

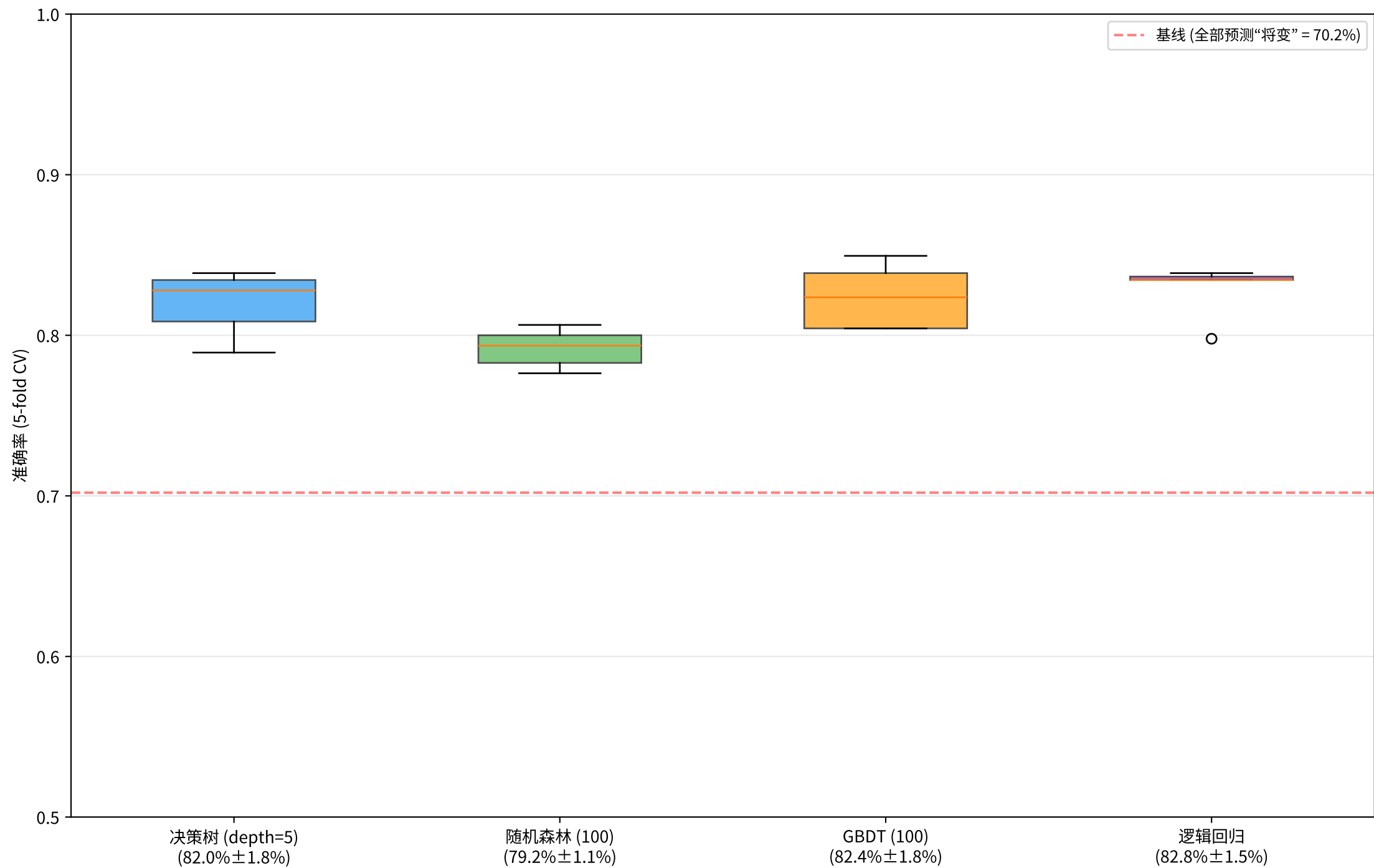
AI 冲击扫描 — 预测建模分析报告

Phase E: 什么因素决定了变化状态？

