## 人工智能赛道----信用风险识别

### 比赛题目：信用风险识别

#### 赛题简介

消费金融有限公司A通过市场调研，针对市场上用户的需求推出一款消费金融产品，如果一个用户申请A公司的产品，A公司需要对这个用户进行资质审查，审查的目的是预测用户是否会出现违约逾期的风险。对公司而言，如果一个客户没有发生逾期，公司将会盈利，相反，公司将会亏损。目前，有一批用户申请该消费金融产品，提供这批用户在近一年内的交易数据，参赛队伍需要通过大数据和算法构建用户的风险识别模型，预测这批用户是否会发生预期风险。

#### 数据说明

本场比赛提供训练集一份和测试集一份，训练集包括：用户的首次申请日期、交易数据和标签，测试集包括：用户首次申请时间和交易数据。

**额外提供以下数据：**

**行政区划代码表**

**mcc分类表**

**交易渠道表**

**交易应答码表**

**交易类型表**

**卡性质及卡介质表**

**详细的信息参考码表。**

首次申请时间是指用户首次申请消费金融产品的时间、用户的身份标识信息：

|  |  |
| --- | --- |
| **字段名** | **解释** |
| certid | 脱敏后的身份证号 |
| firstloandate | 首次申请日期 |
| Age | 年龄 |
| Gender | 性别 |
| Divisioncode | 行政区划代码（身份证前6位） |

交易数据是指用户的交易行为数据，主要包括以下字段，如下所示：

|  |  |
| --- | --- |
| **字段名** | **解释** |
| certid | 脱敏后的身份证号 |
| card\_no | 脱敏后的银行卡号 |
| Settle\_dt | 清算日期 |
| Trans\_tm | 交易时间 |
| **Trans\_at** | **交易金额（以分为单位）** |
| Sys\_tra\_no | 系统追踪号 |
| iss\_ins\_cd | 脱敏后的发卡机构 |
| trans\_chnl | 交易渠道 |
| mchnt\_cd | 脱敏后的商户id |
| mcc\_cd | 商户类型（mcc分类 |
| card\_accprt\_nm\_loc | 商户中文名称 |
| resp\_cd | 交易应答码 |
| trans\_id\_cd | 交易类型 |
| orig\_trans\_st | 原交易状态 |
| trans\_st | 交易状态 |
| trans\_curr\_cd | 交易币种 |
| **fwd\_settle\_at** | **发送方清算金额** |
| fwd\_settle\_cruu\_cd | 发送方清算币种 |
| fwd\_settle\_conv\_rt | 发送方清算汇率 |
| **rcv\_settle\_at** | **接收方清算金额** |
| rcv\_settle\_curr\_cd | 接收方清算币种 |
| rcv\_settle\_conv\_rt | 接收方清算汇率 |
| **cdhd\_at** | **持卡人账户扣帐金额** |
| cdhd\_curr\_cd | 持卡人张惠扣帐币种 |
| cdhd\_conv\_rt | 持卡人账户扣帐汇率 |
| term\_cd | 终端号 |
| card\_attr\_cd | 卡性质 |
| card\_media\_cd | 卡介质 |
| retri\_ref\_no | 检索参考号 |
| pos\_cond\_cd | 服务点条件代码 |
| pos\_entry\_md\_cd | 服务点输入方式 |
| auth\_id\_resp\_cd | 授权标识应答 |

标签数据标识了用户是否违约逾期，包含两个字段，如下所示：

|  |  |
| --- | --- |
| **列名** | **解释** |
| certid | 脱敏后的身份证号 |
| label | 用户标签，1表示没有发生逾期，0表示发生预期 |

#### 评分数据格式

参赛选手完成对用户违约预期风险的预测之后，需要将结果整理成如下格式的文件：

身份证号 标签

110110199201010001 1

第一列是身份证号，第二列式预测结果

#### 评估指标

笔袋采用经典的F1值作为评估指标，具体的计算公式如下：

P=TP/(TP+FP)

R=TP/(TP+FN)

F1=2\*P\*R/(P+R)