

Q/XYZQ

兴业证券股份有限公司企业标准

Q/XYZQ001-2026

大数据资产评分模型

The Big Data Asset Scoring Model

2026 年 3 月 6 日发布 2026 年 3 月 31 日实施

兴业证券股份有限公司发布

目 录

目 录 1

前 言 2

引 言 3

兴业证券大数据资产评分模型 4

1 范围 4

2 规范性引用文件 4

3 术语和定义 4

4 缩略语 5

5 总体原则 5

5.1 客观性原则 5

5.2 可操作性原则 5

5.3 价值优先原则 5

5.4 合规底线原则 6

5.5 持续更新原则 6

6 评分模型框架 6

6.1 评分模型结构 6

6.2 模型计算公式 6

6.3 模型评分范围 6

7 评分指标体系设计 7

7.1 质量评分指标 7

7.2 资源评分指标 7

7.3 安全评分指标 7

7.4 价值评分指标 7

8 算法选择与实施 7

8.1 算法选择 7

8.2 实施流程 7

9 模型动态调整机制 8

9.1 模型验证与校准 8

9.2 模型动态调整因素 9

9.3 模型动态调整触发机制 9

9.4 模型动态调整实施流程 9

10 评分模型应用参考 9

10.1 数据全生命周期价值运营 9

10.2 建立供需双向评价机制 10

10.3 安全评分与动态访问控制 10

10.4 支持可信数据集持续运营 10

附 录 A （资料性附录） 大数据资产评分体系设计参考 11

附 录 B （资料性附录） 模型使用算法介绍 14

附 录 C （资料性附录） 大数据资产评分模型动态调整机制流程图 16

参考文献 17

前 言

本标准按照GB/T 1.1-2020《标准化工作导则第1部分：标准化文件的结构和起草规则》给出的规则起草。

本标准由兴业证券股份有限公司提出。

本标准由兴业证券股份有限公司归口。

本标准起草单位：兴业证券股份有限公司。

本标准主要起草人：陈锋、李细杰、卢丹、周仪、马浩、薛鸣。

本标准为首次制订。

引 言

随着金融科技加速发展与数据要素市场化进程推进，证券期货行业正从经验驱动向数据智能驱动转型。大数据资产作为行业核心生产要素，已成为提升机构竞争力、优化资源配置的关键抓手。《推动数字金融高质量发展行动方案》等政策明确要求夯实数据治理与融合应用能力，而行业实践显示，大数据已深度重构业务流程。为科学评估数据资产的业务价值与合规属性，需建立统一、可量化的评分标准。本标准基于兴业证券实践与理论研究，规范大数据资产评分的维度、指标、方法及实施流程，为公司提供数据资产管理的量化工具，推动数据要素价值释放与合规风险防控。

兴业证券大数据资产评分模型

1 范围

本标准规定了兴业证券大数据资产的评分模型框架、维度划分、指标定义、评分规则及实施要求。

本标准适用于兴业证券大数据平台、数据仓库、数据集市等数据资产的评分管理，以及评分驱动的数据生命周期管理、可信数据集持续运营、数据安全管控等应用参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 36073-2018 数据管理能力成熟度评估模型

GB/T 40685-2021 信息技术服务 数据资产 管理要求

JR/T 0236-2021 金融大数据 术语

JR/T 0158-2018 证券期货业数据分类分级指引

3 术语和定义

GB/T 36073-2018、GB/T 40685-2021、JR/T 0236-2021、JR/T 0158-2018界定的以及下列术语和定义适用于本文件。

3.1

大数据 big data

海量的数据集，其数据在本质上具有体量大、种类多、变化快、变量多的特性，需要一种易扩展的技术来有效存储、处理、管理和分析。

[来源:JR/T 0236-2021, 3.1]

3.2

数据资产 data asset

合法拥有或者控制的，能进行计量的，为组织带来经济和社会价值的数据资源。

[来源:GB/T 40685-2021, 3.1]

3.3

大数据平台 big data platform

采用分布式存储和计算技术，提供大数据的访问和处理，支持大数据应用安全高效运行的软硬件集合，包括监视大数据的存储、输入输出、操作控制等大数据服务软硬件基础设施。

[来源:JR/T 0236-2021, 6.11]

