

Python 数据分析编程 项目报告要求

一、选题

要求从如下题目中选择其中一个题目，撰写项目报告。数据来源见学习通平台。

1. 电商用户购买行为分析

根据"十四五"数字经济发展规划及商务部《关于推动电子商务高质量发展的指导意见》，聚焦电商用户行为数据（浏览、加购、购买等），通过多维度分析挖掘消费动机与流失风险，助力企业精准营销，响应政策关于"深化大数据应用，促进消费升级"的要求。

- **项目描述：**使用电商平台的用户行为数据集（如浏览记录、购买记录、用户画像等），分析用户的购买习惯、偏好特征以及影响购买决策的关键因素。通过构建分类模型预测用户是否会产生购买行为，并为电商平台提供个性化推荐策略建议。
- **涉及知识：**数据清洗、特征工程、分类模型（如逻辑回归、决策树）、数据可视化。
- **提示：**
 - 数据探索与清洗：理解数据集结构，自行判断并处理数据质量问题。
 - 用户行为模式洞察：从时间（如每日/每周趋势）、商品类别等维度，发现用户浏览与购买的主要规律。绘制时间序列图展示用户活跃度的变化趋势；绘制条形图展示热门商品类别。
 - 用户分群与画像：基于行为（如活跃度）或消费能力（如 RFM 模型），将用户划分为不同群体，并描述其独特特征。绘制饼图/条形图展示用户群体分布；使用雷达图或并列条形图对比不同用户群体的特征差异。
 - 购买行为预测：构建一个分类模型（如逻辑回归、随机森林），预测用户是否会发生购买行为。绘制 ROC 曲线评估模型性能；绘制特征重要性条形图。
 - 策略建议：根据上述分析，为精准营销或个性化推荐提出具体建议。
- **提交形式：**Python 代码、报告文档（包含相关性分析图表）。
- **数据集地址：**[Kaggle: E-commerce Behavior Data](#)

2. 城市交通流量预测

近年来，党和政府高度重视智慧交通建设，出台《交通运输领域新型基础设施建设行动方案（2021-2025 年）》等文件，明确提出推动交通流量预测智能化，提升精准感知与精细化管理能力。

- **项目描述：**使用纽约市出租车出行数据，任选其中连续三个月的数据，分析不同时间、地点的交通出行模式，预测未来的交通流量。学生可以通过建立回归模型，预测交通高峰期和低谷期，讨论影响交通流量的关键因素（如天气、节假日等）。
- **涉及知识：**回归模型、时间序列分析、数据可视化。
- **基本要求：**

- 数据清理：处理纽约出租车数据集，去除异常值和重复数据。
- 特征提取：提取影响交通流量的关键因素（如时间、地点、天气）。
- 模型建立：构建回归模型（如线性回归、随机森林）预测未来的交通流量。
- 数据可视化：用时间序列图展示交通流量的变化，用热力图展示不同区域的交通密度。
- 模型评估：通过均方误差（MSE）或其他指标评估模型的准确性。
- **提交形式：Python 代码、报告文档（包含可视化图表）。**

3. 通过润滑进行故障检测

在《国家智能制造标准体系建设指南（2024 版）》中，明确提出了智能装备的生产标准包括设备运行状态监控与优化、故障诊断与设备健康评估等设备运维标准。

- **项目描述：**通过分析齿轮箱润滑油的各种物理化学特性参数，建立模型预测齿轮箱的工作状态（良好/故障）。探索不同润滑参数与齿轮箱状态的关系，为设备维护提供数据支持。
- **涉及知识：**数据清洗与预处理、特征工程、分类模型构建、模型评估、数据可视化。
- **提示：**
 - 数据预处理：处理缺失值 (NaN) 和异常值、对数据进行必要的标准化/归一化处理、分析各特征间的相关性。
 - 数据分析：分析各润滑参数（温度、粘度、压力等）在齿轮箱良好状态和故障状态下的分布差异、确定对齿轮箱状态影响最大的关键参数、探索参数间的相互作用关系。
 - 模型构建：选择合适的分类算法（如逻辑回归、随机森林、SVM 等）、划分训练集和验证集、调参优化模型性能。
 - 模型评估：使用准确率、精确率、召回率、F1 分数等指标评估模型、绘制 ROC 曲线和混淆矩阵、分析模型误判案例的特征
 - 实践建议：基于分析结果提出设备维护建议、讨论模型在实际工业应用中的可行性、分析项目局限性及改进方向
- **提交形式：Python 代码、分析报告（包含模型和可视化图表）。**
- **数据集地址：**[GEARS: FAULT DETECTION THROUGH LUBRICATION | Kaggle](#)

4. 室内机器人导航数据分析

国家《“十四五”机器人产业发展规划》和《“机器人+”应用行动实施方案》明确将智能服务机器人列为重点发展方向，研发智能室内导航系统，重点突破多传感器融合与自主决策技术，助力养老、教育等场景的智能化升级，达到 2025-2027 年智能服务机器人试点方向要求。

- **项目描述：**通过分析机器人在不同类型地面上（平滑表面与粗糙表面）的导航数据，建立模型预测机器人当前行驶的表面类型。探索不同传感器参数与地面类型的关系，为机器人自适应控制提供数据支持。
- **涉及知识：**数据清洗与预处理、特征工程、分类模型构建、模型评估、数据可视化、时序数据分析。
- **提示：**

- 数据预处理：处理缺失值和异常值、进行必要的标准化/归一化处理、分析各传感器特征间的相关性。
- 数据分析：分析机器人导航参数在不同表面类型下的分布差异、确定对表面类型分类影响最大的关键传感器参数、探索机器人运动方向（顺时针/逆时针）与表面类型的关联性、比较平滑表面与粗糙表面的数据特征。
- 模型构建：选择合适的算法构建分类模型或时序模型，特别关注特征选择对分类准确率的影响。
- 模型评估：使用准确率、精确率、召回率、F1 分数等指标评估模型性能、绘制 ROC 曲线和混淆矩阵
- 实践建议：基于分析结果提出机器人控制策略优化建议、讨论模型在实际机器人导航应用中的可行性和局限性、分析项目存在的局限性并提出改进方向
- 提交形式：Python 代码、报告文档（包含分析结果和可视化图表）。
- 数据集地址：[Indoor Robot Navigation Dataset \(IRND\)](#)

二、综合评分考核要点

	评分维度	考核内容
1	数据处理与清理	处理缺失值、异常值，数据预处理是否合理。
2	数据分析与洞察	数据分析的全面性和深度，是否提取有价值的信息。
3	模型选择与构建(如适用)	模型选择合理性、构建与评估（无模型则分配给数据分析）。
4	数据可视化	图表类型是否合适，是否清晰、美观，有效展示数据。
5	创新性与思考	是否有独立思考或创新分析角度。
6	报告结构与表达	报告条理清晰，表达准确流畅。

三、提交要求

- 报告模版：见学习通平台
- 截止时间：第十周 周二（2025.11.14）23:59 分截止
- 提交形式：
 - 电子版报告+代码 提交到学习通平台
 - 纸质版报告由课代表收齐后交到办公室（学院 418）