基于 2,325 节点 × 18 维度 numeric_profile

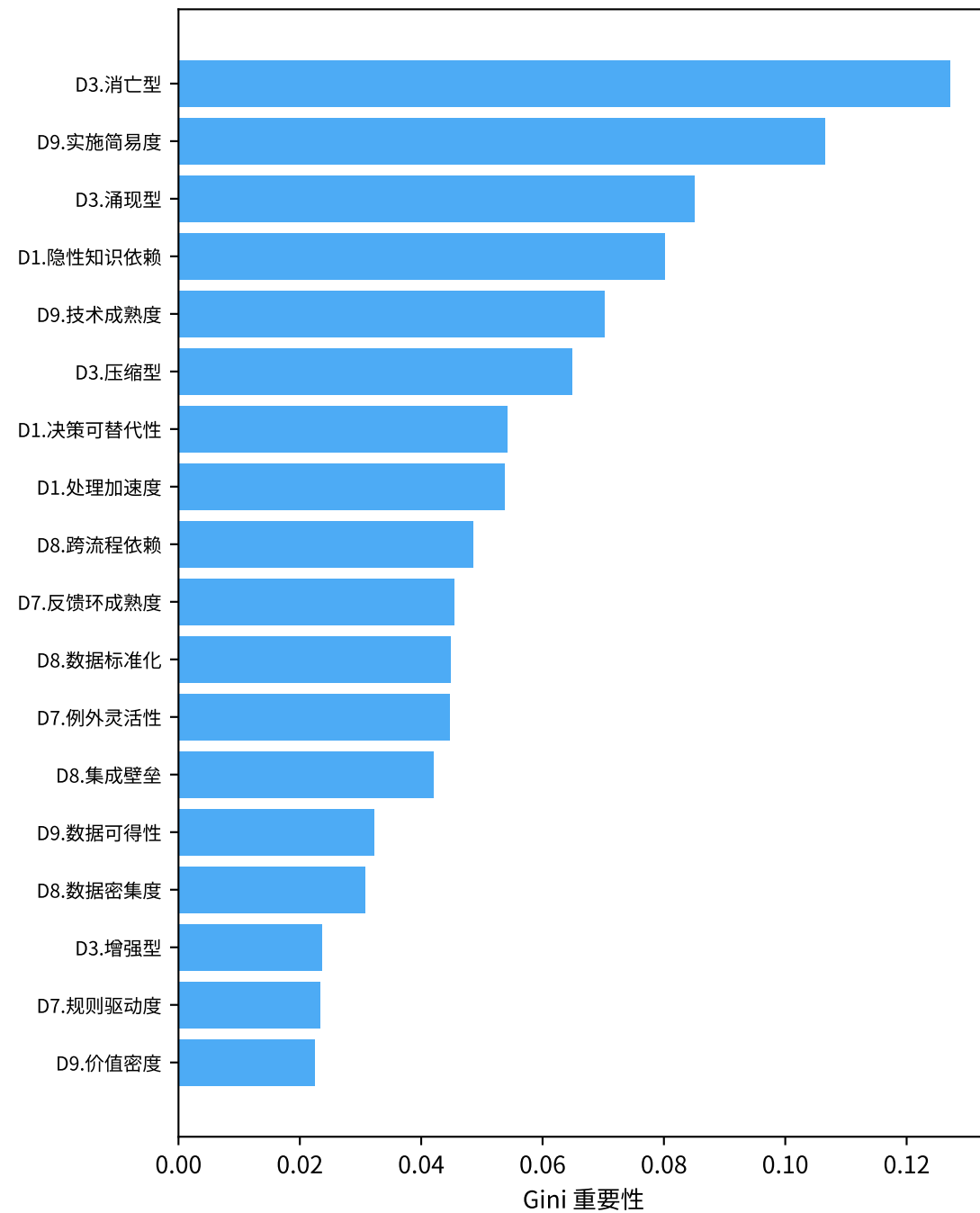
生成日期: 2026-03-02

