Statistics and Application 统计学与应用, 2023, 12(6), 1696-1706
Published Online December 2023 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2023.126173

Hans汉斯

基于Logistic模型的小微贷不良用户画像

孙　璐，王小英，马锐欣

华北电力大学数理学院，北京

收稿日期：2023年11月5日；录用日期：2023年12月20日；发布日期：2023年12月28日

摘　要

随着互联网金融的发展及个人消费需求的日益增长，小微贷，特别是基于互联网的P2P借贷得到了较快的发展。但由于平台风险识别能力的缺失，部分平台产生大量违约情况，致使投资人遭受损失。为辅助平台及投资人有效识别不良用户，减少坏账带来的损失，本文基于国内外较有代表性的P2P平台：Prosper及拍拍贷上的数据，采用逻辑回归、决策树及支撑向量机三类模型对借款人进行信用评估，并依据模型结果得到小微贷不良用户画像。结果表明，逻辑回归模型时间复杂度低，具有优越的可解释性，更加适用于违约因素的研究。并且不良用户借款通常具有高利率、长期限的特点；同时用户本身没有稳定工作，收入较低。而常被我们关注到的性别、年龄以及学历反而影响较低。

关键词

个人信用评估，用户画像，数据不均衡，网络小微贷

Portrait of Non-Performing Users in Small and Micro Loans Based on Logistic Model

Lu Sun, Xiaoying Wang, Ruixin Ma

College of Mathematics and Physics, North China Electric Power University, Beijing

Received: Nov. 5th, 2023; accepted: Dec. 20th, 2023; published: Dec. 28th, 2023

Abstract

With the development of Internet finance and the increasing demand for personal consumption, small and micro loans, especially P2P lending based on the Internet, have developed rapidly. However, due to the lack of platform risk identification ability, some platforms have experienced a large number of default situations, resulting in losses for investors. To assist the platform and investors in effectively identifying non-performing users and reducing losses caused by bad debts, based on data from representative domestic and foreign P2P platforms such as Prospector and PaiPaiDai, this

文章引用: 孙璐, 王小英, 马锐欣. 基于Logistic模型的小微贷不良用户画像[J]. 统计学与应用, 2023, 12(6): 1696-1706.
DOI: 10.12677/sa.2023.126173

article uses three types of models, namely logistic regression, decision tree, and support vector machine to evaluate the credit of borrowers. Based on the model results, a portrait of non-performing users of small and micro loans is obtained. The results indicate that the logistic regression model has low time complexity and superior interpretability, making it more suitable for studying default factors. Non-performing user loans usually have the characteristics of high-interest rates and long-term limits; at the same time, users themselves do not have stable jobs and have lower incomes. The gender, age, and educational background, which we often pay attention to, have a lower impact.

Keywords

Personal Credit Assessment, User Portrait, Unbalanced Data, Online Small and Micro Loans

Copyright © 2023 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

随着互联网的发展及人们消费观念的改变，我国居民消费发生了转型和升级，个人消费的金融需求也日益增加，而小微贷款也因此得到了较快的发展[1]。互联网金融下的小微贷以 P2P 小微贷(peer-to-peer-lending)、银行小微贷、众筹小微贷以及电商小微贷等为主[1]。相较于传统信贷，这类基于网络的小微贷款大多具有金额较小、申请方便、申请门槛低等特点，可以使借款人在更短的时间内筹集到资金，也因此在小微企业融资和个人借贷方面起到了重要的作用。

我国网络信贷体系仍不成熟，在行业监管上许多规范仍未完善，也缺少征信系统信息共享，同时网络借贷过程中通常缺少抵押品，一旦产生违约投资人的权益将无法得到保障[2]。回顾我国 P2P 发展史，发现网络小微贷在为诸多借款人提供便利的同时，也为信贷市场带来了许多风险与不确定性。首先，由于借贷双方信息不对称，易产生违约情况[2]。其次，部分投资人风险识别能力较低，无法有效识别潜在的违约用户，这些使得网络小微贷款违约率进一步上升[2]。当大量的资金违约出现时，会使得贷款催收更加困难，资金流停滞，导致各小微贷平台出现问题甚至停业，进而使得投资人及平台遭受损失[3]。

