



# 中华人民共和国国家标准

GB/T XXXXX—XXXX

## 数据产品 质量评价通用要求

Data product—General requirements for quality evaluation

(点击此处添加与国际标准一致性程度的标识)

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

国家市场监督管理总局  
国家标准化管理委员会 发布



目 次

前 言 ..... III

1 范围 ..... 6

2 规范性引用文件 ..... 6

3 术语和定义 ..... 6

4 数据产品质量评价框架 ..... 8

5 数据产品质量评价方法 ..... 11

6 数据产品质量特性及测量 ..... 13

7 数据产品质量评价过程 ..... 22

附录 A （规范性附录） 面向机器学习的数据产品质量特性扩展 ..... 24

附录 B （资料性附录） 数据产品质量评价示例 ..... 25

参考文献 ..... 27



## 前 言

本文件按照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》的规定起草。

请注意本文中的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国数据标准化技术委员会（SAC/TC 609）提出并归口。

本文件起草单位：中国南方电网有限责任公司、中国电子技术标准化研究院、清华大学、国家数据发展研究院、国家工业信息安全发展研究中心、中国信息通信研究院、哈尔滨工业大学、中国移动信息技术有限公司、清华四川能源互联网研究院、国家能源投资集团有限责任公司信息技术分公司、四川大学华西医院、中国石油天然气股份有限公司西南油气田分公司、暨南大学、西安电子科技大学、广州数据集团有限公司、广州数据交易所有限公司、广州信安数据有限公司、广州芳禾数据有限公司、工业和信息化部电子第五研究所、中电数据产业集团有限公司、安阳数智科技集团有限公司、上海观安信息技术股份有限公司、浙商证券股份有限公司、北京华宇软件股份有限公司、重庆市质量和标准化研究院、中国信息协会、北京大数据先进技术研究院、北京数字理政规划设计研究院有限公司、商业信用中心、农业农村部大数据发展中心、中国质量认证中心、南京银行股份有限公司、吉林省吉林祥云信息技术有限公司、广州网盈信息科技有限公司、北京软件和信息服务交易所有限公司、国网四川省电力公司、中国铁道科学研究院、上海合合信息科技股份有限公司、中国华电集团有限公司、广东电网有限责任公司、成都飞机工业（集团）有限责任公司、贵阳大数据交易所有限责任公司。

本文件主要起草人：王晨、陈彬、王为中、宋亮、禚亮、白玉真、杨靖世、王宏志、杨秋勇、汪睿棋、丁小欧、崔瑞媛、胡清源、翁健、杨安家、李晖、李成博、赵兴文、刘国庆、高翔、郑彬彬、张昊天、曾立波、穆帅先、任博忱、严宇平、梁维铿、张茵翠、郑丹辉、王文君、王一华、冯四凤、魏丽丽、李红、韩金丽、刘天成、郭远洋、曾新科、孙丽、张智超、杜潇霖、吴蓉蓉、刘爽、葛振兴、袁庆然、周艺、李迎、张亚欣、王岩、顾延甲、王全胜、邓水平、徐欢、李学武、董贇、游绍华、李文科、陈诗、麦竣朗、杨春、柴雁欣、马一宁、罗育林、萧展辉、杨光、张世良、陈开杰、郭彦楠、曹熙、林荣政、王敏、汤平瑜、刘晓天、孙韵、王佳、李赋欣、龚燕、舒意、殷晋、杨旭、李晓洋、崔博洋、王少锋、刘翔锋、杨吉飞、欧阳森山、唐飞、毛伟凯。

# 数据产品 质量评价通用要求

## 1 范围

本文件规定了通用的数据产品质量评价的指标体系、计算方法和评价过程。

本文件适用于数据交易流通业务场景中数据产品的质量评价，包括但不限于数据提供方、需求方、第三方评估机构及数据交易平台在数据产品质量管控、验证、评测、争议解决等场景中的应用。其他业务场景的数据产品质量评价可参考使用。

## 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 25000.1-2021 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第1部分：指南

GB T 25000.10-2016 系统与软件工程系统与软件质量要求和评价(SQuaRE) 第10部分 系统与软件质量模型

GB/T 25000.12-2017 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第12部分：数据质量模型

GB/T 36344 信息技术 数据质量评价指标

GB/T 42381.8-2023 数据质量 第8部分：信息和数据质量 概念和测量

GB/T XXX—XXX 数据产品 数据产品描述

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**数据产品** data product

基于数据加工形成的，可满足特定需求的数据加工品和数据服务，包括：数据集、数据报告、数据接口等。

[来源：GB/T XXX—XXX，3.1]

### 3.2

**数据产品基底数据** data product foundation data

指加工数据产品所利用的核心原料数据。

### 3.3

**要求** requirement

明示的，通常隐含的或者必须履行的需求或期望。

[来源：GB/T 42381.8—2023，3.5]

### 3.4

**数据产品质量 data product quality**

数据产品满足要求的程度。

### 3.5

**数据产品质量特性 data product characteristic**

对数据质量有影响的数据质量属性或属性类别，数据产品质量特性又称为数据产品质量指标。

[来源：GB/T 25000.12-2017 4.4]

### 3.6

**数据产品质量测度**

变量，为其赋值作为数据质量特性的测量结果。

[来源：GB/T 25000.12-2017 4.5]

### 3.7

**测量 measurement**

以确定测度的值为目的的一组操作。

[来源：GB/T 25000.12-2017 4.5]

### 3.8

**数据产品质量模型 data product quality model**

为规定数据产品质量需求和评价数据产品质量提供一种框架而定义的一组特性及特性间的关系集合。

[来源：GB/T 25000.10-2016 3.18]

### 3.9

**使用质量 quality in use**

在指定使用周境中，数据产品达到其规定目标并满足用户要求的程度。

注：用户的实际要求可能与需求中所期望的要求不一定完全一致。

[来源：GB/T 25000.1-2021，3.31]

### 3.10

**数据产品质量评价 data product quality evaluation**

对数据产品能够满足要求的能力的系统性考查。

[来源：GB/T 25000.1-2021 3.44]

## 4 数据产品质量评价框架

### 4.1 总体框架

数据产品是基于数据加工形成的，可满足特定需求的数据加工品，其质量评价框架为面向数据产品全生命周期质量评价提供一个结构化方法论，涵盖数据产品质量模型、数据产品质量评价方法、数据产品质量评价过程，见图1。

