佰仟金融评分卡开发项目文档

风控部

Risk Management and Control Department

*Confidential*

**版本说明**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本号 | 作者 | 批准日期 | 执行日期 | 备注 |
| V1.0 |  |  |  |  |

目 录

[业务综述](#_Toc463710148)

[业务驱动 4](#_Toc463710149)

[ 现状 4](#_Toc463710150)

[ 挑战 4](#_Toc463710151)

[项目时间表](#_Toc463710152)

[项目参数和定义](#_Toc463710153)

[目标客户定义 6](#_Toc463710154)

[好坏客户定义 6](#_Toc463710155)

[ 坏客户定义： 6](#_Toc463710156)

[ 好客户定义： 6](#_Toc463710157)

[ 灰色客户定义： 6](#_Toc463710158)

[采样窗口和表现窗口 6](#_Toc463710159)

[ 表现窗口 6](#_Toc463710160)

[ 采样窗口 6](#_Toc463710161)

[被排除样本 7](#_Toc463710162)

[数据准备](#_Toc463710163)

[数据来源 8](#_Toc463710164)

[ 基本数据 8](#_Toc463710165)

[ 行为数据 8](#_Toc463710166)

[ 其他数据 8](#_Toc463710167)

[数据清理 8](#_Toc463710168)

[ 必填项数据的处理 8](#_Toc463710169)

[训练和验证数据 9](#_Toc463710170)

[建模](#_Toc463710171)

[同质性删除（Homogeneity Exclusions） 10](#_Toc463710172)

[粗分（Variables Pre Binning） 10](#_Toc463710173)

[变量相关性分析（Correlation Analysis） 11](#_Toc463710174)

[变量细分（Variables Binning） 12](#_Toc463710175)

[算法 14](#_Toc463710176)

[评分卡（Score Card） 15](#_Toc463710177)

[评分卡的有效性](#_Toc463710178)

[K－S检验（Kolmogorov-Smirnov） 18](#_Toc463710179)

[提升度（Lift） 18](#_Toc463710180)

[不同分数段的坏客户比例 19](#_Toc463710181)

[区分度（Divergence Chart） 19](#_Toc463710182)

[评分卡分数转换](#_Toc463710183)

[转换方法 20](#_Toc463710184)

[ODDS分数对应表 20](#_Toc463710185)

[评分卡实施前的分析](#_Toc463710186)

[用最新的样本验证 22](#_Toc463710187)

业务综述

业务驱动

* 现状

随着佰仟业务迅速的发展，销量爆发式增长，催收人力无法满足逾期案件的催收需求。目前对所有的案件采用无差异化的催收策略，造成对低风险案件的人力浪费和对高风险案件的催收力度不足，为实现催收案件的精细化管理，以便于合理分配人力资源和建立多样化催收策略，开发本套催收评分卡，期望能够对客户进行风险等级的划分，提高催收效率，降低人力成本。

* 挑战
* 首期进入催收的客户没有还款行为数据。
* 申请表数据的正确性和可靠性有待提高。
* 业务发展迅速，时间窗口选的太远则样本表现比较稳定但是样本数量不够，时间窗口选的太近则样本数量充足但是样本表现不稳定，选取样本时需综合考虑样本数量及稳定性，使得评分卡模型能够充分体现业务特色，反映人群风险。
* 不同产品的业务模式不同，直接影响到评分卡的稳定性。
* 催收团队发展迅速，在不同时期的可用催收人力不同，回收率具有不稳定性。
* 业务发展变化过快，不同时期申请合同的逾期表现不同。
* SA由于风控绩效考核，会私下对客户进行催收，此部分行为并未记录在案，所以部分催收数据不够精确。

项目时间表

表1 项目时间表

|  |  |  |
| --- | --- | --- |
| Item | Descriptions | Schedule |
| (i) | Define Business Problem | 10.17 – 10.21 |
| 1 | Identify and confirm project objectives | 10.17 – 10.21 |
| 2 | Select data required for analysis exercise | 10.17 – 10.21 |
| (ii) | Collect and Understand Data | 10.21 – 11.3 |
| 1 | Detail Data Requirement definition | 10.21 – 10.22 |
| 2 | Collect Data | 10.21 – 11.3 |
| 3 | Data Cleaning and preparation | 10.21 – 11.3 |
| (iii) | Scorecard Development | 11.3 – 11.22 |
| 1 | Determine Sample & performance windows | 10.17 – 10.21 |
| 2 | Good/Bad Definition | 10.17 – 10.21 |
| 3 | Fine & Coarse Classing | 11.3-11.11 |
| 4 | Logistic Regression, First Build of Model | 11.14-11.18 |
| 5 | Model Refining | 11.21 – 11.22 |
| (iv) | Pre/Implementation Analysis |  |
| 1 | Validate on recent samples |  |
| 2 | Characteristic Analysis |  |
| (v) | Documentation | 10.17 – 11.22 |

