**国安社区 用户活跃度模型**

1. **模型背景**

对用户建立活跃度模型，既可以对不活跃用户进行及时激活，也可以对活跃用户开展忠诚度计划，加强用户粘性。

注意：建模过程排除微超账号和虚拟账号（t\_customer.short\_name like ‘%虚拟%’）

统计口径：全量用户，成功订单口径

1. **数据源及模型指标**
   * 用户最近一次下单距今时长（天数）：R1
   * 最近两次不同日期下单间隔（天数）：R2
   * 用户近30天消费次数（0,30]：F30
   * 用户近30-90天消费次数（30,90]：F90
   * 用户近90-180天消费次数（90,180]：F180
2. **模型介绍**

* 模型主要通过聚类分析模型判断各指标对最后分类的影响程度，然后人为赋予指标权重进行活跃度得分计算。
* 模型具体模型参数如下：

1. 若R1>360，则活跃度得分Y=0；注册未成功下单用户，Y=0；
2. 若180<R1<=360，则活跃度得分Y=30
3. 若F180=1 and F90=0 and F30=0，则活跃度得分Y=40+5\*（180-R1）/90
4. 若F180=0 and F90=1 and F30=0，则活跃度得分Y=45+5\*（90-R1）/60
5. 若F180=0 and F90=0 and F30=1，则活跃度得分Y=50+5\*（30-R1）/30
6. 若F30>30，则活跃度得分Y=100
7. 针对R1<=180 and F30<=30 and F180+F90+F30>=2的用户，使用以下模型判断活跃度。

注：模型结果得分为55<=Y<=100

**第一步：使用R1、R2、F30、F90、F180五个指标进行聚类分析，得出N组客户特征，判断各指标的影响权重。**

**建模过程：**

（一）模型概述：

建模时间为20181113-20181126，共筛选出用户286,512人，排除虚拟及微超用户后，共计286,506人。

用户口径：全量用户

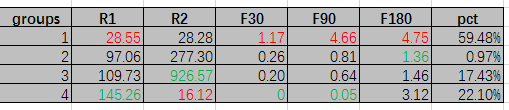
时间口径：未限制时间

R2表示用户最近**两次下单的日期间隔**，如果用户在同一天下单或仅下过以此单，我们将其替换为总体结果的最大值。（如果替换为平均值则会影响最终结果，并且替换为平均值并不符合此处的逻辑）

（二）聚类分析结果：

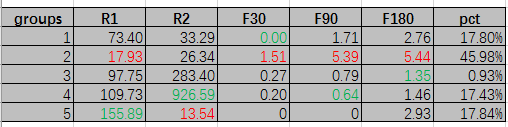
根据以上建模思路进行聚类分析实验，得到以下三种模型结果。

* + 1. 聚为四类：groups=4



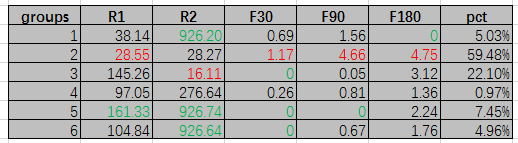
注：表中标红色为该项最优，标绿色为该项最差。

* 1. 第1类：近期很活跃，且有稳定的复购率——忠诚用户；
  2. 第2类和第3类：近期不太活跃，复购率虽然不高，但是有连续消费习惯——老用户，需要激励措施重点激活；
  3. 第4类：最近最不活跃度，90-180天的消费次数较高，但是近90天几乎没有消费——即将流失，需要实行客户挽留策略；
     1. 聚为五类：groups=5



注：表中标红色为该项最优，标绿色为该项最差。

1. 第2类：近期很活跃，且有稳定的复购率——忠诚用户；
2. 第1类：近期不太活跃，30天以前有较好的复购习惯，但是近30天几乎没有消费——老客户，需要激活；
3. 第3类和第4类：跟第1类相比，更加不活跃，但是近180天有连续消费习惯——老用户，需要重点激活；
4. 第5类：最近最不活跃度，90-180天的平均消费次数为3次，但是近90天几乎没有消费——即将流失，需要实行客户挽留策略；
   * 1. 聚为六类：groups=6



