

# 证券期货业用户事件标签标准研究

## 【摘要】

本文介绍了在证券期货行业中建立用户行为埋点统计规范，并建立标准化的用户事件规范和 one ID 原则，以建设一个多维度的画像标签体系。在用户账户统一识别方面，通过将设备 ID 和登录 ID 关联，分配唯一的用户 ID，实现跨设备、登录账户的用户行为无缝追踪。在终端事件体系方面，通过埋点设计原则将用户行为划分为启动、曝光、点击、浏览和停留，并设计了与之对应的五个事件。在全生命周期事件标签体系方面，建立了四大层级结构化的方式，以业务视角构建多维度的标签，实现对用户各类事件行为的分析。最后，提出了标签安全体系的设计，对标签数据进行分类，并确定了不同标签的安全等级。

## 关键词：

用户行为埋点；终端事件；画像标签

## 正 文

### 一、多重账户统一识别

准确地标识用户和建立用户关联是用户行为分析的基础，同时也是用户账户统一识别的基础，任何细微的遗漏都可能对后续的统计和分析结果造成质疑。传统分析通过用户的设备信息或账号信息对用户进行标识，但由于移动终端迭代及证券期货行业存在的多重账户情况，传统分析无法准确识别用户。因此，需对用户设备信息及账号信息进行体系化智能化的研究，实现精准识别及有效融合，从而实现对用户的正确识别。

在真实世界中，我们通常通过身份证号来准确标识一个自然人，但在互联网业务中，这种方式已经不再适用。因此，出现了两种常见的方案。第一种是通过设备 ID 来唯一标识用户，这在一定程度上解决了用户标识的问题，但也存在缺陷。例如，同一台手机被多个用户使用，它们的行为被标记为同一个人；而老用户更换新手机后也会被识别为一个全新的用户。第二种方案是通过用户账号或客户号来识别用户，这种方案在业务后台系统中比较常见，但无法识别未登录状态下的用户行为，因此只能准确记录业务数据，在业务数据仓库的构建中充当用户标识的角色。

综上所述我们选择了关联设备 ID 和登录 ID 的方案，在一定程度上融合设备 ID 和登录 ID 来实现更准确的用户追踪。通过将设备 ID 和登陆 ID 关联起来，分配一个唯一的用户 ID，建立设备 ID、登录 ID 和用户 ID 一一对应的关系，成功关联

设备 ID 和登录 ID 之后，用户在该设备 ID 下或该登录 ID 下的行为就会贯通，被认为是一个用户 ID 发生的，在进行事件、漏斗、留存等用户相关分析时也会算作一个用户。该方案实现贯通一个用户在一个设备上注册前后的行为，同时也贯通一个注册用户在不同设备上登录之后的行为。

具体实现方案如下：

维护一张用户表(users),用于记录设备 ID(first\_id)、登陆 ID(second\_id)和用户 ID(id)的绑定关系。在事件表(events)维护 distinct\_id(缓存 ID)字段和 user\_id(用户 ID)字段，确保通过 user\_id 能查找到 users 表里的 id，映射到具体的设备 ID 和登录 ID，以下将通过具体场景说明改方案的实现逻辑：

events 表	users 表
device_id (事件设备 id)	first_id (设备 ID)
distinct_id (缓存 ID)	second_id (登陆 ID)
user_id (用户 ID)	id (用户 ID)

场景 1：某用户在手机上新安装了某券商 App，并进行了一系列操作，SDK 生成的设备 ID 为 X，distinct\_id 记为 X，对应分配的用户 ID 为 1。

events 表		users 表	
device_id	X	first_id(设备 ID)	X

distinct_id	X	second_id (登陆 ID)	为空, 因为没有登陆
user_id	1	id (用户 ID)	1

场景 2: 该用户进行了注册并登录, 登录之后继续进行一系列操作, 其登录 ID 为 A, 此时设备 ID X 和登录 ID A 关联成功, 并将登录 ID A 存入 users 表的 second\_id 字段, 用户 ID 仍为 1。该用户登录之后继续进行一系列操作, 发送的 distinct\_id 为 A。

events 表		users 表	
device_id	X	first_id(设备 ID)	X
distinct_id	A	second_id (登陆 ID)	A
user_id	1	id (用户 ID)	1