项目参数和定义

目标客户定义

在2016/1/1 – 2016/5/31申请且通过的POS贷合同，曾经进入过M1/M2/M3（CPD=4，CPD=31，CPD=61）。

好坏客户定义

* 坏客户定义：

在2016/1/1 – 2016/5/31申请且通过的POS贷合同，在介入催收后30天内逾期金额不能够100%回收（即回到正常的状态）。

* 好客户定义：

在2016/1/1 – 2016/5/31申请且通过的POS贷合同，在介入催收后30天内逾期金额能够100%回收（即回到正常的状态）

采样窗口和表现窗口

* 表现窗口

观察日期：2016/1/1-2016/8/31

观察窗口：合同在进入M1/M2/M3时间点（CPD=4，CPD=31，CPD=61）前。

表现窗口：介入催收后的未来30天内。

* 采样窗口

目标客户采样窗口：2016/1/1 – 2016/5/31申请且通过的POS贷客户。

被排除样本

以下合同/客户被排除在分析样本之外：

1. 排除元旦和春节期间申请的合同。（1/1-1/3 2/6-2/14）。
2. 提前委外的合同。
3. 排除非POS贷、教育贷、学生贷及商户贷。（loan\_type = ‘030’ and sub\_product\_type = ‘0’）。
4. 排除部分cpd异常的数据。
5. 排除已冲销合同（434个合同，LoanStatus =‘6’）

详细数据，请参见附件。



数据准备

数据来源

* 基本数据

所有的客户申请表数据来自于安硕提单系统。

* 行为数据

客户拖欠、还款情况和催收情况的统一基于底层合同还款计划表、合同实还表和客户催收记录表等。

* 其他数据

销售、门店和销售经理数据来自于数据仓库的风控总表和佰仟员工记录表。

数据清理

* 必填项数据的处理

下列字段是必填项。每个字段的选项如下表所示。建模前需进行数据清理，任何字段的值只能是表中所列的选项。否则，此数据将被剔除。详情见数据字典表。

表2 数据清洗明细表（部分）

|  |  |  |  |
| --- | --- | --- | --- |
| Scorecard 字段 | 字段名称 | 选项内容 | 备注 |
| INNER\_CODE | 内部代码 | 1：正常客户；  2：资质稍差客户；  3：疑似欺诈客户； |  |
| INSURE\_STATUS | 是否购买保险 | 0投保待发送；  1投保已发送；  2投保成功；  3投保失败；  4为申请退保；  5退保待发送；  6退保已发送；  7退保成功；  8退保失败； |  |
| IS\_WORK\_HR | 是否工作日的工作时间申请 | 1：是；  0：否 | 根据申请时间及时 |
| OPERATE\_MODE | 运作方式 | ‘01’-‘05’ |  |
| POS\_TYPE | POS门店类型 | ‘01’-‘04’ |  |
| IS\_SUIXINHUAN | 是否购买随心还 | 1：是；  0：否；  2：取消；  3：业务未开始； |  |
| IS\_WECHAT\_SUB | 是否关注佰仟分期购公众号 | 1：未关注；  2：关注；  3：取消关注 |  |
| PERSON\_SEX | 性别 | 1：男；  0：女 |  |
| PERSON\_APP\_AGE | 年龄 | [18, 55] |  |
| FAMILY\_STATE | 婚姻状态 | 1：未婚；  2：已婚；  3：离异； |  |
| EDUCATION | 教育程度 | [‘1’-‘7’]; ‘0’missing |  |
| OTHER\_PERSON\_TYPE | 其他联系人类型 | [‘6’-‘8’]; ‘0’missing，异常 |  |
| PUTOUT\_SAGROUP | SA分组\_申请时间 | [‘0’-‘5’]; |  |