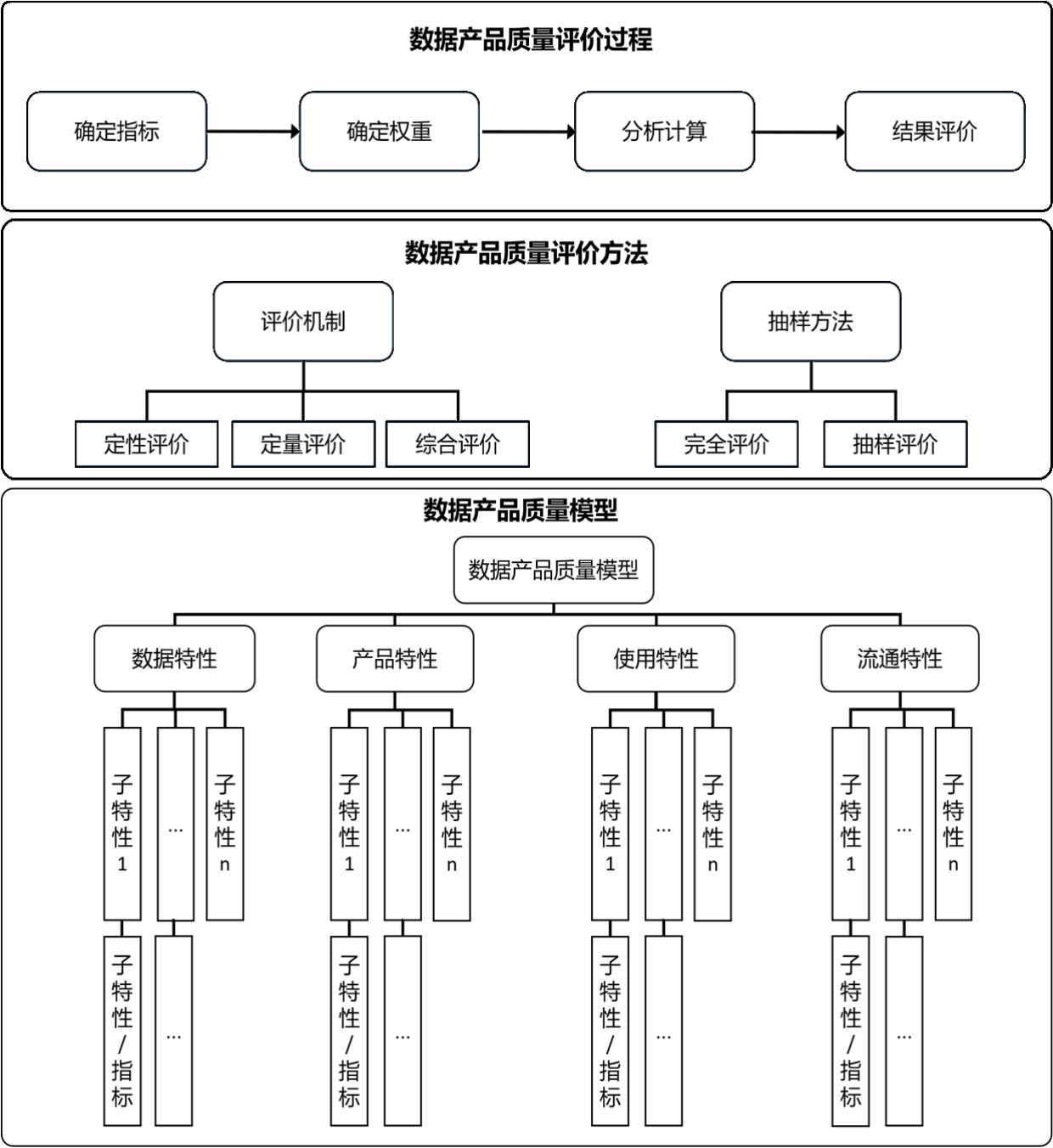


图1 数据产品质量评价框架



4.2 数据产品质量模型

4.2.1 概述

数据产品质量模型是数据产品质量的表达，是数据产品质量评价的基础。数据产品质量模型由一组数据产品质量特性构成（即数据特性、产品特性、使用特性和流通特性），质量特性可进一步分解为若干子特性，子特性又可进一步被分解为若干子子特性，以此类推。

数据产品质量模型的详细内容见第6章。  
面向机器学习的数据产品质量特性扩展见附录A。

4.2.2 数据特性

数据特性是指构成数据产品核心价值的基底数据所固有的质量属性。数据产品质量模型的数据特性的分解结构框架如下图所示，和GB/T 36344相关指标保持一致，详细内容参见第6.2节。

