**厦门国际银行**

**数信互融场景分评估报告**

**2020年09月**

目录

[一、 评估目标 4](#_Toc48053868)

[二、 数据准备 4](#_Toc48053869)

[三、 验证方案 4](#_Toc48053870)

[四、 验证结果 9](#_Toc48053871)

[五、 综合结论 16](#_Toc48053872)

**文档版本**

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 版本号 | 修改细节 | 作者 |
| 202009 | V1.0 |  | 人工智能与模型开发处 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## 评估目标

本次评估验证主要为了验证数信互融的场景分产品，其主要数据来源为友盟数据，并根据其使用数据的场景分为借呗1场景分、借呗2场景分、京东白条场景分、京东金条场景分等4个场景分。本次评估目标是验证4种场景分是否能作为我行通用分的参考，因此选取我行产品中查过征信的好E贷-美团点评和好E贷-新网两个产品。通过我行产品对各个场景分进行已验证，验证各个场景分对于我行逾期客户的识别能力并观察各个场景分是否能作为通用分的参照标准。

## 数据准备

1. **我行数据准备情况**

本次验证，我行准备了15000条数据记录，其中分为好E贷-美团点评（10000）、好E贷-新网（5000）。好E贷-美团点评和好E贷-新网使用的是建模数据，逾期客户标签以M2+（逾期天数大于30天）进行定义，好坏比均为4:1，即逾期客户占比为20%。

1. **场景分数据**

数信互融通过友盟300+行业数据体系，根据友盟底层的数据特征（包括用户行为信息，登陆信息，设备指纹等），定制了适合几大主流互联网资产的场景分。根据已有项目经验，对外部数据应用流程及应用策略进行设计，辅助风险决策。主要分为借呗1、借呗2（高信用）、京东白条、京东金条、信用卡等五个场景。本次数据验证主要验证非信用卡的场景分，以下4表为数信互融根据全量建模客户设定的分数应用区间。

**表1. 借呗1场景分应用分数区间**



**表2. 借呗2场景分应用分数区间**



**表3.京东白条场景分应用分数区间**



**表4. 京东金条场景分应用分数区间**



表1、2、3、4的结果是数信互融在构建场景分的全量数据的情况，包含了分数段、处于该分数段客户的违约率和通过率，其中通过率为拒绝该分数以下后仍通过人数占比，即客户累计占比。同时也给出的了后续使用的相关建议。

## 验证方案

本次验证主要的目标是观测数信互融-友盟的场景分是否能作为我行通用分的标准。根据数据情况，本报告主要从以下几个维度选择合适的指标对厂商的数据进行验证：

1. **KS值(Kolmogorov-Smirnov)值**

KS检验全名为Kolmogorov-Smirnov检验，是一应用非常广泛的非参数分布检验。KS原本设计为检验一个样本是否来自于假定的分布，或两个样本是否来自于同一分布。KS对所检验样本的位置和形状非常敏感，对二值变量作为目标变量的数据挖掘模型来说，是一个非常直观有效的检验指标。

一般意义上的KS统计量是找到多个分类样本分布与合并样本分布的最大偏差。运用KS检验来验证模型能否区别出违约户与正常户，当两组样本的累积相对次数分配非常接近，且差异为随机时，则两组样本的评级分配应为一致；反之当两组样本的评级分配并不一致时，样本累积相对次数分配的差异会很显。

两样本的KS统计量由以下公式计算：

其中，是样本1（好/坏）累计比率，是样本2（坏/好）累计比率。

KS值的计算步骤如下：

1. 将分析对象（可以是评级、违约概率、变量值）的所有纪录按照分析对象的值从低到高[[1]](#footnote-1)进行排序
2. 计算累计好账户数占好账户总数的比率和累计坏账户数占坏账户总数的比率
3. 对每条记录，计算累计坏账户占比与累计好账户占比的差，并取绝对值
4. 取所有绝对值的最大值

根据行业经验，KS值所对应的区分能力可参照下表进行评估：

**表5. KS值评价标准**

|  |  |
| --- | --- |
| **KS值(%)** | **模型区分能力** |
| 20以下 | 不建议采用 |
| 20-40 | 中等 |
| 41-50 | 好 |
| 51-60 | 强 |
| 61-75 | 非常强 |
| 75以上 | 能力高但疑似有误 |

