

1. 核心任务与分工

1.1. 完成最小可行产品，搭建好核心架构，进行概念验证

流程：数据采集 - 模型融合 - 结果展示

- 1. 数据层：开发自动爬取数据的爬虫程序和清洗程序
- 2. 数据层：数据清洗、数据标注算法框架搭建：开发自动标签化数据的算法进行数据标注。
- 3. 算法层：搭建模型框架，完成基础调优，进行先一轮的时序模型训练和FinBERT模型微调。
- 4. 应用层：简易 Web 展示预测结果，无需集成 Backtrader 回测或实时预警功能。

1.2. 交付物

- 1. 自动爬虫系统
- 2. 数据清洗、标注系统
- 3. 数据集
- 4. 模型算法框架
- 5. 调优模型+自己训练的时序模型
- 6. web 展示系统

1.3. 里程碑

最小可行产品(MVP) 的完成可作为项目第一阶段里程碑：

- 阶段 1：完成单股票 MVP 验证；
- 阶段 2：扩展至沪深 300 或者成本限制其他典型过股票池，引入多模态动态融合；
- 阶段 3：集成回测与预警功能。

1.4. 目的

- 降低试错成本
- 为写论文，加速学术产出提供初步的数据支撑
- 预测的结果好可以申请到更多经费

2. 分工

数据层	模型层	应用层
杨佳冉		
李敏慧		
杨诗言		

- 1. 数据层：文本抓取的爬虫+行情数据的爬虫
- 2. 数据层：数据清洗、数据标注算法框架搭建
- 3. 模型层：时序模型的算法（脚本）+ FinBERT 模型的微调
- 4. 应用层：简易 Web 的搭建和模型接入

5. 测试：数据集的生成、模型效果的验证

6. 调整：若 FinBERT 情感分析准确率低于 70%，李敏慧需重新标注关键新闻，杨佳冉调整融合权重，杨诗言优化数据输入格式

分工占比：40%+30%+30%