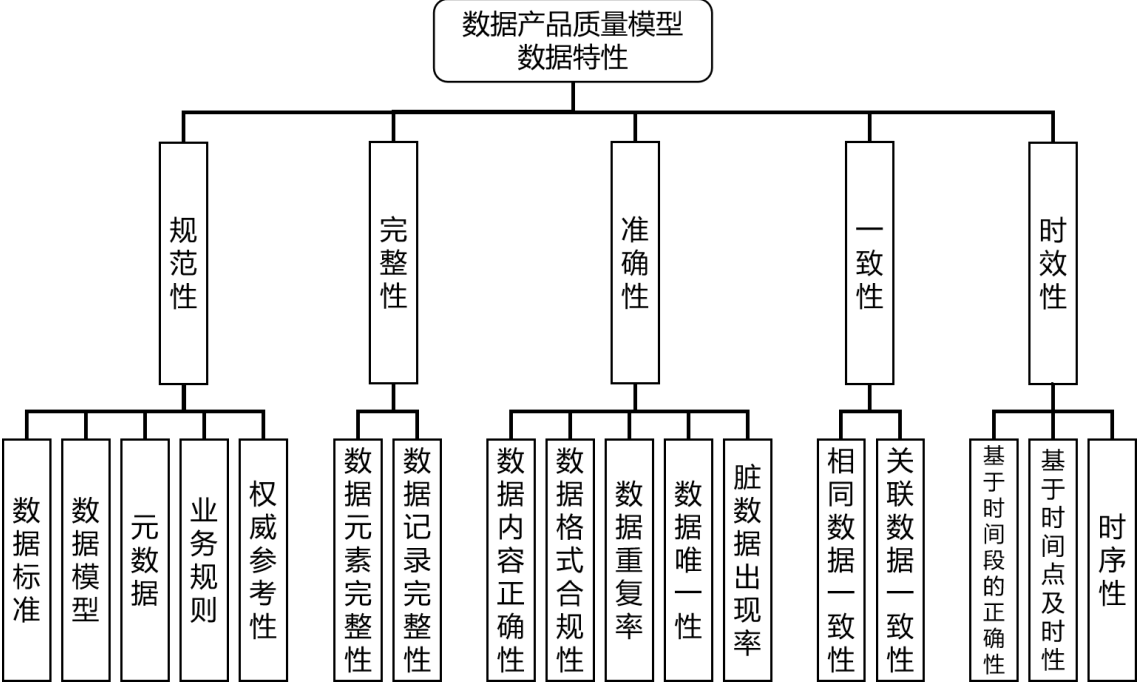


图2 数据特性分解结构

4.2.3 产品特性

产品特性是指数据产品被封装、加工后，作为一种可供流通的标准化商品所固有的质量属性。数据产品质量模型的产品特性的分解结构如下图所示，详细内容参见第6.3节。

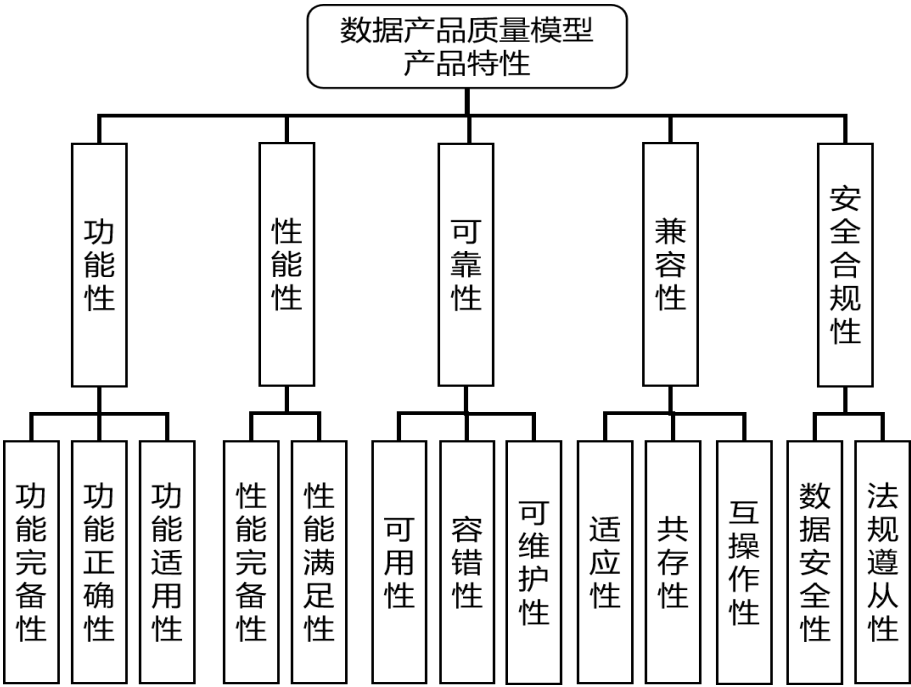


图3 产品特性分解结构

4.2.4 使用特性

使用特性是指数据产品在用户的特定应用场景与使用过程中，为满足其业务目标而呈现出的效用性质量属性，关注的是用户的实际体验与效果。数据产品质量模型的使用特性的分解结构如下图所示，详细内容参见第6.4节。

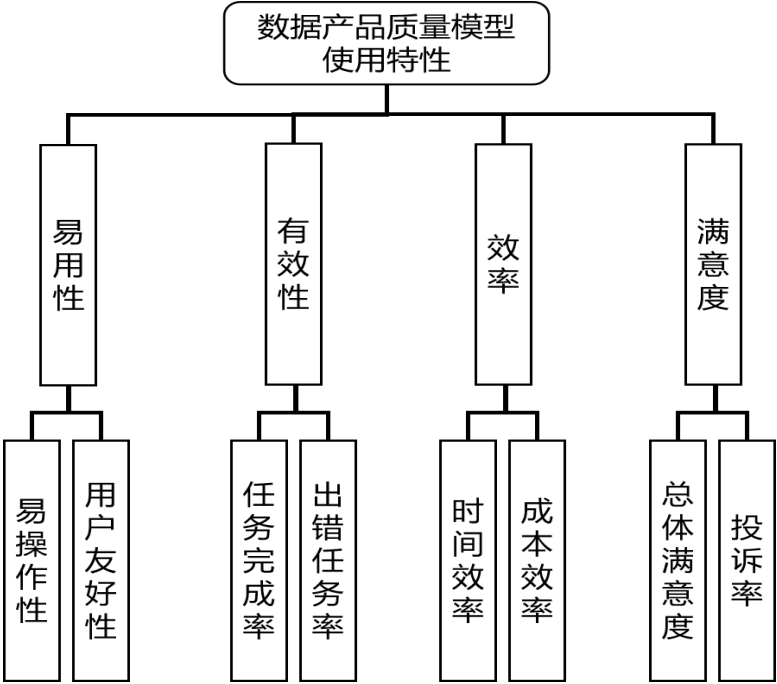


图4 使用特性分解结构

4.2.5 流通特性

数据产品流通特性，是指数据产品在交易、共享、交付与应用等流通过程中，所固有的、用以决定其能否高效、安全、合规且可持续地实现价值交换的一系列质量属性。数据产品质量模型的流通特性的分解结构如下图所示，详细内容参见第6.5节。

