**一、流失预警简易步骤流程图：**

历史数据通过计算得到训练数据

根据算法确定训练数据特征和格式

按照训练好的算法计算用户流失概率

使用训练数据优化算法模型参数

**二、输出：用户的流失概率**

**三、数据准备：**

1、会员年龄性别类型，八大消费属性标签，是否医保卡，地区，消费频次，客单价，客户积分，最后一次消费时间等特征数据；

2、会员月度活跃数据；

**四、流失预警算法简要说明：**

1、基于统计的流失预警模型：

取最近一个月流失标签用户作为流失初始化，取每个用户最后一次消费时间为起始时间，往前推一年，定义月活跃用户为在该月消费次数month\_sale\_times>=1次，得到每个用户每个月的活跃度如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| 用户A | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| 用户B | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 |

通过上表可以建立用户不活跃档案：

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| 用户A | 0 | 0 | 1 | 2 | 3 | 4 | 0 | 0 | 1 | 2 | 3 | 0 |
| 用户B | -1 | -1 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 0 |

通过用户不活跃档案表，可以对连续不活跃月份数尽心统计得到月份-连续不活跃月分数表

如下表：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 12月 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| 0 | 4815 | 4919 | 5042 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 1 |  | 574 | 649 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 2 |  |  | 425 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 3 |  |  |  | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 4 |  |  |  |  | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 5 |  |  |  |  |  | 504 | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 6 |  |  |  |  |  |  | 504 | 504 | 504 | 504 | 504 | 504 | 504 |
| 7 |  |  |  |  |  |  |  | 504 | 504 | 504 | 504 | 504 | 504 |
| 8 |  |  |  |  |  |  |  |  | 504 | 504 | 504 | 504 | 504 |
| 9 |  |  |  |  |  |  |  |  |  | 504 | 504 | 504 | 504 |
| 10 |  |  |  |  |  |  |  |  |  |  | 504 | 504 | 504 |
| 11 |  |  |  |  |  |  |  |  |  |  |  | 504 | 504 |
| 12 |  |  |  |  |  |  |  |  |  |  |  |  | 504 |

比如列1月行1表示已经不活跃1个月的用户数量，该数量来源于之前12月的4815个活跃用户。

利用公式：

2月连续不活跃2个月用户占比=2月连续不活跃2个月用户数/2月连续不活跃1个月用户数

可以得到每个月份连续不活跃n个月用户用户占比。

利用贝叶斯公式：

在流失概率模型中为：

分子第二项没有统计意义，舍去

得到：

分母之前已经求得，现在看分子

已经不活跃1个月用户流失的概率为：

同理可得，

假设用户每个月的购买行为为独立事件

则

2、基于RFM会员价值模型的流失预警模型：

RFM模型是衡量客户价值和客户创利能力的重要工具和手段：其中，R表示最近一次消费(Recency)，F表示消费频率(Frequency)，M表示消费金额(Monetary)。

参考已有会员价值模型（参考文档《会员价值模型.ppt》），可以得到一个用户的会员价值为：

Value=

其中R\_为最近一次消费的倒数，L为生命时长

使用生命周期会员标签，得到会员价值分布与生命周期对应表如下：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 新客（人数） | 老客（人数） | 流失（人数） |
| 会员价值最低5% | 1024 | 10 | 500 |
| 会员价值最低10% | 1024 | 10 | 500 |
| 会员价值最低15% | 1024 | 10 | 500 |
| 会员价值最低20% | 1024 | 10 | 500 |
| 会员价值最低25% | 1024 | 10 | 500 |
| 会员价值最低30% | 1024 | 10 | 500 |
| 会员价值最低35% | 1024 | 10 | 500 |
| 会员价值最低40% | 1024 | 10 | 500 |
| 会员价值最低45% | 1024 | 10 | 500 |
| 会员价值最低50% | 1024 | 10 | 500 |
| 会员价值最低55% | 1024 | 10 | 500 |
| 会员价值最低60% | 1024 | 10 | 500 |
| 会员价值最低65% | 1024 | 10 | 500 |
| 会员价值最低70% | 1024 | 10 | 500 |
| 会员价值最低75% | 1024 | 10 | 500 |
| 会员价值最低80% | 1024 | 10 | 500 |
| 会员价值最低85% | 1024 | 10 | 500 |
| 会员价值最低90% | 1024 | 10 | 500 |
| 会员价值最低95% | 1024 | 10 | 500 |

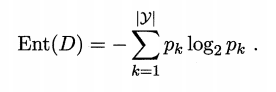
流失率计算公式为：

为了保证效果，可以将区间继续细分。

3、基于决策树和逻辑回归的流失预警模型：

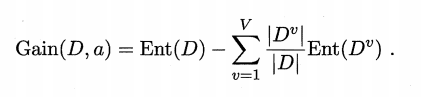
准备用户数据特征（年龄性别类型，八大消费属性标签，是否医保卡，地区，消费频次，客单价，客户积分，最后一次消费时间等）。

计算每个特征的信息熵：



信息熵越小，纯度越高

使用ID3决策树，信息增益为：



分别计算每个特征的信息增益，得到特征的信息增益信息如下：

|  |  |
| --- | --- |
|  | 信息增益 |
| 年龄性别类型 | 0.526 |
| 地区 | 0.123 |
| 医保 | 0.100 |
| 消费频次类型 | 0.688 |
| 会员等级 | 0.798 |

本例中，因为会员等级信息增益最大，取会员等级为根节点，分别取不同会员等级情况为第二级的根节点，迭代计算每个根节点下各个特征的信息增益，最后得到整个决策树。

该决策树可保存在一张表中，判断一个用户是否会流失只要在这个表中一层层往下找，如果没有下层，就可以根据当层做出判断。

结论：

算法1可以得出统计意义上的流失连续不活跃月份判断，算法2可以得出会员价值与流失的关系，算法3可以给出基于用户特征的流失判断