训练和验证数据

根据以上的各种条件，在2016/1/1 – 2016/5/31这一采样窗口内，共有约60万个样本在2016/1/1 – 2016/8/31内进入M1，训练集、测试集和样本外验证集情况如下。

Sample Window

(60W)

2016/5/1-2016/5/31 APP\_DATE (8W)

2016/4/1-2016/4/30 APP\_DATE (11W)

2016/1/1-2016/3/31 APP\_DATE (40W)

TEST (3W)

BADRATE: 21.09%

TRAIN (3W)

BADRATE: 21.09%

TEST (3W)

BADRATE: 21.09%

VALIDATION (3W)

BADRATE: 21.37%

VALIDATION (3W)

BADRATE: 20.96%

图1 训练和验证数据

注：训练集、测试集1和测试集2来自同一样本，各随机抽取3万条。验证集1和验证集2来自另一采样窗口，各随机抽取3万条。

建模

同质性删除（Homogeneity Exclusions）

删除空缺值比例>=90%和单一值比例>=90%的变量。

细分（Variables Pre Binning）

筛选出IV值>=0.01的有效变量，如下：

表3 变量显著性分析表

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN\_NAME** | **COLUMN\_CN** | **gini** | **iv** |
| AVG\_DAYS | 平均每次逾期停留天数 | 0.2853 | 0.3198 |
| MAX\_CPD | 历史最大逾期cs\_cpd | 0.2760 | 0.3025 |
| MAX\_SEQ\_DUEDAYS | 延滞天数（每期最大（本金、费用可能延滞天数不同）的最大（所有期次求最大）） | 0.2702 | 0.2814 |
| PAYFINE | 累计应还滞纳金 | 0.2306 | 0.2290 |
| ACTUALPAYFINE | 累计实还滞纳金 | 0.2278 | 0.2228 |
| DELAY\_DAYS | 处于逾期状态的天数（cpd5天起算） | 0.2241 | 0.1927 |
| PAYFINECNT | 累计应还滞纳金次数 | 0.2246 | 0.1908 |
| BPTP\_RATIO | BPTP比率 | 0.2269 | 0.1903 |
| MAX\_CONDUE10 | 历史最大连续逾期10天的期数 | 0.2200 | 0.1870 |
| SEQ\_DUEDAYS | 延滞天数（每期最大（本金、费用可能延滞天数不同）的总和（所有期次相加）） | 0.2203 | 0.1867 |
| ACTUALPAYFINECNT | 累计实还滞纳金次数 | 0.2219 | 0.1863 |
| KPTP\_RATIO | KPTP比率 | 0.2277 | 0.1847 |
| CS\_TIMES | 历史总催收次数 | 0.2248 | 0.1767 |
| DUE\_DELAY\_RATIO | 当前欠款金额/延滞天数 | 0.2059 | 0.1628 |
| BPTP | BPTP次数 | 0.1978 | 0.1532 |
| RECENT\_CONTACT\_DAY | 最近一次可联距今时间 | 0.2143 | 0.1513 |
| AVG\_ROLLSEQ | 历史回退平均期数 | 0.1451 | 0.1445 |
| MAX\_ROLL\_SEQ | 最大回退期数 | 0.1422 | 0.1428 |
| DUE\_CSTIME\_RATIO | 当前欠款金额/总催收天数 | 0.2028 | 0.1411 |
| LOST | 历史完全失联次数 | 0.1357 | 0.1309 |
| MAX\_OVERDUE | 历史最大逾期金额 | 0.1710 | 0.1079 |
| MOST\_CONTACT\_3M | 最近三个月出现最多的可联催收代码 | 0.1555 | 0.1077 |
| CON10\_DUE\_TIMES | 历史连续逾期10天的次数 | 0.1197 | 0.1046 |
| CONTACT | 历史可联次数 | 0.1751 | 0.1023 |
| F\_SCORE | 申请最终评分 | 0.1647 | 0.0895 |
| OTHER\_PERSON\_TYPE | 其他联系人类型 | 0.1102 | 0.0828 |
| DUE\_CONTACT\_RATIO | 当前欠款金额/可联次数 | 0.1570 | 0.0805 |
| CITY\_CLUSTER | 城市 | 0.1349 | 0.0686 |
| STATE\_SAGROUP | SA分组\_评分时间 | 0.1460 | 0.0679 |