场景 3: 该用户把手机送给朋友了, 朋友用自己的账号 B 登录设备 X, 登录 ID 为 B, 此时尝试将设备 ID X 与登录 ID B 进行关联, 由于 X 已与 A 关联, 所以此次关联失败, 同时会分配一个新的用户 ID 2 来标识此用户, 并将登录 ID B 同时存入 users 表的 first\_id 和 second\_id 字段。该用户的朋友一直使用账号 B 在设备 X 上进行了一系列操作, 发送的 distinct\_id 为 B (用户的朋友账号上之前未关联过别的设备, 且首次登录设备关联失败, 则将登录 ID 同时记录到 first\_id 上)

events 表		users 表	
device_id	X	first_id(设备 ID)	B
distinct_id	B	second_id (登陆 ID)	B
user_id	2	id (用户 ID)	2

场景 4: 该用户更换了一个新的手机, 并进行一系列操作, 由于尚未登录, 此时使用全新的设备 ID Y 来标识此设备, 发送的 distinct\_id 为 Y, 对应分配的用户 ID 为 3。

events 表		users 表	
device_id	Y	first_id(设备 ID)	Y
distinct_id	Y	second_id (登陆 ID)	为空
user_id	3	id (用户 ID)	3

场景 5: 该用户在该手机上使用原先的账号 A 进行登录, 此时将尝试将设备 ID Y 与登录 ID A 进行关联, 由于 A 已经与 X 关联, 因此会关联失败, 但是依然会切换到以 A 为登录 ID 的用户, 其对应的用户 ID 依然为 1。该用户登录之后的后续操作, 发送的 distinct\_id 为 A, 所以仍以用户 ID 1 标识。且该用户即使在该手机上退出登录 ID A, distinct\_id(缓存 ID) 仍会上报为 A, 直到该设备登录其他账号。

events 表		users 表	
device_id	Y	first_id(设备 ID)	X

distinct_id	A	second_id (登陆 ID)	A
user_id	1	id (用户 ID)	1

综合上述方案，通过关联设备 ID 和登录 ID，我们实现了跨设备、登录账户的用户行为的无缝追踪，这种方法为我们提供了更全面、更准确的数据洞察力，有助于深化对用户行为的理解，进一步优化产品和服务，为进一步的埋点标签标准化建设打下基础。

## 二、标准化的终端事件体系

证券期货相关软件具有业务复杂性高，页面功能丰富，产品更新迭代快等特点，为实现对用户各类事件行为客观、准确地分析，应立足证券期货业务特性，从动作类型，页面位置，业务分类及具体功能展位等多维度对用户行为进行界定，形成事件埋点统计规范。为了标准标准化的终端事件，我们建立了四个埋点设计原则：

1、埋点事件抽象原则：相同用户行为抽象为同一事件，如：不同模块下的点击应上报为同一点击事件，通过属性区分，而不是上报成行情点击事件，资讯点击事件等。

2、事件属性设计原则：相同含义的属性应复用同一命名，降低理解和查询成本，方便后续可以多维度对比分析。如：资讯文章的作者和短视频作者应统一为 author，

而不能一个为 author，一个为 writer。

### 3、信息采集遵循 4W1H 原则：

WHO：记录行为主体，谁发起动作，例如谁在点击，谁在浏览；

WHAT：记录动作对象，例如点击的是什么按钮，浏览的是什么页面；

WHEN：记录动作时间，例如在哪一时刻点击的，浏览了多久；

WHERE：记录所处环境，例如按钮所在页面，文章所在模块，端口是 H5 还是 APP，操作系统等；

HOW：记录来源，怎么到达这一步，例如是通过行情进入的文章详情，还是通过资讯进入的文章详情等。

4、 新增原则：新增事件或新增属性时，应优先复用已有事件和属性，如不能满足需求，才提新增事件或属性的需求，以确保事件和属性数量收敛。

在这四个设计原则下，我们将用户行为划分为五个动作：分别是启动，曝光，点击，浏览和停留，并设计了五个事件：APP 启动事件、元素曝光事件、用户点击事件、用户浏览事件和用户停留事件与之一一对应。同时规范了每个事件需要上报的属性以记录每个用户动作的主体

(who)、动作对象(what)、时间(when)、动作环境(when)、来源 (how)

