

罗柏荣, 周杰

算法工程师, 数据开发工程师
兴业数字金融

基于机器学习的欺诈风险识别

- 近年来，跨境赌博、电信网络诈骗、黑产等外部欺诈违法犯罪形势日益严峻，呈现线上化、产业化、团伙化等特征，国家层面高度重视反欺诈治理工作，执法和监管机构对反欺诈管理要求日趋严格；
- 此背景下，兴业银行逐步构建和完善符合电信网络诈骗活动特征的异常账户和可疑交易的监测机制；
- 数据仓库应用研发团队从业务场景出发，基于英特尔 oneAPI 工具进行数据预处理和建模分析，构建涉赌涉诈账户算法识别模型，助力排查存量风险。本方案具有数据加工效率高和模型分类效果好的优点。



罗柏荣毕业于华侨大学，统计学硕士，热爱数据挖掘竞赛，在传统机器学习的特征工程模块有比较丰富实战经验，曾经获得过2022科大讯飞人岗匹配挑战赛的冠军以及2022联合利华新品市场销量预测冠军。



周杰毕业于上海对外经贸大学应用统计专硕。毕业后一直从事于数据类的相关工作，2022年在曼伦商贸有限公司担任数据分析师一职，期间获得高级商业数据分析师证书。

英特尔® oneAPI 人工智能黑客松大赛

团队名称：数金之光

成员：彭广亮（数据仓库团队负责人）、金轩亦（数据开发工程师）、周杰（数据开发工程师）王怡丁（数据分析师）、罗柏荣（算法工程师）

所选赛道：赛道三

所在公司：兴业数字金融服务（上海）股份有限公司

参赛经历：2023兴业银行“安康杯”反欺诈识别算法大赛三等奖

2022联合利华新品市场销量预测全国冠军

2022科大讯飞人岗匹配挑战赛冠军

2021科大讯飞家庭环境识别挑战赛季军

2021字节跳动安全AI挑战赛最具潜力奖



intel®

方案简介

获奖证书



- **赛道三：基于机器学习的欺诈风险识别**

- **方案背景：**

近年来，跨境赌博、电信网络诈骗、黑产等外部欺诈违法犯罪形势日益严峻，呈现线上化、产业化、团伙化等特征，国家层面高度重视反欺诈治理工作，执法和监管机构对反欺诈管理要求日趋严格。党的二十大的报告中发布了“以人民为中心”“统筹发展与安全”的重要指示。银行业金融机构需筑牢反欺诈“防火墙”，为业务的健康发展保驾护航；践行金融为民，坚决守护人民群众的财产安全，提高人民群众的幸福感和获得感；牢牢把握金融工作的政治性、人民性，提升专业性。同时，监管机构要求各家银行开发和运用有效的金融科技和大数据系统，分析涉赌涉诈资金交易新方式，总结资金风险新特征，持续优化风险监测模型，全面排查存量风险。于2022年12月正式施行的《中华人民共和国反电信网络诈骗法》把反欺诈工作上升到了法律层面。该法案亦指出，银行业金融机构应当建立完善符合电信网络诈骗活动特征的异常账户和可疑交易的监测机制。

- **数据说明：**

兴业银行部分账户的交易流水信息表（百万级别）、账户静态信息表和涉诈黑名单，数据经过脱敏处理。

- **方案简介：**

从业务场景出发，基于英特尔OneAPI工具进行数据预处理和建模分析，构建涉赌涉诈账户算法识别模型，全面排查存量风险。本方案具有数据加工效率高和模型分类效果好的优点，并在2023年兴业银行“安康杯”涉诈涉赌账户识别算法竞赛中荣获三等奖的佳绩。



intel®

工具使用



oneAPI工具使用情况

oneAPI工具	功能描述
Intel® Distribution of Modin*	数据读取和数据加工
Intel® Extension for Scikit-learn*	数据加工和建模准备
XGBoost Optimized for Intel® Architectures	模型训练和模型推理



建模流程

建模流程图



特征工程

优化技术1：围绕关键信息进行特征构建和加工

基础特征	特征加工说明
对方账号	计算用户的交易对方账户个数、平均交易次数以及进行频率编码
对方行号	计算用户的交易对方行号个数、平均交易次数以及进行频率编码
交易渠道	计算用户的交易渠道个数、平均交易次数以及进行频率编码
交易代码	计算用户的交易代码个数、平均交易次数以及进行频率编码
交易金额	进行均值、标准差等统计特征计算
交易时间	计算一阶差分并围绕一阶差分进行均值、标准差等统计特征计算
	等等.....