因此，如何更为有效地识别出小微贷不良用户，以降低小微贷违约风险，对于网络小微贷平台及其投资者乃至整个小微贷行业日后的良性发展是十分重要的。本文将基于互联网小贷平台的实际数据，搭建逻辑回归模型，找出违约影响因素，并刻画小微贷不良用户画像。从而为小微贷平台及投资人在选取借贷用户时，提供一个更为简洁直观的参考，提升其风险辨别能力，进而降低贷款违约发生的概率，对网络小微贷的可持续性发展提供一些帮助。

2. 数据来源与预处理

2.1. 数据来源

本文使用的第一个数据集来自 Prosper。Prosper 是美国的纯中介类型小额借贷平台，经营始于 2006 年，通过向借贷双方收取管理费以及借款人逾期费用进行营利，运营时间较长，数据维度丰富[4]。该数据集包含 Prosper 平台 2006~2014 年间的 113,937 条数据，包含借款人相关信息、贷款相关信息等 81 个特征，且大部分特征存在不同程度的缺失情况。

本文使用的第二个数据集来自拍拍贷。拍拍贷是我国首家 P2P 网络借贷平台，成立于 2007 年，同样

通过贷款活动中借款人及投资人的手续费以及贷款逾期产生的费用获得收益[5]。作为我国最早成立的P2P网络借贷公司，积累了大量、丰富且连续性较好的数据。该数据集包含平台2015~2017年间的292,539条数据，共37个变量。数据较Prosper数据更为规整，仅少量特征存在数据缺失情况。

拍拍贷与Prosper的收益方式，平台性质都较为相近，同是网络小微借贷头部平台，具有较高代表性，因而将其应用于后续研究中。

2.2. 数据预处理

由于数据集中存在数据缺失、特征过多等情况，不利于后续的建模分析。为了将数据集更好地应用于最后的模型中，本文先通过Python对其进行了异常值排除、缺失值处理、特征选择、特征编码等预处理，流程如图1所示。

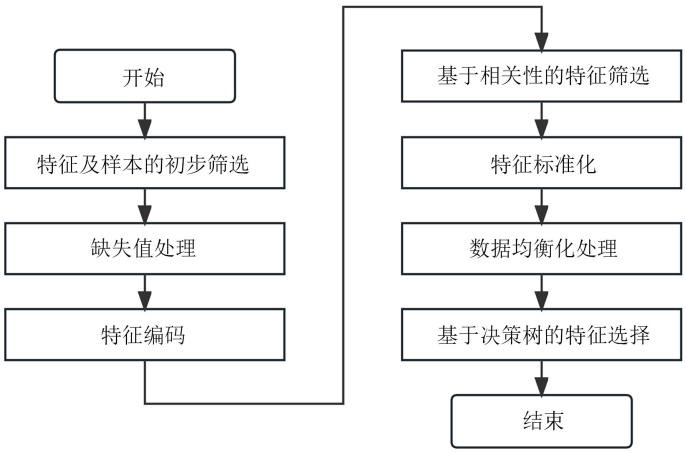

Figure 1. Data preprocessing flowchart
图 1. 数据预处理流程图

2.2.1. 特征及样本的初步选择

本文先删除掉其中一些无意义的编号类变量(贷款编号等)或与研究内容无关的变量(如平台方收益信息等)。筛选后Prosper数据集变量数目共37个，拍拍贷数据集剩余特征20个。

本文仅对用户最终是否违约进行预测，所以对于处于未完成状态的贷款样本做删除处理。此外，由于Prosper平台2009年7月重新运营后机制发生改变，于是仅选取2009年7月后的28,203条数据进行后续研究。对拍拍贷数据集，样本仅有三种状态：“逾期中”、“已还清”以及“正在还款中”。于是对正在还款中的变量进行删除，处理后样本量变为118,767。