注：表中标红色为该项最优，标绿色为该项最差。

（三）模型结果分析：

根据聚类分析三种结果，对比后**选择使用groups=5的聚类结果**。

通过聚类分析结果，可以判断R1和F30能够非常明确地区分出用户特征，在模型计算时会赋予最大的权重；其次F90、R2、F180将会按从大到小的顺序依次赋予权重。

**第二步：根据第一步模型结果，对R1、R2、F30、F90、F180五个指标进行归一化处理，并赋予权重，用于计算每位客户最终的活跃度。**

针对R1<=180 and F30<=30 and F180+F90+F30>=2的用户，使用以下公式进行活跃度判断。

**R1=**

**R2=**

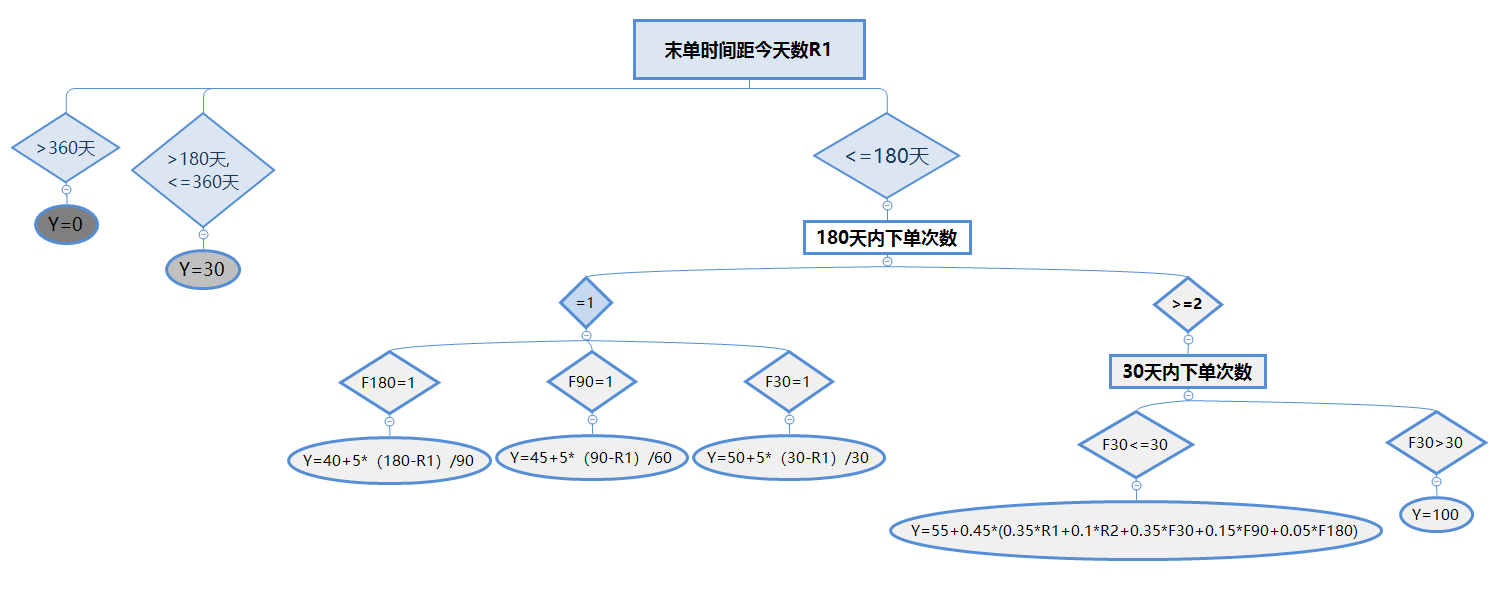
**F30=**

**F90=**

**F180=**

**Y=55+0.45\*(0.35\*R1+0.1\*R2+0.35\*F30+0.15\*F90+0.05\*F180)**

1. **模型结果**



1. **模型程序和代码**

《附件1：国安社区-用户活跃度模型-指标生成HQL》

《附件2：模型得分计算R程序》

由附件1得到每位用户的5项模型指标R1、R2、F30、F90、F180。

由附件2程序得到每位用户的活跃度得分Y以及分段值Y-cut。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| customer\_id | R1 | R2 | F30 | F90 | F180 | Y | Y-cut |
| 用户id | 天数 |  | 30天内消费次数 | 30-90天内消费总次数 | 90-180天内消费总次数 | 得分原值 | 分段值 |

1. **用户活跃度得分分布（现状）**

数据统计时间：20181123（更新用户标签，建议每周更新结果）

**把活跃度分成以下分段：**

0，（0-30】，（30-45】，（45-50】，（50-55】，（55-60】，（60-70】，（70-80】，（80-90】，（90-100），100

