

实时反欺诈风控体系

🔒 脱敏示意样本

Real-time Anti-Fraud · 全自动投放闭环的入口质量闸 · 风控策略分析师(AI Staff) 出品

出品	风控策略分析师 · AI Staff	版本	v2.1 · 演示
定位	全自动投放闭环的入口质量闸:拦截诈 + 回灌渠道质量信号	状态	方案定稿 / 待评审
数据口径	脱敏演示数据(信号与数值均为示意)	保密	公开可见样本

01 背景与目标 · 在全自动投放闭环中的位置

本命题是**全自动投放**:投放系统按「渠道 → 注册 → 授信 → 放款 → 回款」的真实结果自动分配预算。反欺诈是这条闭环的**入口质量闸**——它决定「投进来的量干不干净」。互金欺诈已**团伙化、工具化、规模化**:设备农场、群控改机、身份冒用、中介养号、多头套现,靠黑名单 + 人工的事后拦截,响应慢、打不住团伙、还误杀好客户。

所以反欺诈是**双目标**:①毫秒级拦住欺诈申请,压欺诈坏账;②把**渠道级质量信号**(每个投放渠道来的量:欺诈率/授信通过率/团伙浓度)实时回传给投放,让投放掐掉劣质渠道、下调出价。第 ② 条是它区别于孤立反欺诈系统、真正接进 team work 的关键。

维度	目标(观察窗内)
识别能力	团伙/设备农场/身份冒用识别率显著提升,欺诈坏账占比下降
回灌投放	渠道级质量信号实时回传投放,劣质渠道自动降预算/熔断
时延	实时决策毫秒级,不影响正常用户体验
误杀	好客户误杀率可控,配套申诉复核回路
非目标	不替代信用风险模型(反欺诈解决「是不是坏人」,信用模型解决「还还得起」)

02 接入全自动投放:渠道级质量回灌

反欺诈不止拦单个申请——它给投放的每个渠道打质量分,让「自动出价」吃得到风险信号,这是它在 team work 里的角色。

回传给投放的信号	口径	投放据此的动作
渠道欺诈率	该渠道进件中确认欺诈占比	超阈 → 自动降预算 / 暂停该渠道投放
渠道授信通过率	该渠道进件的授信通过占比	偏低 → 判为劣质量,压该渠道出价
渠道团伙浓度	该渠道命中团伙/设备农场比例	集中 → 拉黑渠道 + 溯源投放来源
实时熔断信号	单渠道欺诈突刺	秒级触发投放侧该渠道熔断

为什么这条是闭环的命门

投放若只优化「注册最便宜」,黑产会把最便宜的欺诈量喂给你——自动化反而放大风险。只有把反欺诈的**渠道质量信号**并进出价目标,投放买的才是「能授信的真人」,而不是「便宜的量」。反欺诈不接投放,就只是个孤立的拦截器;接了,才是自动投放的进水口滤网。

03 实时决策架构

事件驱动:每一次关键动作都触发一次实时评分与处置,毫秒级返回。

层	职责	输入 → 输出
事件接入	申请/登录/绑卡/支用等事件实时接入	原始事件 → 标准事件流
实时特征	设备/行为/身份/关系/历史 五维特征在线计算	事件 → 特征向量
模型层	欺诈评分模型 + 图神经网络团伙识别	特征 → 欺诈分 + 团伙标记
决策层	按分数与团伙标记分层处置	评分 → 放行/加验/降额/拒绝/人工
反馈回流	确认欺诈/误杀申诉回灌 + 渠道质量回传投放	处置结果 → 模型迭代 & 投放调价

04 信号与特征体系

维度	代表信号	识别的欺诈形态
设备	设备指纹、模拟器/改机、群控一致性、IP 聚集	设备农场、群控批量作案
身份	活体、OCR 一致性、多头身份交叉、证件复用	身份冒用、盗用他人资料
行为	填写速度、复制粘贴、操作序列异常、夜间集中	脚本自动化、代填中介
关系	图谱共设备/共联系人/共收款/资金闭环	中介团伙、养号产业链
历史	历史欺诈关联、黑样本邻居度	惯犯、黑产复用资源

05 团伙识别(图计算)

把用户、设备、联系人、收款账户、WiFi/IP 建成**关系图谱**,用社区发现 + 图神经网络找出高密度可疑子图——共享设备/中介代填/资金回流闭环,一次性圈出整个团伙,而不是逐个拦。

为什么单点规则打不住团伙

单点规则看的是「这个人像不像坏人」,团伙作案每个个体都可能「看起来正常」,是**关系结构**暴露了它们——同一批设备、同一个中介、资金最终流向同一处。不上图计算,就只能等团伙把额度撸完之后在坏账里复盘。

06 决策与处置分层

欺诈评分	团伙标记	处置动作	加验方式
低	无	直通放行	—
中	无	加验后放行	活体 / 短信 / 二次核身
中	命中团伙	降额 + 加验或转人工	人工外呼核实
高	—	拒绝	—(记录并回流样本)

07 误杀与申诉护栏

反欺诈的反向成本是**误杀好客户**——拒绝一个真实优质用户,损失的是终身价值和口碑。体系必须自带纠偏:

护栏	机制
申诉复核	被拒用户可申诉,进人工/二次模型复核回路
白名单	历史优质、已核验用户降低误伤概率
阈值 A/B	分段灰度调阈值,监控误杀率与漏杀率的平衡点
误杀回流	确认误杀的样本回灌,修正模型

08 关键指标与验收

指标	类型	口径
欺诈识别率	效果	确认欺诈中被拦截占比
团伙一网打尽率	效果	团伙成员被同批识别占比
渠道质量信号覆盖	闭环	接入投放的渠道中有实时质量回传的占比
好客户误杀率	护栏	优质用户被误拒占比(越低越好)
决策时延	体验	事件到处置返回时间(毫秒级)
欺诈坏账占比	北极星	坏账中欺诈成因占比下降

09 风险与预案

风险	预案
黑产对抗升级(绕过特征)	持续对抗迭代 + 新特征快速上线,保留人工研判回路
模型漂移	日度校准 + 回滚点,漂移超阈自动回退
误杀引发投诉	申诉回路 + 阈值收敛,优先保护存量优质客户
数据源断流	降级到保守规则策略,断流期收紧直通

附:脱敏说明 本文为对外展示的脱敏示意样本,其中主体名称、比例、金额、通过率、指标等均为演示数据,不对应任何真实客户或真实业务结果。实际项目的口径、数值与结论以对应负责人与客户共同确认为准;相关业务与决策权始终在客户与相应负责人,AI 只做辅助与留痕。