



数据背景



晋级规则

个贷违约预测

中国计算机学会 & 中原银行

智能算法 异常检测 金融

队伍 / 人数

2534 / 3059

奖金

¥ 50,000

开赛

评分方式

A 榜

B 榜

初赛 09.22 ~ 11.22

决赛 11.23 ~ 12.05

2021/09/23 14:59:15 赛题数据

训练集 - MD5: 192D3D30A9252E849C600C0931C8C1FB

2021/09/23 14:59:57 赛题数据

测试集 - MD5: D4A2AB57C4B6188CC606949501750420

2021/09/23 15:00:29 提交样例

提交样例 - MD5: 5F272DFEE05DF1C950009A6E5700E509

2021/09/24 17:40:31 示例程序

个贷违约预测baseline - MD5: 0D4D985CB15DB705E89A3DD8CFCDEEC5

数据背景

数据来源：个人贷款数据和网络信贷数据脱敏仿真。

文件清单和使用说明

训练数据

train_public.csv 个人贷款违约记录数据
train_internet_public.csv 某网络信贷产品违约记录数据



测试数据

test_public.csv 用于测试的数据，获取榜单排名



训练数据说明

train_public.csv



| 字段 | 字段描述 |
|-----------------|----------|
| loan_id | 贷款记录唯一标识 |
| user_id | 借款人唯一标识 |
| total_loan | 贷款数额 |
| year_of_loan | 贷款年份 |
| interest | 当前贷款利率 |
| monthly_payment | 分期付款金额 |

赛制规则

数据与评测

排行榜

参赛队伍

作品提交

参赛交流

常见问题

邀请有礼

数据背景

文件清单和使用说明

作品提交要求

提交格式

提交样例

评分与晋级

晋级规则

评分方式

| | |
|--------------------------|------------------------------|
| grade | 信用等级 |
| employment_type | 所在公司类型（世界五百强、国有企业、普通企业...） |
| industry | 工作领域（传统工业、商业、互联网、金融...） |
| work_year | 工作年限 |
| home_exist | 是否有房 |
| cancel_status | 审核情况 |
| issue_month | 贷款发放的月份 |
| use | 贷款用途类别 |
| post_code | 贷款人申请时邮政编码 |
| region | 地区编码 |
| debt_loan_ratio | 债务收入比 |
| del_in_18month | 借款人过去18个月逾期30天以上的违约事件数 |
| scoring_low | 借款人在贷款评分中所属的下限范围 |
| scoring_high | 借款人在贷款评分中所属的上限范围 |
| known_outstanding_loan | 借款人档案中未结信用额度的数量 |
| known_dero | 贬损公共记录的数量 |
| pub_dero_bankrup | 公开记录清除的数量 |
| recircle_bal | 信贷周转余额合计 |
| recircle_util | 循环额度利用率 |
| initial_list_status | 贷款的初始列表状态 |
| app_type | 是否个人申请 |
| earlies_credit_mon | 借款人最早报告的信用额度开立的月份 |
| title | 借款人提供的贷款名称 |
| policy_code | 公开可用的策略_代码=1新产品不公开可用的策略_代码=2 |
| f系列匿名特征 | 匿名特征f0-f4，为一些贷款人行为计数特征的处理 |
| early_return | 借款人提前还款次数 |
| early_return_amount | 贷款人提前还款累积金额 |
| early_return_amount_3mon | 近3个月内提前还款金额 |

train_internet.csv

| 字段 | 字段描述 |
|-----------------|--------------|
| loan_id | 网络贷款记录唯一标识 |
| user_id | 用户唯一标识 |
| total_loan | 网络贷款金额 |
| year_of_loan | 网络贷款期限（year） |
| interest | 网络贷款利率 |
| monthly_payment | 分期付款金额 |

