

AI Hackathon 项目完整说明书

项目名称：ML Trading Benchmark（可复现的机器学习量化交易评测系统）

一、产品简介

本项目是一个面向量化交易 AI 模型的工程化评测平台，核心目标是用统一协议衡量不同模型在真实交易约束下的有效性。系统覆盖数据获取、特征工程、模型训练、回测评估与报告生成，帮助团队快速回答“模型是否可落地”这一关键问题。

项目聚焦解决四类行业痛点：一是只比模型、不比被动基准；二是忽略交易成本导致纸面收益失真；三是前瞻偏差和标签泄漏影响可信度；四是评估协议不统一导致结果不可横向比较。

二、核心功能

1) 一站式流水线：支持“数据下载 特征工程 时间切分 训练预测 回测 指标计算 报告输出”的全流程自动执行。2) 可配置实验协议：通过配置文件管理数据区间、模型启停、回测参数、成本场景与输出格式。3) 结果可复现：支持跳过重复步骤快速复跑，并输出 CSV/LaTeX/PDF/JSON 多格式结果。

默认实验覆盖 50 只美国 ETF、约 20 年日频数据、13 个技术特征与 9 类基线模型（传统机器学习、深度学习、策略基线）。

三、产品亮点

亮点1：真实交易约束评估。系统在 0/5/10/15/25 bps 成本场景下评估策略，直接识别“高收益但不可交易”的模型。亮点2：严格防泄漏机制。采用滚动标准化 + Walk-forward + Embargo 组合，确保训练与验证测试时间严格隔离。

亮点3：收益与信号双重评估。除 Sharpe、CAGR、Max Drawdown 外，增加 IC/ICIR 评估信号质量，可以区分“策略执行问题”和“预测能力不足”。亮点4：自动化论文级输出，可直接用于答辩与报告材料。

四、技术方案特色

技术架构采用“数据层-特征层-建模层-策略层-评估层-报告层”分层设计。数据层负责统一 OHLCV 数据结构；特征层生成收益率、波动率、动量、RSI 等 13 项特征；建模层统一 fit/predict 接口以支持多模型横向对比。

策略层默认使用 Top-K 多空组合（前10做多、后10做空、等权调仓），并显式扣除交易费用与滑点。评估层提供绩效指标、IC/ICIR、Bootstrap 置信区间及 Diebold-Mariano 显著性检验（含 FDR 校正），保证结果具备统计解释力与工程可信度。

五、项目价值与竞赛展示建议

对评委：本项目强调“可复现、可审计、可落地”，不仅展示模型效果，更展示严谨评估能力。

对团队：可快速扩展新模型与新市场，形成长期迭代的量化AI基准平台。

30秒答辩口径：我们做的不是单一模型，而是一套面向真实交易约束的AI评测基础设施。它将数据、建模、回测、统计检验和报告生成打通，能够稳定判断“模型在真实世界是否有效”。

版本日期：2026-02-24