

## 数据科学实践课程 - 中期报告作业要求

### 作业目的

中期报告旨在帮助学生系统总结项目进展，展示对数据集的深入理解、初步模型结果的分析，以及未来改进计划的规划。通过撰写报告，学生将培养数据分析、模型评估和学术表达能力，为最终项目报告奠定基础。

### 报告提交截止日期

2025年4月26日

提交方式：通过Blackboard上传PDF格式文件，文件命名格式为“姓名\_学号\_中期报告.pdf”。

### 报告结构与内容要求

中期报告应包含以下部分，建议总长度为3-4页（不含参考文献和附录）。请确保语言清晰、精炼、逻辑严谨，必要时可使用图表辅助说明。

#### 1. 数据集理解 (约0.5-1页)

- **数据来源与描述**：说明数据集规模、特征数量和类型信息。
- **探索性数据分析 (EDA)**：
  - 描述数据的基本统计特性（如均值、方差等）。
  - 识别数据中存在的 key 问题及对预测效果的影响。
- **数据预处理**：简述已执行的预处理步骤（如图像处理、标准化等）及理由。

#### 2. 初步模型结果 (约0.5-1页)

- **模型选择**：说明已尝试的模型（如resnet-18等）及选择依据。
- **实验设置**：
  - 描述模型训练超参（如batch\_size，学习率，优化器等等）。
- **结果分析**：
  - 展示初步模型的性能（如表格或图表形式）。
  - 分析模型的优劣，包括可能的过拟合、欠拟合或其他问题。
  - 比较不同模型或参数设置的表现（如有）。

#### 3. 计划提升方向 (约1页)

- 基于初步结果，识别当前数据集、模型或方法的不足。
- 提出具体的改进计划，例如：
  - 数据层面：创造更多数据、修复数据缺陷、处理数据不平衡等。
  - 模型层面：尝试更复杂的模型、引入正则化等。

- 说明改进计划的优先级和预期效果。

#### 4. 结论 (约0.5页)

- 总结当前项目进展和主要发现。
- 强调项目的下一步重点和预期成果。

#### 5. 参考文献

- 列出报告中引用的所有资料（如数据集、算法论文、工具文档等），采用APA或IEEE格式。

#### 6. 附录 (可选)

- 包含代码片段、额外图表或其他辅助材料，需简洁并与正文相关。

#### 评分标准

中期报告总分100分，撰写合理即给满分，各部分占比如下：

- 数据集理解（30%）：分析深入，预处理合理。
- 初步模型结果（30%）：实验设计科学，结果分析透彻。
- 计划提升方向（30%）：改进方案具体且可行。
- 报告结构与表达（10%）：逻辑清晰，语言专业，格式规范。

#### 注意事项

1. 确保数据分析和模型结果的真实性，切勿伪造数据或结果。
2. 图表需清晰标注标题、轴标签和单位。
3. 报告中如使用外部资源（如开源代码、数据集），需在参考文献中注明出处。
4. 逾期提交将扣除10%的分数/天，超过3天不予接受。
5. 如有疑问，请在实验课联系助教或老师。

#### 附加说明

- 鼓励在报告中展示创造性，例如提出独特的方法或深入的数据洞察。
- 报告语言为[中文/英文]，请保持专业术语准确。
- 提交前请检查报告的完整性，确保无拼写或格式错误。

期待看到你们在数据科学项目中的深入思考和创新成果！