project\_name/

│

├── 00\_docs/

│ ├── project\_notes.md # 研究笔记/会议纪要/决策记录

│ ├── data\_dictionary.pdf # CHARLS/KLOSA 数据字典

│ ├── analysis\_plan.md # 分析方案 (Methods 对应的SOP)

│ └── references.bib # 文献库 (RIS/BibTex/EndNote)

│

├── 01\_raw\_data/

│ ├── CHARLS\_2018.dta

│ ├── KLOSA\_2018.dta

│ └── readme.txt # 数据来源/下载日期/版本号

│

├── 02\_processed\_data/

│ ├── baseline\_clean.csv # 清洗后数据

│ ├── imputed.csv # 缺失值插补数据

│ ├── analysis\_ready.csv # 建模数据 (标准化/哑变量化)

│ └── logs/

│ ├── missing\_report.xlsx # 缺失率/删除样本记录

│ └── baseline\_distribution.pdf # 基线分布可视化

│

├── 03\_scripts/

│ ├── 00\_utils.py # 公共函数 (绘图、评估、CI计算)

│ ├── 01\_cleaning.py # 数据清理 (异常值/逻辑冲突)

│ ├── 02\_imputation.py # 缺失值插补 (MICE/众数/中位数)

│ ├── 03\_feature\_engineering.py # 特征工程 (标准化/哑变量化)

│ ├── 04\_baseline\_analysis.py # 基线描述 (Table 1)

│ ├── 05a\_hyperparam\_search.py # 超参数调优 (CV/随机/贝叶斯搜索)

│ ├── 05\_model\_training.py # 最终模型训练 (训练+内部验证)

│ ├── 06\_validation.py # 外部验证 (KLOSA)

│ ├── 07\_metrics\_ci.py # 各类指标+95%CI计算

│ ├── 08\_sensitivity.py # 敏感性分析入口

│ ├── shap\_plots.py # SHAP解释性分析 (全局+个体)

│ └── run\_all.sh # 一键运行pipeline

│

├── 04\_results/

│ ├── tables/

│ │ ├── training\_metrics.xlsx # 训练集指标 (均值+95%CI)

│ │ ├── cv\_metrics\_charls.xlsx # 内部验证指标 (均值+95%CI)

│ │ ├── external\_metrics\_klosa.xlsx # 外部验证指标 (点估计+95%CI)

│ │ ├── model\_compare\_charls.xlsx # 多模型对比 (含95%CI)

│ │ ├── threshold\_metrics.xlsx # 不同cutoff下的阈值指标

│ │ └── tuning\_summary.xlsx # 调优过程汇总 (Top-k参数)

│ │

│ ├── tuning/

│ │ ├── search\_log.csv # 调优日志 (参数+分数)

│ │ └── search\_space.yaml # 搜索空间定义

│ │

│ ├── figures/

│ │ ├── roc\_curves.pdf # ROC曲线 (训练/内部/外部)

│ │ ├── pr\_curves.pdf # PR曲线

│ │ ├── calibration\_curves.pdf # 校准曲线

│ │ ├── dca\_curves.pdf # 决策曲线 (DCA)

│ │ ├── shap\_beeswarm.png # SHAP beeswarm图

│ │ ├── shap\_bar.png # SHAP条形图

│ │ └── patient\_explanations/ # 个体化SHAP解释 (force/waterfall)

│ │

│ └── reports/ # 自动生成的结果报告

│ └── model\_report.html

│

├── 05\_supplement/

│ ├── table\_s1\_missingness.xlsx # 缺失值表

│ ├── table\_sx\_metrics.xlsx # 全指标+95%CI汇总 (训练/验证/外部)

│ ├── figure\_s1\_flowchart.png # 样本筛选流程图

│ ├── tripid\_ai\_checklist.docx # TRIPOD-AI检查表

│ ├── methods\_details.docx # 方法学细节 (CI算法等)

│ └── extra\_figures/ # 补充图表 (敏感性/消融/非线性)

│

├── 06\_lit\_review/

│ ├── search\_strategy\_pubmed.txt # 文献检索式

│ ├── reference\_list.ris # RIS文件 (EndNote/Zotero导入)

│ └── top\_papers\_notes.md # 核心文献笔记

│

├── 07\_config/

│ ├── requirements.txt # Python包版本

│ ├── config.yaml # 模型参数配置

│ └── seeds.txt # 随机种子

│

├── 08\_logs/

│ ├── training\_log.txt # 训练日志

│ └── validation\_log.txt # 验证日志

│

├── 09\_deployment/

│ ├── app.py # Streamlit主程序

│ ├── requirements\_app.txt # 部署环境

│ ├── screenshots/ # 工具截图 (Supplement)

│ └── README\_app.md # 部署说明

│

├── 10\_experiments/

│ ├── sensitivity/

│ │ ├── cesd\_cutoffs.py # CES-D cutoff敏感性

│ │ ├── missing\_data\_methods.py # 缺失值处理对比

│ │ ├── subgroup\_analysis.py # 亚组分析 (性别/年龄/城乡/教育)

│ │ └── tautology\_check.py # 排除与结局高度相关变量

│ │

│ ├── ablation/

│ │ ├── single\_feature\_ablation.py # 单变量消融

│ │ ├── domain\_ablation.py # 域级消融

│ │ └── feature\_ranking\_compare.py # SHAP vs permutation 排名

│ │

│ ├── fairness/ # 公平性分析

│ │ └── subgroup\_calibration.py # 不同人群的校准/性能

│ │

│ ├── threshold/ # 阈值优化

│ │ └── cutoff\_analysis.py # Cutoff下的敏感性/特异性/PPV/NPV

│ │

│ └── results/

│ ├── sensitivity\_tables.xlsx

│ ├── ablation\_tables.xlsx

│ ├── fairness\_tables.xlsx

│ ├── threshold\_tables.xlsx

│ ├── sensitivity\_figures.pdf

│ ├── ablation\_figures.pdf

│ ├── fairness\_figures.pdf

│ └── threshold\_figures.pdf

│

└── README.md # 项目说明文档