

基于 LightGBM 分类的上市公司财务报表舞弊识别研究

答辩人: 会计 1610 王淙烨
指导老师: 孙景翠

目录

1. 背景介绍
2. 论文结构
3. 主要内容
4. 结论

研究背景

目的

- 定量分析
- 辅助判断
- 解释原因

意义

- 避免误导投资者
- 维护债权人利益
- 提高市场出清效率

创新点

- 样本容量大
- 解释变量多
- 模型效果好

论文结构

第一部分 前言

第二部分 相关概念综述

第三部分 样本选取与分析

第四部分 财务舞弊识别分类器构建

第五部分 结论

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ 🔍 ↻

模型结构

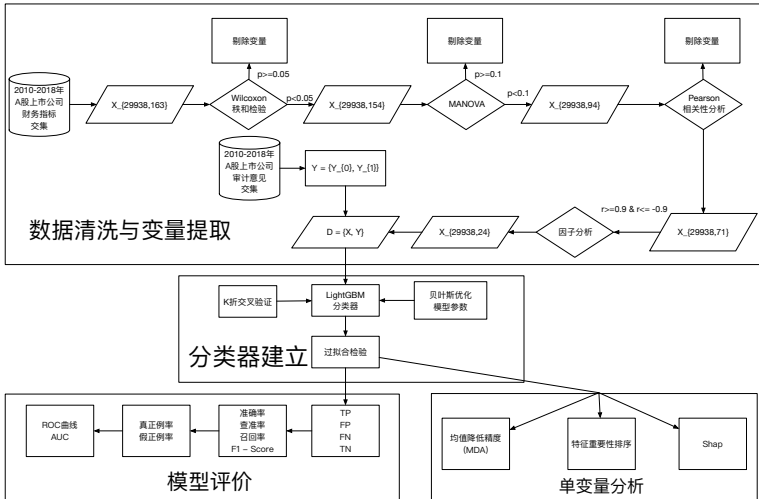


图: 模型结构

模型评价

表: 模型判别预测准确率

观测值	预测类型			准确率 (%)
	异常类型			
	$Y_i = 0$	$Y_i = 1$		
异常类型	$Y_i = 0$	TP=28947	FP=0	100
	$Y_i = 1$	FN=14	TN=977	98.59
整体准确率		TPR=99.95	FPR=0	99.95

模型评价

$$ACC = \frac{TP + TN}{P + N} = 99.95\%$$

$$PRE = \frac{TP}{TP + FP} = 100\%$$

$$REC = \frac{TP}{TP + FN} = 98.59\%$$

$$F_1 - Score = \frac{2}{\frac{1}{PRE} + \frac{1}{REC}} = 99.47\%$$

模型评价

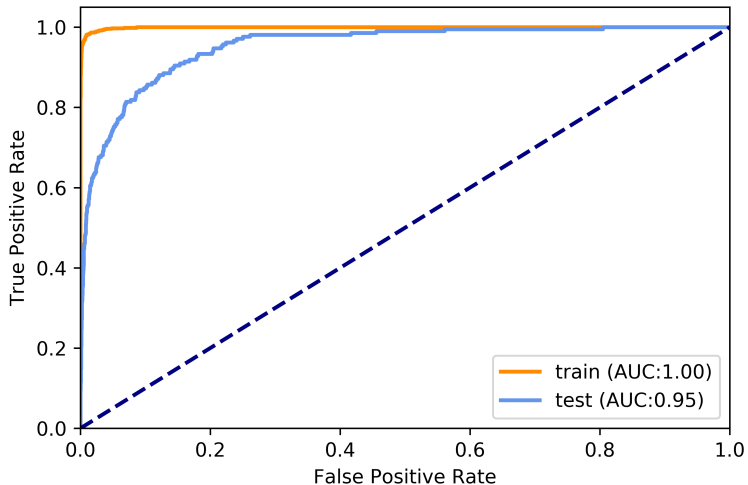


图: ROC-AUC 曲线

变量分析

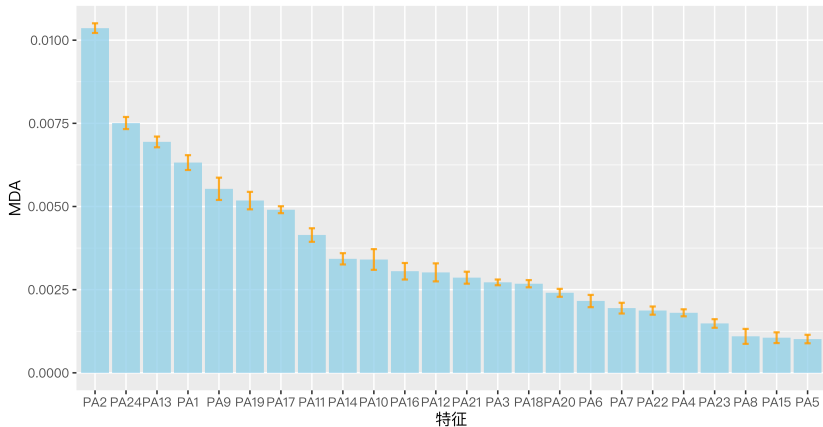


图: 均值降低精度

变量分析

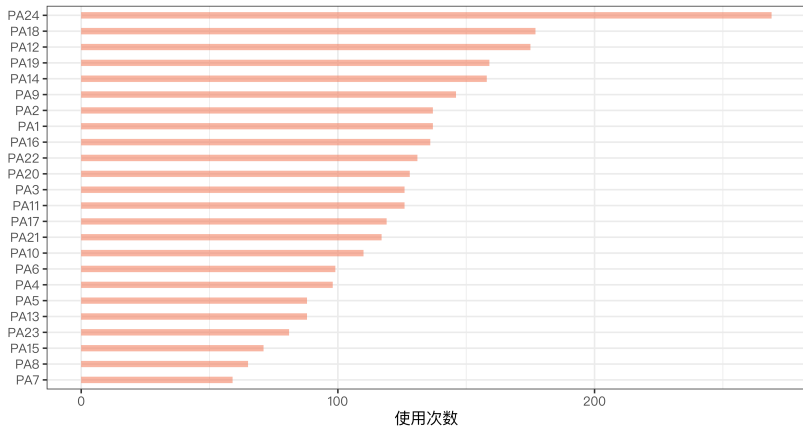


图: 特征重要性排序

变量分析

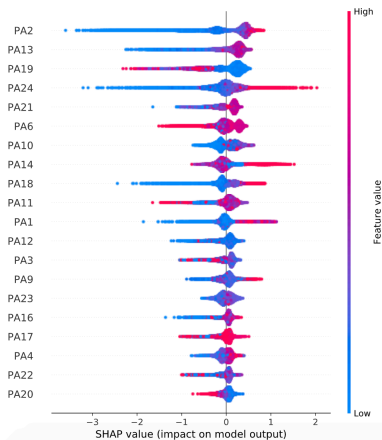


图: 特征 Shap 值分布

结论

- 舞弊强行业聚集性
- 企业现金流 & 每股息税前利润
- 模型简单 & 识别效果好

谢 谢