2.2.2. 缺失值处理

为了后续模型的训练，本文针对存在缺失的变量做出了不同的处理。

Prosper数据集缺失情况如表1所示。

Table 1. Missing data situation on Prosper
表 1. Prosper 数据缺失情况表

变量名	是否缺失	数量	缺失数量	缺失率(%)
Prosper Rating (numeric)	True	28,072	131	0.464
Prosper Score	True	28,072	131	0.464

Continued

Occupation	True	28,176	27	0.096
Employment Status Duration	True	28,194	9	0.032
Debt To Income Ratio	True	25,010	3193	11.321

对于缺失率较低的 Prosper Rating (numeric)、Prosper Score、Occupation、Employment Status Duration 变量，直接删除相关的样本。对于缺失率为 11%的数值型变量 Debt To Income Ratio，使用均值进行填补。

拍拍贷数据集较为完整。在去除历史成功借款次数及历史成功借款金额存在的少量缺失样本后，对缺失率为 33%的历史逾期率特征，使用平均值进行填补。

2.2.3. 特征编码

本文使用的模型无法处理字符串类型变量，于是需要进行特征编码将字符串类型转化为数值类型。

对于取值较多的无序字符串特征，为保持取值之间的无序性，并使编码后维度尽量小，本文采取一种适用于二分类型目标变量的编码方式 WOE 编码(Weight of Evidence)。

对于 Prosper 数据集，需要 WOE 编码的变量有 5 个：Borrower State、Occupation、Income Range、Employment Status、Listing Category。

拍拍贷数据集，存在 11 个字符串型变量。9 个(视频认证、淘宝认证、学历认证、是否首标、征信认证、性别、手机认证、标的状态、户口认证)取值仅为两个，于是对这九个特征进行二进制编码。借款类型及初始评级，相较于其他变量取值较多，因此使用 WOE 编码。

2.2.4. 基于相关性的特征选择

为了进一步减少变量，提升模型训练速度。本文通过特征相关性对变量进行筛选，消除了一些可以被替代变量。对于 Prosper 数据集，筛选出了高相关性(0.7~1.0)特征共 6 组，结合整体情况及特征解释删除每组特征中的一个，最终删除了 5 个特征：Income Verifiable、Credit Score Range Lowe、Borrower APR、Prosper Score、Revolving Credit Balance，剩余特征 31 个。对于拍拍贷数据集，同样地，删除了 2 个特征：借款利率、历史成功借款次数，剩余特征 18 个。

接着，为减轻量纲不同造成的影响，对自变量进行了标准化。

2.2.5. 数据均衡处理

在信贷领域，违约用户通常占比较少，这种数据上的不平衡会使得在模型训练倾向于让多数类更容易被判断正确，牺牲掉少数类，影响模型效果。本文应用的 Prosper 数据中逾期样本量为 8392，还清样本量为 19,644，拍拍贷数据集逾期样本为 9599，已还清样本为 109,268，均存在一定程度上的数据不平衡。

本文尝试采用不同采样方法对数据进行平衡，所选取的采样方式共 4 种：随机下采样、随机上采样，SMOTE [6]以及 ADASYN [7]方法。采用这四种方式对数据进行平衡后，分别训练逻辑回归、决策树、支撑向量机模型，通过 10 折交叉验证检验各模型在不同采样方式下的模型效果，选出不同数据集适用于三类模型的采样方式。

对于 Prosper 数据集，其结果如表 2 所示，逻辑回归及支撑向量机模型均在 SMOTE 采样方式上表现更好，决策树则在随机上采样方式下表现更好，且其模型预测效果最好。

拍拍贷数据集不均衡状况较重。本文使用过采样与欠采样结合方式平衡数据，先通过随机欠采样方式选取 25,000 个还清样本，再对逾期样本进行过采样处理。得到结果如表 3 所示，逻辑回归及支撑向量机模型均在 SMOTE 平衡方式下表现更好，而决策树则是在随机上采样方式下表现更好。

