

## 大作业评分依据

---

- 报告 (50%)
- 答辩 (30%)
- 性能 (20%)

# 大作业报告结构参考

---

- 项目标题
    - 作者
  - 摘要
  - 项目背景介绍
  - 问题概述
    - 数据描述
    - 所要研究的问题1描述
    - 所要研究的问题2描述
    - 所要研究的问题3描述
    - .....
  - 具体问题分析
  - 项目总结
  - 参考文献
  - 补充材料（代码、图片、视频等）
- 具体问题分析
    - 问题1分析
      - 数据分析
        - 数据特性分析（类型，值域，分布等）
        - 针对问题的数据处理
      - 数学模型选择
        - 使用了什么学模型
        - 原理是什么
      - 针对问题的建模
        - 该问题的建模解析
        - 具体建模步骤
      - 结果报告与结论分析
        - 图表形式结果展示
        - 不同模型之间结果的比较
        - 对结果进行分析得出结论
    - 问题2分析
    - 问题3分析

# 团队成员职责列表

---

- 描述每位团队成员在项目中承担的职责
  - 总体设计
    - 项目选择
    - 总体方案设计
  - 材料收集
    - 数据
    - 文献
    - 其他
  - 项目管理
    - 进度管理
    - 人员协调
  - 报告撰写
    - 总体章节设计
    - 章节负责人
- 负责的问题1
  - 问题分析
  - 方案设计
  - 报告撰写
  - 数据处理
  - 编码
  - 实验
  - 其他
- 负责的问题2

# 项目一：公司经营状况分析

---

## 数据来源：

<https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction>

## 简述：

The data were collected from the Taiwan Economic Journal for the years 1999 to 2009. Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange.

## 研究问题参考：

1. 根据所提供的的数据和属性，对公司的经营状况进行分析  
使用的参考数学模型：因子分析、相关分析、对应分析、Bootstrap方法、方差分析等
2. 对公司的经营状况进行评价与排名  
使用的参考数学模型：主成分法排名、因子分析法排名、秩和比综合评价等
3. 对公司进行归类，确定适合的类别个数，以及关键的因子  
使用的参考数学模型：R型-Q型聚类分析等
4. 分析导致公司破产的因素  
使用的参考数学模型：主成分分析，因子分析、相关分析等
- 5. 通过历史数据，对将要破产公司进行预测**  
使用的参考数学模型：时间序列分析、判别分析
6. 其他问题（自行决定）

注：需要对所假设的数据分布、模型有效性，实验结果进行假设检验和显著性分析

## 项目二：个人贷款分析

---

### 数据来源：

<https://www.kaggle.com/competitions/loan-default-prediction>

### 简述：

This data corresponds to a set of financial transactions associated with individuals. The data has been standardized, de-trended, and anonymized. You are provided with over two hundred thousand observations and nearly 800 features. Each observation is independent from the previous.

### 研究问题参考：

1. 根据所提供的的数据和属性，对个人贷款状态进行分析  
使用的参考数学模型：主成分分析、因子分析、相关分析、对应分析、Bootstrap方法、方差分析等
2. 对个人贷款风险进行评价与排名  
使用的参考数学模型：主成分法排名、因子分析法排名、秩和比综合评价等
3. 对客户进行归类，确定适合的类别个数，以及关键的因子  
使用的参考数学模型：R型-Q型聚类分析等
4. 分析个人贷款拖欠的因素  
使用的参考数学模型：主成分分析、因子分析、相关分析等
5. 对个人贷款是否发生拖欠进行检测，并预测造成的损失比例  
使用的参考数学模型：判别分析、回归分析
6. 其他问题（自行决定）

注：需要对所假设的数据分布、模型有效性，实验结果进行假设检验和显著性分析

## 项目三：信用卡诈骗分析

### 数据来源：

<https://www.kaggle.com/datasets/kartik2112/fraud-detection>

### 简述：

This is a simulated credit card transaction dataset containing legitimate and fraud transactions from the duration 1st Jan 2019 - 31st Dec 2020. It covers credit cards of 1000 customers doing transactions with a pool of 800 merchants.

### 研究问题参考：

1. 对交易中与发生信用卡诈骗相关的因子进行分析  
使用的参考数学模型：主成分分析、因子分析、相关分析、对应分析、Bootstrap方法、方差分析等
2. 对容易发生信用卡诈骗的群体（年龄、地区、职业、性别等）进行评价与排名  
使用的参考数学模型：主成分法排名、因子分析法排名、秩和比综合评价等
3. 对容易发生信用卡诈骗的商家类型进行分析与排名  
使用的参考数学模型：主成分分析、因子分析、相关分析、对应分析主成分法排名、因子分析法排名、秩和比综合评价等
4. 对诈骗种类进行归类，确定适合的类别个数以及关键的因子  
使用的参考数学模型：R型-Q型聚类分析等
5. **对交易是否发生诈骗进行检测**  
使用的参考数学模型：判别分析、回归分析
6. 其他问题（自行决定）

注：需要对所假设的数据分布、模型有效性，实验结果进行假设检验和显著性分析

## 项目四：影视作品分析和预测

### 数据来源：

<https://www.kaggle.com/datasets/ashishgup/netflix-rotten-tomatoes-metacritic-imdb>

### 简述：

This dataset combines data sources from Netflix, Rotten Tomatoes, IMBD, posters, box office information, trailers on YouTube, and more using a variety of APIs.

### 研究问题参考：

1. 找出影响影视作品票房 (Boxoffice), 评分 (Rating) 最相关的因素 (如: 演员、导演、体裁等)  
使用的参考数学模型: 因子分析、相关分析、对应分析等
2. 对影视作品的质量进行综合评价 (通过体裁、语言、演员等), 列出年度最佳10部作品  
使用的参考数学模型: 主成分法排名、因子分析法排名、秩和比综合评价等
3. 分析不同导演和演员擅长的体裁风格  
使用的参考数学模型: 对应分析、聚类分析等
4. 分析不同地区的电影 (通过第一语言) 受欢迎的因子之间的差异性  
使用的参考数学模型: 回归分析、对应分析、聚类分析等
5. **对电影票房 (Boxoffice) 进行预测**  
使用的参考数学模型: 主成分分析、回归分析、时间序列分析等
6. 其他问题 (自行决定)

注: 需要对所假设的数据分布、模型有效性, 实验结果进行假设检验和显著性分析