3.4

大数据资产 big data asset

大数据平台中可管控、符合合规要求、具有业务价值的数据资产集合，包括大数据平台采集的源系统原始数据、经数据仓库加工整合的模型化数据，以及数据集市向下游业务系统输出的应用化数据。

3.5

元数据 metadata

关于数据或数据元素的数据（可能包括其数据描述），以及关于数据拥有权、存取路径、访问权和数据波动性的数据。

[来源:JR/T 0236-2021, 5.12]

3.6

数据质量 data quality

在特定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源:JR/T 0236-2021, 5.7]

3.7

数据安全 data security

通过采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力。

[来源:JR/T 0236-2021, 5.19]

3.8

数据价值 data value

从分析学角度考虑的数据对组织的重要性。

[来源:JR/T 0236-2021, 4.24, 有修改]

4 缩略语

DCMM - 数据管理能力成熟度评估模型

AHP - 层次分析法 (Analytic Hierarchy Process)

BIA - 业务影响分析 (Business Impact Analysis)

K-Means - K均值聚类算法

5 总体原则

评分原则是模型运行的“基本准则”，确保评分结果公平、客观、可用，核心包括四大原则。

5.1 客观性原则

以数据为依据，减少主观干预。指标得分需基于可量化的系统数据，避免人工主观打分；仅在专家权重设置阶段引入主观经验，且需通过一致性检验。

5.2 可操作性原则

指标数据可获取，评分流程可落地。所设置指标需满足“数据可从现有系统获取、计算逻辑可自动化”，避免设置“无法量化”或“数据不可得”的指标。

5.3 价值优先原则

优先评价数据的业务价值。在多维度评分中，优先保障“应用价值”维度的权重与影响力，确保评分结果能反映数据资产的核心竞争力。

5.4 合规底线原则

合规不达标则限制使用。将合规性作为评分的“底线”，安全评分不达标的资产，无论其他维度得分多高，均判定为“不可用资产”，限制业务使用。

5.5 持续更新原则

评分模型应能持续更新，以动态反映数据资产自身、应用场景及市场环境的变化。

6 评分模型框架

6.1 评分模型结构

评分模型采用“三层递阶结构”，由上至下包含：目标层、准则层、指标层。

6.1.1 目标层

目标层明确模型最终输出，反映数据资产的综合价值与合规性。核心输出为大数据资产总评分，满分100分。

6.1.2 准则层

准则层拆解目标层为可评价的核心维度，覆盖大数据资产的“质量、成本、安全、价值”属性。核心内容包括质量评分、资源评分、安全评分、价值评分的总分值及权重占比。

示例：质量评分（100分，权重20%）、资源评分（100分，权重20%）、安全评分（100分，权重20%）、价值评分（100分，权重40%）。

6.1.3 指标层

指标层将准则层细化为可量化的操作单元，是评分计算的基础。每个准则层下含有1~n个指标。

6.2 模型计算公式

总评分计算公式 = 质量评分×质量权重占比 + 资源评分×资源权重占比 + 安全评分×安全权重占比 + 价值评分×价值权重占比

各维度评分满分为100分，各维度权重占比总和为100%。

6.3 模型评分范围

6.3.1 评价对象范围

评价对象为可管控的业务数据资产，包含大数据平台采集的源系统数据、数据仓库整合模型数据、数据集市服务下游数据等。同时，排除临时测试数据、个人办公数据、冗余日志数据等不可管控或无业务价值数据。

6.3.2 业务覆盖范围

业务范围为全业务线覆盖，重点聚焦核心业务。包括但不限于传统经纪、承销、自营业务，衍生品交易、跨境投资、做市业务等创新领域，以及期货、资产管理等子公司协同业务。对“高价值、高风险”的核心业务数据资产优先评分。