Table 2. Prosper: Evaluation effect of each model under different data balancing methods
表 2. Prosper：不同数据平衡方式下各模型评估效果

模型	采样方式	Accuracy	Precision	Recall	F1 Score	Auc
LR	Random Undersampling	0.675	0.662	0.714	0.687	0.739
	Random Oversampling	0.677	0.663	0.718	0.689	0.740
	SMOTE	**0.679**	**0.665**	**0.723**	**0.693**	**0.742**
	ADASYN	0.658	0.641	0.688	0.664	0.716
DT	Random Undersampling	0.594	0.596	0.587	0.591	0.594
	Random Oversampling	0.831	0.773	0.936	0.846	0.831
	SMOTE	**0.717**	**0.704**	**0.726**	**0.708**	**0.717**
	ADASYN	0.707	0.691	0.707	0.692	0.707
SVM	Random Undersampling	0.681	0.661	0.743	0.699	0.74
	Random Oversampling	**0.712**	**0.687**	**0.776**	**0.729**	**0.782**
	SMOTE	0.723	0.697	0.789	0.740	0.795
	ADASYN	0.703	0.670	0.777	0.720	0.769

Table 3. PaiPaiDai: Evaluation effect of each model under different data balancing methods
表 3. PaiPaiDai：不同数据平衡方式下各模型评估效果

模型	采样方式	Accuracy	Precision	Recall	F1 Score	Auc
LR	Random Undersampling	0.712	0.682	0.796	0.735	0.800
	Random Oversampling	0.711	0.682	0.790	0.732	0.800
	SMOTE	**0.712**	**0.683**	**0.794**	**0.734**	**0.801**
	ADASYN	0.692	0.655	0.783	0.713	0.774
DT	Random Undersampling	0.867	0.839	0.908	0.872	0.871
	Random Oversampling	0.948	0.917	0.985	0.950	0.954
	SMOTE	**0.909**	**0.901**	**0.919**	**0.910**	**0.913**
	ADASYN	0.894	0.892	0.893	0.892	0.899
SVM	Random Undersampling	0.745	0.697	0.867	0.773	0.825
	Random Oversampling	**0.762**	**0.719**	**0.863**	**0.784**	**0.840**
	SMOTE	0.763	0.717	0.870	0.786	0.842
	ADASYN	0.740	0.684	0.876	0.768	0.812

2.2.6. 基于决策树的特征选择

前文中数据集在决策树模型下表现最好，预测结果最为准确。于是使用随机上采样方式下的决策树特征重要性来进行特征选择，令对标签影响较大的特征进入最终的模型。需要找到一个特征重要性的阈值，令大于该阈值的变量入模时预测效果最好。阈值的选取过程如下：

1). 使用决策树模型计算特征的重要性，对模型中 30 个变量使用特征重要性进行降序排列；

2). 每次删除一个特征重要性最低的变量，使用剩余变量建立决策树模型，记录模型评估结果；

3). 绘制不同特征数量下，各模型的评估指标曲线，选择最优模型自变量数量。

通过 Python 实现以上任务，训练得到在 Prosper 及拍拍贷数据集上，不同阈值下模型的准确率和 AUC 值分别如图 2 及图 3 所示。

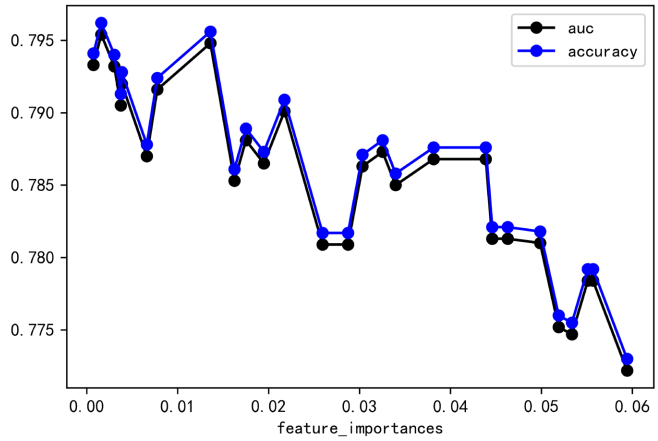

Figure 2. Prosper: Model performance under different number of features
图 2. Prosper：不同特征数量下模型表现