E1. 四种模型的交叉验证性能对比

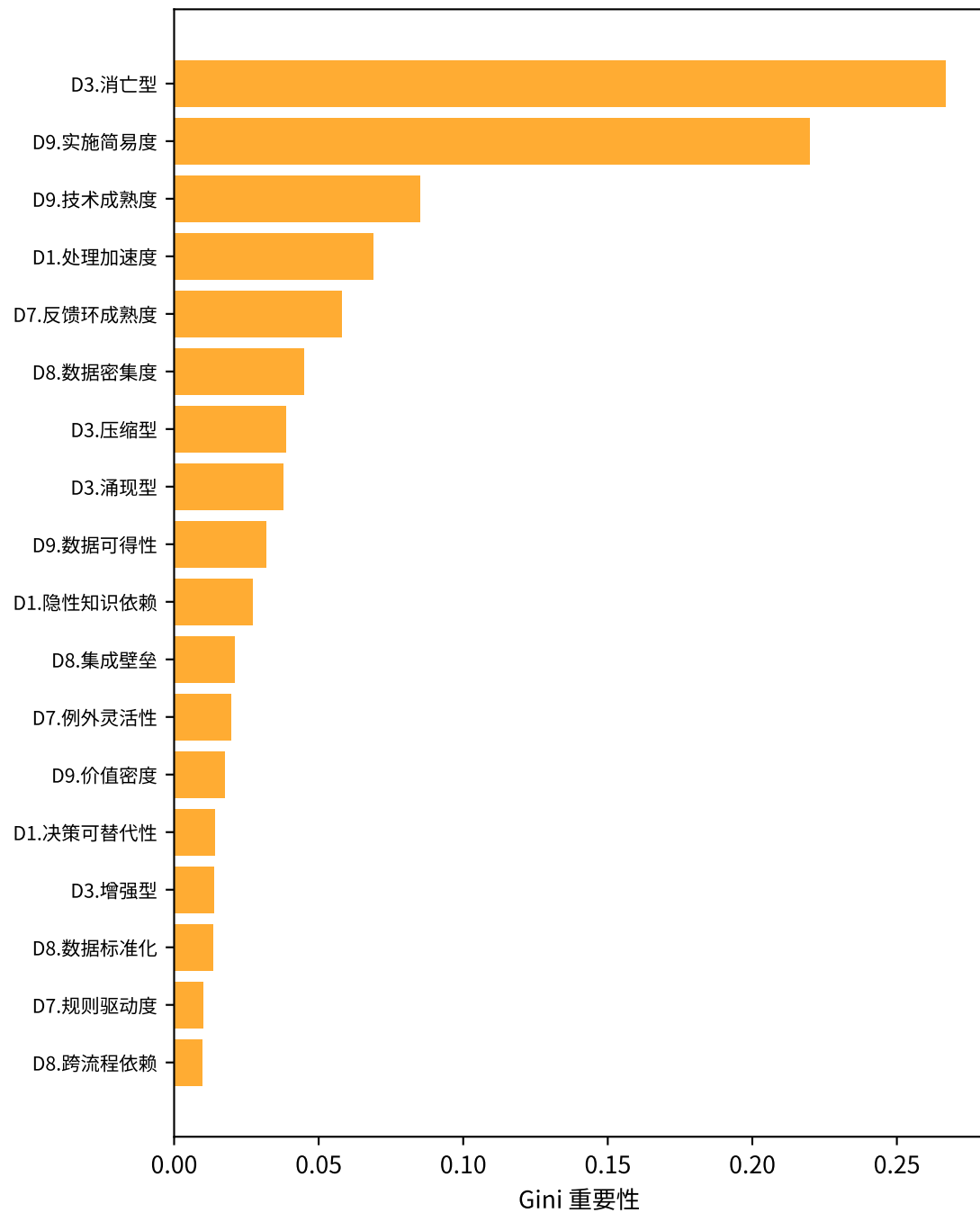


E2. 特征重要性分析

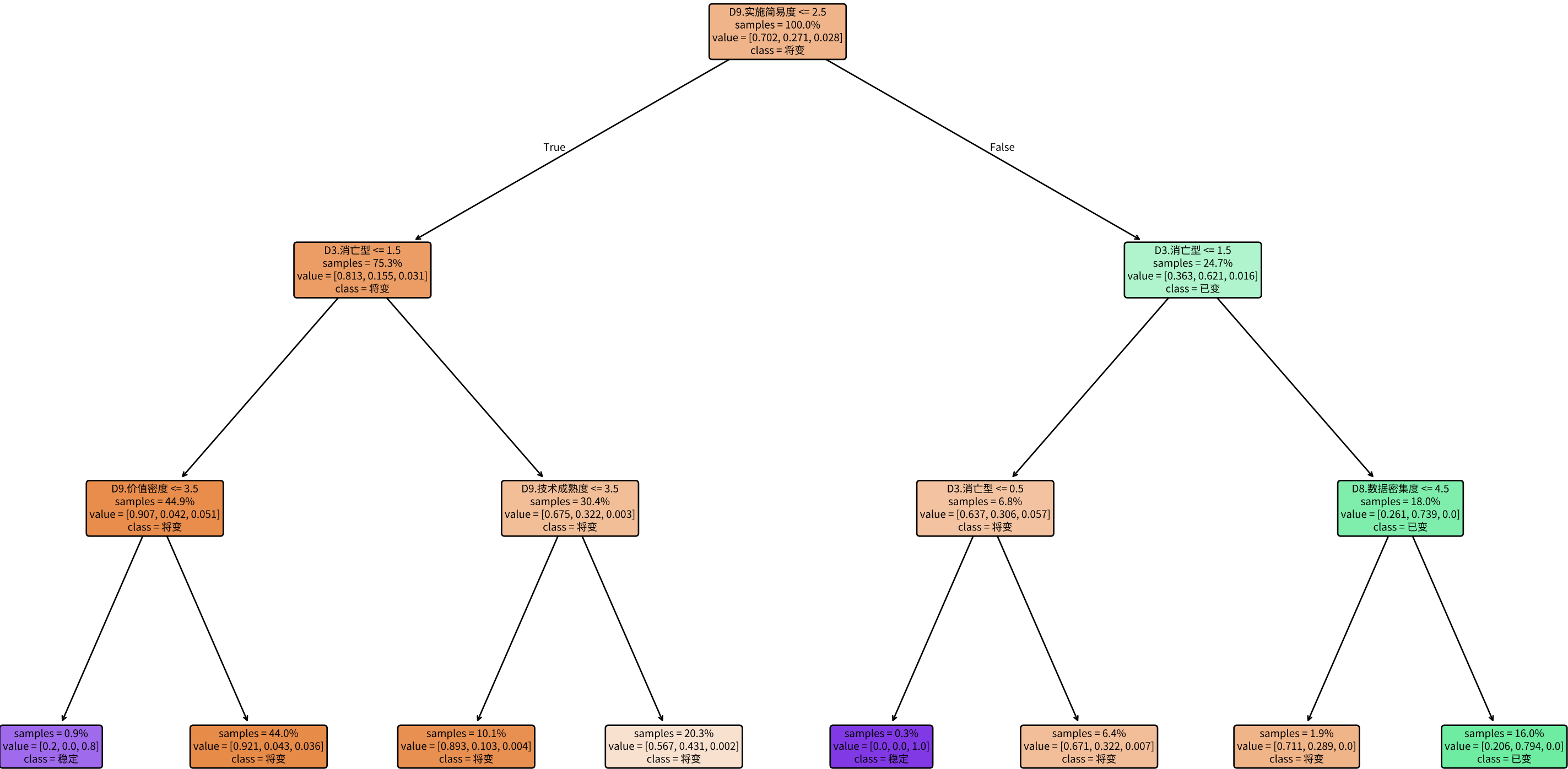