6.3.3 数据类型范围

数据类型包含结构化数据和非结构化数据，且需能通过指标量化评分。非结构化数据难以量化指标，暂不纳入。如非结构化数据具有较高业务价值，可转换为结构化数据纳入评分。

7 评分指标体系设计

7.1 质量评分指标

质量评分聚焦“数据的准确性、完整性、可用性”，包含“元数据质量”“数据内容质量”2个大类，具体指标设置参考附录A表A.1。

7.2 资源评分指标

资源评分聚焦“数据资产的存储成本与计算性能”，包含“计算性能”“存储成本”2个大类，具体指标设置参考附录A表A.2。

7.3 安全评分指标

安全评分聚焦“数据的分类分级、访问控制、审计追溯”，包含“安全分类分级”“访问控制”“审计追溯”“导出管控”4个大类，具体指标设置参考附录A表A.3。

7.4 价值评分指标

价值评分聚焦“数据的业务支撑能力、使用热度、血缘影响”，包含“资产盘点价值”“数据使用热度”“数据血缘”“无效数据”4个大类，具体指标设置参考附录A表A.4。

8 算法选择与实施

8.1 算法选择

证券公司数据具有数据量大、实时性高、合规要求严、业务场景复杂等特点，算法选择需贴合特性，避免通用算法直接套用。本文所列算法的详细说明见附录B。

8.1.1 多准则决策算法选择

公司数据资产评分涉及证券期货业务、安全合规等专业领域，需依赖专家经验，因此选择“层次分析法+德尔菲法”组合。

德尔菲法通过匿名征求数据专家、业务专家的意见，确保指标与权重符合专业认知。层次分析法将专家的“定性判断”转化为“定量权重”，避免“纯经验权重”的主观偏差。

8.1.2 统计分析算法选择

公司数据资产的核心指标多符合正态分布或偏态分布，且有明确的业务阈值。指标与得分的关系为线性或分段线性，使用描述性统计加线性映射已能满足需求，且计算效率更高。

8.1.3 无监督学习算法选择

公司数据资产规模达PB级，数据分类分级和异常资产识别标注成本极高，采用孤立森林和K-Means聚类等无监督学习算法可大幅提高异常数据识别和分类分级效率。

8.2 实施流程

8.2.1 指标权重确定

a) 德尔菲法初筛指标与权重范围。采用德尔菲法组建专家库，通过多轮匿名意见征求与反馈，迭代达成共识，初筛指标和权重范围。

b) 层次分析法计算精确权重。采用层次分析法，构建层次结构（目标层、准则层、指标层），比较形成判断矩阵，计算权重与一致性检验，最终确定指标和权重值。

8.2.2 常规指标得分计算

a) 数据采集与预处理。从元数据管理平台、数据质量监控系统、大数据平台监控系统、数据分类分级系统、数据服务平台等数据管理相关系统采集评价指标的原始数据。剔除异常值，平滑数据，确保数据质量。

b) 统计量计算与阈值设置。对连续型指标计算近30天/180天均值、分位数，参考行业标准与公司实际，结合统计量设置阈值；对分类指标计算占比，直接按规则得分。

c) 计算准则层得分。将各准则层下所有指标得分按权重乘加。

8.2.3 异常资产调整

a) 孤立森林识别异常资产。选择3个异常特征，对特征数据标准化；用历史数据训练孤立森林模型，设置树数量和异常得分阈值；计算新增资产异常得分，小于阈值的判定为异常资产，质量/安全评分进行相应下调。

b) K-Means聚类分级数据资产。确定K值为3，分为“优质、一般、劣质”3级；对4个准则层得分进行聚类，得到3个簇中心；根据簇中心得分为数据资产进行分级标识。

8.2.4 总评分计算

a) 归一化准则层得分，按100分值归一化；

b) 计算总评分，根据6.2模型计算公式计算各项数据资产最终得分；

c) 失效红线调整，若资产符合失效资产判定的3个条件之一，总评分置为0。

9 模型动态调整机制

9.1 模型验证与校准

9.1.1 验证模型准确性的方法

模型准确性验证采用多维度方法，包括：

a) 回溯验证。通过对比模型评分结果与历史数据资产的实际价值表现，检验高评分资产是否确实被高频访问并有效支持下游应用，低评分资产是否已失效或无实际应用。

b) 专家评审。定期组织数据开发与应用人员对评分结果进行抽样评审，综合业务直觉与开发经验判断模型结果的合理性。

c) A/B测试。针对模型的小版本更新，在小范围资产内实施测试，对比新旧模型的评分结果差异，科学评估新模型的改进效果。

9.1.2 校准模型参数流程

当出现评分分布严重偏离、用户反馈与实际情况明显背离、关键指标权重失衡导致评分失真等情况，应对模型参数进行校准。模型校准流程分为四个步骤：

- a) 首先进行根因分析，明确问题根源；
- b) 其次对权重、阈值、归一化方法等进行针对性微调；
- c) 然后使用历史数据模拟运行调整后的模型，全面评估效果；
- d) 最后完成测试后，部署至生产环境。