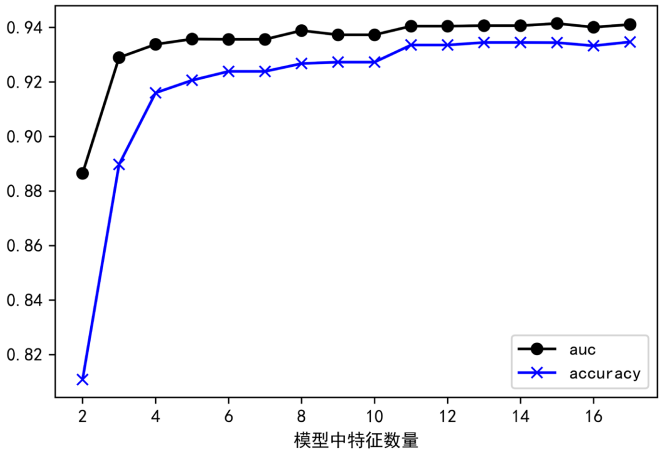

Figure 3. PaiPaiDai: Model performance under different number of features
图 3. 拍拍贷：不同特征数量下模型表现

本文希望通过特征选择提升模型效果，并尽可能地减少特征以缩短模型训练时间。综合考虑这两点因素，来选取阈值。

对于 Prosper 数据集，最终选取了 feature_importance = 0.013602 作为阈值，对该阈值下变量(Is Borrower Homeowner、Current Delinquencies、Income Verifiable、Income Range、Currently In Group、Public Records Last 12 Months、Investment From Friends Amount、Investment From Friends Count)进行删除，剩余变量 22 个。

对于拍拍贷数据集，随着模型中特征数量增加，模型效果是呈上升趋势的，并且在 17 个特征全部进入模型时模型效果最佳，因此不对模型变量进行删除。

3. 模型训练及结果

在本节中，将基于前文中验证得到的适用于各类模型的平衡方法，平衡相应模型下的数据集。将数据集以 3:7 的比例划分为测试集与训练集。通过 Python 的 sklearn 库中的逻辑回归、决策树和支撑向量机

函数对模型进行训练。

3.1. 基于 Prosper 平台数据的模型建立

基于模型准确率对模型部分参数进行调优，得到模型参数调整情况如表 4 所示。

Table 4. Model tuning parameters based on the Prosper dataset
表 4. 基于 Prosper 数据集的模型调优参数

模型	采样方式	参数
逻辑回归	SMOTE	C = 0.071
决策树	Random Oversampling	random_state = 0, max_depth = 32
支撑向量机	SMOTE	Kernel = ‘rbf’, cache_siz e= 3000

模型调优后，得到模型在测试集上的效果如表 5 所示。

Table 5. Model effect based on the Prosper dataset
表 5. 基于 Prosper 数据集的模型效果

模型	采样方式	Accuracy	Precision	Recall	F1 Score	Auc	Time (s)
逻辑回归	SMOTE	0.682	0.671	0.723	0.696	0.742	0.23
决策树	Random Oversampling	**0.796**	**0.758**	**0.873**	**0.811**	**0.795**	**0.65**
支撑向量机	SMOTE	0.712	0.677	0.798	0.732	0.784	74.78

模型评估指标方面，可以看到，无论从 Accuracy、F1，还是 AUC 值几个指标上，决策树模型的表现都是最优的，其次为支撑向量机模型，逻辑回归模型相比而言较低。模型的解释性方面，逻辑回归模型优于决策树模型性，更优于支撑向量机模型。逻辑回归模型返回的特征系数不仅可以反映特征重要性，还可以表现各自变量如何对因变量产生影响。模型的时间复杂度上，逻辑回归及决策树模型计算时间均较短。