| MAX\_CONDUE5 | 历史最大连续逾期5天的期数 | 0.1384 | 0.0667 |
| PUTOUT\_SAGROUP | SA分组\_申请时间 | 0.1342 | 0.0650 |
| CSFQ | 催收频次 | 0.1380 | 0.0615 |
| ROLL\_BACK\_SEQ | 累计回退期数 | 0.1292 | 0.0564 |
| KPTP | KPTP次数 | 0.1211 | 0.0509 |
| CON5\_DUE\_TIMES | 历史连续逾期5天次数 | 0.1001 | 0.0494 |
| PAYFEE | 累计应还费用 | 0.1206 | 0.0465 |
| HIS\_PTPDAYS | PTP复核天数总和 | 0.1185 | 0.0459 |
| DELAY\_TIMES | 进入逾期状态的次数（cpd5天起算） | 0.1074 | 0.0438 |
| ROLL\_BACK | 回退次数 | 0.1081 | 0.0437 |
| PTP | PTP次数 | 0.1127 | 0.0414 |
| ACTUALPAYFEE | 累计实还费用 | 0.1109 | 0.0397 |
| CERT4\_CLUSTER | 客户身份证前4位 | 0.0781 | 0.0364 |
| DUE\_PTP\_RATIO | 当前欠款金额/PTP次数 | 0.1008 | 0.0340 |
| PERSON\_SEX | 性别 | 0.0772 | 0.0332 |
| INTIME\_PAY | 及时还款期次 | 0.0900 | 0.0319 |
| MANAGEMENTFEESRATE | 管理费率 | 0.0920 | 0.0286 |
| CREDIT\_AMOUNT | 贷款金额 | 0.0952 | 0.0285 |
| EFFECTIVEANNUALRATE | EIR 有效年利率 | 0.0919 | 0.0284 |
| CUSTOMERSERVICERATES | 客户服务费率 | 0.0903 | 0.0278 |
| PRICE | 商品价格 | 0.0897 | 0.0260 |
| OVER\_DUE\_VALUE | 当前逾期金额 | 0.0897 | 0.0247 |
| QQ\_LENGTH | QQ号码长度 | 0.0836 | 0.0247 |
| SALESCOMMISSION | 提成比例 | 0.0826 | 0.0233 |
| QQNO\_INIT | QQ号码第一位 | 0.0808 | 0.0230 |
| VALUE\_BALANCE\_RATIO | 应还金额比贷款余额 | 0.0825 | 0.0229 |
| TOT\_CREDIT\_AMOUNT | 总贷款金额（客户） | 0.0829 | 0.0224 |
| APR\_CREDIT\_AMOUNT | 通过合同贷款总额 | 0.0827 | 0.0223 |
| PROVINCE\_CLUSTER | 省份 | 0.0764 | 0.0210 |
| INIT\_PAY | 首付金额 | 0.0770 | 0.0199 |
| FINISH\_PERIODS\_RATIO | 实还期数比 | 0.0716 | 0.0199 |
| DUE\_PERIODS\_RATIO | 应还期数比 | 0.0724 | 0.0196 |
| ONTIME\_PAY | 按时还款期次 | 0.0706 | 0.0170 |
| RELATIVETYPE | 亲属关系人类型 | 0.0575 | 0.0168 |
| INSURE\_STATUS | 是否购买保险 | 0.0528 | 0.0166 |
| VALUE\_INCOME\_RATIO | 应还金额比家庭月收入 | 0.0696 | 0.0157 |
| FAMILY\_STATE | 婚姻状态 | 0.0601 | 0.0155 |
| MIN\_INIT\_RATE | 产品最低首付比例 | 0.0553 | 0.0152 |
| INDUSTRY | 工作单位所属行业 | 0.0664 | 0.0143 |
| SELF\_INCOME | 个人收入 | 0.0645 | 0.0134 |
| HOUSE\_TYPE | 住房类型 | 0.0607 | 0.0133 |
| ACTUALPAYPRINCIPALNFEE | 累计实还本金+费用 | 0.0647 | 0.0131 |
| PAYINTEAMT | 累计应还利息 | 0.0628 | 0.0122 |
| PAYPRINCIPALNFEE | 累计应还本金+费用 | 0.0614 | 0.0121 |
| EMAIL | 电子邮箱类型 | 0.0483 | 0.0109 |
| EDUCATION | 教育程度 | 0.0517 | 0.0108 |
| PERIODS | 分期还款期数 | 0.0520 | 0.0104 |