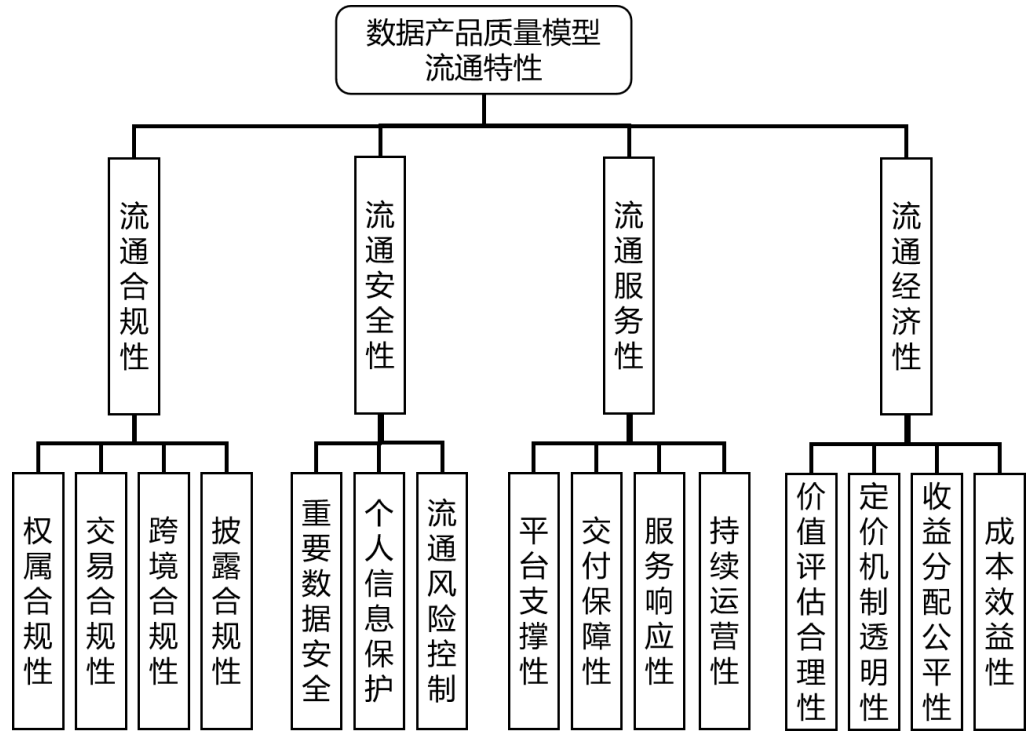


图5 流通特性分解结构

4.3 数据产品质量评价方法

数据产品质量评价方法是开展数据产品质量评价，对产品质量模型中质量指标进行赋值的机制和方法，包括评价机制（如定量、定性或综合）、抽样方法（如完全评价、抽样评价）等，详情参加第5章。

4.4 数据产品质量评价过程

数据产品质量评价是为获得对数据产品质量评价结果而实施的一组相互关联或相互作用的活动，包括确定指标、确定权重、分析计算、结果评价等，详情参见第6章。

5 数据产品质量评价方法

5.1 评价机制

5.1.1 评价机制的类型

评价分为定性评价和定量评价形式进行。

- a) 定性评价。定性评价通过对用户的访谈、观察、行业分析等对数据产品质量特性进行描述性分析，以非数值的形式来评数据产品的质量。
- b) 定量评价。定量评价通过测量数据、统计指标等客观数据和统计数据来刻画数据产品质量特性，对数据产品进行量化评价。

- c) 综合方法。综合方法将定性和定量两种方法有机的集合起来，从客观和主观两个方面对数据产品质量特性进行评测。例如，采用层次分析法先主观的将评估对象进行优劣和层级排序，再利用定量的形式对评估主体进行评测，最后根据层级结构进行加权，得到最终的评测结果。

### 5.1.2 定性评价的适用情况

数据产品质量定性评价机制的可适用于下列情况。

- a) 性质特征判断。当关注的數據产品质量方面侧重于性质、特征等难以用具体数值直接衡量的属性时，例如数据产品的合规性、用户满意度等，宜采用定性评价。
- b) 初步评价需求。在数据产品质量评价的初期阶段，旨在快速获取一个大致的、方向性的判断时，可运用定性评价方式来初步了解数据产品质量概况，如对新收集的数据来源的可靠性进行初步核查，宜通过查看数据提供方的资质、收集方法的合理性等定性因素来形成初步判断。

### 5.1.3 定量评价的适用情况

数据产品质量定量评价机制的可适用于下列情况。

- a) 精确衡量需求。若能够获取可量化的数据指标，且需要精确衡量数据产品质量各维度的水平，例如准确统计数据记录中的错误率来衡量准确性、计算缺失值数量占总数据量的比例以评估完整性、通过平均更新周期量化数据的时效性等情况时，宜选用定量评价方式。
- b) 对比与跟踪需求。在需要对不同基底数据集的数据产品质量进行横向对比分析，或者跟踪同一基底数据集的数据产品质量随时间变化的纵向情况，以及依据具体数值设定数据产品质量改进目标并衡量改进成效时，宜选用定量评价。

## 5.2 抽样方法

### 5.2.1 抽样方法的类型

抽样方法分为完全评价和抽样评价。

- a) 完全评价。完全评价适用于评价对象规模相对较小时，对评价对象的所有元素进行评价。
- b) 抽样评价。抽样检查适用于评价对象规模相对较大时，先对评价对象进行采样，然后评价数据产品质量，以样本基底数据产品质量代表整体数据产品质量。评价团队可根据基底数据总体规模以及评价精度的要求，选择简单随机抽样、分层抽样、系统抽样、整群抽样等方式。

### 5.2.2 完全评价适用情况

在数据质量评价的过程中，完全评价可适用以下几种情况。

- a) 基底数据规模较小。当待评价的数据总体规模相对较小，在人力、时间、计算资源等条件允许的情况下，对全部基底数据进行评价不会造成过高成本和过大负担，适合采用完全评价方式。
- b) 高风险决策依据。若数据产品质量的评价结果将作为高风险决策的核心依据，需要确保对数据产品质量有绝对精准、全面的把控，不容许存在因抽样可能带来的偏差，此时宜采用完全评价。

### 5.2.3 抽样评价适用情况

在数据产品质量评价的过程中，抽样评价可适用以下几种情况。

- a) 大规模基底数据。面对海量基底数据，若进行完全评价将耗费大量的人力、物力、时间以及计算资源，且从实际操作可行性角度来看几乎难以实现全面检测，此时抽样评价是更为可行的选择。
- b) 资源与时间限制。当在人力、时间、资金等资源方面存在明显限制，无法支持对全部基底数据进行逐一评价，而又需要在规定时间内获取数据产品质量的大致结论时，可采用抽样评价方式。

6 数据产品质量特性及测量

6.1 特性编码

6.1.1 编码规则

特性代码是数据产品质量特性的唯一性编号，采用层次编码方法，编码位数为6位，按照一级特性、二级特性和三级特性的从属关系顺序编码。每一级特性代码分别用2位阿拉伯数字表示。编码规则如图6所示。