3.2. 基于“拍拍贷”平台数据的模型建立

基于模型准确率对模型部分参数进行调优，得到模型参数调整情况如表 6 所示。

Table 6. Model tuning parameters based on the PaiPaiDai dataset
表 6. 基于拍拍贷数据集的模型调优参数

模型	采样方式	参数
逻辑回归	SMOTE	C = 0.03
决策树	Random Oversampling	random_state = 0, max_depth = 35
支撑向量机	SMOTE	Kernel = ‘rbf’, cache_size = 3000

模型调优后，得到模型在测试集上的效果如表 7 所示。

Table 7. Model effect based on the PaiPaiDai dataset
表 7. 基于拍拍贷数据集的模型效果

模型	采样方式	Accuracy	Precision	Recall	F1 Score	Auc	Time (s)
逻辑回归	SMOTE	0.711	0.679	0.800	0.735	0.800	0.89
决策树	Random Oversampling	**0.935**	**0.901**	**0.977**	**0.938**	**0.942**	**0.20**
支撑向量机	SMOTE	0.761	0.718	0.865	0.785	0.837	77.46

与 Prosper 数据表相似，在准确性方面，仍是决策树优于支撑向量机、逻辑回归模型，决策树模型在测试集上的准确性甚至准确率达到了 0.935。在时间复杂度上，支撑向量机模型则远超逻辑回归和决策树模型。在可解释性上，逻辑回归则优于其他两种。

3.3. 违约要素分析

本节基于上文逻辑回归模型的回归系数及决策树模型训练后返回的特征重要性，分别得到两模型下的特征重要性排序，从而寻找到对借款人违约行为影响较大的因素。于是将相应特征分别按回归系数和特征重要性降序排列，得到 Prosper 数据集上两模型中特征重要性排名前十位的变量如表 8 所示。

Table 8. Prosper: Comparison of feature importance levels
表 8. Prosper：特征重要程度对比

逻辑回归	决策树
Borrower APR	**Borrower APR**
Term	**Stated Monthly Income**
Occupation	Employment Status Duration
Loan Original Amount	Available Bankcard Credit
Stated Monthly Income	Investors
Credit Score Range Lower	**Total Trades**
Employment Status	**Borrower State**
Borrower State	Revolving Credit Balance
Listing Category (numeric)	Debt To Income Ratio
Total Trades	**Occupation**

在 Prosper 数据集中，尽管两模型给出特征重要程度较高的变量有所差别，但有 5 个变量都起到了重要作用。

其中，在两类模型中重要性最高的特征均为借款利率(Borrower APR)，在 Prosper 平台中借款利率由 Prosper 评级给出，评级更安全的用户可以享受更低的借款利率。也因此借款利率和 Prosper 评级这两个变量对最终违约预测十分重要。

而在用户信息方面，用户所在的州(Borrower State)以及当前职业(Occupation)这两类变量也是预测过程中要被重点关注的。用户偿还能力方面，月收入(Stated Monthly Income)以及当前开设的贷款数量(Total Trades)，这两类变量也应成为考察的重点。

拍拍贷数据集上两模型中特征重要性排名前十位的变量如表 9 所示。

Table 9. PaiPaiDai: Comparison of feature importance levels
表 9. 拍拍贷：特征重要程度对比

排序	逻辑回归模型	决策树模型
1	**借款利率**	**借款利率**
2	**借款期限**	借款金额
3	**借款类型**	**历史成功借款金额**
4	手机认证	年龄
5	**历史成功借款金额**	**借款期限**
6	户口认证	**借款类型**