9.1.3 评估模型稳定性

模型稳定性评估指标包括评分稳定性、榜单稳定性和指标一致性。

- a) 评分稳定性指同一数据资产在相邻周期内评分波动的标准差，波动应处于合理区间；
- b) 榜单稳定性指数据资产的名单变化率不应频繁剧烈变动；
- c) 指标一致性指不同业务域间，同等级资产的平均得分应相对一致，避免系统性偏差。

9.2 模型动态调整因素

动态调整主要受四方面因素驱动：

- a) 模型自身缺陷暴露；
- b) 技术架构演进；
- c) 数据治理水平提升；
- d) 外部合规要求变化。

9.3 模型动态调整触发机制

动态调整触发机制包括定期评估、异常事件触发、特征监测三种方式。

- a) 定期评估机制。每半年进行一次综合的模型评估与调优，确保模型持续优化。
- b) 异常事件触发机制。当新增数据源或数据应用需集成到模型中、多次收到重要部门对评分结果的意见反馈、发生因数据质量问题引发的重大故障时，启动调整程序。
- c) 特征监测机制。通过资产评分模型监控看板持续监测核心输入数据的分布变化，当发现数据质量分显著跳变、资产平均下游服务应用数量大幅波动、存储计算成本分布发生结构性变化等情况时，开展模型异常分析并评估是否对模型调优。

9.4 模型动态调整实施流程

动态调整实施流程包含四个关键环节：

- a) 需求评估与决策阶段。收集来自定期评估、异常触发或监控预警的调整需求，组织技术、业务代表进行分析评估，明确调整的必要性、范围和优先级，形成模型调整方案。
- b) 指标体系更新与权重重估阶段。根据决策结果，引入新指标、废弃过时指标或修改现有指标计算逻辑，并采用专家打分法结合机器学习方法重新计算指标权重。
- c) 逻辑优化与模型迭代部署阶段。重点优化现有计算逻辑以提升评估精度，在测试环境完成开发与测试，进行全面的回溯验证和影响分析，并对模型代码、参数、指标定义实施严格的版本控制。
- d) 调整效果评估与反馈阶段。密切监控新模型运行稳定性、评分分布变化和业务反馈，经过1-2个评分周期后评估调整是否达到预期目标，并将评估结果反馈至决策流程，形成持续改进的完整闭环。

10 评分模型应用参考

10.1 数据全生命周期价值运营

a) 数据采集阶段：建立数据来源可信度评分机制，评估数据采集渠道的合法性、完整性及时效性，确保原始数据质量符合合规要求。

b) 数据存储与处理阶段：通过自动化评分系统对数据存储、数据调度、数据任务资源消耗情况进行分级，量化数据存储与处理成本，为资源节流提供决策依据。

c) 数据应用阶段：结合业务场景需求，动态评估数据资产的使用效率、应用潜力及合规风险，为优质资产（红榜）的优化配置与劣质资产（黑榜）的清退提供决策依据。

d) 数据退出阶段：制定数据资产退役评分标准，评估数据生命周期终止后的残余价值及潜在风险，确保资产退出风险可控。

10.2 建立供需双向评价机制

a) 需求驱动评分：基于业务需求特征（如实时性、准确性、合规性）建立评分维度，量化数据资产与需求的匹配度，避免资源错配。

b) 资产反哺需求：通过资产评分结果反馈至需求侧，优化数据采集与处理策略，例如针对高需求场景优先提升数据质量或扩展数据覆盖范围。

c) 资源精准投入：以评分结果为依据，动态调整数据治理资源分配，例如对高价值资产增加标注、清洗投入，对低效资产减少冗余处理成本。

d) 闭环反馈优化：建立评分模型迭代机制，结合需求变化与资产表现数据，持续优化评价结果的闭环应用，驱动数据团队与业务部门共同协作，实现提好数据与用好数据的良性循环。

