

数据集描述

本次数据共分为两个数据集，主表.rar(x_train,y_train 、 x_test)和数据表.rar。其中主表.rar 内含训练集的主表 x_train、训练集的标签 y_train 和测试集的主表 x_test；数据表.rar 包含可供训练集和测试集用来加工衍生特征的数据表。建模的目标即根据训练集对模型进行训练，并对测试集进行预测。

训练集为 7、8 月的抽样数据，测试集为 9 月的抽样数据。其他数据表是 2021.1 到 2021.8 的数据。下面各表可以通过 core_cust_id 和 prod_code 进行关联。以下为数据集中每张表的字段描述：

- a) x_train 表：x_train 是训练主表，x_train 中的 id 字段和 y_train 中的 id 字段需要关联，作为一条带标签的样本。

字段	含义或阐述
id	主键
core_cust_id	用户唯一 id
prod_code	产品代码
a2	产品类别
a3	交易日期

- b) y_train 表：y_train 是训练标签集，id 字段是主键，y 字段是标签，y 取值为 0 或 1，0 表示未购买，1 表示购买。

字段	含义或阐述
id	主键
y	标签

- c) x_test 表：x_test 是 A 榜测试主表。

字段	含义或阐述
id	主键
core_cust_id	用户唯一 id
prod_code	产品代码
c2	产品类别
c3	交易日期

d) D 表：客户信息表

字段	含义或阐述
core_cust_id	用户唯一 id
d1	性别
d2	客户等级
d3	年龄

e) E 表：客户风险表

字段	含义或阐述
core_cust_id	用户唯一 id
e1	客户风险等级
e2	评估日期

f) F 表：资产信息表

字段	含义或阐述
core_cust_id	用户唯一 id
f1	客户编号创建日期
f2	定期季日均
f3	大额存单季日均

f4	A 理财产品季日均
f5	代销资管季日均
f6	代销基金季日均
f7	定期时点余额
f8	大额存单时点余额
f9	A 理财产品时点余额
f10	代销资管时点余额
f11	代销基金时点余额
f12	定期月日均
f13	大额存单月日均
f14	A 理财产品月日均
f15	代销资管月日均
f16	代销基金月日均
f17	定期年日均
f18	大额存单年日均
f19	A 理财产品年日均
f20	代销资管年日均
f21	代销基金年日均
f22	数据日期

g) G 表：A 理财产品产品表 1

字段	含义或阐述
prod_code	产品代码
g1	计价类型
g2	周期类型
g3	产品模式

g4	产品风险等级
g5	是否滚存型产品
g6	是否允许变更分红方式
g7	产品品种
g8	产品持有天数
g9	数据日期

h) H 表：B 理财产品表 1:

字段	含义或阐述
prod_code	产品代码
H1	计价类型
H2	周期类型
H3	产品模式
H4	产品风险等级
h5	是否允许变更分红方式
h6	产品品种
h7	产品模式 2
h8	数据日期

i) I 表：C 理财产品表 1:

字段	含义或阐述
prod_code	产品代码
i1	计价类型
i2	周期类型
i3	产品模式
i4	产品风险等级
i5	是否允许变更分红方式

i6	份额冻结的比率
i7	产品品种
i8	产品持有天数
i9	数据日期

j) J 表：D 理财产品表 1:

字段	含义或阐述
prod_code	产品代码
j1	计价类型
j2	周期类型
j3	产品模式
j4	产品风险等级
j5	是否允许变更分红方式
j6	收益计息基数
j7	预期最低收益率
j8	预期最高收益率
j9	产品面值
j10	发行价格
j11	收益率
j12	持有天数
j13	数据日期

k) K 表：A 类理财产品产品表 2:

字段	含义或阐述
prod_code	产品代码

k1	募集方式
k2	管理方式
k3	业务模式
k4	收益特点
k5	产品期限
k6	产品投资模式
k7	预期收益率
k8	最高收益率
k9	最低收益率
k10	产品展示等级
k11	日期

l) L 表：B 类理财产品产品表 2:

字段	含义或阐述
prod_code	产品代码
l1	募集方式
l2	管理方式
l3	业务模式
l4	收益特点
l5	产品期限
l6	产品投资模式
l7	日期

m) M 表：C 类理财产品产品表 2:

字段	含义或阐述
----	-------

prod_code	产品代码
m1	募集方式
m2	管理方式
m3	业务模式
m4	收益特点
m5	产品期限
m6	产品投资模式
m7	预期收益率
m8	产品展示等级
m9	数据日期

n) N 表：A 类理财产品交易流水表

字段	含义或阐述
n1	交易流水号
n2	业务代码
n3	渠道标识
core_cust_id	用户唯一 id
prod_code	产品代码
n6	净值
n7	申请金额
n8	资金状态
n9	交易状态
n10	总金额
n11	交易日期

o) O 表：B 类理财产品交易流水表

字段	含义或阐述
o1	交易流水号
o2	业务代码
o3	渠道标志
core_cust_id	用户唯一 id
prod_code	产品代码
o6	净值
o7	申请金额
o8	交易状态
o9	资金状态
o10	总金额
o11	超额管理费金额
o12	交易日期

p) P 表：D 类理财产品交易流水表:

字段	含义或阐述
p1	交易流水号
p2	业务代码
p3	渠道标志
core_cust_id	用户唯一 id
prod_code	产品代码
p6	净值
p7	申请金额
p8	折扣率
p9	交易状态

p10	资金状态
p11	费率
p12	交易日期

q) Q 表：C 类理财产品交易流水表：

字段	含义或阐述
q1	交易流水号
q2	业务代码
q3	渠道标志
core_cust_id	用户唯一 id
prod_code	产品代码
q6	净值
q7	申请金额
q8	资金状态
q9	交易状态
q10	交易日期

r) R 表：app 点击行为表

字段	含义或阐述
r1	id 编码
core_cust_id	用户唯一 id
r3	点击事件类型
prod_code	产品代码
r5	点击时间

s) S 表：账户交易流水表，其中 s3 和 s6 是客户编号，可以和其他表中的 core_cust_id 进行关联。

字段	含义或阐述
s1	事件编号
s2	交易类型代码
s3	借方客户编号
s4	借方金额
s5	借方交易日期
s6	贷方客户编号
s7	处理日期