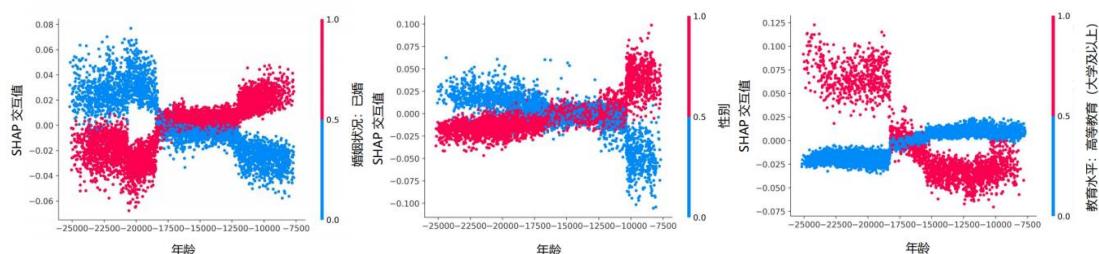


图 9 年龄与婚姻状况、性别、教育水平的 SHAP 交互影响图

根据图 9, 对于年龄与婚姻状况的交互作用, 已婚和其他类型借款人的 SHAP 交互值随年龄变化走势完全相反, 已婚借款人随着年龄增大 SHAP 交互值逐渐降低, 预期其信用风险会降低。其变动可大致分为三个阶段, 借款人年龄在 32 岁 (约 11500 天) 以下时, 已婚借款人相比其他类型借款人 SHAP 交互值更高, 违约预测概率增加更多。此阶段借款人较为年轻, 工作经验和社会经历较少, 心理相对不够成熟和理性, 物质和财富积累也相对不足, 组建家庭后, 可能会面临更多方面的开销以及更大的生活压力, 更可能发生信用风险。借款人年龄在 32 到 47 岁 (约 11500 天至 17500 天) 时, 已婚借款人的 SHAP 交互值均基本在 0 附近小幅摆动, 差距不明显, 违约预测概率变动较为接近。此时借款人在社会中大多工作时

间也较长，资金有了持续的积累，自我约束能力和风险控制能力也有了更大的提高，抵抗风险能力大体相当。借款人年龄在 47 岁以上时，已婚借款人的 SHAP 交互值均为负值且最低到 -0.06 左右，违约预测概率可降低程度更多，此阶段借款人随着年龄增长带来的心理成熟和理性以及家庭的稳定性作用更为凸显。

对于年龄与性别的交互作用，男性和女性随年龄增加其 SHAP 交互值变动趋势也相反。具体变动可大致分为三个阶段，借款人年龄在 28 岁(约 10500 天)以下时，女性借款人 SHAP 交互值比男性借款人高，最高可到 0.075 左右，违约预测概率可增加的程度较大。在借款人较为年轻时，女性借款人的消费观念理性程度和自我控制能力相比男性较弱，风险事项发生后，女性的融资渠道相对有限，对风险抵抗能力相对较弱，更容易发生信用风险。借款人年龄在 28 到 43 岁(约 10500 天至 16000 天)时，SHAP 交互值的性别差异不大，违约概率无明显的增加或降低情况，此阶段年龄与性别的交互作用影响不明显。借款人年龄在 43 岁以上时，借款人又出现较为明显的性别差异，此时女性借款人的违约预测概率会降低，年长女性在各方面的稳定性可能开始发挥作用，预期此类借款人信用风险会相应降低。

对于年龄与高学历的交互作用，整体分为两个阶段，借款人年龄在 47 岁及以下时，借款人的 SHAP 值为负值且最大可达 -0.07 左右，无明显变动趋势，预期信用风险水平会存在不同程度的降低。第二阶段借款人年龄在 47 岁以上时，借款人的 SHAP 值迅速转变为正值且大致分布在 0.03 至 0.1 之间，预期信用风险水平增加，可能是因为借款人到了一定年龄之后，由于身体原因日常健康管理的费用和治疗费用会大幅上涨，同时借款人的收入基本已经定型，造成其资产下降，信用风险容易上升。

五、结论与政策建议

(一) 研究结论

本文运用捷信消费金融的公开信贷数据集，比较了树模型机器学习方法的分类预测效果，选取 LightGBM 模型作为分类基准模型，并运用 SHAP 模型进行归因分析，可视化各因素对消费信贷借款人信用风险的影响。实证结果发现：