10.3 安全评分与动态访问控制

a) 安全评分维度：定义数据敏感性、访问权限、安全防护等评分指标，量化数据资产的潜在安全风险。

b) 动态访问通行证：基于安全评分生成动态访问控制策略，通过代码固化核验规则，实现访问过程的动态风险评估与分级授权。

c) 风险自适应控制：根据评分结果调整访问阈值，例如对高风险操作触发多因素认证或限制访问频率，确保数据使用合规性。

d) 全链路审计：记录访问控制决策过程，形成可追溯的审计日志，满足监管要求并支持事后风险分析。

10.4 支持可信数据集持续运营

围绕可信数据集的核心特质，将评分结果深度嵌入“准入-监控-优化-迭代-应用”闭环全流程，形成“以评分驱动可信度提升、以运营巩固评分价值”的正向循环。具体实施路径包括如下步骤：

a) 基于评分结果构建可信数据集准入标准：根据数据集的业务重要性（如核心交易数据集），制定差异化的评分准入标准，避免一刀切。

b) 实时监控、定期复盘：针对评分模型中对可信度影响最大的维度（如安全、质量），设置实时监控规则，一旦触发阈值立即预警；每月基于评分结果开展可信数据集健康度复盘。

c) 评分短板实施精准优化：通过分析各维度评分，针对性解决数据集的可信度缺陷，实现评分提升到可信度增强的闭环。

d) 联动业务迭代更新可信标准与数据集：业务需求、监管政策、技术环境的变化会导致可信定义的变化（如新规要求客户数据需额外脱敏，原安全评分标准不再适用），需基于评分结果联动业务迭代，确保可信数据集的时效性。

e) 应用反馈驱动运营优化：可信数据集的运营需落地到业务应用，通过业务反馈反向验证评分结果的准确性，进一步优化运营策略，形成闭环。

附 录 A
(资料性附录)
大数据资产评分体系设计参考

表 A.1 质量评分指标参考

评分大类	评分指标	指标定义	评分规则参考	数据来源
元数据质量 (50分)	表中文名完整性 (10分)	表中文名不为空,且符合公司命名规范	存在表中文名或注释,得10分	元数据管理平台
	字段中文名完整性 (30分)	字段中文名是否明确含义(无英文缩写)、无重复、无缺失(核心字段必选)	0-10%,不得分; 11%-30%,得15分; 31%-60%,得20分; 61%-100%,得30分	元数据管理平台
	资产目录完整性 (10分)	数据资产是否完成“业务分类(如经纪业务)”分类	有目录得10分	数据资产目录系统
数据内容质量 (50分)	是否存在运行的质量检核任务 (40分)	是否配置自动化质量检核任务(如缺失率、错误率检核),且每日运行正常	配置且每日运行:40分; 未配置:0分	数据质量监控系统
	近1个月是否有异常告警 (10分)	近30天内是否因数据质量问题(如数据重复、更新延迟)产生非统计类告警	无告警:10分; 有告警:0分	数据质量监控系统

表 A.2 资源评分指标参考

评分大类	评分指标	指标定义	评分规则参考	数据来源
计算性能 (60分)	调度任务平均耗时 (20分)	近30天内,每天最近一次查询时调度任务的平均耗时(单位:秒)	0,不得分 (0,10],得20分 (10,30],得15分 (30,50],得10分 (50,100],得5分 超过100,不得分	大数据平台任务监控系统
	成功任务CPU占用 (20分)	近30天内,每天最近一次成功查询的任务CPU平均占用率(单位:%)	为0,不得分 在(0,60],得20分 在(60,200],得15分	大数据平台任务监控系统

			在(200,500], 得10分 在(500,1000], 得5分 超过1000, 不得分	
	成功任务内存占用 (20分)	近30天内,每天最近一次 成功查询的任务内存平 均占用率(单位: %)	为0, 不得分 在(0,20万], 得20分 在(20万,50万], 得15分 在(50万,400万], 得10分 在(400万,8000万], 得5分 超过8000万, 不得分	大数据平台任 务监控系统
存储成本(40分)	HDFS已使用容量(20 分)	数据资产在HDFS分布式 文件系统中占用的实际 容量(单位: TB)	0, 不得分 (0,0.1G], 得20分 (0.1G,1G], 得15分 (1G,100G], 得10分 (100G,1T], 得5分 超过1T, 不得分	HDFS存储监控 系统
	表所用文件数(20 分)	数据资产对应的HDFS文 件数量(含分区文件)	0, 不得分 [1,20], 得20分 [21,50], 得15分 [51,200], 得10分 [201,500], 得5分 超过500, 不得分	HDFS存储监控 系统

