# 8月5日商业数据建模核心技能实训-实战作业

**背景**

在银行业务中，信用卡客户流失是一个重要的挑战.客户流失对银行业务造成负面影响，不仅意味着失去潜在的收入来源，还可能增加客户获取和留存的成本.如果能够预测哪些客户即将流失,并主动向这些客户提供更好的服务,能够挽回这些即将流失的客户.

**本例数据字段：**

|  |  |
| --- | --- |
| 字段名称 | 字段描述 |
| CLIENTNUM | 客户编号。客户持有账户的唯一标识符。 |
| Attrition\_Flag | 内部事件（客户活动）变量 - 如果账户关闭则为1，否则为0。 |
| Customer\_Age | 人口统计变量 - 客户年龄（以年为单位）。 |
| Gender | 人口统计变量 - M=男性，F=女性。 |
| Dependent\_count | 人口统计变量 - 受抚养人数。 |
| Education\_Level | 人口统计变量 - 账户持有者的教育程度（例如：高中、大学毕业等）。 |
| Marital\_Status | 人口统计变量 - 婚姻状况（已婚、单身、离异、未知）。 |
| Income\_Category | 人口统计变量 - 账户持有者的年收入类别（< $40K、$40K - 60K、$60K - $80K、$80K - $120K、> $120K、未知）。 |
| Card\_Category | 产品变量 - 卡片类型（蓝色、银色、金色、白金）。 |
| Months\_on\_book | 与银行建立关系的时间（以月为单位）。 |
| Total\_Relationship\_Count | 客户持有的产品总数。 |
| Months\_Inactive\_12\_mon | 过去12个月内不活跃的月份数。 |
| Contacts\_Count\_12\_mon | 过去12个月内的联系次数。 |
| Credit\_Limit | 信用卡的信用额度。 |
| Total\_Revolving\_Bal | 信用卡的总循环余额。 |
| Avg\_Open\_To\_Buy | 可用信用额度（过去12个月的平均值）。 |
| Total\_Amt\_Chng\_Q4\_Q1 | 交易金额变化（Q4相对于Q1）。 |
| Total\_Trans\_Amt | 总交易金额（过去12个月）。 |
| Total\_Trans\_Ct | 总交易次数（过去12个月）。 |
| Total\_Ct\_Chng\_Q4\_Q1 | 交易次数变化（Q4相对于Q1）。 |
| Avg\_Utilization\_Ratio | 平均卡片使用率。 |
| Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_1 | Naive Bayes分类器输出，不属于字段描述信息。 |
| Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_2 | Naive Bayes分类器输出，不属于字段描述信息。 |

**任务要求：**

1. **任务1：采用 Pyhton 语言对数据集进行数据预处理，具体包括缺失值检测与处理，数据特征描述，数据样例数据查看，5数查看(最大值、最小值等)**
2. **任务2：探索年龄分布情况**
3. **任务3: 探索客户受教育水平**
4. **任务4: 婚姻状况探索**
5. **任务5: 完成特征选择**
6. **任务6: 基于决策树建模**