

人工智能小组作业

软件工程系

提交日期：第 15 周（6.6）随堂提交

题目：5G 用户预测

目标

本作业旨在巩固理论知识，通过使用 python 编程环境实现相应的人工智能解决方案对现实生活中的问题进行探究并解决。本次作业能更好地帮助学习者熟悉掌握对数据集的操作、相关 AI python 库的使用、对实际问题中数据特性的目标的思考、以及不同模型间的比较和分析。该作用分组完成，**可跨教学班级组队**，每队严格限制在 **8~12 人**，组长将**分组名单**于 **4.14 日 23 时 59 分**前发送至邮箱 `jy.lin@fjnu.edu.cn`。

4.1

问题背景

在 2022 世界互联网大会乌镇峰会期间发布的《世界互联网发展报告 2022》显示，2022 年第一季度，全球 5G 用户数增加 7000 万人，总数达到 6.2 亿人左右，5G 人口覆盖率超 25%。据爱立信及全球移动通信系统协会 (GSMA) 预测，到 2022 年底，全球 5G 用户数量将突破 10 亿人。对于通信运营商来说，面对如此庞大的 5G 市场，如何基于一些用户侧的信息进行用户画像，再进一步对于潜在的 5G 使用者进行精准的推销也是非常有帮助的。

任务目标

本次作业任务是根据用户基本信息和通信相关数据、比如用户话费信息、流量、活跃行为、套餐类型、区域信息等特征字段，然后通过训练数据训练模型，预测测试集中每个样本是否属于 5G 用户。

数据说明

本次作业数据包含 60 个字段，其中 `target` 字段为预测目标，主要特征字段分为两类 `cat` 和 `num`，分别是离散型特征和数值型特征。**数据集已上传云平台，文件名为 `train.csv`。**

特征字段	字段描述
<code>id</code>	样本标识 <code>id</code>
<code>cat_0 ~ cat_19</code>	离散型特征
<code>num_0 ~ num_37</code>	数值型特征
<code>target</code>	目标字段，是否为 5G 用户

评估指标

本次作业的评价标准采用 AUC，即分数越高，效果越好。评估代码参考：

```
from sklearn.metrics import roc_auc_score
y_pred = [1, 0, 1, 0]
y_true = [0, 0, 0, 0]
roc_auc_score(y_true, y_pred)
```

要求

1. 实现不少于两种人工智能算法，对 5G 用户进行预测。
2. 撰写相应分析报告（**双面打印，不多于 6 页**）。

作业提交要求

1. 每个小组打包好的源码、电子版报告、成员分工与小组自评表（1 份），发送至邮箱：jy.lin@fjnu.edu.cn
2. 源码压缩包应至少包含一份 jupyter notebook 文件, 该文件展示大致建模过程和实验结果，其余自定义文件可以以普通 python（.py）文件包含在压缩包内。
3. 报告纸质版文件于第 15 周随堂提交。

提示

1. 用户预测问题可抽象为二分类问题或 0~1 的回归问题。
2. 报告中可包含对任务的思考、数据的分析、模型选择的原因、对查阅相关资料的思考和总结、不同模型的比较、结果的分析、以及可改进思路的思考等等。
3. 报告中**不能**包含任何实验过程的源码，如有需要可以包含**简洁**的伪代码。
4. 撰写报告时建议图文并茂，在分析时可适当加入图表。
5. 可以使用未学过的人工智能模型或算法，但必须在报告中给出模型原理、选择原因、结果分析。
6. 可使用任意实验课中未涉及的 python 人工智能库（library），不必给出详细说明。

加分项

小组作业上传 GitHub 开源，需要撰写相关**英文** ReadMe 文件, 并进行必要**英文**注释。作业总成绩最多加 5 分，依据组员分工不同，小组成员所加分数会有差异（会参考成员分工与小组自评表）。**开源地址请于实验报告中标明。**

小组作业期中检查

5 中旬（暂定 5.23），各小组实验课随堂进行作业进度汇报。

小组作业 ppt 汇报

16 周实验课（最后一节实验课），组长汇报展示（每组 10~15 分钟）