

# Research on the Factors Influencing Successful Targets on P2P Platform

## —A Case Study of Ren Ren Dai

Mengling Liu

School of Finance and Statistics, Shanghai University of Finance and Economics Zhejiang College, Jinhua Zhejiang  
Email: liumengling90@163.com

Received: Jan. 7<sup>th</sup>, 2016; accepted: Jan. 24<sup>th</sup>, 2016; published: Jan. 27<sup>th</sup>, 2016

Copyright © 2016 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

### Abstract

In this paper, 5718 target recordings on Ren Ren Dai (a Peer-to-Peer lending platform) are selected for research. We take the variable which can reflect the target status as dependent variable, and lend credit scores, loan amount, etc. as independent variables. In order to help borrowers predict the success ratio, we build the Logistic model; the accuracy rate of the model is verified to be 93.7%. In addition, the results of the decision tree classification show a failure rate of 0.05. Finally, a comprehensive of the two methods comes to a conclusion that: credit rating, lending credit score and overdue times have the greatest impacts on the target status, and the variable target type has no significant effect.

### Keywords

Peer-to-Peer, Logistic Model, Decision Tree, Failure Rate

---

## P2P平台满标影响因素研究

### —以人人贷为例

刘梦玲

上海财经大学浙江学院, 金融与统计学院, 浙江 金华  
Email: liumengling90@163.com

收稿日期：2016年1月7日；录用日期：2016年1月24日；发布日期：2016年1月27日

## 摘要

本文选取以人人贷为例的P2P平台上5718条标的信息记录，以能反映标的状况的变量作为因变量，以借出信用分、借款金额等5个变量作为自变量，建立Logistic回归模型以帮助借款人对借款成功(满标)概率进行预测，最终模型预测的正确率达到93.7%。此外，以标的状况为因变量的决策树分类的结果显示误判率为0.05，综合两种方法最终得出信用等级、借出信用分数和以往逾期次数对标的状态影响最大，而标的类型的影响不显著。

## 关键词

P2P, Logistic模型, 决策树, 误判率

## 1. 引言

随着互联网微金融的发展，P2P (peer to peer)网络借贷作为一种新的借贷形式，凭借其低门槛，高收益等优点越来越广泛的被人们所接受。P2P 成立的思想起源于穆罕默德·尤努斯的“乡村银行”这一理念，旨在为那些资质不足以在银行获取贷款的群体提供小额贷款帮助，平台本身不参与交易，只是作为中介提供借贷双方交易的平台。P2P 平台作为金融中介的网络借贷平台能有效地提高借款人的信用状况，能减少借款人与出借人之间信息不对称的问题[1]。P2P 平台交易的方式类似于竞拍：首先是借款人发布借款列表，然后是投资人给出自己的投资金额和投资利率，与其他投资人进行竞标。最后按照投资利率高低筛选出竞标成功的投资人，标的生效。如果过了标期，仍未筹满借款，则宣布流标。基于此第一家 P2P 平台 Zopa 2005 成立于英国，随后在 2006 年美国第一家 P2P 平台 Prosper 成立了，国内第一家 P2P 平台拍拍贷成立于 2007 年，借鉴了 Prosper 运营模式。具不完全统计，截止到 9 月底，我国有 1438 家 P2P 平台。

