**B题 基于餐厅消费数据的隐形资助研究**

近年来，随着精准资助的开展，如何准确判定高校家庭经济困难学生、切实完善精准资助手段，对高校资助工作具有重要意义。隐性资助是通过大数据挖掘的形式，找准家庭经济困难学生的行为或经济状况特征，隐形认定（识别）出困难学生，并通过隐形实施的方式（不评比，不公示）给予适度的资助补偿，以保护家庭经济困难学生隐私，助力实现教育公平。随着信息化水平的提高，学生部分消费数据（如食堂三餐消费）有效记录并留存。通常认为学生经济情况可以通过餐厅消费金额、消费品类与消费次数等信息间接反映。某管理部门仿真了一组学生的消费数据。请建立数学模型解决如下问题：

1. 附件0是性别信息。附件1-3是该组学生不同学年的日三餐餐厅消费金额数据记录（部分），附件4-7同时给出了其中部分同学的饮食种类信息。请你的团队针对这些数据建立模型，挖掘不同代表性群体，并定量分析该群体三学年的主要消费行为特征变化规律、饮食种类变化规律等。注意建模前需对数据进行必要预处理（如删除不相关数据、缺失补全、特征提取等）。

**（第一问我用聚类+描述统计）**

1. 除以上信息外，附件8给出部分同学第一学年后经其它方式认定的贫困程度等级（粗粒度），其中等级2准确（可能不全）、其它等级认定可能有少量偏差。请建立数学模型依据消费行为（附件1-3）预测贫困程度，补全附件9（不要改动附件9的已有数据及顺序）并作为附件提交；

进一步结合第1问研究结论预测该组同学第二、第三学年的贫困程度隐形认定等级，分析相关变化。

**（第二问第一个，我用了机器学习，这个预测没啥问题。**

**后面标红的，让结合第一问再讨论，没思路了）**

1. 在第2问基础上，结合附件4-7饮食种类数据，改进你们的预测模型，比较分析相关同学的预测结果变化情况。

**（第三问，附件4-7的数据是部分同学的，直接把这些特征添加到第二问的机器学习的模型里面？ 第二问的模型是全体同学的数据，这样不是很多缺失值？）**

1. 通过以上贫困生本质特征挖掘，构建差异化（细粒度）资助额度分配算法，并以第三学年为例给出具体结果：对象为附件4-7中涉及的同学、资助总金额10万、资助人员80名，并对资助结果的公平合理性进行评估。

这问没思路

附件0 性别标签；

附件1 第一学年消费数据；

附件2 第二学年消费数据；

附件3 第三学年消费数据；

附件4-7 部分同学的三学年饮食种类信息；

附件8 已知贫困标签；

附件9 贫困标签待补全名单（问题2）。

注1：提供的所有数据表中同一序号是代表同一个人。

注2：数据中金额的单位为“分”。