

2025 年上海市专项职业技能大赛“人工智能训练师”项目技能操作考核
试题单

准考证号：

试题代码：B2

试题名称：基于股票市场指数预测算法训练集构建与测试方案设计

考核时间：25min

1. 场地设备要求

计算机：CPU（Intel i5 及以上）、内存（不少于 8GB）、操作系统（Windows10 及以上）

2. 工作任务

某金融科技公司正在开发一个股票市场指数预测系统，目标是根据历史指数走势、个股表现及宏观经济指标预测指数短期涨跌趋势。然而，在初步测试中发现，不同市场板块或交易时间段的指数波动存在显著差异。如果不加控制，模型上线后可能在特定板块或时间段表现不稳定。

为了保证模型上线后的预测效果稳定，作为人工智能训练师，需要构建高质量训练集、黄金测试集，并评估模型在不同板块或交易时间段的表现。

目前已收集了部分历史指数数据（文件名：B2.csv），需完成以下操作：

数据字段说明：

➤ 数据字段包括：

字段名	中文名称	说明
date	日期	交易日期
time	时间	交易时间（可精确到分钟）
index_id	指数 ID	指标唯一标识，如“上证指数”、“创业板指数”等
pct_change	涨跌幅	指数相对于前一时间点的涨跌幅，单位%
sector	市场板块	指数所属市场板块，如“科技”、“金融”等
market_phase	市场阶段	市场阶段标签，如“震荡期”、“上涨期”、“下跌期”

任务一：构建并保存高质量训练集（相应的实现代码填写在 B2.ipynb 框架

中), 要求从时间字段中提取日期和小时, 并自主分析数据分布与业务逻辑, 设计并实现以下 3 类特征:

- 指数在每个时间段的平均涨跌幅
- 不同板块的指数在不同市场阶段的平均涨跌幅
- 不同板块的指数在过去 5 个交易日的涨跌趋势

任务二: 构建黄金测试集并进行偏差检测, 确保测试集能够代表整体数据分布, 尤其是确保不同板块的分布一致(相应的实现代码填写在 B2.ipynb 框架中):

(1) 黄金测试集构建

按照时间顺序将数据划分为训练集与测试集, 保证测试集能够代表整体市场分布, 并符合市场时序特性。

(2) 偏差分析

使用 Kolmogorov-Smirnov (KS) 检验分析测试集中不同板块及市场阶段的涨跌幅分布是否与整体数据一致, 避免模型在特定板块或阶段表现偏差。

任务三: 结合本题业务场景, 针对近期波动幅度显著高于平均水平的热点板块指数, 设计一套完整测试方案, 包括测试目标、测试方法及评估指标。

桌面创建考生文件夹, 命名为“参赛号/B2”。请将 B2.ipynb 中的代码及运行结果导出为 HTML 格式并命名为 B2.html, 同时将自行创建的“答题卷.doc”文件(内含答案)一并保存至该考生文件夹中。此外, 按照考试平台的操作规程, 将所有作答结果录入系统, 并在完成后最终提交。

3. 技能要求

- (1) 能结合业务特征构建算法的高质量训练集, 并成为算法的核心竞争力;
- (2) 能结合业务特征构建算法的黄金测试集, 并作为算法上线前的质量保障;
- (3) 能结合业务特性设计合理的测试方案。

4. 质量指标

- (1) 高质量训练集构建正确;
- (2) 黄金测试集构建正确;
- (3) 测试方法合理可行。