由于在 P2P 平台上借款人信用等级较低，平台为了控制风险，往往会对借款人的个人信息进行审核，并制定出一个较高的利率。同时将满标率控制在一定的范围内[2]。就人人贷而言，截止到 2013 年 8 月 15 日的所有 14446 条借款记录中，仅有 2307 条记录显示借款成功，这部分仅占到 15.97%。分析满标的主要影响因素一方面能够帮助借款者提高成功筹借资金的概率；另一方面对于平台来说，在借款人资质符合要求的前提下提高满标率，可以增加平台收益。国内外研究 P2P 平台的文献较少，特别是对满标影响因素的分析，丁婕在对借款满标的影响因素分析时，以拍拍贷数据为研究对象，将历史成功/失败次数、朋友数目加入 Logistic 回归模型中，来辅助借款人预测满标概率，最终预测率为 89.6% [3]。国外学者利用 Prosper 网站的数据研究了不同种类的变量对借贷结果的影响，通常包括借款人的信用水平、财务状况以及借款特征信息[4]。Puro 针对 Prosper 平台数据，建立了借款人决策辅助工具，根据研究对象的特征，分析了 2 万条借贷需求进行分析，最终筛选出信用等级、当前债务拖欠期、借款金额、债务收入比和借款利率作为主要变量，预测借款人的借款成功率[5]。Scott 等人研究了社会资金对出借人决策的影响，通过实验数据证明了出借者会根据借款者的借款金额来决定是否进行投资[6]。

本文拟前人研究的基础上，针对人人贷平台数据，加入借款人逾期次数以及标的基本信息(如借款金额、借款期限等)，希望对借款人满标概率进行预测，进一步得出对满标概率影响较为重要的因素来帮助借款人提高借款成功率。

## 2. 理论背景

本文拟采用 Logistic 回归对满标影响因素进行建模分析，希望通过最终建立的模型解释各个变量对满标具体的影响程度。然后结合决策树分类的方法，借鉴其划分类别的标志变量，找出影响满标的重要因素。

### 2.1. Logistic 回归模型

如果令二分类变量  $Y=1$  的概率为  $p$ ，则有  $Y=0$  的概率为  $(1-p)$ 。有

$$\text{logit}(p) = \ln \frac{P}{1-P} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n。$$

这种  $p$  与自变量之间的回归关系式就是 Logistic 回归模型。将  $p$  变换为  $\ln \left[ \frac{p}{(1-p)} \right]$  成为 logit 变换，记为  $\text{logit}(p)$ ，所以也称为 logit 模型。logit 变换使得在  $[0,1]$  范围内取值的  $p$  变换到  $(-\infty, +\infty)$ ，当  $p$  趋向于 0， $\text{logit}(p)$  趋向于  $-\infty$ ，当  $p$  趋向于 1， $\text{logit}(p)$  趋向于  $+\infty$ 。 $\frac{P}{1-P}$  称为“机会”或“优势”(odds)。

### 2.2. 决策树

决策树(Decision Tree)又称为判定树，是运用于分类的一种树结构。其中的每个内部节点代表对某个属性的一次测试，每条边代表一个测试结果，叶节点代表某个类或者类的分布，最上面的结点是根结点。决策树提供了一种展示类似在什么条件下会得到什么值的这类规则的方法。

使用决策树进行分类分为两个步骤：

1) 利用训练集建立并精化一颗决策树，建立决策树模型。这个过程实际上是一个从数据中获取知识，进行机器学习的过程。

2) 利用生成完毕的决策树对输入数据进行分类。对输入的记录，从根结点依次测试记录的属性值，直到到达某个叶结点，从而找到该记录所在的类。

## 3. 实证研究

### 3.1. Logistic 回归结果

#### 3.1.1. 数据描述

本文所选取的是利用网络爬虫从人人贷平台网站上抓取的自 2013 年 5 月至 12 月的 5718 条借款记录，其中有 1758 条记录显示已满标(用还款状态标记，若已还清或已垫付，则表明该标的已满标；已流标表明借款未能及时筹到所需金额的款项)。另有 3960 条记录显示已流标，流标率占 69.26%。

本文选择标的还款状态来表示满标与否，人人贷平台上标的还款状态有已还清、已垫付和已流标。已还清和已垫付说明标的列表在标期内已筹借到足够的资金，我们将其视为标的满标；已流标代表着该标的由于平台审核等原因而未能在标期内及时抽满借款，借款宣布失败。然后选择了标的类型(信用认证标、机构担保标、实地认证标)、借款金额、借款期限、利率、信用等级、借出信用分这六个变量作为自变量来研究满标的影响因素。