表 A.3 安全评分指标参考

评分大类	评分指标	指标定义	评分规则参考	数据来源
安全分类分级 (20分)	分类分级打标率 (20分)	已完成业务分类+安全分 级的字段数占总字段数的 比例	≥90%: 20分; 80%-89%: 15 分; 70%-79%: 10分; <70%: 0分	公司数据分类分级 系统
访问控制(20 分)	数据隔离(10分)	是否按角色(如投研、风 控)+业务条线(如经纪、 资管)实现数据权限隔离	双维度隔离: 10分; 仅角色 隔离: 5分; 无隔离: 0分	数据权限管理系统
	查询水印(10分)	页面查询数据时, 是否自 动添加含用户ID、查询时 间的动态水印	含动态水印: 10分; 含静态 水印(无用户ID): 5分; 无水印: 0分	各业务系统数据查 询日志
审计追溯(30 分)	数据库审计(10 分)	数据资产是否纳入数据库 审计系统, 记录所有增删 改查操作	纳入且审计日志保存≥1年: 10分; 纳入但保存<1年: 5 分; 未纳入: 0分	数据库审计系统
	导出日志(10分)	是否记录数据导出的用	记录完整且可追溯: 10分;	各业务系统数据导

		户、时间、文件类型、导出量，日志可追溯	记录不完整（缺导出量）：5分；无日志：0分	出日志
	访问日志(10分)	是否记录页面访问的用户、时间、访问内容，日志保存≥6个月	记录完整且保存≥6个月：10分；记录完整但保存<6个月：5分；无日志：0分	各业务系统数据查询日志
导出管控（30分）	导出审批(10分)	客户敏感数据导出是否需双人审批（业务岗+合规岗），且审批记录留痕	双人审批+留痕：10分；单人审批+留痕：5分；无审批：0分	各业务系统数据导出管控
	导出加密(10分)	客户敏感数据导出文件是否采用国密算法（如SM4）加密	国密算法加密：10分；非国密算法（如AES）加密：5分；未加密：0分	各业务系统数据导出管控
	销毁确认(10分)	导出的敏感数据文件销毁后，是否提交纸质/电子回执，确认销毁完成	提交回执+记录保存：10分；提交回执无记录：5分；无回执：0分	各业务系统数据导出管控

表 A.4 价值评分指标参考

评分大类	评分指标	指标定义	评分规则参考	数据来源
资产盘点价值（10分）	资产元数据查询频次（2分）	近30天内，业务部门查询该资产元数据的人次	0次，不得分 (0,5]人次，得1分 超过5人次，得2分	元数据管理平台
	涉及系统标签个数（5分）	资产关联的系统标签数量	数量0，不得分 数量>=1，得5分	数据资产管理系统
	是否是指标落地表（3分）	资产是否为业务指标的最终存储表	不是指标落地表，不得分； 是指标落地表，得3分	数据资产管理系统 /指标管理平台
	下游系统数（文件下载）（15分）	近30天内，通过文件下载服务使用该资产的下游业务系统数量	0个，不得分 (0,3]个，得10分 3个以上：15分	数据服务平台
	表下载频率（12分）	近30天内，业务部门下载该资产的次数	0人次：不得分； (0,5]人次，得9分； 超过5人次，得12分	数据服务平台
	下游系统数（接口服务）（10分）	近30天内，通过API接口调用该资产的下游业务系统数量	0个，不得分； (0,2]个，得6分； (2,6]个，得8分； 6个以上，得10分	数据服务平台