(1) 在处理大数据、高维度的信贷数据集时，树模型机器学习算法的分类性能显著优于逻辑回归模型。在各种树模型中，XGBoost、LightGBM 和 CatBoost 三种模型分类表现最好且性能相似，LightGBM 的运算速度更快且分类预测表现更优。

(2) 特征重要性方面，借款人的个人基本信息和在消费金融平台的历史贷款信息两类特征的重要性相对更高。本文构造的时序特征和借款人个人基本信息的比率特征重要性排名也较高，消费信贷信用评估中信息的时效性分类和借款人情况的相对性较为重要。

(3) 特征影响方面，借款人个人基本特征的 SHAP 绝对值普遍高于其他类型特征，引起的借款人预期信用风险变化更大；借款人的标准化信用评分、贷款年金收入比率、最近 5 笔分期付款中提前还款次数和 1 年内征信记录贷款次数的 SHAP 值走势呈现明显的阶段性；教育水平、性别、年龄等特征对借款人信用风险的交互影响较为明显。

(二) 政策建议

基于本文中各因素对消费信贷借款人信用风险的影响结果，本文提出以下三方面建议：

第一，增强借款人信用影响因素信息统计的精细化。一方面，可以增加借款人收入来源渠道统计，涵盖主业、副业、房租收入等多种形式，并对职业所属的行业类型、公司规模、行业地位等划分类别和等级，从而更全面的评估借款人收入的稳定性和社会地位。另一方面，可以根据征信报告，以贷款类别统计贷款笔数和剩余应还款金额等特征指标，对住房贷款、个人经营性贷款、助学贷款等类型的贷款进行细分，有助于掌握借款人贷款参与度，更细致的评估借款人的其它方面的资金压力。

第二，消费金融公司可根据因素影响优化各阶段信用风险控制流程。贷前信用评估阶段，消费金融公司需对接多家征信机构，提升信用评估模型性能。目前，我国除官方的中国人民银行征信系统，还有百行征信、朴道征信等具有个人征信业务资质的市场化征信机构，以及芝麻信用、鹏元征信等个人信用评分机构。他们所用数据涵盖的传统信贷数据和互联网数据（如电商平台消费交易记录、水电煤缴费、人脉社交数据等）的比例、评估标准不同，信用评分结果存在差异。消费金融公司不仅需考虑征信机构的权威性和可靠性，还需保证其涵盖不同类型底层数据，从而增强信用风险管理。贷后信用风险管理阶段，消费金融机构需加强

高风险借款人的管理追踪。本文发现此类借款人包括一年内贷款次数超过 4 次，贷款年金收入比超 40%，高学历、28 岁以下的女性，以及 32 岁以下、高学历的已婚人群等。消费金融机构需持续追踪借款人的新增贷款情况、贷款额度变化以及贷款资金流向等，定期评估信用风险，增加信用评估频率，避免借款人出现套现、以贷养贷等行为。

第三，拓宽征信体系中借款人征信替代因素的范围和贷款表现因素的可查询年限。对于“征信白户”或“薄信用群体”，人民银行征信系统可通过拓宽征信替代数据范围补充其信用信息。资产负债类方面，可以增加对股票、债券等金融资产的统计；支付结算类方面，可以增加对房租缴纳、保险、公共事业缴费（水电气话费等）等信息的统计；对于经营行为信息，可以增加对用电用水用工、租赁、众筹等信息的统计；对于社交行为信息，可以增加对各社交平台中言论不当处理记录、手机使用率等信息的统计。此外，征信系统还可以扩大可查询贷款的年限范围。目前我国消费贷款年限最长可达 10 年，个人信用报告展示近 5 年还款记录和近 2 年的逾期金额。本文发现消费金融机构 2 年以上历史贷款的贷款状态和逾期天数都是借款人信用风险评估的重要特征，若能在征信报告中呈现借款人在其他消金机构较长时间内的贷款表现，一方面可以为消费金融机构评估借款人风险提供更多信息，另一方面也可以激励借款人中长期的良好贷款行为养成，有利于我国消费信贷行业后续发展。