#### 3.1.2. 数据整理

$$\text{btype} = \begin{cases} 1 & \text{信用认证标} \\ 0 & \text{其它} \end{cases} \quad \text{credit1} = \begin{cases} 1 & \text{AA, A} \\ 0 & \text{其它} \end{cases} \quad \text{credit2} = \begin{cases} 1 & \text{B, C, D} \\ 0 & \text{其它} \end{cases}$$

$$\text{credit3} = \begin{cases} 1 & E \\ 0 & \text{其它} \end{cases}$$

$\text{btype} = 1$  表示信用认证标， $\text{btype} = 0$  代表机构担保标和实地认证标； $\text{credit1} = 1$ ， $\text{credit2} = 0$ ， $\text{credit3} = 0$  代表信用状况高(AA,A)； $\text{credit1} = 0$ ， $\text{credit2} = 1$ ， $\text{credit3} = 0$  代表信用状况良好(B,C 或 D)； $\text{credit1} = 0$ ，

credit2 = 0, credit3 = 1 代表信用状况中等(E); credit1 = 0, credit2 = 0, credit3 = 0 代表信用状况差(HR)。amount 为借款金额, period 为借款期限, interest 为借款利率, lcredit 为借出信用分, overdue 表示历史逾期次数。

在所有的 5117 条信用标中, 1157 条记录显示为满标占 22.61%, 3960 条记录显示为流标, 占 77.39%; 其他标的类型的 601 条记录均满标。信用状况高的 645 条记录中 99% 的借款满标, 只有 1% 的流标; 信用状况良好的 324 条记录中 85.5% 的借款满标, 14.5% 的借款流标; 信用状况中等的 395 条记录中 71.9% 的借款满标, 28.1% 的借款流标; 信用状况差的 12.8% 的借款满标, 87.2% 的借款流标。

### 3.1.3. 模型建立

我们将要建立的模型形式为:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \text{btype} + \beta_2 \text{amount} + \beta_3 \text{period} + \beta_4 \text{interest} + \beta_5 \text{credit1} + \beta_7 \text{credit2} + \beta_8 \text{credit3}$$

其中  $P$  为待求的满标概率。

用 spss 对数据进行向后步进(似然比)方法进行回归后的结果如表 1 所示。

由表 1 可以看出, 模型中 btype 这个变量并不显著, 我们将其强制去掉后再次进行回归, 得到的结果如表 2 所示。

据此建立如下模型:

$$\ln\left(\frac{P}{1-P}\right) = -8.137 - 0.000009\text{amount} - 0.065\text{period} - 4.502\text{interest} \\ + 2.129\text{credit2} + 2.001\text{credit3} + 0.081\text{lcredit}$$

模型结果表明: 随着借款金额、借款期限和借款利率的提高, 满标概率逐渐降低; 信用等级越高, 借出信用分数越高, 满标概率大。

### 3.1.4. 模型评价

该模型分类的效果如表 3 所示。

由表 3 我们可以看出, 满标的正确率为 94.0%, 流标的正确率为 93.2%。模型整体拟合的正确率为 93.7%, 误判率较低。

表 4 显示了最终模型的 Cox&Snell R 方为 0.595, Nagelkerke R 方为 0.840。说明模型整体上拟合效果较好。

为了进一步验证模型的推广作用, 我们用五折交叉验证对该数据集进行检验, 用 R 软件操作的结果显示, 训练集的误判率为 0.0624, 测试集的误判率为 0.0633。再次说明了所建立的该模型效果较好, 可以用以推广。

## 3.2. 决策树分类结果

为了更进一步的得出对满标有重要影响的因素, 我们采用决策树来找出能够较程度上区分满标和流标的关键变量。用 R 软件操作的结果为:

运行的结果如图 1 所示。

图 1 显示了决策树分类的结果, 我们可以看出决定满标与否的两个重要变量为借出信用分(lcredit)。运用决策树分类的误判率为 0.055。分类效果较好, 说明依据 lcredit 这个变量进行分类的结果是可靠的。