随机森林



GBDT

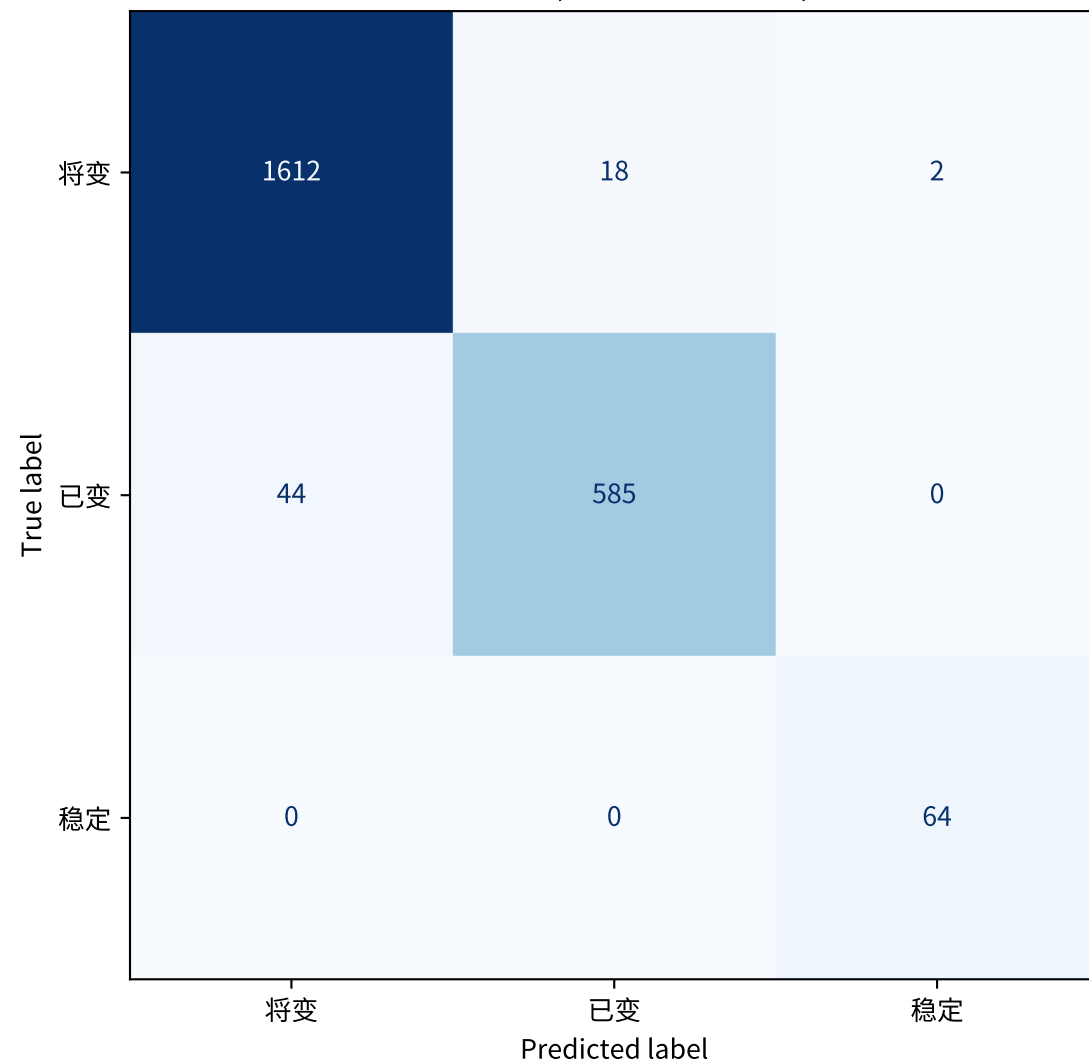


E3. 决策树可视化 (max_depth=3)