Modin

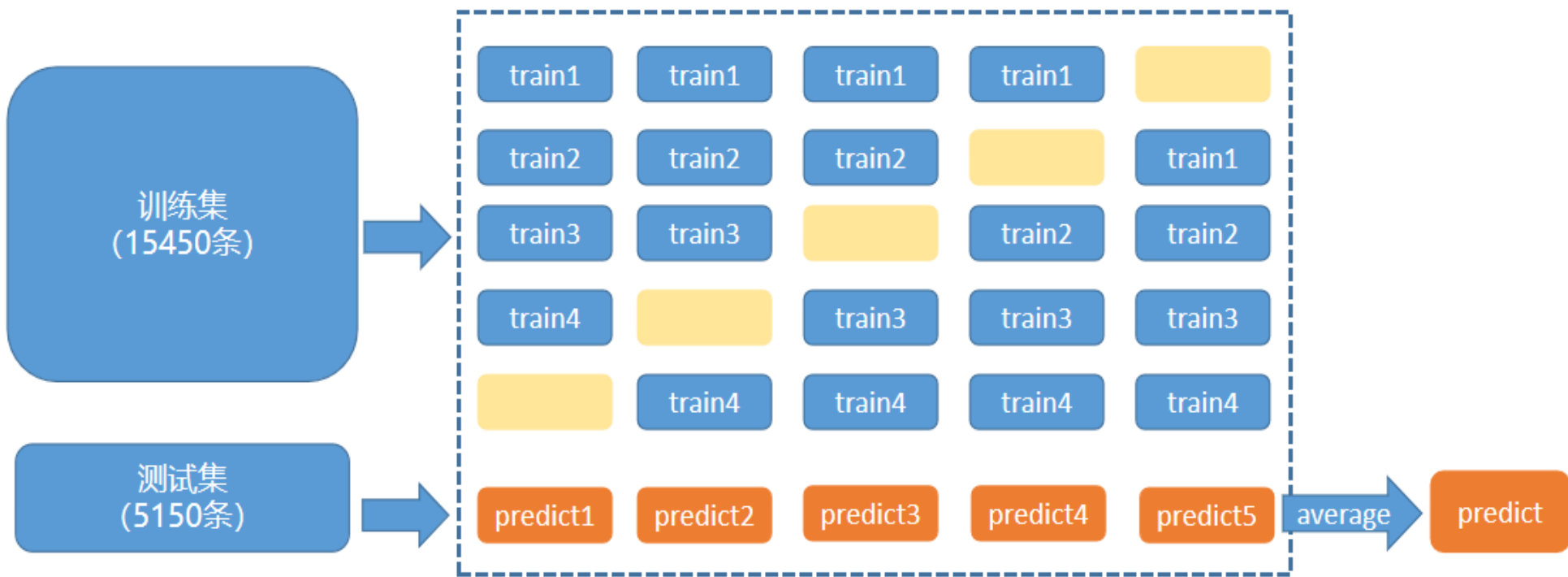
Pandas 是一个功能强大的数据处理工具，但在面对大型数据集和复杂加工运算时，其性能可能变得较慢。Modin 致力于解决这个问题，使得在大数据处理方面能够更加高效和快速。以下常用的 pandas (DataFrame) 的API可以得到Modin的加速优化。

API	备注
apply	
diff	差分
drop_duplicates	去重
groupby	聚合计算
join	连接
shift	
rolling	滑动窗口函数