Table 1. Regression results for the first time

表 1. 第一次回归的结果

	B	S.E.	Wals	df	Sig.	Exp (B)
btype(1)	21.380	1599.984	0.000	1	0.989	1928037172.82
Amount	0.000	0.000	31.552	1	0.000	1.000
Period	-0.088	0.012	55.426	1	0.000	0.916
Interest	-4.112	1.628	6.382	1	0.012	0.016
credit1(1)	4.151	0.674	37.969	1	0.000	63.497
credit2(1)	2.040	0.252	65.803	1	0.000	7.694
credit3(1)	1.940	0.193	101.506	1	0.000	6.957
lcredit	0.078	0.003	894.283	1	0.000	1.082
常量	-12.003	0.979	150.301	1	0.000	0.000

Table 2. Regression results for the second time

表 2. 第二次回归的结果

	B	S.E.	Wals	df	Sig.	Exp (B)
Amount	0.000	0.000	39.994	1	0.000	1.000
Period	-0.065	0.010	42.566	1	0.000	0.937
Interest	-4.502	1.630	7.633	1	0.006	0.011
credit2(1)	2.129	0.236	81.384	1	0.000	8.407
credit3(1)	2.001	0.181	122.475	1	0.000	7.393
lcredit	0.080	0.002	1122.611	1	0.000	1.083
常量	-8.137	0.495	270.191	1	0.000	0.000

Table 3. Classification results

表 3. 分类结果表

已观测	已预测			百分比校正
	finished			
	0	1		
finished	0	3721	239	94.0
	1	119	1639	93.2

Table 4. Model summary

表 4. 模型汇总

步骤	-2对数似然值	Cox & Snell R方	Nagelkerke R方
1	1884.689	0.595	0.840
2	1886.340	0.595	0.840

#### 4. 结论

综合实证部分的分析我们可以得到，在决定满标状态的因素中，借款人的信用等级、借出信用分起

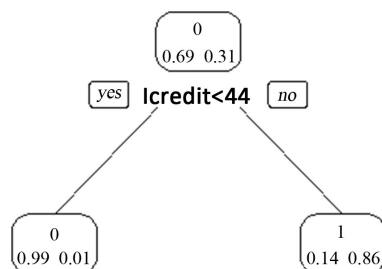


Figure 1. Decision tree classification results

图 1. 决策树分类结果

着重要的作用，具体来说信用级别为 B、C 和 D 的借款人借款成功机率高于其他级别的 8.4 倍；信用级别为 E 的借款人借款成功机率高于其他级别的 7.4 倍；随着借出信用分数的增加，每增加一分信用分，就增加了 1.08 倍的成功几率。

由于本文只是针对人人贷的数据进行的建模，而每个 P2P 平台运营模式不完全相同，数据都各有自己的特色，我们所建立的模型的适用性有待进一步研究。因此在接下来的研究中，我们希望通过搜集更多平台的标的的数据，找出各个平台的数据共性，最终能建立一套普适的模型，对广大的借款者进行发布借款列表起到一定的参考作用。

## 参考文献 (References)

- [1] 陈冬宇, 李伟军, 丁婕. 网络借贷借款人决策辅助模型[J]. 吉林大学学报, 2012.
- [2] 辛硕. 关于 p2p 的风险控制的探讨[J]. 金融观察, 2014.
- [3] 丁婕. 我国 P2P 网络借贷平台及借款人行为研究——以拍拍贷为例[D]. 四川: 西南财经大学, 2012.
- [4] 莫易娴. P2P 网络借贷国内外理论与实践研究文献综述[J]. 金融讲坛, 2011.
- [5] Kumar (2007) Bank of One: Empirical Analysis of Peer-to-Peer Financial Marketplaces. *13th Americas Conference on Information Systems*, Association for Information System, Keystone, 1-8.
- [6] Puro, L. (2010) Borrower Decision Aid for People-to-People Lending. *Decision Support Systems*, **49**, 52-60. <http://dx.doi.org/10.1016/j.dss.2009.12.009>