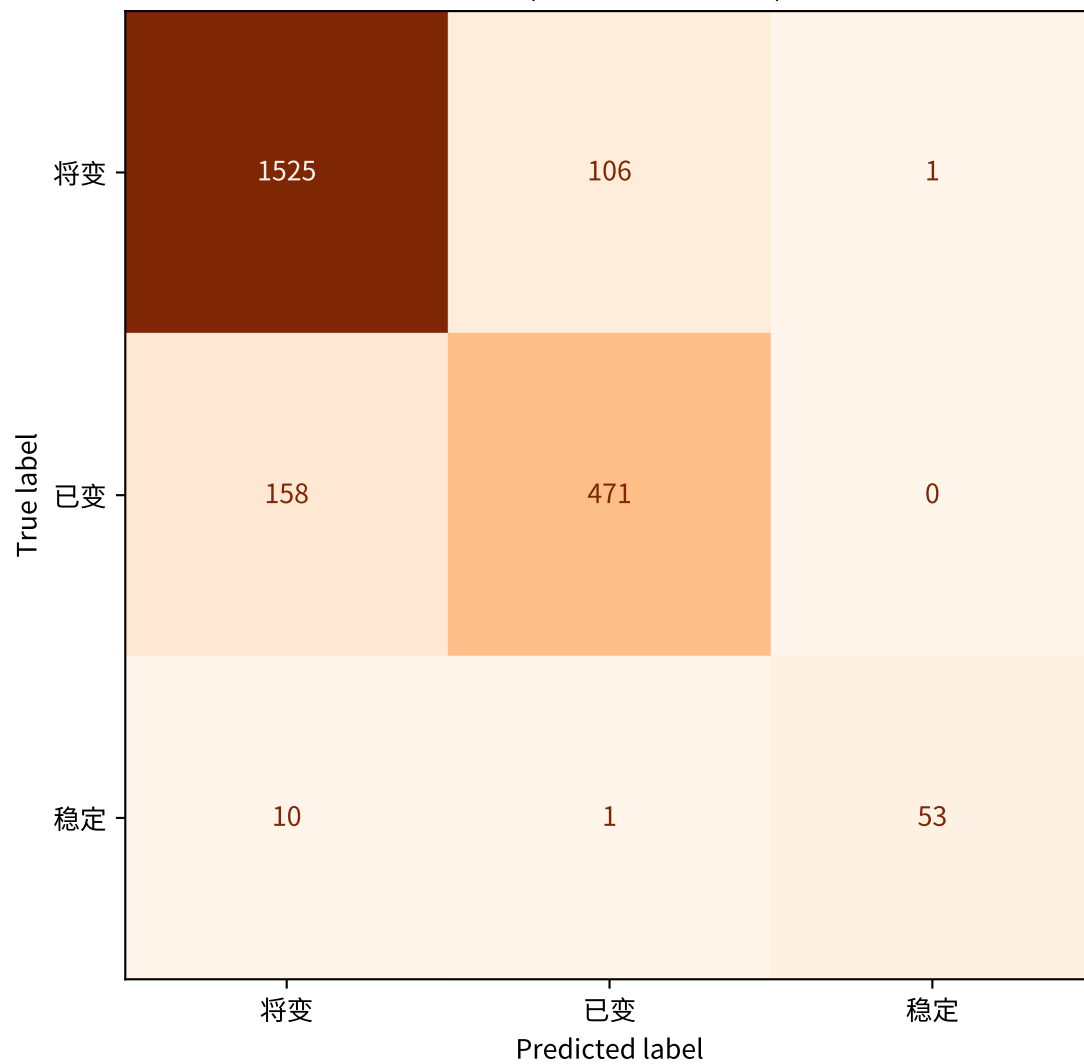


E4. 混淆矩阵 (训练集)

随机森林 (train acc=97.2%)