事件	动作时间	动作主体	动作对象	动作环境	来源	公共属性
用户点击事件 (用户点击时上报)	记录发生点击的时间	记录用户信息, 如客户编号、广发通、ip地址等	记录点击对象的属性, 如点击的按钮名称, 点击的产品id, 标题, 产品分类等	记录所在页面内容的属性, 如当前页面文章 ID, 标题, 文章所属内容分类, 文章作者等	记录当前页面上步来源	记录设备号, app 版本, 手机型号, 操作系统等公共信息
用户浏览事件 (页面加载时上报)	记录发生浏览的时间		记录浏览页面的属性, 如文章 ID, 视频标题, 文章所属内容分类, 文章作者等	对于浏览事件, 动作对象也是动作环境	记录当前页面上步来源	
元素曝光事件 (元素展示时上报)	记录发生曝光的时间		记录被曝光对象的属性, 如视频 ID, 标题, 视频所属内容分类, 视频作者等	记录所在页面内容的属性, 如当前页面文章 ID, 标题, 文章所属内容分类, 文章作者等	记录当前页面上步来源	
用户停留事件 (页面退出时上报)	记录停留结束的时间 记录停留的时长		记录停留页面的属性, 如文章 ID, 标题, 文章所属内容分类, 文章作者等	对于停留事件, 动作对象也是动作环境	记录当前页面上步来源	
APP 启动事件 (APP 启动时上报)	记录 APP 启动的事件		无需	无需	无需	

定义好五个基础事件之后, 我们设计了一套属性库,



用于统一上报规范。

中文名称	英文名称	说明	类型
动作对象描述			
点击类型	click_type	记录点击类型或含义,如点赞、关注等	STRING
点击按钮名称	click_button_name	记录按钮上文字	STRING
点击模块	click_module	记录点击的模块,如 banner、楼宇广告	STRING
点击模块序号	click_module_no	记录点击的模块序号,如推荐 tab 下运营卡片的顺序,点击第二个运营位,上报“2”	NUMBER
点击位置序号	click_no	记录点击的位置序号,如点击了列表页的第 5 篇资讯,上报“5”,或如点击了某个运营位从左往右的第 2 个卡片,上报“2”	NUMBER
点击的内容类型	click_ob_type	记录点击对应内容的类型,如短视频、文章	STRING
点击的内容名称	click_ob_name	记录点击对应内容的标题,如文章标题	STRING
点击的内容代码	click_ob_code	记录点击对应内容的代码,如文章 ID	STRING

点击的内容分类	click_ob_label	记录点击对应内容的分类，如内部资讯、外部资讯	STRING
点击的内容二级分类	click_ob_sublabel	记录点击对应内容的二级分类，如ETF资讯、研报	STRING
点击的内容作者	click_ob_producer	记录点击对应内容的作者(生产者)，如蔡劲、财联社直播	STRING
点击的内容作者id	click_ob_producerid	记录点击对应内容的作者id，如点击用户发帖的内容记录用户id	STRING
点击的内容作者oa	click_ob_produceroa	记录点击对应内容的作者oa	STRING
点击的内容作者erp	click_ob_producERP	记录点击对应内容的作者erp	STRING
点击的内容提供方	click_ob_supplier	记录点击对应内容的供应方，如财联社、中国基金报等	STRING

点击的内容相关金额	click_ob_amount	记录点击内容相关的金额，例如点击支付时支付的金额	STRING
点击的内容相关方式	click_ob_method	记录点击内容相关的方式属性，例如点击支付时支付的方式	STRING
点击的内容相关时间	click_ob_time	记录点击内容时相关的时间属性，如点击视频时，视频的播放时间进度	STRING
点击的内容相关进度	click_ob_percent	记录点击内容时相关的进度属性，如点击视频时，视频的播放的百分比	STRING
页面名称	page_name	记录所处页面的名称，如资讯详情页，个股详情页	STRING
页面所处模块	page_module	记录所处页面的模块，形式为“业务线_功能模块”，如“理财模块 ETF 专区”	STRING
页面停留时长	event_duration	记录页面停留时长（秒）	NUMBER

页面的内容类型	page_ob_type	记录页面对应内容的类型，如资讯文章，直播	STRING
页面的内容名称	page_ob_name	记录页面对应内容的标题，如文章标题，直播标题	STRING
页面的内容代码	page_ob_code	记录页面对应内容的代码，如文章 ID，直播 ID	STRING
页面的内容分类	page_ob_label	记录页面对应内容的分类，如内部资讯、外部资讯	STRING
页面的内容二级分类	page_ob_sublabel	记录页面对应内容的二级分类，如 ETF 资讯、研报	STRING
页面的内容作者	page_ob_producer	记录页面对应内容的作者(生产者)，如蔡劲、财联社直播	STRING
页面的内容作者 id	page_ob_producerid	记录页面对应内容的作者 id，如用户帖子详情页的用户 id	STRING
页面的内容作者 oa	page_ob_produceroa	记录页面对应内容的作者 oa	STRING