数据使用热度 (80 分)	接口请求次数(7分)	近 30 天内, API 接口调用该资产的总次数	0 人次, 不得分; (1, 5]人次, 得 5 分; (5, 100]人次, 得 6 分; 超 100 人次, 得 7 分	数据服务平台
	直接下游任务数(10分)	近 30 天内, 依赖该资产的 ETL 任务、分析任务数量	0 个, 不得分; (0, 1]个, 得 8 分; 1 个以上, 得 10 分	大数据平台任务监控系统
	180 天分区平均访问次数(12 分)	180 天内, 该资产各分区(如按日分区)的平均访问次数	0 次, 不得分; (0, 0.04]次, 得 6 分 (0.04, 1]次, 得 9 分 超过 1 次, 得 12 分	大数据平台任务监控系统
	推送任务运行次数(14 分)	近 30 天内, 该资产自动推送至业务系统(如投研终端)的任务运行次数	0 次, 不得分 (0, 3]次, 得 10 分 超过 3 次, 得 14 分	数据推送平台
数据血缘(10 分)	血缘上下游关系数量(10 分)	该资产的上游数据源数量+下游依赖资产数量(反映数据流转范围)	0, 不得分 (0, 3], 得 6 分 (3, 10], 得 8 分 超过 10, 得 10 分	数据管控平台 血缘管理模块
无效数据(0 分)	来源系统已技术下线	贴源层数据的来源系统已下线	直接置为 0	大数据平台任务监控系统
	HDFS 容量为 0	该表存储容量为 0	直接置为 0	
	数据停止更新超过一年	该表数据已经停止更新超过一年	直接置为 0	

附 录 B

(资料性附录)

模型使用算法介绍

B.1 层次分析法

原理：将复杂目标拆解为“目标层、准则层、指标层”的递阶结构，通过专家对同层元素的“两两比较”构建判断矩阵，计算最大特征值与权重向量，并通过一致性检验（ $CR < 0.1$ ）确保权重逻辑无矛盾。

特点：将主观经验转化为数学逻辑，适用于“多维度、无明确量化标准”的权重设置；但依赖专家判断，若专家意见分歧大，需多次迭代。

适用场景：模型准则层（质量、资源、安全、价值）与指标层的权重计算。

B.2 德尔菲法

原理：通过“匿名征求专家意见、汇总反馈、多轮迭代、达成共识”的闭环流程，整合领域专家（数

据治理、数据安全、大数据、业务）的判断，确定指标清单、权重范围或评分规则。

特点：匿名性可避免专家受他人影响；迭代性可多轮反馈修正；一般耗时较长（需 3-4 轮迭代），依赖专家专业性。

适用场景：模型指标体系初筛、权重范围初步确定。

B.3 描述性统计

原理：计算数据资产指标的基础统计量，如均值、分位数、占比等，反映数据的集中趋势与离散程度，为评分阈值设置提供依据。

特点：纯客观数据驱动，无主观干预。

适用场景：适用于“可量化、数据分布规律明确”的指标得分计算。

B.4 线性映射

原理：将指标的原始统计量通过线性公式映射为评分，公式为“得分 = (最大值-原始值) / (最大值-最小值) × (满分-最低分) + 最低分”。

特点：计算简单，得分随指标值线性变化。

适用场景：适用于指标值与得分正相关/负相关的指标得分转化。

B.5 孤立森林

原理：通过构建多棵“孤立树”，随机选择特征与分割阈值，将数据点隔离在叶节点；异常数据因特征特殊，被隔离的路径长度更短，通过路径长度得分识别异常资产（得分<0.3 为异常）。