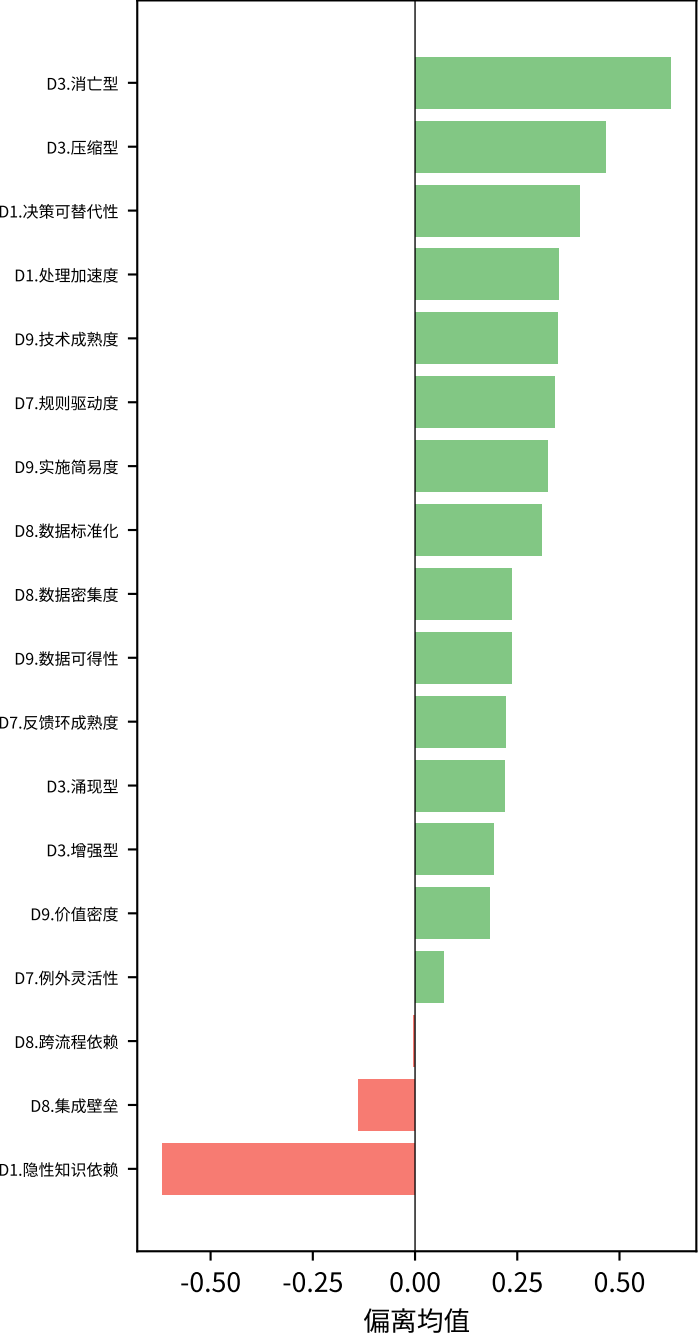


GBDT (train acc=88.1%)

