欺诈风险是借款人恶意利用金融规则的漏洞以非法占有为目的，采用虚构事实或者隐瞒事实真相的方法，骗取借款的风险。

反欺诈解决方案经历了从简单的黑名单规则，到反欺诈规则引擎，再到有监督的机器学习，再到无监督的大数据欺诈检测，而其数据特征提取依然是反欺诈能力的基础。

总结一些常用的反欺诈数据特征类型，使用该类的欺诈特征数据，可以进一步搭建反欺诈模型：或得到欺诈评分，或得到欺诈规则。

1.用户身份信息交叉验证规则

验证类型

输出结果

银行卡四要素

手机号码、银行卡、姓名、身份证号一致/不一致

银行卡三要素

银行卡、姓名、身份证号一致/不一致

手机号码三要素

身份证号、姓名、手机号码一致/不一致

二要素身份认证

姓名、身份证号一致/不一致

人像比对&人脸识别

图像比对是否一致，活体检测等

2.用户手机号及运营商数据

2.1 手机号码特征

· 手机号前缀是否相同

· 手机号归属地是否相同

· 是否是虚拟运营商

· 流量卡 or 通话卡

· 手机号码注册多平台验证

2.2 运营商数据

· **运营商数据匹配性：**手机通讯录联系人电话与运营商通话记录联系人电话匹配度

· **申请人手机有效性：**申请手机运营商状态异常、接听个人手机次数占联系手机总次数比重较低、关机时间过长、月均被叫次数增长率过高、当月主叫次数过低、电话使用时长（月）、申请手机月均账单异常

· **申请人通讯录关联情况：**申请人通讯录高度重合（申请手机通讯录名单和最近 N 个月内其他申请手机通讯录重合度≥70%）

· **申请人注册手机与异常号码通话情况：**例如过去 N 个月与贷款类号码话大于等于 X 次

· **手机号码在网时长**

2.3 联系人信息交叉验证

· **联系人有效性（非真实）：**联系人手机不在通讯录内、联系人号码近 N 天无>30s 的主叫通话记录

3.用户基本信息特征交叉验证

· **文字类信息关联类对比：**正则、字节拆分、关键字提取、相似度计算、模糊匹配、错别字/同音字识别等方法

字段

衍生特征（频度、关联性）

昵称

昵称符合固定的规律（中文+数字）

出生日期

年纪、星座、生肖

性别

同一时段申请人的性别是否失衡（集中为男性或女性）

密码

同时段用户设置的密码是相似程度

邮箱

是否是一次性邮箱,邮箱名是否满足特定规律、是否同一邮箱服务商

地址信息

工作地址，住房及租房地址是否雷同

公司名称

是否雷同，或关联逾期客户

公司电话

是否雷同，或关联逾期客户（公司电话异常关联：同公司名称有逾期 X 天以上客户数≥N；同公司电话最近 N 天申请人≥X）

联系人信息及电话

名字及手机号是否关联多个账户

4.设备环境数据

数据

详细（用于交叉验证关联性）

设备类

设备(名称、品牌）

手机品牌、手机型号是否相同

操作系统

每次打开操作系统是否都相同

操作系统是否都相同

版本是否太旧

设备指纹

设备 imei 号是否关联多个账户

MAC 地址信息是否关联多个账户

设备 ID

每次登录 device id 号是否都相同

是否使用模拟器

屏幕分辨率信息

手机型号和屏幕分辨率是否匹配

app 列表

是否含有多个借贷软件

环境类

IP 地址

IP 网络类型及运营商

IP 粗略/精确地理位置

IP 是否是同一个号段

每次登录 ip 地址是否相同

IP 异常：是否境外 IP，是否 3G/4G 等基站类 IP

代理信息

每次打开是否是同一个 user agent

GPS 信息

交叉验证经纬度相似性分析如 ip 和 gps 是否能对的上

基站定位

根据基站编号查询对应的基站地理位置经纬度

WIFI 定位

根据 WIFI 的 BSSID 查询 WIFI 热点的地理位置经纬度

wifi list 贷款前的几分钟有没有切换过 wifi

渠道 ID

Ssid 渠道 ID 属于违规渠道

app 版本

是否有可能利用老版本的 APP 的 bug 做攻击

可以运用设备及环境信息的交叉验证例如：

· **同硬件设备网络聚集性**：IP 近 X 天聚集多个设备值大于一定数量

· **设备异常操作：**使用代理注册、设备相邻两次注册时间间隔极短、设备或平台账号短时间移动的位置距离异常

· **设备指纹关联风险（机构代办风险）：**7 天内设备上提交借款的个人信息极多（例如：7 天内同一设备上提交借款的个人信息极多≥3）

设备信息核查规则：设备关联手机数较多

· **多次尝试：**失败后变更环境尝试--同一账户最近 4 小时关联 ip 数≥2

· **借款反欺诈规则-设备异常操作：**使用代理借款、登录与借款间隔时间极短、设备借款次数过多

5. 用户行为数据

数据类型

详细

注册、申请、时间

申请时间在凌晨 1-5 点等

注册申请信息行为数据

注册用了多长时间，文本输入时长

键盘敲击时长

注册、登录、集中频次

一共申请了几次

同一时间登录做一个校验（同一时间多人登录）

6. 其他第三方数据

包括但不限于用户的电商消费数据、社交网络数据、银行及信用卡数据、司法数据等等

数据类型

详细

学历信息授权采集

学历信息字段

人行征信授权采集

公司信息是否一致

学历是否一致

居住地址是否一致

手机号码是否一致

逾期数据

各类黑、灰名单

设备欺诈库，IP 欺诈库，身份欺诈库电话欺诈库、司法不良名单

银行卡授权采集

半年银行卡流水

关系网络数据

查询个人 QQ 是否加入过异常 QQ 群组

微信是否关联异常用户

1 度关系或二度关系（直系或非直系亲属）命中其他机构逾期账龄或命中黑中介社交网

消费收支数据

线上电商和线下消费、银联消费、银行卡收支、航旅出行数据等

以及工资、社保、公积金等数据

多头借贷数据

近 3 个月用户（身份证号、手机号等基本信息关联在）在其他借贷机构申请次数

以上仅仅是例举部分反欺诈特征数据，当然不同的金融应用场景，有着不同的业务流程和环节，需要设计不同的风险检查环节和风控策略，构建基于场景、事件和规则驱动下的欺诈风险判别功能，