特点：无需标注数据，训练速度快（适用于 PB 级数据）；但对“局部异常”识别精度较低，需结合业务场景调整参数。

适用场景：模型异常数据资产的识别，调整相应评分。

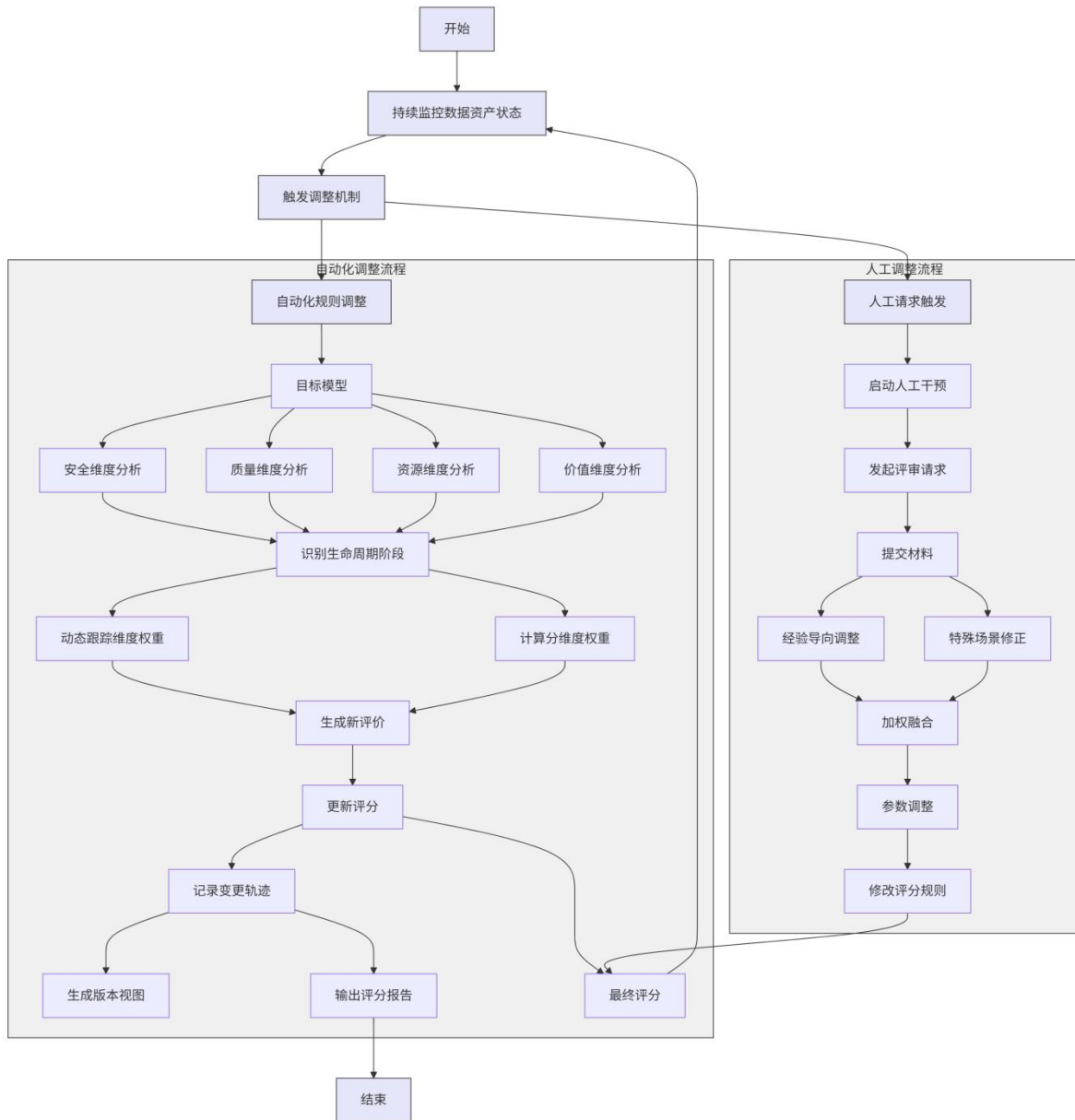
B.6 K-Means 聚类

原理：基于欧氏距离将数据资产聚为 K 类，使同一类内资产相似度高、不同类间相似度低；通过“肘部法则”（计算不同 K 值的误差平方和，拐点处为最优 K）确定 K 值（例如，模型 K=3：优质/一般/劣质）。

特点：聚类速度快，可批量处理海量资产；但对初始聚类中心敏感，需多次随机初始化。

适用场景：模型数据资产评分的分级，支撑数据资产分级运营。

附录 C
(资料性附录)
大数据资产评分模型动态调整机制流程图



参考文献

- [1] GB/T 36073-2018 数据管理能力成熟度评估模型
- [2] GB/T 40685-2021 信息技术服务 数据资产 管理要求
- [3] JR/T 0158-2018 证券期货业数据分类分级指引
- [4] 数据资产管理实践指南（7.0版），CCSA TC601大数据技术标准推进委员会