Continued

7	**历史逾期率**	**历史成功借款次数**
8	视频认证	**历史逾期率**
9	历史逾期还款期数	性别
10	**历史成功借款次数**	学历认证

在拍拍贷数据集中，尽管两模型返回的特征重要性较高变量有些差别，但其中有 6 个特征都在违约预测过程中起到了重要作用。“借款利率”这一变量无论是在逻辑回归还是在决策树模型当中都是最重要的，通过拍拍贷的运营模式我们知道借款利率是通过“初始评级”来确定的，也就是说用户的初始评级对违约行为的预测起着重要作用，而“借款类型”、“借款期限”同样需要关注。“历史成功借款金额”、“历史逾期率”及“历史成功借款次数”这三类历史表现也应被列为考察的重点。

3.4. 构建用户画像

本节将基于上文逻辑回归模型结果进行用户画像的构造。由于在数据预处理阶段对特征进行了标准化，消除了特征间量纲差异，这使得逻辑回归模型中各变量的回归系数具有可比性。而其中绝对值较大的变量则对因变量违约预测影响较大。

本文希望选择对违约情况影响较大的变量，进行模型的构建。于是对于 Prosper 数据集选取了回归系数绝对值大于 0.15 的 10 个变量，通过词云的方式对基于 Prosper 建立的用户画像进行展示，得到不良用户画像如图 4 所示，优质用户画像如图 5 所示。