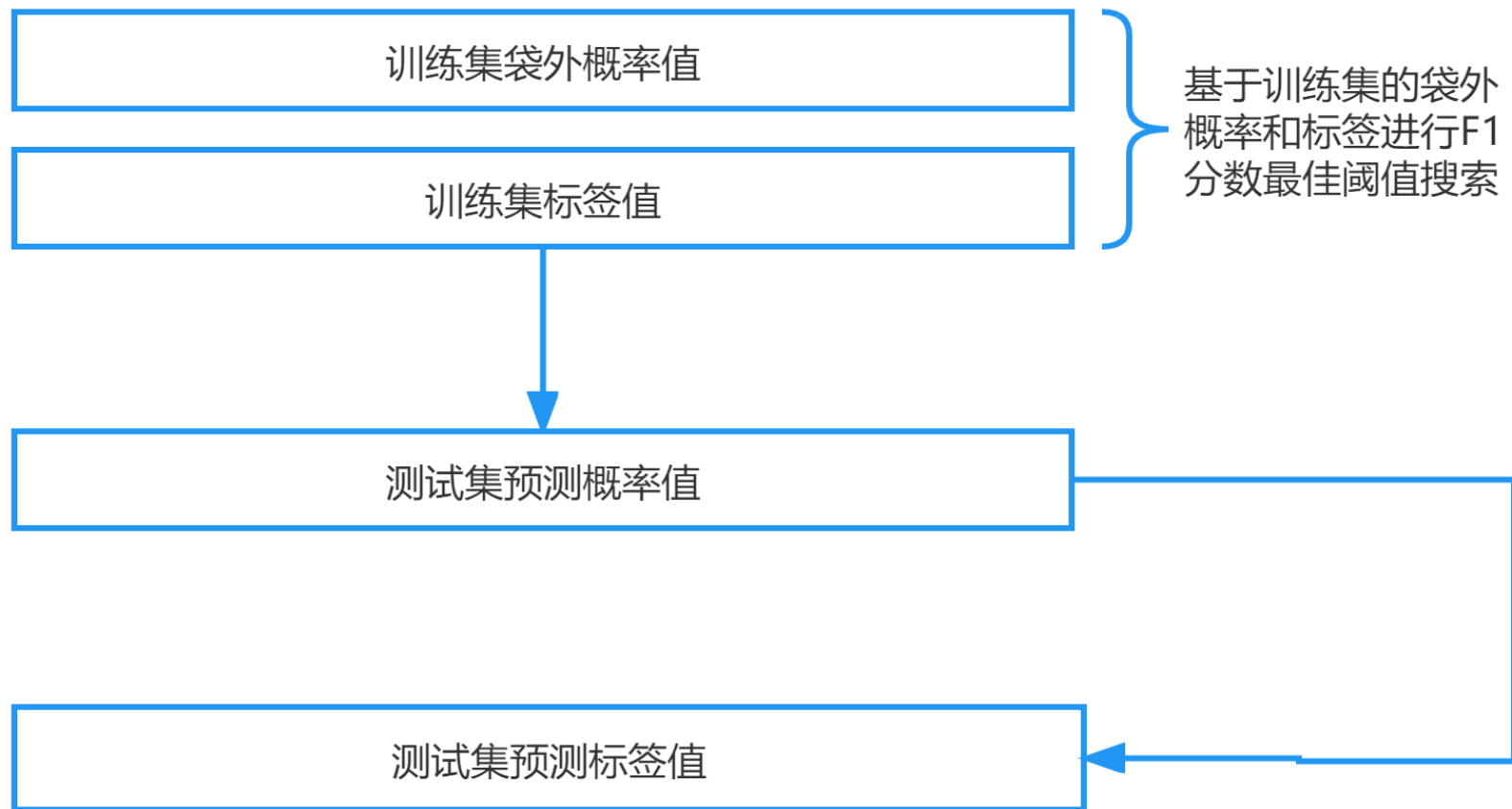


建模预测

优化技术2: 模型5折交叉训练与平权预测



建模预测



模型对比



优化技术2: 模型5折交叉训练与平权预测

96.85%

97.12%

97.99%

F1-Score



结果展示

加工效率与模型精度分析

建模环节	数据读取和数据加工	模型精度
优化前	145.37秒	97.99%(F1)
优化后	115.78秒	

经过英特尔Modin*工具优化，数据读取与数据加工时间减少约30s，效率提高20.68%。同时，XGBoost模型可在3min26s内实现15450条数据的五折交叉训练，并在测试集取得高达97.99%的F1分数，兼顾训练速度与分类精度。更多建模细节详见附件。

intel®