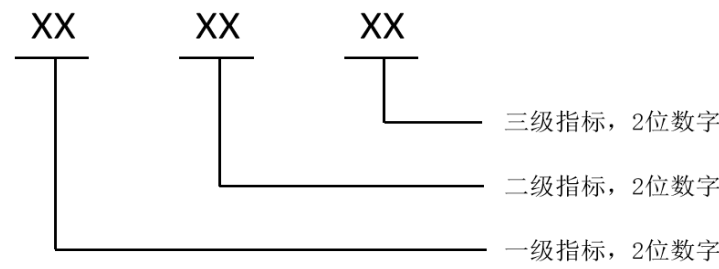


图6 编码规则

6.1.2 一级特性代码

第1位和第2位一级特性代码由2位阿拉伯数字组成。一级特性代码及含义见表1。

表1 一级特性代码及含义

代码	含义
01	数据特性
02	产品特性
03	使用特性
04	流通特性

6.1.3 二级特性代码

第3位和第4位二级特性代码由2位阿拉伯数字组成，编码范围为01～99。

6.1.3 三级特性代码

第3位和第4位二级特性代码由2位阿拉伯数字组成，编码范围为01～99。

6.2 数据特性

6.2.1 规范性

规范性指数据符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度。规范性评价特性及测量方法定义见表2。

表2 规范性评价特性

特性编号	特性名称	特性描述	测量方法
010101	数据标准	数据符合数据标准的度量。	$X=A/B$ , 式中： A=满足数据标准要求的数据集中

表2 规范性评价特性（续）

特性编号	特性名称	特性描述	测量方法
			元素的个数； B=被评价的数据集中应该满足数据标准的数据集中元素的个数。
010102	数据模型	数据符合数据模型的度量。	$X=A/B$ , 式中： A=满足数据模型要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010103	元数据	数据符合元数据定义的度量。	$X=A/B$ , 式中： A=满足元数据定义的数据集中元素的个数； B=被评价的数据集中元素的个数。
010104	业务规则	数据符合业务规则的度量。	$X=A/B$ , 式中： A=满足业务规则的数据集中元素的个数； B=被评价的数据集中元素的个数。
010105	权威参考性	参考数据是系统、应用软件、数据库、流程、报告及交易记录 and 主记录用来参考的数值集合或分类表。	$X=A/B$ , 式中： A=满足参考数据规则的数据集中元素的个数； B=被评价的数据集中元素的个数。

6.2.2 完整性

完整性指按照数据规则要求,数据元素被赋予数值的程度。完整性评价特性及测量方法定义见表3。

表3 完整性评价特性

特性编号	特性名称	特性描述	测量方法
010201	数据元素完整性	按照业务规则要求,数据集中应被赋值的数据元素的赋值程度。	$X=A/B$ , 式中： A=被赋值的数据集中元素的个数； B=预期被赋值的数据集中元素的个数。
010202	数据记录完整性	按照业务规则要求,数据集中应被赋值的数据记录的赋值程度。	$X=A/B$ , 式中： A=被赋值的数据集中记录的个数； B=预期被赋值的数据集中记录的个数。

6.2.3 准确性

准确性指数据准确表示其所描述的真实实体(实际对象) 真实值的程度。准确性评价特性及测量方法定义见表4。

表4 准确性评价特性

特性编号	特性名称	特性描述	测量方法
------	------	------	------

表4 准确性评价特性（续）

特性编号	特性名称	特性描述	测量方法
010301	数据内容正确性	数据内容是否为预期数据。	$X=A/B$ , 式中： A=满足数据正确性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010302	数据格式合规性	数据格式(包括数据类型、数值范围、数据长度、精度等)是否满足预期要求。 示例：性别一栏不能出现男/女以外的内容；身份证号不能出现标点符号；以及对字符编码的一些限制,都需要通过规定内容的格式来实现。	$X=A/B$ , 式中： A=满足格式要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010303	数据重复率	特定字段、记录、文件或数据集重复的度量。	$X=A/B$ , 式中： A=重复的数据集中元素的个数； B=被评价的数据集中元素的个数。
010304	数据唯一性	特定字段、记录、文件或数据集唯一性的度量。	$X=A/B$ , 式中： A=满足唯一性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010305	脏数据出现率	正确字段、记录、文件或数据集之外无效数据的度量。 示例：事务发生回滚时由于回滚机制不健全或不完善导致可能出现脏数据。	$X=A/B$ , 式中： A=有脏数据出现的数据集中元素的个数； B=被评价的数据集中元素的个数。

6.2.4 一致性

一致性指数据与其他特定上下文中使用的数据无矛盾的程度。一致性评价特性及测量方法定义见表5。

表5 一致性评价特性

特性编号	特性名称	特性描述	测量方法
010401	相同数据一致性	同一数据在不同位置存储或被不同应用或用户使用时,数	$X=A/B$ 式中： A=满足一致性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010402	关联数据一致性	根据一致性约束规则检查关联数据的一致性。	$X=A/B$ 式中： A=满足一致性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。

6.2.5 时效性

时效性指数据在时间变化中的正确程度。时效性评价特性及测量方法定义见表6。

表6 时效性评价特性

特性编号	特性名称	特性描述	测量方法
010501	基于时间段的正确性	基于日期范围的记录数或频率分布符合业务需求的程度	$X=A/B$ 式中： A=满足有效性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010502	基于时间点及时性	基于时间戳的记录数、频率分布或延迟时间符合业务需求的程度	$X=A/B$ 式中： A=满足及时性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。
010503	时序性	数据集中同一实体的数据元素之间的相对时序关系	$X=A/B$ 式中： A=满足时序性要求的数据集中元素的个数； B=被评价的数据集中元素的个数。

6.3 产品特性

6.3.1 功能性

功能性指在指定条件下使用时，数据产品提供满足明确和隐含要求的功能的程度。功能性评价特性及测量方法定义见表 7。

表7 功能性评价特性

特性编号	特性名称	特性描述	测量方法
020101	功能完备性	数据产品功能集对指定的任务 and 用户目标的覆盖程度。	$X=1-A/B$ A=缺少的功能数量； B=指定的功能数量。
020102	功能正确性	数据产品提供具有所需精度的正确的结果的程度。	$X=1-A/B$ A=功能不正确的数量； B=考虑的功能数量。
020103	功能适用性	功能促使指定的任务和目标实现的程度。	$X=1-A/B$ A=为实现特定使用目标所需的功能中缺少或不正确的功能的数量； B=为实现特定使用目标所需的功能数量。

6.3.2 性能性

性能性在指定条件下使用时，数据产品提供满足明确和隐含要求的性能的程度。性能性评价特性及测量方法定义见表 8。



表8 性能性评价特性

特性编号	特性名称	特性描述	测量方法
020201	性能完备性	数据产品性能指标对概率产品性能指标集的覆盖程度。	$X=1-A/B$ A=缺少的性能指标数量； B=指定的性能指标数量。
020202	性能满足性	数据产品性能指标满足标准、约定或法规规定要求的程度。	$X=1-A/B$ A=不合要求的性能指标数量； B=总的性能指标数量。

### 6.3.3 可靠性

可靠性指数据产品在指定条件下、指定时间内执行指定功能的程度。可靠性评价特性及测量方法定义见表9。

表9 可靠性评价特性

特性编号	特性名称	特性描述	测量方法
020301	可用性	产品在设定有效生存周期内的可使用程度	$X=A/B$ A=产品实际提供的可用时间； B=产品计划中规定的可用时间。
020302	容错性	对抗异常输入、网络波动、数据异常不崩溃的能力程度	$X=A/B$ A=异常出现时避免失效的数量； B=出现的异常总数。
020303	可维护性	数据产品能够被预期的维护人员配置、修改的有效性和效率的程度。	$X=A/B$ A=在评价中证实可提供的维护性 依从项数； B=总的可维护性要求项数。

