

订阅DeepL Pro以编辑此演示文稿。  
访问[www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document)，了解更多信息。

**数据分析基础 - ITC-5203**

**最终课程项目**

报告截止日期：2024 年 4 月 8 日星期五

银行营销

该问题与葡萄牙一家银行机构的直销活动（电话）有关。**分类目标**是预测客户是否订阅了定期存款。目标类别是最后一个属性（*订阅*），有两个值（*是*和*否*）。

训练集 (trainset.csv) 包含 3,196 条订阅记录和 26,076 条未订阅记录。测试集 (testSet.csv) 包含 1,444 条订阅记录和 1,047 条未订阅记录。

**属性信息：**

1. **年龄**
2. **job**：工作类型（分类："行政"、"蓝领"、"企业家"、"女佣"、"管理"、"退休"、"自营职业"、"服务"、"学生"、"技术人员"、"失业"、"未知"）。
3. **marital**：婚姻状况（分类："离婚"、"已婚"、"单身"、"未知"；注："离婚 "指离婚或丧偶）
4. **教育**：（分类："基础 4 年"、"基础 6 年"、"基础 9 年"、"高中"、"文盲"、"专业课程"、"大学学位"、"未知"）。
5. **住房**：是否有住房贷款？ （分类："否"、"是"、"未知）
6. **贷款**：是否有个人贷款？ （分类："否"、"是"、"未知）
7. **contact**：联系人通信类型（分类："手机"、"电话）
8. **月份**：当年最后一个联系月份（分类："1 月"、"2 月"、"3 月"......、"11 月"、"12 月"）。
9. **day\_of\_week**：一周中最后一个联系日（分类："周一"、"周二"、"周三"、"周四"、"周五"）。
10. **持续时间**：最后一次联系的持续时间，以秒计（数字）。
11. **活动**：在此活动期间为该客户执行的联系次数（数字，包括最后一次联系）。
12. **pdays**：上次活动联系客户后的天数（数值；999 表示以前未联系过客户）
13. **poutcome**：上次营销活动的结果（分类："失败"、"不存在"、"成功"）。
14. **nr.employed**：雇员人数--季度指标（数值）
15. **目标属性**：已认购 - 客户是否已认购定期存款？ （二进制："是 "或 "否）

# 步骤

该项目包括以下步骤：

1. **数据探索**：尝试了解数据，并对特征和目标属性中的重要特征进行统计。
2. **预处理数据**。这一步的目的是从训练集中的记录中提取特征，并将这些特征用于测试数据集。请注意，数据中的 "**未知** "值需要清理。
3. **使用 python** 提供的**分类学习方法**从训练示例集中学习模型。为此，您可以使用任何一种分类方法。
4. 在测试集上测试**学习到**的**模型**，并报告测试结果。

# 交什么？

1. 您为本项目编写的程序（**Python**）。
2. 一份包含以下内容的**报告**
   1. 项目的目标（可使用引言部分进行描述）。
   2. 您使用什么学习方法来学习模型，以及学习到的模型在测试数据上的测试结果（如分类准确率或 ROC 曲线）。
   3. 您在项目过程中发现的任何讨论和结论。
3. **10 分钟演示**的幻灯片
   1. 最后一堂课将有一个演讲，每个小组都必须介绍自己的项目。

# 如何上交：

* **演讲：**您必须在 4 月 8 日星期一th 下午 11:59 之前提交演示文稿。您将在课堂上展示您的项目。演示时间稍后公布。
* **报告和 Python 程序：**您必须在 4 月 8 日星期五之前将 Python 程序和报告（PDF 文件）上传到 Blackboard。

# 评分标准（25 分）

您的项目标记将由以下部分组成：

* 您的陈述（7 分）
* 报告的清晰度和条理性（10 分）
* 解决方案的合理性和正确性（如报告所述并在程序中实现（如有））（8 分）。

**不得迟交。**