# 他真的会续约“超级套餐”吗？

目前，通信行业波涛汹涌，某移动通信公司，给潜在客户打电话，希望其能续约某超级套餐。请根据所提供的一些（31188名）用户信息（包括，性别、年龄、状况等）建立模型，并对给定的客户（10000名）进行预测，判断其是否会续约该套餐服务。

*Will he really renew the “super package”?*

*At present, a mobile communication company calls potential customers, hoping that they wil renew the “super package”. Based on some(31,188) of the user information provided (including gender, age, status and so on), the project models and predicts whether the given customers (10,000) will renew the service or not.*

# 数据介绍

下面详细介绍数据（6-数据集 文件夹中）：

Train.csv 31188行， 21列， 最后一列（第21列）为用户是否续约该套餐的状态： 1-为续约 0-为没有续约

Test.csv 10000行， 20列

相关文件：

Train.csv：

用于训练和构建预测模型的数据集（31188条数据）。每条记录包含了21个特征属性，包括了客户数据、社会和经济背景属性等其他属性。特征21是目标。

Test.csv:

用于测试的数据集（10000条数据）。除了目标特征缺失以外，数据集与Train.csv拥有相同格式。

**各个特征的含义表格：**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据项（特征） | 数据意义 | 数据取值 | 说明 |
| 1 | 年龄 | numeric | （数字） |
| 2 | 工作 | categorical: "admin.","blue-collar","entrepreneur","housemaid","management","retired","self-employed","services","student","technician","unemployed","unknown" | 工作类型（分类：“行政”、“蓝领”、“企业家”、“女佣”、“管理”、“退休”、“自雇”、“服务”、“学生”、“技术员”、“失业”、“未知”）。 |
| 3 | 婚姻 | categorical: "divorced","married","single","unknown" | 婚姻状况（分类：“离婚”、“已婚”、“单身”、“未知”；注：“离婚”指离婚或丧偶） |
| 4 | 教育 | categorical: "basic.4y","basic.6y","basic.9y","high.school","illiterate","professional.course","university.degree","unknown" | （分类：“基础.4y”，“基础.6y”，“基础.9y”，“高中”，“文盲”，“专业.课程”，“大学.学位”，“未知”）。 |
| 5 | 违约 | categorical: "no","yes","unknown" | 信用违约？（分类：“否”、“是”、“未知”） |
| 6 | 住房 | categorical: "no","yes","unknown" | 有住房贷款吗？（分类：“否”、“是”、“未知”） |
| 7 | 贷款 | categorical: "no","yes","unknown" | 有个人贷款吗？（分类：“否”、“是”、“未知”） |
| 8 | 联系人 | categorical: "cellular","telephone" | 联系人通信类型（分类：“蜂窝”、“电话”）。 |
| 9 | 月 | last contact month of year, categorical: "jan", "feb", "mar", ..., "nov", "dec" | 每年的最后一个联系月（分类：“1月”、“2月”、“3月”、…“11月”、“12月”）。 |
| 10 | 天 | last contact day of the week (categorical: "mon","tue","wed","thu","fri") | 每周最后一个联系日（分类：“周一”、“周二”、“周三”、“周四”、“周五”）。 |
| 11 | 持续时间 | last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y="no"). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model. | 上次接触持续时间，以秒为单位（数字）。重要提示：此属性高度影响输出目标（例如，如果持续时间=0，则y=“no”）。然而，在执行呼叫之前，持续时间还不知道。而且，在通话结束后，Y显然是已知的。因此，这个输入应该只包含在基准测试中，如果想要有一个现实的预测模型，那么应该丢弃它。 |
| 12 | campaign | number of contacts performed during this campaign and for this client (numeric, includes last contact) | 在此活动期间为此客户端执行的联系人数（数字，包括最后一个联系人） |
| 13 | pdays | number of days that passed by after the client was last contacted from a previous campaign | 上次从上一个活动联系客户后经过的天数（数字；999表示以前未联系客户） |
| 14 | previous | number of contacts performed before this campaign and for this client (numeric) | 在此活动之前为此客户端执行的联系人数（数字） |
| 15 | poutcome | outcome of the previous marketing campaign (categorical: "failure","nonexistent","success") | 上一次营销活动的结果（分类：“失败”、“不存在”、“成功”）。 |
| 16 | EMP.var.rate | employment variation rate- quarterly indicator (numeric) | 就业变化率,季度指标（数字） |
| 17 | cons.price.idx | consumer price index- monthly indicator (numeric) | 消费者价格指数,月度指标（数字） |
| 18 | cons.conf.idx | consumer confidence index-monthly indicator (numeric) | 消费者信心指数月度指标（数字） |
| 19 | 同业拆借利率 | 3 month rate - daily indicator (numeric) | 同业拆借利率3个月利率每日指标（数字） |
| 20 | 雇佣人数 | number of employees - quarterly indicator (numeric) | 员工人数-季度指标（数字） |
| 21 | 客户是否已认购某超级套餐 | has the client subscribed a term deposit? (binary: "yes","no") | （分类：“是”，“否”） |