### 6.3.4 兼容性

兼容性指在共享相同的硬件或软件环境的条件下，数据产品能够与其他产品、系统或组件和/或执行其所需的功能的程度。兼容性评价特性及测量方法定义见表10。

表 10 兼容性评价特性

特性编号	特性名称	特性描述	测量方法
020401	适应性	数据产品能够有效地、有效率地适应不同的或演变的硬件、软件、或者其他硬性环境下的程度。	$X=A/B$ A=该产品可适应的软硬件环境数量； B=该产品需要适应的软硬件环境数量。
020402	共存性	在与其他产品共享通用的环境和资源的条件下，数据产品能够有效执行其所需的功能并且不会对其他产品造成负面影响的程度。	$X=A/B$ A=与该产品可共存的其他规定的其他产品的数量； B=在运行环境中，该产品需要与其他软件共存的数量。

表 10 兼容性评价特性（续）

特性编号	特性名称	特性描述	测量方法
020403	互操作性	两个或多个系统、产品或组件能够交换数据并使用已交换的数据的程度。	$X=A/B$ A=与该产品可互操作的其他规定的其他产品的数量； B=在运行环境中，该产品需要与其他产品互操作的数量。

6.3.5 安全合规性

安全合规性指数据产品满足数据安全保护需求以及法规要求的程度。安全合规性评价特性及测量方法定义见表11。

表11 安全合规性评价特性

特性编号	特性名称	特性描述	测量方法
020501	数据安全性	数据访问控制、脱敏、加密、隐私保护等需求得到满足的程度。	$X=1-A/B$ A=与该产品不满足的数据安全要求的数量； B=与该产品应满足的数据安全要求的数量。
020502	法规依从性	数据产品遵行法规、约定、相关标准以及类似规定的程度。	$X=1-A/B$ A=与该产品不满足的合规要求的项数； B=与该产品应满足的合规要求的项数。

6.4 使用特性

6.4.1 易用性

易用性指在指定的使用周境中，数据产品在有效性、效率和满意度特性方面为了指定的目标可为指定用户使用的程度。易用性评价特性及测量方法定义见表12。

表12 易用性评价特性

特性编号	特性名称	特性描述	测量方法
030101	易操作性	数据产品易于操作和控制的程度。	$X=A/B$ A=易操作性调查问卷总得分； B=问卷数×满分； 注：问卷调查法。
030102	用户友好性	产品提供令人舒适的交互的程度。	$X=A/B$ A=用户友好性调查问卷总得分； B=问卷数×满分； 注：问卷调查法。

6.4.2 有效性

有效性指用户实现制定目标的准确性和完备性。有效性评价特性及测量方法定义见表13。

表13 有效性评价特性

特性编号	特性名称	特性描述	测量方法
030201	任务完成率	正确完成任务的比例	$X=A/B$ A=已完成的任务数量； B=所尝试的任务总数。
030202	出错任务率	出现用户所造成差错的任务比例	$X=A/B$ A=出错任务数量； B=任务总数。

#### 6.4.3 效率

效率指与用户实现目标的准确性和完备性相应的资源消耗。效率评价特性及测量方法定义见表14。

表14 效率评价特性

特性编号	特性名称	特性描述	测量方法
030301	时间效率	用户在使用数据产品时实现目标的效率	$X=A/T$ A=实现的目标数量； T=时间； 注：用户表现测量法。
030302	成本效率	用户的成本效率	$X=A/B$ A=执行任务的总成本； B=实现的目标数量； 注：用户表现测量法。

#### 6.4.4 满意度

满意度指数据产品在指定的使用周境中使用时，用户的要求被满足的程度。满意度评价特性及测量方法定义见表15。

表15 满意度评价特性

特性编号	特性名称	特性描述	测量方法
030401	总体满意度	用户总体的满意度。	$X=A/B$ A=问卷总得分数； B=问卷数×满分； 注：问卷调查法。
030402	投诉率	发出投诉的用户比例	$X=1-A/B$ A=投诉用户数； B=总用户数。

### 6.5 流通特性

#### 6.5.1 流通合规性

流通合规性指数据产品在流通过程中符合数据交易流通法律、法规与交易规则的程度。流通合规性评价特性及测量方法定义见表16。

表16 流通合规性评价特性

特性编号	特性名称	特性描述	测量方法
040101	权属合规性	数据产品持有权、使用权、经营权的清晰界定与合法合规的程度	$X=A/B$ A=合法、清晰界定的权属项数； B=全部权属合规项数。
040102	交易合规性	符合数据交易场所规则、合同备案规程、价格形成机制等交易规范的程度	$X=A/B$ A=实际符合的交易合规要求项数； B=应遵守的交易合规要求的总项数。
040103	跨境合规性	满足数据跨境传输相关法规要求的程度	$X=A/B$ A=实际满足的跨境传输合规要求项数； B=应满足的跨境传输合规要求总数。
040104	披露合规性	产品描述、质量证书、外部性说明等信息的真实完整披露的程度	$X=A/B$ A=真实、完整披露信息的项数； B=应披露信息的总项数。

6.5.2 流通安全性

流通安全性指保障数据产品在交易流通过程中的安全可控与隐私保护的程度。流通安全性评价特性及测量方法定义见表17。

表17 流通安全性评价特性

特性编号	特性名称	特性描述	测量方法
040201	重要数据安全	重要数据交易流通安全要求落实程度	$X=A/B$ A=重要数据交易流通实际符合的安全要求项数； B=重要数据交易流通应满足的安全要求总项数。
040202	个人信息保护	个人信息交易流通过标识化、匿名化处理、最小必要原则等落实的程度	$X=A/B$ A=个人信息交易流通实际满足的保护要求项数； B=个人信息交易流通应满足的保护要求总项数。
040203	流通安全风险控制	数据泄露、滥用、篡改等安全风险的识别与防范机制	$X=A/B$ A=实际识别并有效防范的安全风险项数； B=应识别和防范的安全风险总项数。

### 6.5.3 流通服务性

流通服务性是衡量支撑数据产品高效交付与使用的技术保障与服务能力提供的程度。流通服务性特性及测量方法定义见表18。

表18 流通服务性评价特性

特性编号	特性名称	特性描述	测量方法
040301	平台支撑性	交易平台健壮性、吞吐量、接口稳定性等技术保障	$X=A/B$ A=交易平台实际提供的能满足要求的技术保障项目数； T=交易平台应提供的技术保障项目总数。
040302	交付保障性	多种交付方式（API、介质、服务化）的支持与可靠性	$X=A/B$ A=交易平台实际可提供的可靠的交付方式数； B=交易平台应提供的交付方式总数。
040303	服务响应性	问题处理、技术支持、售后服务的及时性与有效性	$X=A/B$ A=数据产品提供方实际及时有效提供服务的项目数； B=数据产品提供方应提供的服务总项目数。
040304	持续运营性	数据更新、版本迭代、长期维护的承诺与执行	$X=A/B$ A=数据产品提供方实际有效提供的持续运营项目数； B=数据产品提供方应提供的持续运营总项目数。

