二选一。鼓励选择题目2。不限制编程语言和编译环境。

提交截止时间为2023年1月10日。建议在2022年12月31日之前提交。

请点击以下链接提交。

https://workspace.jianguoyun.com/inbox/collect/7357710cac2248bba7a56ada0667fb26/submitv2

1.《药物》数据集：

* **介绍：**该数据集包含影响预测的各种信息，如年龄、性别、血压、胆固醇水平、钠钾比，最后是药物类型。搭建一个模型，
* **要求：**文件中train.csv为训练集，test.csv为测试集，搭建模型，以确定哪种药物可能适合未来患有相同疾病的患者。结果提交文件实例见example.csv，撰写报告，提交代码和结果文件。

2.《恒星》数据及介绍：

* **介绍：**这些数据包括100,000 次空间观测。每个观察结果都由 17个特征列和 1个类别列描述，将其标识为恒星、星系或类星体。

obj\_ID = Object Identifier，在CAS使用的图像目录中标识对象的唯一值

alpha = 赤经角（在 J2000 纪元）

delta = 偏角（在 J2000 纪元）

u = 光度系统中的紫外线过滤器

g = 光度系统中的绿色过滤器

r = 光度系统中的红色滤光片

i = 光度系统中的近红外滤光片

z = 光度系统中的红外滤光片

run\_ID = 用于识别特定扫描的运行编号

cam\_col = 相机列，用于识别运行中的扫描线

field\_ID = 标识每个字段的字段编号

spec\_obj\_ID = 用于光学光谱对象的唯一 ID（这意味着具有相同 spec\_obj\_ID 的 2 个不同观测必须共享输出类）

redshift = 基于波长增加的红移值

plate = 板 ID，标识 SDSS 中的每个板

MJD = Modified Julian Date，用于指示何时获取给定的 SDSS 数据

fiber\_ID = 光纤 ID，用于标识在每次观察中将光指向焦平面的光纤

label = 对象类别（星系、类星体对象或恒星,分别对应证书0,1,2）

* **要求：**文件中train.csv为训练集，test.csv为测试集，搭建模型，实现恒星的分类，结果提交文件实例见sample\_submission.csv，撰写测试报告，提交代码和结果文件。