

期末作业

题目（二选一）

- A: 通过 **Python** 对数据集 `orders.csv` 进行数据分析。这个数据集包含某商店关于多种产品的销售数据，以及与地理、产品类别、销售收入和利润、消费者细分等相关信息。希望通过数据分析从数据中获取到有意义的见解，以改善市场和销售策略，其中必须包含但不限于可视化分析。
- B: 提供一个与某金融机构的营销活动有关的数据集，这些营销活动是基于电话的。通常情况下，需要对同一客户进行一次以上的联系，以便了解该金融产品是否会被认购（“yes”）或不（“no”）。数据集中的各字段描述如下：
 - 客户数据：
 - `age`：年龄(数值型);
 - `job`：工作情况(分类: ‘admin.’, ‘blue-collar’, ‘entrepreneur’, ‘housemaid’, ‘management’, ‘retired’, ‘self-employed’, ‘services’, ‘student’, ‘technician’, ‘unemployed’, ‘unknown’);
 - `marital`：婚姻状况(分类: ‘divorced’, ‘married’, ‘single’, ‘unknown’; 注: ‘divorced’ 指离婚或丧偶);
 - `education`：教育水平(分类: ‘basic.4y’, ‘basic.6y’, ‘basic.9y’, ‘high.school’, ‘illiterate’, ‘professional.course’, ‘university.degree’, ‘unknown’);
 - `default`：信用是否已经违约(分类: ‘no’, ‘yes’, ‘unknown’);
 - `housing`：是否有住房贷款(分类: ‘no’, ‘yes’, ‘unknown’);
 - `loan`：是否有个人贷款(分类: ‘no’, ‘yes’, ‘unknown’);
 - 与当前活动的最后一次联系有关的数据：
 - `contact`：联系时通信类型(分类: ‘cellular’, ‘telephone’);
 - `month`：一年中最后一次联系的月份(分类: ‘jan’, ‘feb’, ‘mar’, ..., ‘nov’, ‘dec’);
 - `dayofweek`：一周中最后一次联系是周几(分类: ‘mon’, ‘tue’, ‘wed’, ‘thu’, ‘fri’);
 - `campaign`：在本次活动中与该客户的联系次数(数值型，包括最后一次联系);
 - `passed_days`：上次联系客户后的天数(数值型，999 表示以前没有联系过客户);
 - `previous`：在此活动之前与该客户的联系次数(数值型);
 - `pre_outcome`：前期营销活动的结果(分类: ‘failure’, ‘nonexistent’, ‘success’);
 - 与社会和经济背景有关的一些指标：
 - `emp_rate`：就业变化率，季度指标(数值型);
 - `cpi`：消费者价格指数，月度指标(数值型);
 - `cci`：消费者信心指数，月度指标(数值型);
 - `r3m`：银行的 3 个月利率，每日指标(数值型);
 - `employed`：雇员人数，季度指标(数值型);

- 输出变量

- `y`：客户是否认购该金融产品(分类: 'yes', 'no').

该题包含两个文件，分别是完整的数据集(`data.xlsx`)，以及将要预测的结果(`学号.csv`)。要求根据 `data.xlsx` 构建合适的机器学习模型，然后对 `学号.csv` 中的数据进行预测，并填入预测结果(`yes` 或 `no`)。

书写格式与提交要求

- 报告在所提供的模板上进行书写，要符合常规的论文格式规范与行文要求；
- 报告主要呈现自己的分析方法、思路、结论等，不要出现 Python 代码；
- 所有的 Python 代码写在一个 notebook 文件中，加上必要的注释与编号，以与 word 文章中出现的数据、表格、图相一致的顺序组织，提交时删去不必要的代码，并有文中一致的执行结果；
- **notebook 文件在提交时请导出为 html 格式再提交；**
- **B 题在提交时，务必还要将 学号.csv 文件填好后，并改为自己的学号再提交；**
- word 格式的报告不需要在线提交，A 题仅需提交一个 html 格式的代码，**B 题除要提交一个 html 格式的代码文件外，还要提交 csv 文件；**
- 提交邮箱：realycf@vip.163.com；
- 邮件标题：**Python 语言+第几节+题号+学号**，
如： `Python语言+67+A+21020040000` 或 `Python语言+89+A+21020040000`，邮件收到后会有自动回复，如果没有收到，很有可能是标题不对被当成垃圾邮件了。

截止时间

- 在线提交：2023 年 1 月 3 日 22:00；
- word 报告纸质提交：2023 年 1 月 4 日 10:00-11:00，行政楼 716