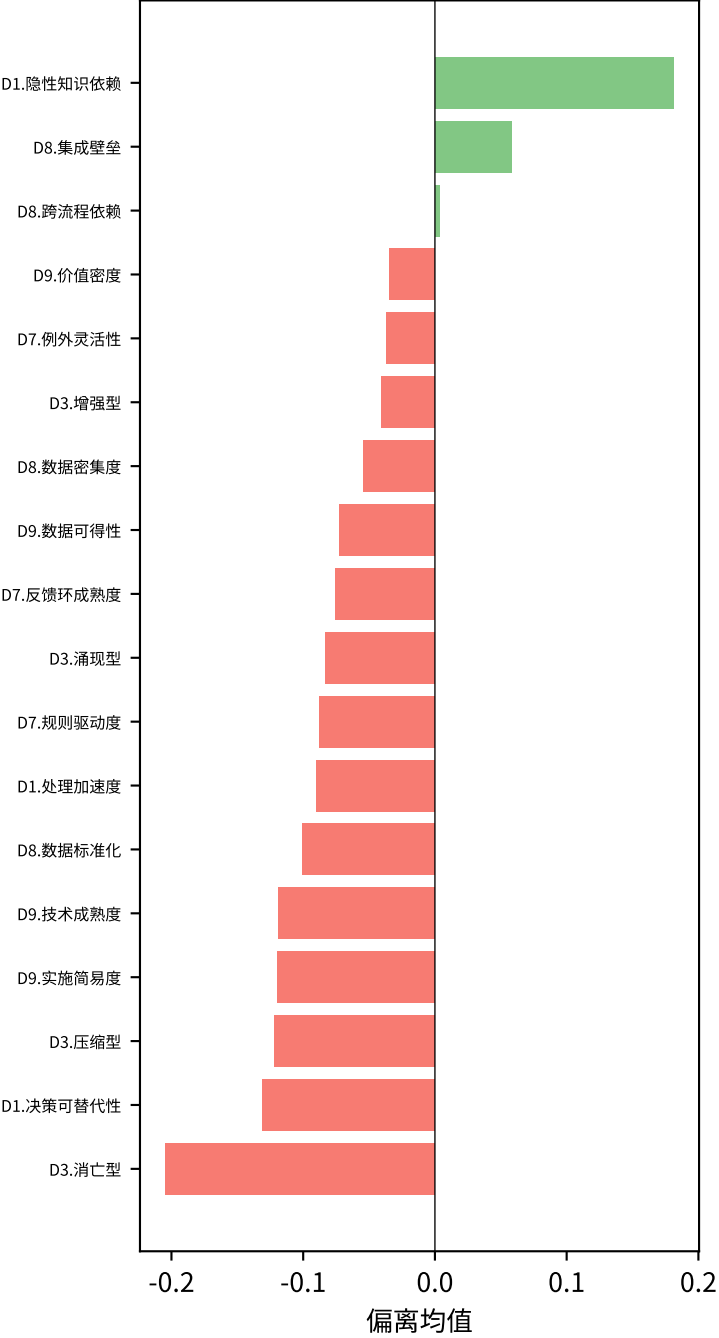


E5. 三类变化状态的特征差异

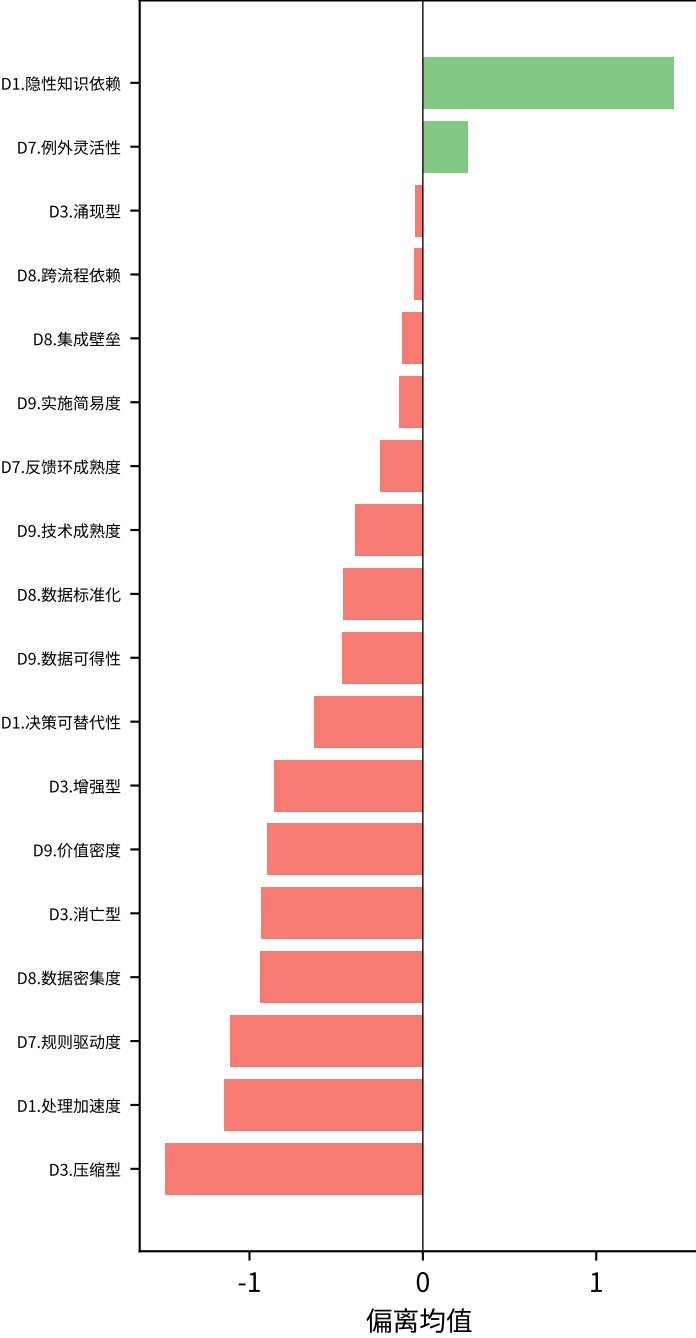
已变 (N=629)



将变 (N=1632)



稳定 (N=64)