1. **模型的排序分布能力**

一般适用于评级结果数据的验证，主要验证其评级结果排序在我行标签下的排序能力。

ROC曲线及AUC系数主要用来检验模型对客户进行正确排序的能力。ROC曲线描述了在一定累计好客户比例下的累计坏客户的比例，模型的分别能力越强，ROC曲线越往左上角靠近。AUC系数表示ROC曲线下方的面积。AUC系数越高，模型的风险区分能力越强。

在下图中，AUC系数表示ROC曲线下方的面积。



1. **ROC曲线示例**

或者也能使用我行标签和他人的评级结果（概率最佳，分数也可）进行测算，通过AUC计算公式进行计算。公式如下：

其中 M 代表好样本数，N代表坏样本数，需要先按得分从最大到最小进行排序，rank分从（M\*N）往下排，最后计算好样本的rank分之和。

根据行业经验，AUC所对应的稳定性可参照下表进行评估：

**表6. AUC值评价标准**

|  |  |
| --- | --- |
| **AUC值（%）** | **模型区分能力** |
| 50 | 不具备区分能力 |
| 50-60 | 较差 |
| 60-70 | 一般 |
| 70-80 | 良好 |
| 80-90 | 强 |
| 90以上 | 优异 |

1. **特征/模型稳定性**

特征稳定性，反映了该客群的特征随时间的变化程度。客群稳定，模型的预测能力才能得到保障。模型稳定性是为了验证在客群变化小的情况下，模型判别后的客群（标签）分布是否发生过大变化。在满足客群稳定的情况下，如若模型稳定性发生重大变化，则模型出现问题需要更新迭代。

特征/模型稳定性通常使用PSI（Population Stability Index，稳定性指标）进行评估。PSI的计算公式如下：

其中 代表第i个箱中第一个时点的占比, 代表第i个箱中第二个时点的占比。PSI可根据模型结果（评分结果分箱）或者特征（特征分箱）进行评估，所以分为模型稳定性和特征稳定性。

根据行业经验，PSI值所对应的稳定性可参照下表进行评估：

**表7. PSI值参照标准**

|  |  |
| --- | --- |
| **PSI** | **评价** |
| <0.1 | 无关紧要的差距，稳定性较好 |
| 0.1-0.25 | 有一定差距，稳定性可接受 |
| >0.25 | 差距较大，模型或需要进行重建 |

综上，结合本次数据情况，本报告将从**模型排序/区分能力、数据的稳定性进行对各个场景分进行验证。**

## 验证结果

1. **总体情况**

本次验证客户两个时点的情况，包括进件时间以及2020年3月31日。数信互融也返回查询结果，其中未查得的结果中包含客户近180天行为表现不足的客户。在本次验证中，将该情况的客户定义为缺失。因此在查验后，仅存在一个时点的客户也定义为未查得。最终、客户的结果的查得率如表8所示:

**表8. 客户查得情况表**

|  |  |  |
| --- | --- | --- |
| **我行产品** | **产品** | **查得率** |
| 好E贷-美团点评 | 场景分 | 90.85% |
| 好E贷-新网 | 场景分 | 90.6% |

最终，好E贷-美团点评使用9085条客户记录，逾期客户1798位，逾期客户占比19.79%；好E贷-新网使用4530条客户记录，逾期客户896位，逾期客户占比19.80%。两个产品客户合计13615条客户记录，逾期客户2694位，逾期客户占比为19.79%。

1. **模型排序/区分能力验证**
2. 借呗1场景分

**表9 借呗1场景分验证（好E贷-美团点评）**

**表10. 借呗1场景分验证（好E贷-新网）**

**表11. 借呗1场景分验证（两种我行产品）**

如表9,10所示，借呗1场景分在好E贷-美团点评上能更好的对逾期客户进行区分，AUC为64%，KS在20%以上。而在好E贷-新网上表现不佳。同时参考表1，观测我行数据和数信互融全量用户数据进行比对，发现我行客户均有一定程度上的下沉，好E贷-美团点评和好E贷-新网在最低一档的人数占比均有增加。

如表11所示，单纯使用场景分做为我行的通用分的参照标准，虽然能有比较好的坏好比下降趋势，但各个区间之间区分不明显的，也存在有偏差。整体上无法有效的区分逾期客户。不建议做为当做通用分的使用标准。

1. 借呗2场景分

**表12.借呗2场景分验证(好E贷-美团点评)**

**表13. 借呗2场景分验证(好E贷-新网)**

**表14. 借呗2场景分验证(两种我行产品)**



如表12,13所示，借呗2场景分在好E贷-美团点评上对于逾期客的识别能力更优于好E贷-新网。借呗2场景分对于好E贷-新网的逾期客户识别能力较差，不能有效的识别出好坏客户，每个区间之间的坏好比相对混淆，呈现波动变化，不能有效地划分区间从而对客户进行区分。