详细数据，请参见附件。



变量相关性分析（Correlation Analysis）

计算变量的相关性矩阵，剔除相关性超过0.7的变量，相关性分析详细结果见附件。



变量粗分（Variables Binning）

BPTP\_RATIO

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N1 | N | BAD\_RATE | WOE | DEFAULT |
| 1 | (0.6,+∞] | 1349 | 4053 | 0.33284 | 0.613756 | 3 |
| 2 | (0.4,0.6] | 805 | 3270 | 0.246177 | 0.19002 | 3 |
| 3 | missing | 2644 | 10990 | 0.240582 | 0.159635 | 3 |
| 4 | (0.2,0.4] | 397 | 2178 | 0.182277 | -0.19187 | 3 |
| 5 | (-∞,0.2] | 1184 | 9509 | 0.124514 | -0.64124 | 3 |

DELAY\_TIMES

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N1 | N | BAD\_RATE | WOE | DEFAULT |
| 1 | (2,+∞] | 990 | 4000 | 0.2475 | 0.197134 | 4 |
| 2 | (1,2] | 1187 | 4841 | 0.245197 | 0.184731 | 4 |
| 3 | (0,1] | 2111 | 8956 | 0.235708 | 0.132768 | 4 |
| 4 | (-∞,0] | 2091 | 12203 | 0.171351 | -0.26696 | 4 |

FAMILY\_STATE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | 3 | 432 | 1586 | 0.272383 | 0.326561 | 4 |
| 2 | 1 | 4000 | 18066 | 0.22141 | 0.051659 | 4 |
| 3 | 2 | 1947 | 10348 | 0.188152 | -0.15294 | 4 |

FINISH\_PERIODS\_RATIO

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N1 | N | BAD\_RATE | WOE | DEFAULT |
| 1 | (-∞,0.2] | 4308 | 19152 | 0.224937 | 0.072003 | 1 |
| 2 | (0.2,0.35] | 1601 | 7849 | 0.203975 | -0.05251 | 1 |
| 3 | (0.35,+∞] | 470 | 2999 | 0.156719 | -0.37372 | 1 |

INTIME\_PAY

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | (4,+∞] | 876 | 3439 | 0.254725 | 0.235557 | 3 |
| 2 | (0,4] | 4777 | 21874 | 0.218387 | 0.034035 | 3 |
| 3 | (-∞,0] | 726 | 4687 | 0.154897 | -0.38758 | 3 |

MAX\_CPD

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | (26,+∞] | 1344 | 2810 | 0.478292 | 1.222237 | 5 |
| 2 | (18,26] | 865 | 2832 | 0.305438 | 0.487589 | 5 |
| 3 | (12,18] | 690 | 2757 | 0.250272 | 0.211963 | 5 |
| 4 | (7,12] | 778 | 4447 | 0.174949 | -0.24182 | 5 |
| 5 | (0,7] | 2702 | 17154 | 0.157514 | -0.36772 | 5 |

MAX\_OVERDUE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | (550,+∞] | 1173 | 3326 | 0.352676 | 0.701827 | 5 |
| 2 | (400,550] | 1358 | 5933 | 0.228889 | 0.094531 | 5 |
| 3 | (350,400] | 1090 | 4846 | 0.224928 | 0.071948 | 5 |
| 4 | (200,350] | 2348 | 13146 | 0.178609 | -0.21667 | 5 |
| 5 | (-∞,200] | 410 | 2749 | 0.149145 | -0.4322 | 5 |

PERSON\_SEX

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | 1 | 5113 | 22223 | 0.230077 | 0.101248 | 1 |
| 2 | 0 | 1266 | 7777 | 0.162788 | -0.32851 | 1 |

MAX\_OVERDUE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | (-∞,0.25] | 1481 | 4735 | 0.312777 | 0.521957 | 3 |
| 2 | (0.25,0.35] | 491 | 1771 | 0.277244 | 0.350954 | 3 |
| 3 | missing | 1834 | 8307 | 0.220778 | 0.047984 | 3 |
| 4 | (0.35,0.65] | 1264 | 5955 | 0.212259 | -0.00224 | 3 |
| 5 | (0.65,0.8] | 462 | 2819 | 0.163888 | -0.32046 | 3 |
| 6 | (0.8,+∞] | 847 | 6413 | 0.132075 | -0.57361 | 3 |