### 6.5.4 流通经济性

流通经济性是评估数据产品在流通中的价值合理性、定价透明性与收益公平性的特性。流通经济性评价特性及测量方法定义见表19。

表19 流通经济性评价特性

特性编号	特性名称	特性描述	测量方法
040401	价值评估合理性	数据产品价值评估方法科学、参数合理	$X=A/B$ A=实际符合规范的价值评估项数； B=应评估项总数。
040402	定价机制透明性	价格形成机制公开透明、可解释	$X=A/B$ A=公开、可解释的定价要素数； B=定价要素总数。
040403	收益分配公平性	数据贡献者、加工者、经营者收益分配合理	$X = 1 - \frac{\sum  r_i - w_i }{2}$

表19 流通经济性评价特性（续）

特性编号	特性名称	特性描述	测量方法
			$X$ : 公平性指数, 采用基尼系数简化版, 0-1, 越接近1越公平; $r_i$ : 某方实际收益占比; $w_i$ : 某方贡献权重(按投入、成本、价值贡献确定)。
040404	成本效益性	流通成本可控、效益可量化评估	$X=A/B$ $A$ =预期总收益; $B$ =实际总成本。

7 数据产品质量评价过程

7.1 评价准备

数据产品质量了评价前, 应确定评价对象, 结合具体的应用领域和应用场景, 明确数据产品质量需求, 确立评价目的, 确认评价方法。

7.2 方案编制

7.2.1 概述

开展评价方案编制和确认, 评价方案主要包括根据评价需求选取合适的评价指标, 确定相关指标的权重, 明确质量等级评定规则。

7.2.2 指标选取

质量评价指标宜从第6章规定的数据产品通用质量评价特性和附录A规定的面向机器学习的数据产品质量特性选取, 应结合评价对象、业务领域、应用场景和评价需求, 确定采用的质量评价特性。可结合评价的实际需要, 增加个性化指标。

7.2.3 权重分析

对于选定的质量评价特性指标, 可根据数据产品应用场景中的重要程度、数据产品质量导致问题的影响程度等维度确定各级指标权重。

7.2.4 等级划分

可根据需要, 划分质量等级标准(例如, 产品质量是否达标、属于哪一级、是否可使用等), 给出相应等级划分阈值。

7.3 评价实施

结合数据产品质量特性测量和权重进行综合质量评分计算。

综合评分按公式(1)通过一级特性和相应权重计算得出。

$$Q = \sum_{i=1}^4 w1_i \times q1_i \dots\dots\dots (1)$$

式中:

$Q$ ——为数据产品质量综合评价得分;

