

# 期末项目要求

Academic Year 2025-2026 Semester 1

420405 - 统计分析与建模

## 1. 项目背景与目标

本项目旨在通过统计分析与建模的技术，解决某一特定领域中的实际问题。学生需根据所选领域的问题，运用统计学方法进行数据分析、建模，并开发一个原型系统用于展示所构建模型的实际应用与效果。项目最终应能够帮助用户更好地理解该领域问题，并提供切实可行的解决方案。

## 2. 项目内容

### 1. 选题与问题定义

学生需根据兴趣选择一个领域（如医疗健康、金融分析、社会行为、气候变化、市场营销等），并明确该领域中的一个具体问题或挑战。问题定义需简洁明确，能够通过数据分析和统计建模来进行探索和解决。

### 2. 项目开源要求

- 必须通过Github或Gitee公开，包括项目数据、代码、项目说明文档、项目小视频/Video Abstract(5分钟以内)。

### 3. 数据分析与建模

- 数据收集与准备：**学生需收集相关领域的数据，支持后续的统计分析与建模。统计建模不同于机器学习（Machine Learning）或深度神经网络（DNN）方法，通常对数据规模要求较低，更关注数据的质量与代表性，数据规模请控制在 10 MB以内。
- 数据可视化分析：**对数据进行可视化探索，识别数据中的潜在模式、异常值、相关性等。通过原型系统进行可视化展示。
- 数据预处理：**对数据进行清理和预处理，包括缺失值处理、异常值检测与处理、特征选择与转换等，确保数据符合建模需求。通过原型系统展示预处理成果。
- 建模与算法选择：**根据问题的特性，选择合适的统计模型进行建模。模型的选择应与具体问题相匹配。至少包括2类模型：回归模型，分类模型。通过原型系统展示建模结果。
- 模型评估与优化：**使用适当的评估指标（如准确率、精确率、召回率、AUC值等）对模型性能进行评估，并讨论模型优化效果。通过原型系统展示指标评估结果及优化效果。

- **模型解读与应用：**对所构建的模型进行详细解读，包括模型的核心参数及意义，将模型应用到具体场景中，使其具有实际意义。通过原型系统展示模型的具体应用。

#### 4. 报告与总结

学生需提交一份详细的项目报告文档，内容包括，答辩PPT需要包括报告核心内容：

- 项目GitHub/Gitee链接
- 问题背景与研究目标的描述
- 数据收集与处理过程
- 建模、优化及性能评估
- 模型的解读与实际应用分析
- 团队分工及合作情况（提供数据支撑）
- 项目总结与反思，指出可能的改进方向
- 项目中AI工具使用情况介绍

### 3. 项目评估标准

满分100分，占总成绩比例为30%

评分依据	权重
<b>数据可视化</b> (需要对可视化结果进行准确分析、给出明确结论)	10%
<b>数据预处理与数据建模</b> (重点考察课程介绍建模方法的应用、选择适配应用的建模方法完成任务、并能够准确评价模型性能)	20%
<b>模型解读与应用</b> (基于模型对研究问题进行定量分析、能够正确解读模型参数、研究结论应该合理、是否适配应用场景等)	20%
<b>系统创新性、实用性以及系统改进方向</b> (解决方案是否具备实际价值与应用潜力，思辨能力和文献查阅等)	10%
<b>系统演示与现场答辩</b> (PPT内容完整度、系统演示效果、表达是否准确流畅、是否能正确回答问题等)	20%
<b>通过Github/Gitee展示团队合作</b> (全员参与、分工协作等)	10%

(正确、完整、按时提交等)

## 4. 分组答辩要求

- **小组项目：**人数不超过4人，资料以小组为单位进行提交
- **提交内容：** (1) 项目报告：Word或PDF格式 (2) 答辩PPT (3) 完整数据 (4) 完整源代码 (5) 项目小视频（不超过5分钟）
- **提交方法：**通过 Canvas 提交作业。每组仅需一名成员提交。请将数据文件放入 `data` 目录，源代码放入 `code` 目录，并将所有相关资料打包为一个 `.zip` 文件进行提交。须在答辩前完成提交。
- **答辩要求：**时间为16周上课时间，按照小组编号进行答辩，要求全员到场，介绍10分钟，问答5分钟。