参考文献

- [1] 廖理,吉霖,张伟强. 借贷市场能准确识别学历的价值吗?——来自 P2P 平台的经验证据[J]. 金融研究, 2015, 41(3): 146-159.
- [2] 马源聪,洪灏琪,林丽琼.网络借贷中借款人履约的逆向激励——基于“人人贷”数据的经验证据[J].财经研究,2020,46(5):66-80.DOI:10.16538/j.cnki.jfe.2020.05.005.
- [3] 陆燕春,周浩然. P2P 网贷中借款用途对借贷行为的影响——基于“人人贷”平台的实证研究[J]. 财会通讯, 2019, 810(10): 32-36.
- [4] Hertzberg A, Liberman A, Paravisini D. Screening on loan terms: evidence from maturity choice in consumer credit[J]. The Review of Financial Studies, 2018, 31(9): 3532-3567.
- [5] Hibbeln M, Norden L, Usselmann P, et al. Informational synergies in consumer credit[J]. Journal of Financial Intermediation, 2020, 44: 100831.
- [6] 张成洪,肖帅勇,陆天,卢向华,黄丽华.基于校园消费数据分析大学生网络借贷行为:借款倾向、消费变化与违约风险[J].系统工程理论与实践,2021,41(3):574-586.
- [7] 陈林,谢彦妩,李平,李强.借款陈述文字中的违约信号——基于 P2P 网络借贷的实证研究[J].中国管理科学,2019,27(4):37-47.DOI:10.16381/j.cnki.issn1003-207x.2019.04.004.
- [8] Duarte J, Siegel S, Young L. Trust and credit: The role of appearance in peer-to-peer lending[J]. The Review of Financial Studies, 2012, 25(8): 2455-2484.
- [9] 孙武军,樊小莹.从业经历和教育背景是否能提高借贷成功率?——来自 P2P 平台的经验证据[J].中央财经大学学报,2016(3):33-41.
- [10] 彭红枫,林川.言之有物:网络借贷中语言有用吗?——来自人人贷借款描述的经验证据[J].金融研究,2018(11):133-152.
- [11] 王正位,周从意,廖理,张伟强.消费行为在个人信用风险识别中的信息含量研究[J].经济研究,2020,55(1):149-163.
- [12] Ma X, Sha J, Wang D, et al. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning[J]. Electronic Commerce Research and Applications, 2018, 31: 24-39.

- [13] Xia Y, Liu C, Li Y Y, et al. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring[J]. Expert Systems with Applications, 2017, 78: 225-241.
- [14] Taha A A, Malebary S J. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine[J]. IEEE Access, 2020, 8: 25579-25587.
- [15] Al Daoud E. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset[J]. International Journal of Computer and Information Engineering, 2019, 13(1): 6-10.
- [16] 马晓君,宋嫣琦,常百舒,袁铭忆,苏衡.基于 CatBoost 算法的 P2P 违约预测模型应用研究[J].统计与信息论坛,2020,35(7):9-17.
- [17] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]. 31st Conference on Neural Information Processing Systems (NIPS 2017) , Long Beach, CA, USA.
- [18] Lundberg S M, Erion G G, Lee S I. Consistent individualized feature attribution for tree ensembles[J]. arXiv preprint arXiv:1802.03888, 2018.
- [19] Mangalathu S, Hwang S H, Jeon J S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach[J]. Engineering Structures, 2020, 219: 110927.
- [20] 李宗敏,张琪,杜鑫雨.基于辟谣微博的互动及热门评论情感倾向的辟谣效果研究——以新冠疫情相关辟谣微博为例[J].情报杂志,2020,39(11):90-95+110.
- [21] Parsa A B, Movahedi A, Taghipour H, et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis[J]. Accident Analysis & Prevention, 2020, 136: 105405.
- [22] Wen X, Xie Y, Wu L, et al. Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP[J]. Accident Analysis & Prevention, 2021, 159: 106261.

The Research on Credit Risk Influencing Factors of Consumer Credit Borrowers Based on SHAP

Sun Chao, Yao Xiaoyi, Liu Yi

(Hunan University, Chang Sha / Hunan, 410006)

Abstract: Based on the Kaggle public data set of Home Credit Consumer Finance, this paper compares the classification and prediction performance of tree model machine learning method, selects LightGBM model as the base model of classification, then uses SHAP model to analyze the attribution of classification results and visualize the impact of credit risk on consumer credit borrowers. The empirical results show that SHAP absolute values of borrower's personal basic characteristics are generally higher than that of other types of characteristics, which causes a greater change in borrower's expected credit risk. Characteristics such as borrowers' standardized credit score and annuity income ratio of loan, the SHAP value trend of them showed obvious stages. Characteristics such as education level, gender and age have obvious interactive influence on the credit risk of borrowers.

Keywords: Consumer Credit; Credit Risk; SHAP; LightGBM Model