赛制规则

数据与评测

排行榜

参赛队伍

作品提交

参赛交流

常见问题

邀请有礼

数据背景

文件清单和使用说明

作品提交要求

提交格式

提交样例

评分与晋级

晋级规则

评分方式

| | |
|--------------------------|---------------------------------|
| class | 网络贷款等级 |
| sub_class | 网络贷款等级之子级 |
| work_type | 工作类型（公务员、企业白领、创业...） |
| employment_type | 所在公司类型（世界五百强、国有企业、普通企业...） |
| industry | 工作领域（传统工业、商业、互联网、金融...） |
| work_year | 就业年限（年） |
| house_ownership | 是否有房 |
| house_loan_status | 房屋贷款状况（无房贷、正在还房贷、已经还完房贷） |
| censor_status | 验证状态 |
| marriage | 婚姻状态（未婚、已婚、离异、丧偶） |
| offsprings | 子女状态(无子女、学前、小学、中学、大学、工作) |
| issue_date | 网络贷款发放的月份 |
| use | 贷款用途 |
| post_code | 借款人邮政编码的前3位 |
| region | 地区编码 |
| debt_loan_ratio | 债务收入比 |
| del_in_18month | 借款人过去18个月信用档案中逾期60天内的违约事件数 |
| scoring_low | 借款人在信用评分系统所属的下限范围 |
| scoring_high | 借款人在信用评分系统所属的上限范围 |
| pub_dero_bankrup | 公开记录清除的数量 |
| early_return | 提前还款次数 |
| early_return_amount | 提前还款累积金额 |
| early_return_amount_3mon | 近3个月内提前还款金额 |
| recircle_bal | 信贷周转余额合计 |
| recircle_util | 循环额度利用率，或借款人使用的相对于所有可用循环信贷的信贷金额 |
| initial_list_status | 网络贷款的初始列表状态 |
| earlies_credit_line | 网络贷款信用额度开立的月份 |
| title | 借款人提供的网络贷款名称 |
| policy_code | 公开策略=1不公开策略=2 |
| f系列匿名特征 | 匿名特征f0-f5，为一些网络贷款人行为计数特征的处理 |

其他数据文件

选手提交 submission.csv

| 字段名 | 字段说明 |
|-----------|--------|
| id | 贷款记录ID |
| isDefault | 是否违约 |

作品提交要求

赛制规则

数据与评测

排行榜

参赛队伍

作品提交

参赛交流

常见问题

邀请有礼



数据背景

[文件清单和使用说明](#)

作品提交要求

提交格式

提交样例

评分与晋级

晋级规则

评分方式

选手提交 submission.csv

| 字段名 | 字段说明 | 格式 |
|-----------|----------|-----------|
| id | 文本ID | int |
| isDefault | 贷款违约预测结果 | int/float |

提交样例

id, isDefault

0,0

1,0

2,1

...

(说明: isDefault可以是小数)

评分与晋级

晋级规则

B榜评测结束后, 按照B榜测试结果进行排名。组委会将对排行榜TOP10参赛队伍最优提交成绩的模型、完整代码(包含数据处理练)、报告、论文进行审核。对于未提交、复现未成功或审核不通过的队伍, 将取消决赛资格和比赛奖励。最终, 审核通过的队伍, 排名TOP5的参赛队伍进入决赛答辩环节。

评分方式

评价指标 AUC

本次挑战的是经典的预测任务, 以AUC作为衡量标准。

评测指标

使用ROC曲线下面积AUC (Area Under Curve) 作为评价指标。AUC值越大, 预测越准确。

产品与服务

关于我们

商务合作

官方公众号

加入竞赛交流群



案例学习

关于DF

商务总监: 郭老师



参与竞赛

隐私协议

邮箱: contact@datafountain.cn



海量数据集

服务协议

添加微信: sophy619641

社区交流

常见问题

咨询电话: (010) 62381637 / 82249592



合作伙伴

中国计算机学会

北京市经信局

国家电投

国家电网

中国气象局

航天科工集团

鹏城实验室

兴业银行

建设银行

中原银

OPPO

华为

蚂蚁金服

腾讯

京东

平安产险

清华大学

中国人民大学

北京科技大学

闽江学