OTHER\_PERSON\_TYPE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | 同学 | 34 | 114 | 0.298246 | 0.453459 | 3 |
| 2 | 同事或朋友 | 5542 | 23533 | 0.235499 | 0.131609 | 3 |
| 3 | 其他 | 803 | 6353 | 0.126397 | -0.62407 | 3 |

EDUCATION

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | 小学；硕士及以上 | 69 | 259 | 0.266409 | 0.296207 | 1 |
| 2 | 初中；高中；中专 | 5662 | 26100 | 0.216935 | 0.025506 | 1 |
| 3 | 专科；本科 | 648 | 3641 | 0.177973 | -0.22102 | 1 |

PUTOUT\_SAGROUP

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | 1 | 753 | 2273 | 0.33128 | 0.606724 | 3 |
| 2 | 2 | 565 | 2147 | 0.263158 | 0.279505 | 3 |
| 3 | 0；3 | 1621 | 7107 | 0.228085 | 0.089969 | 3 |
| 4 | 4 | 1533 | 7350 | 0.208571 | -0.02443 | 3 |
| 5 | 5 | 1907 | 11123 | 0.171447 | -0.26628 | 3 |

PROVINCE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLUS | TESTNAME | N | N1 | BAD\_RATE | WOE | DEFAULT |
| 1 | 省份聚类1 | 265 | 987 | 0.26849 | 0.306829 | 4 |
| 2 | 省份聚类2 | 2643 | 11345 | 0.232966 | 0.117487 | 4 |
| 3 | 省份聚类3 | 2274 | 10842 | 0.20974 | -0.01737 | 4 |
| 4 | 省份聚类4 | 1197 | 6826 | 0.175359 | -0.23899 | 4 |

详细数据请参考附件。



算法

采用SAS的Logistic Regression算法。

建模结果如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 变量标签 | 估计 | 标准误差 | Wald 卡方 | Pr > 卡方 | 估计类型 |
| Intercept | 1 | -1.3055 | 0.0154 | 7189.4621 | <.0001 |
| INTIME\_PAY | 1 | 2.6073 | 0.1049 | 617.4527 | <.0001 |
| FINISH\_PERIODS\_RATIO | 1 | 2.4893 | 0.1367 | 331.4408 | <.0001 |
| MAX\_CPD | 1 | 0.7189 | 0.041 | 307.6537 | <.0001 |
| BPTP\_RATIO | 1 | 0.6246 | 0.0397 | 247.2056 | <.0001 |
| OTHER\_PERSON\_TYPE | 1 | 0.892 | 0.0568 | 246.4707 | <.0001 |
| PUTOUT\_SAGROUP | 1 | 0.8463 | 0.0587 | 208.1164 | <.0001 |
| PERSON\_SEX | 1 | 0.8278 | 0.0855 | 93.691 | <.0001 |
| PROVINCE\_CLUSTER | 1 | 0.9683 | 0.1061 | 83.3627 | <.0001 |
| MAX\_OVERDUE | 1 | 0.4581 | 0.052 | 77.6357 | <.0001 |
| PTP\_RATIO | 1 | 0.4094 | 0.0468 | 76.6725 | <.0001 |
| DELAY\_TIMES | 1 | 0.6243 | 0.0962 | 42.1404 | <.0001 |
| FAMILY\_STATE | 1 | 0.7055 | 0.122 | 33.4638 | <.0001 |
| EDUCATION | 1 | 0.906 | 0.1874 | 23.3842 | <.0001 |

评分卡（Score Card）

最终入选模型一共13个变量，其中包括个人信息模块变量4个，合同及产品信息类模块变量1个，还款信息模块变量5个，催收信息模块变量2个，其他模块信息变量1个，变量分组得分详细情况见下表。

