

银行客户定期存款意向二分类预测

本学期期末考核以竞赛的方式进行。具体要求如下：

1. 任务内容

利用提供的银行客户信息对客户是否会定期存款进行二分类预测。首先使用训练数据集对分类器进行训练，然后用训练好的分类器在测试集上进行二分类预测，预测结果提交给系统进行打分。系统采用 **Mean F-Score** 对预测结果进行评估并排名。

编程语言限制规定

- 只可使用 Python 语言编程
- Python 库的使用没有限制，可使用课堂所讲解的 Python 库以外的 Python 库。

使用账号限制

- 每位同学只能使用一个 kaggle 账号参加考核提交结果，禁止使用多个账号参加考核提交结果。已设置最大参加人数上限，若有同学使用多账号参加，将导致其他同学无法参加考核。

2. 数据

从网站上下载训练集数据和测试集数据。**train.csv** 为训练集数据，**test.csv** 为测试集数据。数据文件内容具体说明参见 **DataDescription.txt**

3. 提交内容

本次考核提交内容包括两部分：

a. 测试集分类预测结果：

测试集分类预测结果以 **csv** 文件格式向系统提交。每天每个账号最多向系统提交两次测试结果。提交的 **csv** 文件包括两列：**SampleId** 和 **y**。其中 **SampleId** 为样本编号，**y** 为预测结果。每一行 **SampleId** 对应的 **y** 值为测试集中同一 **SampleId** 的样本分类预测结果。**y** 的取值为字符串，只可为 **yes** 或 **no**。提交的预测结果文件需包含所有测试集中样本的预测结果，具体格式如下：

```
SampleId,y
0,yes
1,no
2,yes
3,no
```

完整的预测结果文件样例参考 `SampleSubmission.csv`

b. 考核报告：

考核报告内容包括两个文件：

1. `.ipynb` 文件：可以输出和最终提交的分类集测试结果一致的在 `jupyter notebook` 中运行的 `Python` 程序。
2. `pdf` 文件：程序实现的思路、方法和算法等的说明以及中间实验结果等。可以包括在考核期间对各种算法的尝试和各种参数的选择采用的方法和实验结果的说明。同时**在 `pdf` 文件第一页注明在 `kaggle` 提交结果的用户名和昵称**。

将两个文件以 `zip` 或 `rar` 压缩打包，打包后的文件名以各自的姓名学号命名，格式为：“姓名_学号.zip”或“姓名_学号.rar”

将压缩文件以附件形式发送至 `cenfengcourses@163.com`

邮件主题以各自的姓名学号命名，格式为：`[2020_PR_FP]_姓名_学号`

4. 评分

测试集数据竞赛系统内部自动分为 `Public` 和 `Private` 两个子集分别进行评估（子集的具体划分不公开）。`Public leaderboard` 根据 `Public` 子集评估结果进行排名，`Private leaderboard` 根据 `Private` 子集评估结果进行排名。`Public leaderboard` 会显示评估的 `Mean F-Score`，`Private leaderboard` 没有 `Mean F-Score`。`Public` 子集和 `Private` 子集统计特性相近。`Private leaderboard` 的排名结果用于期中考核成绩的评定。

本次期中考核评分分为两部分：

- a. 程序实现部分：70%，对于考核报告中程序可重现的最终提交测试结果将根据提交测试结果在 `Private leaderboard` 中的排名进行评分。`Private leaderboard` 中 `Baseline` 为及格线。不可重现的最终提交测试结果为无效结果。
 - b. 考核报告部分：30%
- 期中考核最终评分为：程序实现部分得分*70%+考核报告部分得分*30%

5. 截止时间

- 测试结果提交截止时间: 01/21/2021 12:00 PM
- 报告提交截止时间: 01/23/2021 12:00 PM

6. 其它注意事项：

- 各位同学需独立完成程序实现和报告撰写。任何抄袭将取消成绩。
- 可在班级微信讨论群中对技术问题进行交流，不可私下交流。任何私下交流（包括但不限于微信群中的私下交流）视为考核违规。

- 微信交流群只可进行技术理论分析交流，不可发表具体程序实现，发表程序代码的行为视为考核违规。