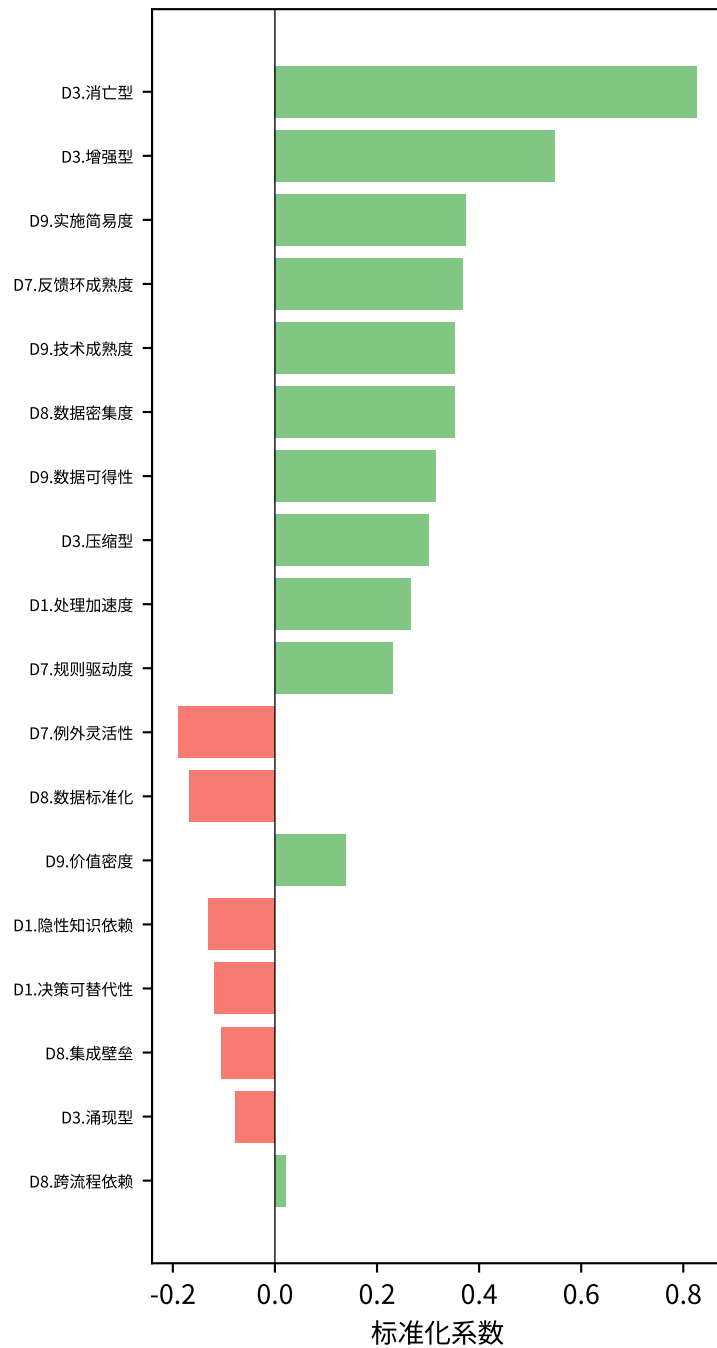


E6. 逻辑回归系数

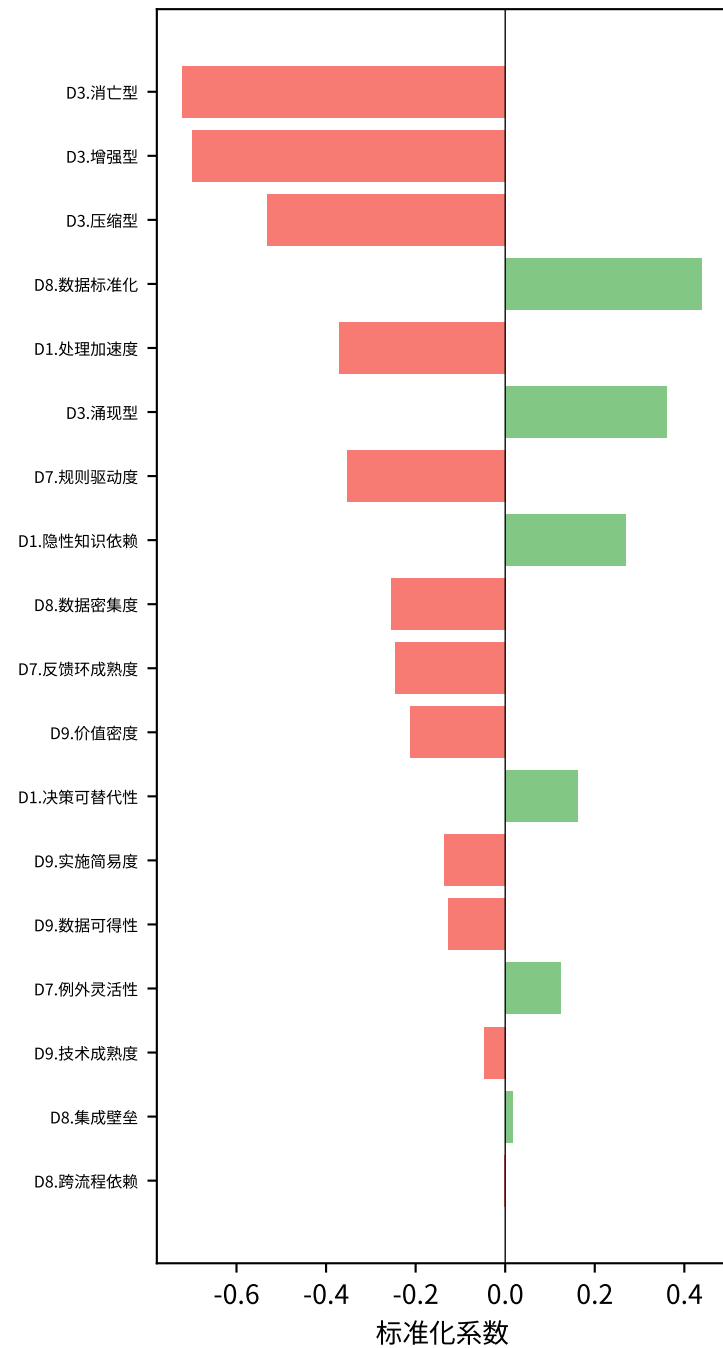
→ “将变”



→ “已变”



→ “稳定”



E7. 预测建模关键发现

1. 预测能力：

随机森林 5-fold CV = 79.2% \pm 1.1%

GBDT 5-fold CV = 82.4% \pm 1.8%

基线(全部预测“将变”) = 70.2%

→ 18 维度比随机猜测提升 9 个百分点

2. 最重要的预测因子(随机森林 Gini)：

#1 D3.消亡型 = 0.127

#2 D9.实施简易度 = 0.107

#3 D3.涌现型 = 0.085

#4 D1.隐性知识依赖 = 0.080

#5 D9.技术成熟度 = 0.070

3. “已变”流程的关键信号：

- D1.决策可替代性 高 + D1.隐性知识依赖 低
- D3.消亡型(type_c) 显著高于平均
- D7.规则驱动度 高 + D8.数据标准化 高

4. “稳定”流程的关键信号：

- D1.隐性知识依赖 极高(依赖人类直觉)
- D8.数据密集度 低(非数据驱动型)
- D3.涌现型(type_d) 偏高(因AI而生)

5. 模型局限性：

- 类别严重不平衡(将变70%:已变27%:稳定3%)
- 18维度均来自Gemini，存在系统性偏差风险
- 预测模型实质是在学习LLM的判断规律