**数据说明**

本次比赛提供包含客户行为、资产、当前产品持有的历史业务数据，格式为UTF8编码格式的excel文件，以逗号分隔。

## 内容

数据集共包含：train.csv，test.csv两个个数据文件。

1.train.csv（feature+label）为78万条，包括70万条负样本和8万条正样本。

2.test.csv(feature)为34万条业务测试数据。

备注：1）提供的数据经过了脱敏；2）null代表缺失值。

## 特征说明

训练集里的feature包括了339个变量，其中cust\_id是定义客户的唯一id，flag字段为是否购买产品的label，其中第3-7个变量为客户基础属性， 第8-14个变量为账户数量， 第15-35个变量为账户交易情况，第36-48个变量为资产情况，第49-58个变量为贷款情况，第59-166及315-330个变量为渠道交易，第167-178个变量为渠道行为，第179-282个变量为第三方交易，第283-314个变量为自助设备交易，第331-338个变量为其他标识，最后一个变量为购买时间。测试集除没有flag外，其余特征与训练集相同。