页面的内容作者 erp	page_ob_producererp	记录页面对应内容的作者 erp	STRING
页面的内容提供方	page_ob_supplier	记录页面对应内容的供应方，如财联社、中国基金报等	STRING
页面的内容状态	page_ob_status	记录页面对应内容的状态，如直播中，预告中，回顾	STRING
曝光所处模块	exposure_module	记录曝光的模块，如 banner、开屏	STRING
曝光标题	exposure_name	记录曝光的标题，记录曝光文案	STRING
曝光 ID	exposure_id	记录曝光的 id	STRING
曝光模块序号	exposure_module_no	记录曝光的模块序号，如推荐 tab 下运营卡片的顺序，曝光第二个运营位研报精华，上报“2”	NUMBER
曝光位置序号	exposure_no	记录曝光的位置序号，如曝光了列表页的第 5 篇资讯，上报“5”，或如曝光某个运营位从左往右的第 2 个卡片，上报“2”	NUMBER

曝光的内容类型	exposure_ob_type	记录曝光对应内容的类型，如资讯文章，直播	STRING
曝光的内容名称	exposure_ob_name	记录曝光对应内容的标题，如文章标题，活动标题	STRING
曝光的内容代码	exposure_ob_code	记录曝光对应内容的代码，如文章 id，活动 id	STRING
曝光的内容分类	exposure_ob_label	记录曝光对应内容的分类，如内部资讯、外部资讯	STRING
曝光的内容二级分类	exposure_ob_sublabel	记录曝光对应内容的二级分类，如 ETF 资讯、研报	STRING
曝光的内容作者	exposure_ob_producer	记录曝光对应内容的作者(生产者)，如蔡劲、财联社直播	STRING
曝光的内容作者 id	exposure_ob_producerid	记录曝光对应内容的作者 id，如发帖用户 id	STRING
曝光的内容作者 oa	exposure_ob_produceroa	记录曝光对应内容的作者 oa	STRING

曝光的内容作者 erp	exposure_ob_producererp	记录曝光对应内容的作者 erp	STRING
曝光的内容提供方	exposure_ob_supplier	记录曝光对应内容的供应方，如财联社、中国基金报等	STRING
环境描述			
内容策略 ID	strategy_id	记录内容分发策略对应的 ID	STRING
上步渠道	refer	记录渠道来源或入口的参数，如从某个运营位调整过来的参数	STRING
广发通	gft_id	记录广发通	STRING
端口	client	记录端口，如 APP、WEB、H5，用于区分打开的终端	STRING
操作系统	os	记录操作系统，如安卓、iOS、linux	STRING
版本	app_version	记录 APP 版本	STRING
IP 地址	ip	记录 IP 地址	STRING
设备号	device_id	记录设备号	STRING
上报时间	time	触发时间	DATETIME
客户编号	distinct_id	记录客户编号	STRING

总结而言，建立针对证券期货相关软件的用户行为埋点统计规范是一项复杂而关键的任务。通过遵循埋点设计原则，我们能够实现对用户各类事件行为的客观、准确分析，同时提高了数据的可理解性和可比性。这一规范不仅有助于深入理解用户在复杂业务环境下的行为，还为产品更新迭代提供了有力支持，使我们能够根据数据洞察来不断优化用户体验，满足用户需求。

### 三、全生命周期事件标签体系

建立画像标签体系是金融行业中重要的任务之一，它通过基于用户终端事件行为的追踪，实现对用户全生命周期的事件标签规范化制定，以实现对用户属性、账户属性、资产交易属性、用户平台行为、用户全生命周期阶段、风险测评，投教服务等各类用户事件性质地全面分析，进而引导金融从业人员更好地识别客户，了解客户，服务客户，实现客户财富有效管理及财富增值。

