第一章：经济问题定义与背景分析

1.根据选题进行经济学理论分析和背景描述；

2.完成问题的技术转化（将经济问题转化为可量化指标/字段指标设计）；

3.设计问题解决的技术路线（或有）；

4.经济问题所用到的模型/算法的理论知识介绍。

第二章：数据获取

选择合适的方法获取数据（选择其中之一）

（1）使用Python爬虫获取公开数据集（基于库的爬虫Requests+BeautifulSoup、基于框架的爬虫Scrapy、特殊技术爬虫Selenium）

（2）直接获取已有数据集（txt、Excel、csv、图像、音频）或者数据库数据（SQL、nosql）、API接口爬取Requests + JSON解析

（3）使用Selenium模拟用户行为生成日志数据

第三章：数据预处理

1.数据清洗：处理缺失值、异常值检测等（缺失值分布热力图、异常值处理前后对比）

2.数据特征构建（特征相关性矩阵、数据标准化）

3.数据存储（SQL数据库：建表、插入数据、提交事务、查询数据、关闭连接）（或有）

4.描述性统计分析：协方差、箱线图、正态检验、单/双样本T检验等。

第四章：模型构建与评估优化

1.对比并选择合适的算法（至少选择两个算法进行对比）

1. 例如分类任务：Logistic Regression vs Random Forest vs XGBoost
2. 例如回归任务：Linear Regression vs LightGBM vs LSTM

2.模型结果评估：是否达到预期、误差可能带来多少损失

第五章：模型分析与政策建议

基于实验结果提出的结论，例如经济政策建议、商业策略等。

第六章：拓展练习

更换研究对象/数据/期限，让学生模仿练习。

# 选题参考

1. 相关性分析
   1. 数字化转型对经济效率的影响
   2. 电影票房收入影响因素分析
   3. 毕业生薪资水平影响因素分析
2. 经济现象预测
   1. 景区游客流量预测
   2. 客户价值预测
   3. 餐饮外卖店铺评分预测
3. 分类
   1. 社媒用户情感分析
   2. 年报文本分类
   3. 客户分群
4. 综合
   1. 新能源汽车充电桩选址规划\*（涉及遗传算法）
   2. 共享经济动态定价策略\*（需采集实时供需数据）
   3. 社交媒体舆情对股价影响分析\*（需分析多模态数据）
   4. 城市交通拥堵经济成本评估\*（要基于GIS进行空间分析）
   5. 金融精准营销
   6. 保险欺诈识别
   7. 信用违约预测与评分卡建立