

项目要求

Identify Fraud from Enron Email

代码质量

标准	符合要求
功能性	代码反映了 writeup 中问题答案的描述，即代码执行 writeup 中记录的功能，且 writeup 清楚指明了最终分析策略。
可用性	可运行 poi_id.py 来导出数据集、特征列表和算法，使最终算法可使用 tester.py 轻松检查。

理解数据集和问题

标准	符合要求
数据探索（相关小项目：第 5 课）	<p>学生的回复解决了数据集最重要的特点，并使用这些特点为分析提供信息。这些重要特点包括：</p> <ul style="list-style-type: none">• 数据点总数• 类之间的分配（POI/非 POI）• 使用的特征数量• 是否有哪些特征有很多缺失值？等。

标准	符合要求
异常值调查 (相关小项目：第 7 课)	学生的回复确定了财务数据中的异常值，并解释了如何消除或以其他方式处理它们。

优化特征选择/工程

标准	符合要求
创建新特征 (相关小项目：第 11 课)	至少实现了一个特征。书面回复中提供了选择该特征的理由，并测试了该特征对最终算法性能的影响。未要求学生将新特征包含在其最终特征集内。
明智地选择特征 (相关小项目：第 11 课)	部署了单变量或递归特征部分，或手动选择特征（尝试了不同的特征组合，并记录了每种组合的性能）。对所选的特征进行了报告并说明了所需特征数量的理由。对于支持获取特征重要性（如：决策树）或特征得分（如：SelectKBest）的算法，也进行了记录。
适当缩放特征 (相关小项目：第 9 课)	若算法要求缩放的特征，则部署特征缩放。

选择和调整算法

标准	符合要求

标准	符合要求
选择算法（相关小项目：第 1 至 3 课）	至少尝试了 2 种不同的算法并比较了它们的性能，最终分析中使用了性能较高的一个。
讨论参数调整及其重要性	回复解决了执行算法调整的意义及重要性。
调整算法（相关小项目：第 2、3、13 课）	<p>至少调整了一个重要的参数，并至少系统调查了 3 个设置，或以下任意为真：</p> <ul style="list-style-type: none"> • 使用 GridSearchCV 进行参数调整 • 调整了多个参数 • 参数调整融入了算法部分（即，为超过一个算法调整了参数，并选择了最佳算法-调整组合用于最终分析）

验证和评估

标准	符合要求
评估度量的使用（相关小项目：第 14 课）	至少使用了两个适当度量来评估算法性能（如：精确度、召回率），学生明确表达了这些度量在项目任务背景下的测量对象。
讨论验证及其重要性	回复解决了什么是验证以及它的重要性

标准	符合要求
验证策略（相关小项目：第 13 课）	所选最终算法通过以下方式进行评估：将数据拆分为培训和测试集或使用交叉验证，并说明所执行验证的具体类型。
算法性能	当使用 tester.py 评估性能时，精确度、召回率均至少为 0.3。

如何让你的项目脱颖而出？

你可以在[这里查看此评审标准的英文版本](#)。