$w1_i$ ——第*i*项一级质量特性计算权重；

$q1_i$ ——第*i*项一级质量特性测量得分；

*i*——取值范围为1~4，分别对应数据产品4项一级质量特性。

每一个一级质量特性按公式（2）通过二级质量特性和相应权重计算得出。

$$q1_i = \sum_{j=1}^n w2_j \times q2_j \dots\dots\dots (2)$$

式中：

$q1_i$ ——第*i*项一级质量特性测量得分；

$w2_j$ ——第*j*项二级质量特性计算权重；

$q2_j$ ——第*j*项二级质量特性测量得分；

*j*——取值范围为1~*n*，分别对应第*i*项一级质量特性分解的*n*项二级质量特性。

类似的，每一个二级质量特性如果进行了分解，可按公式（3）通过三级质量特性和相应权重计算得出。

$$q2_j = \sum_{k=1}^m w3_k \times q3_k \dots\dots\dots (3)$$

式中：

$q2_j$ ——第*j*项二级质量特性测量得分；

$w3_k$ ——第*k*项三级质量特性计算权重；

$q3_k$ ——第*k*项三级质量特性测量得分；

*k*——取值范围为1~*m*，分别对应第*j*项二级质量特性分解的*m*项三级质量特性。

以此类推，上级质量特性测量可以通过分解的下级质量特性测量计算得出。

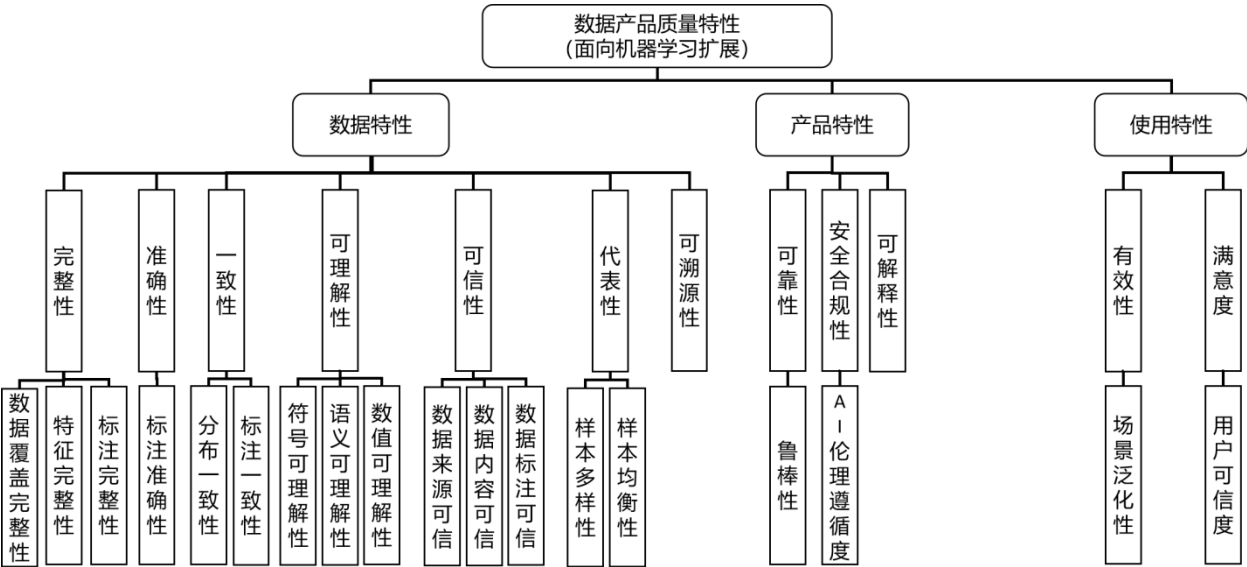
#### 7.4 结果报告

根据得到的数据产品质量综合评分和质量等级划分标准，给出数据产品质量进行评价结论，编制并提交质量评价报告。

附录 A  
(规范性附录)  
面向机器学习的数据产品质量特性扩展

A.1 扩展特性结构

传统的面向分析统计等应用场景的数据产品质量特性不能很好地满足面向机器学习等AI应用场景的质量评价需要，对面向机器学习的数据产品质量评价特性扩展如下图所示。



图A.1 面向机器学习的数据产品质量特性扩展

A.2 扩展特性描述与测量

面向机器学习的数据产品质量特性及测量方法定义见表A.1。

表 A.1 面向机器学习的数据产品质量评价特性扩展

一级特性名称/编号	二级特性名称/编号	三级特性名称/编号	特性描述	测量方法
数据特性 (01)	完整性 (0102)	数据覆盖完整性 (010203)	数据样本覆盖不同领域的程度。	$X=A/B$ A=数据样本实际覆盖的领域数量 B=数据样本计划覆盖的领域数量
		特征完整性 (010204)	具备某一特征数据项在全部数据项中所占的比例	$X=A/B$ A=数据样本实际具备的特征数量 B=数据样本计划覆盖的特征数量
		标注完整性 (010205)	未标注或未完整标注数据样本所占比例。	$X=A/B$ A=数据样本实际完成标注的数量 B=数据样本计划完成标注的数量



表 A.1 面向机器学习的数据产品质量评价特性扩展（续）

一级特性名称/编号	二级特性名称/编号	三级特性名称/编号	特性描述	测量方法
数据特性 (01)	准确性 (0103)	标注准确性 (010306)	对于有监督机器学习，数据集元素被正确标注的程度	$X=1-A/B$ A=数据样本无标注或标注不准确的数量 B=数据样本计划标注的数量
	一致性 (0104)	分布一致性 (010401)	样本分布一致的程度	$X=KL(P  Q)$ KL(P  Q): P和Q之间的KL-散度。 P=被评估数据集的概率分布 Q=参考模型的概率分布
		标注一致性 (010402)	相似数据项标注一致的程度	$X=A/B$ A=满足标注一致性要求的样本数量 B=样本总数
	可理解性 (0106)	符号可理解性 (010601)	数据样本中符号可理解程度	$X=A/B$ A=符号可理解的数量 B=样本总数
		语义可理解性 (010602)	数据样本中语义可理解程度	$X=A/B$ A=语义可理解的数量 B=样本总数
		数值可理解性 (010603)	数据样本中取值可理解程度	$X=A/B$ A=数值可理解的数量 B=样本总数
	可信性 (0107)	数据来源可信 (010701)	数据来源可信的程度	$X=A/B$ A=数据来源可信的数量 B=样本总数
		数据内容可信 (010702)	数据内容可信的程度	$X=A/B$ A=内容可信的数量 B=样本总数
		数据标注可信 (010703)	数据标注可信的程度	$X=A/B$ A=数据标注可信的数量 B=样本总数
	代表性 (0108)	样本多样性 (010801)	被评估的数据样本每个类别带来的信息量	$X=A/B$ A=实际覆盖样本数量 B=理论应具备的样本总数
		样本均衡性 (010802)	数据中不同类别的样本数量是否均衡	$X=A/B$ A=最小类别样本量 B=最大类别样本量
	可溯源性 (0109)	—	指数据来源可查、路径可追、操作可验、责任可定的程度	$X=A/B$ A=能够通过标识符、元数据等信息明确关联到其直接或关键原始数据源的数据记录数 B=被评价的数据产品的总记录数

表 A.1 面向机器学习的数据产品质量评价特性扩展（续）

一级特性名称/编号	二级特性名称/编号	三级特性名称/编号	特性描述	测量方法
产品特性 (02)	可靠性 (0203)	鲁棒性 (020304)	抗漂移、抗噪声等的程度	$X=A/B$ A=异常出现时避免失效的数量 B=出现的异常总数
	安全合规性 (0205)	AI伦理遵循度 (020503)	过程中立，无偏见无歧视，无违法、违规、诱导、暴力、歧视性内容等伦理合规程度	$X=A/B$ A=实际遵循的AI伦理合规项数 B=要求遵循的AI伦理合规项数
	可解释性 (0206)	—	数据产品输出结果可解释的程度。	$X=A/B$ A=可解释结果数量 B=输出结果总数
使用特性 (03)	有效性 (0302)	场景泛化性 (030203)	用户在不同领域场景中使用产品可行性	$X=A/B$ A=实际可使用的领域场景数量 B=计划可使用的领域场景数量
	满意度 (0304)	用户可信性 (030403)	用户对产品输出结果的信任程度	$X=A/B$ A=从可信性问卷得出的心理测量值 B=问卷数×满分 注：问卷调查法

附录 B  
(资料性附录)  
数据产品质量评价示例

**B.1 数据产品与质量评价背景介绍**

以XX数据产品为例，开展数据产品质量评价。数据产品描述如下：

**B.2 质量评价准备**

评价准备工作如下：

- a)
- b)

**B.3 质量评价方案编制**

**B.2.1 质量指标选定**

如下：

**B.2.2 质量权重分析**

如下：

**B.2.3 质量等级划分**

如下：

**B.4 质量评价实施**

如下：

**B.5 质量结果报告**

如下：



## 参 考 文 献

- [1] GB/T 25000.10-2016 (ISO/IEC 25010: 2011) 系统与软件工程 系统与软件质量要求和评价 (SQuaRE) 第10部分 系统与软件质量模型
- [2] GB/T 25000.22-2019 (ISO/IEC 25022: 2016) 系统与软件工程 系统与软件质量要求和评价 (SQuaRE) 第22部分 使用质量测量
- [3] GB/T 25000.23-2019 (ISO/IEC 25023: 2016) 系统与软件工程 系统与软件质量要求和评价 (SQuaRE) 第23部分 系统与软件产品质量测量
- [4] GB/T 25000.24-2017 (ISO/IEC 25024: 2015) 系统与软件工程 系统与软件质量要求和评价 (SQuaRE) 第24部分: 数据质量测量
- [5] ISO/IEC 5259-2: 2024 Artificial intelligence — Data quality for analytics and machine learning (ML) Part 2: Data quality measures
- [6] 国家数据局《数据领域常用名词解释》(第一批)
- [7] YD/T 4522-2023 面向机器学习的电信数据规范 数据质量
- [8] YD/T 6486-2025 面向人工智能的数据集质量通用评估方法 总体要求
- [9] DB31/T 1523-2024 公共数据质量评价要求
-