Figure 4. Non-performing user profile based on Prosper
图 4. 基于 Prosper 的不良用户画像

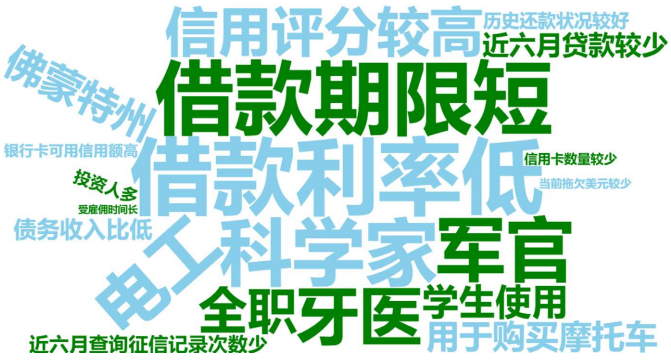

Figure 5. High-quality user profile based on Prosper
图 5. 基于 Prosper 的优质用户画像

相应地，对于拍拍贷数据集，选出回归系数绝对值大于 0.13 的变量。得到基于拍拍贷数据的不良用户画像如图 6 所示，优质用户画像如图 7 所示。

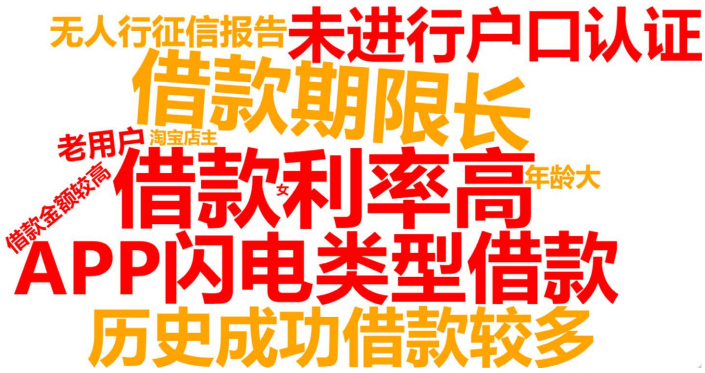

Figure 6. Non-performing user profile based on PaiPaiDai
图 6. 基于拍拍贷的不良用户画像

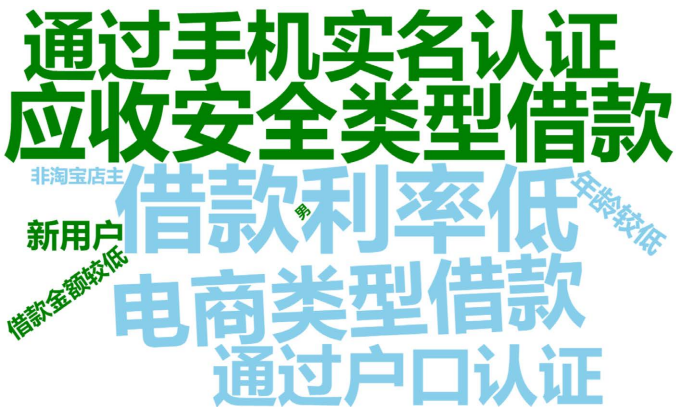

Figure 7. High-quality user profile based on PaiPaiDai
图 7. 基于拍拍贷的优质用户画像

4. 结论

本文分别基于国内外网络小微贷平台：Prosper 及拍拍贷数据进行了实证分析，通过建立逻辑回归、决策树及支撑向量机模型来进行信用评估，对用户违约情况进行预测，通过准确率等模型评价指标进行了模型效果的对比，并依据模型结果得到了违约影响因素，完成了小微贷不良用户画像的刻画。

通过本文研究，可以得出以下三方面结论：

1) 信用评估模型效果比较：

本文构建的三类模型均在拍拍贷数据集上表现较好，而三类模型相比较而言：支撑向量机模型的预测效果略优于逻辑回归模型，但是可解释性较低，计算量大，时间复杂度最高。逻辑回归模型具有较高可解释性，其回归系数不仅可以反映特征重要性，同时也可以通过符号判断各特征如何作用于因变量，模型的时间复杂度低，但模型预测效果低于决策树和支撑向量机模型。决策树模型同样具有时间复杂度较低的特点，同时模型效果是三者中最优的，模型解释性方面略逊于逻辑回归模型。从综合表现来看，决策树模型表现最优。而在实际应用中，逻辑回归尤其优越的可解释性，也常被应用到信用评估模型当中。

2) 违约影响因素：

本文依据决策树和逻辑回归模型返回的结果进行了特征重要性的对比，得到了对违约情况影响较强

的因素。尽管在这两类模型当中并不完全一致，但是通过对比发现：借款信息类的利率或期限等、由第三方数据平台数据得到的信用评级类数据、历史行为类的历史借款及还款情况以及用户收入情况类数据都对违约情况的产生有着较大的影响，而这也是我们在进行信用评估时应加以注意的。

3) 不良用户画像构建：

本文通过逻辑回归模型训练结果，构造了不良用户画像。不良用户画像主要特点如下：

借款信息类：借款情况通常具有高利率、长期限的特点。

用户基本信息类：通常没有收入较高且较为稳定的工作。特别地，在 Prosper 数据集中，不良用户通常来自于经济不景气的州。而常被我们关注到的性别、学历以及年龄对用户的违约行为影响较小，不作为不良用户刻画的指标。

历史行为类：历史逾期比例通常较高，并且由于其对贷款的依赖，其历史的借款次数及金额通常也比较多和高。

另外，不良用户通常具有较高的资金周转需求，这使得他当前具有较多的贷款交易数，而同时各平台给出的信用评级也是重要参考之一。

参考文献

[1] 唐兴红. 浅谈商业银行在互联网金融时代“小微贷”发展模式[J]. 知识经济, 2015(23): 54.

[2] 金虎斌, 张成虎. 网络借贷平台的信息处理与信用评级效率分析——基于人人贷与 Prosper 的实证对比检验[J]. 上海经济研究, 2017(10): 45-58.

[3] 王浩博. 基于二元 Logistic 模型的 P2P 违约分析[J]. 现代商业, 2020(30): 106-109.

[4] 曾雪云, 张舒铭. Prosper 商业模式的“双刃剑”效应——兼论社区自治与风险隐患[J]. 财会月刊, 2022(4): 115-118.

[5] 刘美玲, 王佳. P2P 网贷运行模式及风险分析——以拍拍贷为例[J]. 现代商贸工业, 2018, 39(2): 114-116.

[6] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. https://doi.org/10.1613/jair.953

[7] He, H.B., Bai, Y., Garcia, E.A. and Li, S.T. (2008) ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. 2008 *International Joint Conference on Neural Networks* (*IEEE World Congress on Computational Intelligence*), Hong Kong, 1-8 June 2008, 1322-1328. https://doi.org/10.1109/IJCNN.2008.4633969