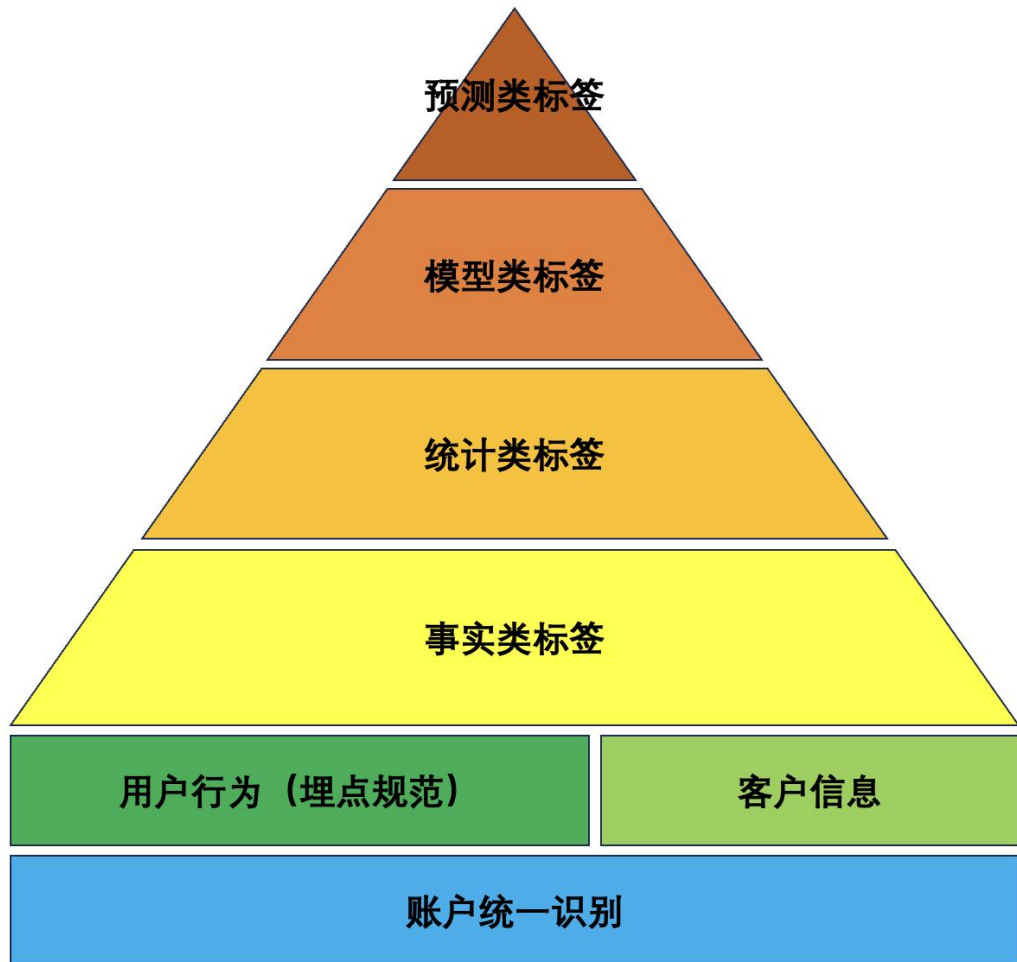
因此，在实现用户账户统一识别和终端事件体系标准化建设的基础上，我们探究设计了全生命周期事件标签体系，基于标准化的用户事件规范和 one ID 原则建设了一套多维度的画像标签体系。

画像标签通过四大层级结构化的方式，基于客户生命周期，以业务视角进行多维度的标签构建，针对海量的数据进行结构化的处理，数据来源包括客户在客户端的行为路径，



客户的基本信息、人口属性、设备属性等，也包括客户的账户属性以及各个业务标的的资产与交易流水情况。

标签的四大层级，包括事实类标签、统计类标签、模型类标签和预测挖掘类标签。事实类标签是基于客户的基础数据，如交易次数、交易频次、年龄、浏览次数等，经过清洗和规整后得到。统计类标签是基于事实类标签，通过一定的统计维度和分段进行统计，例如高频交易型、低活跃客户、年龄分段、行业集中度等。模型类标签是基于分析模型或多个统计类标签的组合，例如 RFM 模型、生命周期模型，用于标识客户的投资偏好、投资迷茫程度等。预测挖掘类标签是基于 AI 算法挖掘，用于预测客户的未来行为，如有效户预测、潜在服务客户预测、资讯推荐、产品推荐等。这些层级覆盖了客户基础数据、统计维度、分析模型和 AI 算法，为不同维度的标签提供了坚实的基础。



最后规定标签的开发需要遵循统一的规范，从最小统计粒度出发构建原子类事实标签。这有助于满足不同维度的标签衍生需求，同时保障数据同源性和可溯源性，减少数据治理的难度。因为标签体系不是一成不变的，它需要持续更新和优化以适应不断变化的市场和客户需求。因此，建立标签体系后，也应建立一个机制来定期审查和更新标签，以确保其有效性和准确性