如表14所示，借呗2场景分能有效的区分出逾期客户，在最低两个分箱中，坏好比比均相对较高的，且区分明显。在剩下的8个分箱中，坏好比虽然变化较小，但仍呈现递减趋势。对于我行两个产品的客户，借呗2场景分已经具备使用的能力。

1. 京东白条场景分

**表15. 京东白条场景分验证（好E贷-美团点评）**

**表16. 京东白条场景分验证（好E贷-新网）**

**表17. 京东白条场景分验证（两种我行产品）**

如表15,16所示，京东白条场景分在好E贷-美团点评上对于逾期客的识别能力依旧优于好E贷-新网，且京东白条场景分对于好E贷-美团点评能够达到使用标准。京东白条场景分对于好E贷-新网的逾期客户识别能力较差，不能有效的识别出好坏客户，每个区间之间的坏好比相对混淆，不能有效地划分区间对客户进行区分。

同时参考表3，观测我行数据和数信互融全量用户数据进行比对，发现我行客户有大幅度的下沉。数信互融使用的分箱方式为等频分箱，客户在各个区间基本上占比在10%左右。但在低分5挡的客户，我行各类产品人数占比达87%，90%，而该部分总体占比为50%左右。同时有一半左右的客户处于最低一档。

如表17所示，整体上能够使用京东白条作为通用评分标准，区分能力良好。可以将514分以下的进行拒绝，但相对客户占比过大，需要拒绝50%的客户（能找到73%的逾期客户）。如果需要使用，需要重新找合适的拒绝分数。

1. 京东金条场景分

**表18. 京东金条场景分验证（好E贷-美团点评）**

**表19. 京东金条场景分验证（好E贷-新网）**

**表20. 京东金条场景分验证（两种我行产品）**

如表18,19所示，京东金条场景分的表现效果和京东白条场景分相似，但相比京东白条更能准确地识别出逾期客户。针对好E贷-美团点评的客群，能在少拒绝10%的情况下仍找到70%的逾期客户。而针对好E贷-新网，能较为粗浅的划分出的客户，但区分/排序能力弱。

同时参考表4，观测我行数据和数信互融全量用户数据进行比对，同样发现我行客户有大幅度的下沉。如表20所示，整体上能够使用京东金条场景分作为通用评分标准，区分能力相对优秀。同比京东白条场景分，京东金条场景分能在拒绝客户中仍有较细粒度一层的分箱。

1. 综合使用场景分

因考虑到我行产品场景均不属于借呗和京东方面的渠道，所以针对两种产品和四个场景分的综合结果进行使用。使用逻辑回归寻求最优的AUC结果。最终，将结果等频10等分。结果如下表所示：

**表21. 综合情况（两种我行产品）**

从综合结果和单独的结果的表现上看，4个场景产综合结果已经能很好的识别出我行逾期客户。

1. **PSI**

目前使用的是两个时点的评分结果，进件时间点、2020年03月31日。参照公式可以进行PSI值计算，计算结果如下所示：

**表22. PSI结果表**



每一个场景分的PSI均小于0.1,对于的我行各个产品的表现都趋于稳定，并无较大的浮动。

## 综合结论

**表23.KS、AUC汇总表**



1. 4个场景分均对好E贷-美团点评的客户群体区分/排序能力好，而对新网识别能力弱，或是由于新网客群和建模客群不匹配所导致。而在四个场景分中京东金条的区分/排序能力最好。
2. 使用综合客群上看，借呗1场景分不能达到使用标准，仅有60%的AUC和15.70%的KS表现。而其他三个场景分均可以作为通用分的使用标准。
3. 场景分的综合结果进行结合去识别我行客户，其最终的AUC和KS表现达到AUC 70%以及KS 30%的水平。如需要再提升，建议可以尝试使用底层字段建模，提升水平应该会比简单使用场景分融合的效果高。
4. 各个场景分PSI数值在观测期内保持稳定。
5. 依托于数信互融提供的建模数据表现，我行客群会在各个场景分中属于下沉部分的人群。尤其以京东金条场景分的表现最为明显，在京东金条逾期率和我行提供数据逾期相互接近的情况下，我行83%的客户处于的京东金条中的50%以下的客户。

1. 如果分析对象的值的排序是从高到低，并不影响KS值的计算，KS公式中使用的是绝对值 [↑](#footnote-ref-1)