|  |  |  |
| --- | --- | --- |
| 变量名 | 区间 | 分数 |
| INTERCEPT |  | -1.3055 |
| BPTP\_RATIO | 1 | 0.383352135 |
| 2 | 0.118686487 |
| 3 | 0.099708234 |
| 4 | -0.119841503 |
| 5 | -0.400518315 |
| DELAY\_TIMES | 1 | 0.123070998 |
| 2 | 0.11532785 |
| 3 | 0.082887132 |
| 4 | -0.166660391 |
| FAMILY\_STATE | 1 | 0.230388745 |
| 2 | 0.03644516 |
| 3 | -0.107896502 |
| FINISH\_PERIODS\_RATIO | 1 | 0.179236614 |
| 2 | -0.130708605 |
| 3 | -0.930305553 |
| INTIME\_PAY | 1 | 0.614168199 |
| 2 | 0.088738221 |
| 3 | -1.010529463 |
| MAX\_CPD | 1 | 0.878666496 |
| 2 | 0.350528084 |
| 3 | 0.152380085 |
| 4 | -0.173846626 |
| 5 | -0.264350869 |
| MAX\_OVERDUE | 1 | 0.321507019 |
| 2 | 0.043304712 |
| 3 | 0.03295935 |
| 4 | -0.099257548 |
| 5 | -0.197989356 |
| PERSON\_SEX | 1 | 0.083813119 |
| 2 | -0.271937208 |
| PTP\_RATIO | 1 | 0.21368933 |
| 2 | 0.143680393 |
| 3 | 0.019644853 |
| 4 | -0.000916926 |
| 5 | -0.131194363 |
| 6 | -0.234834479 |
| OTHER\_PERSON\_TYPE | 1 | 0.404485153 |
| 2 | 0.117394885 |
| 3 | -0.556673733 |
| EDUCATION | 1 | 0.268363754 |
| 2 | 0.023108504 |
| 3 | -0.200240505 |
| PUTOUT\_SAGROUP | 1 | 0.513470873 |
| 2 | 0.236545407 |
| 3 | 0.07614047 |
| 4 | -0.020677871 |
| 5 | -0.225356975 |
| PROVINCE | 1 | 0.297102994 |
| 2 | 0.113762229 |
| 3 | -0.016818535 |
| 4 | -0.23141264 |

表4 佰仟金融催收M1评分卡

评分卡的有效性

K－S检验（Kolmogorov-Smirnov）

图2 KS Chart

ROC图

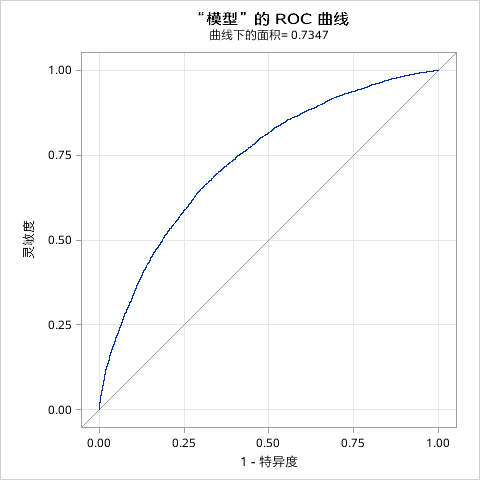


图3 ROC GRAPH

不同分数段的坏客户比例

图4不同分数段的坏客户比例

区分度（Divergence Chart）

图5 Divergence Chart

评分卡分数转换

转换方法

为了得到一个友好的分数，每一变量对应的每个属性的分数可以用一个引子和偏移量的线性方程来表示。



Note:





* A score of P0 (600) corresponds to good/bad odds (take the test dataset as example: )
* An increase in the score of PDO (50) points corresponds to a doubling of the good/bad odds

So in this example:





ODDS分数对应表

表5 Odds分数对应表

|  |  |  |
| --- | --- | --- |
| **Odds** | **Badrate** | **Score** |
| 8.309926524 | 89.26% | 300 |
| 4.154963262 | 80.60% | 350 |
| 2.077481631 | 67.51% | 400 |
| 1.038740815 | 50.95% | 450 |
| 0.519370408 | 34.18% | 500 |
| 0.259685204 | 20.62% | 550 |
| 0.129842602 | 11.49% | 600 |
| 0.064921301 | 6.10% | 650 |
| 0.03246065 | 3.14% | 700 |
| 0.016230325 | 1.60% | 750 |
| 0.008115163 | 0.80% | 800 |

评分卡实施前的分析

用最新的样本验证

采取2016年10月21日进入M1的客户样本（CPD=4），其中包括新增及存量客户（新增定义为2016年1月后申请的客户，存量为2014年和2015年申请的客户）。

使用介入催收后30天内是否回收作为Y值验证评分模型有效性。

图6不同分数段的坏客户比例（新增）

图7 Divergence Chart（新增）

图8不同分数段的坏客户比例（存量+新增）

图9 Divergence Chart（存量+新增）