总之，通过明确定义标签的层级，处理数据来源，遵循开发规范，以及持续更新和优化，我们成功设计了一套完整的画像标签体系。金融机构可以更好地理解客户，提供更精

确的服务，实现客户财富的有效管理和增值。这一标签体系将成为金融从业人员的强大工具，帮助他们更好地满足客户的需求和提升客户体验。

#### **四、标签安全体系**

建立事件标签安全体系的目的是保护用户的隐私和数据安全，提升用户体验和满意度。通过建立完善的安全体系，能够确保用户事件标签的机密性，并根据安全等级的不同，采取数据完整性校验、防篡改技术、数字签名等手段，防止数据被篡改或伪造，提供可信的事件标签数据，并防止用户敏感信息被未经授权的人员获取或滥用。

根据数据的价值、敏感程度、重要程度以及数据发生丢失、泄露、被篡改、被毁损事件的影响范围，可将标签数据划分四个安全等级：

第 1 级（公开数据）：标签数据可被公开或可被公众获知、使用。标签数据的安全属性（完整性、保密性、可用性）遭到破坏后或数据损失后，不造成任何影响。

第 2 级（一般数据）：标签数据用于一般业务使用，针对受限对象公开，一般指内部管理且不宜广泛公开的数据。标签数据的安全属性（完整性、保密性、可用性）遭到破坏或数据损失后，影响范围局限在本机构，可能对公司内部管理造成轻微影响、或导致个别非核心业务短时间无法开展、可能引起较小的声誉风险事件。

第3级（重要数据）：标签数据用于重要业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。标签数据的安全属性（完整性、保密性、可用性）遭到破坏或数据损失后，影响范围局限在本机构，可能影响投资者、员工或公众权益、严重影响公司内部管理工作、或导致部分业务无法开展，造成一定经济损失、或可能引起公司局部范围的声誉风险事件。

第4级（核心敏感数据）：标签数据主要用于核心业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。标签数据的安全属性（完整性、保密性、可用性）遭到破坏后或数据损失后，影响范围不仅局限于本机构，可能对个人隐私造成一定伤害、严重影响公司经营管理、或导致部分或全部业务无法开展，造成重大经济损失、或可能引起全公司范围的、重大社会影响的声誉风险事件。

我们对标签进行了分类，并对各标签的最低安全等级设计如下：

标签大类	大类说明	标签细类	标签举例	最低安全等级
事实类	事实类标签是基于客户的基础数据，如交易次数、交易频次、年龄、浏览次数等，经过清洗和规整后得到	企业客户	是否大股东客户、是否上市公司客户、是否股权激励客户、是否董监高客户	2
		人口属性	出生日期、年龄、联系地址	3
		设备属性	手机号所属城市、手机网络类型、手机操作系统版本	3
		账户属性	开户渠道、开发时长、账户权限	3

		业务提醒	付费状态、到期日期、订单状态	2
		资产与交易	交易日期、交易次数、持仓是市值	3
		生命周期	首次交易日期、历史最高总资产、单笔最大金额、末次活跃日期	3
		服务触点	绑定官微日期、咨询次数、是否登陆 APP	3
		平台行为	点击 tab 天数、点击菜单次数、浏览个股 top	3
		营销属性	优惠券类型、广告点击次数、领取积分日期	2
预测类	预测挖掘类标签是基于 AI 算法挖掘，用于预测客户的未来行为，如有效户预测、潜在服务客户预测、资讯推荐、产品推荐等	预测属性	定投预测、产品购买预测、新开待转有效户预测	2
模型类	模型类标签是基于分析模型或多个统计类标签的组合，例如 RFM 模型、生命周期模型，用于标识客户的投资偏好、投资迷茫程度等	客户分类	客户类型、投资偏好分群、是否有效户	2
统计类	统计类标签是基于事实类标签，通过一定的统计维度和分段进行统计，例如高频交易型、低活跃客户、年龄分段、行业集中度等	资产与交易	当月交易额、日均总资产、总净资产分段	3
		风险测评	金融知识、投资目标、投资资产分段	2
		考核属性	是否分支机构考核、有效户奖励金额、空户激活后股票交易天数	2

课题负责人:	姓名	单位	职务
	辛治运	广发证券股份有限公司	公司副总经理兼信息技术部门负责人

课题成员:	姓名	单位	职务
	黄紫菲	广发证券股份有限公司	总监
	陈鹏	广发证券股份有限公司	普通员工
	潘志龙	广发证券股份有限公司	普通员工
	刘心愉	广发证券股份有限公司	普通员工
	王楚	广发证券股份有限公司	普通员工