

ISO/IEC 5259-2 데이터 품질표준 요약표

- 작성: 페블러스 데이터 커뮤니케이션 팀
- 일자: 2025-11-13
- 버전: 1.0

문서에 대하여 (About this Document)

본 문서는 ISO/IEC 5259-2 표준에 명시된 데이터 품질 측정 기준(QM) 목록을 정리한 표를 담고 있다. 각 목록은 데이터 품질을 평가하는 네 가지 주요 특성 그룹으로 분류되며, 각 측정 기준은 구체적인 데이터 품질 개념과 목적을 한국어와 영어로 병기하여 설명한다. 첫 번째 그룹인 내재적 데이터 품질 특성은 정확성, 완전성, 일관성 등 데이터 자체의 본질적인 속성을 다루며, 두 번째와 세 번째 그룹은 시스템 의존성 여부에 따른 특성들(예: 접근성, 효율성, 가용성)을 포함하고 있다. 특히, 마지막 섹션인 추가 데이터 품질 특성에서는 분석 및 기계 학습(ML) 환경의 데이터셋 충실도를 평가하기 위한 감사 가능성, 균형, 다양성 등의 중요한 기준들을 제시하고 있다.

1. 내재적 데이터 품질 특성 (Inherent Data Quality Characteristics)

소분류 (Sub-classification)	QM ID	QM 항목 (Name) (한/영 병기)	QM 한 줄 설명/개념
정확성 (Accuracy)	Acc-ML-1	Syntactic data accuracy (구문 데이터 정확성)	구문적으로 정확한 데이터 값 집합에 대해 데이터 값이 얼마나 가까운지를 측정합니다.
정확성 (Accuracy)	Acc-ML-2	Semantic data accuracy (의미 데이터 정확성)	의미적으로 정확한 데이터 값 집합에 대해 데이터 값이 얼마나 가까운지를 측정합니다.
정확성	Acc-	Data accuracy assurance (데이	데이터가 정확하다고 보장되는 정도

(Accuracy)	ML-3	터 정확성 보증)	를 측정합니다.
정확성 (Accuracy)	Acc-ML-4	Risk of dataset inaccuracy (데이터셋 부정확성 위험)	데이터셋의 부정확성으로 인해 발생할 수 있는 잠재적 위험을 측정합니다.
정확성 (Accuracy)	Acc-ML-5	Data model accuracy (데이터 모델 정확성)	데이터 모델이 데이터의 실제 특성을 얼마나 정확하게 표현하는지 측정합니다.
정확성 (Accuracy)	Acc-ML-6	Data accuracy range (데이터 정확성 범위)	데이터 값의 정확성이 허용되는 범위를 측정합니다.
정확성 (Accuracy)	Acc-ML-7	Data label accuracy (데이터 라벨 정확성)	데이터셋 내 각 요소에 라벨이 정확하게 할당되었는지 측정합니다.
완전성 (Completeness)	Com-ML-1	Value completeness (값 완전성)	데이터 항목의 전체 수 대비 널(null) 값이 없는 데이터 항목의 비율을 측정합니다.
완전성 (Completeness)	Com-ML-2	Value occurrence completeness (값 발생 완전성)	주어진 데이터 값의 발생 횟수와 데이터 품질 요구사항에 명시된 예상 발생 횟수의 비율을 측정합니다.
완전성 (Completeness)	Com-ML-3	Feature completeness (특징 완전성)	특정 특징(feature)과 관련된 데이터 항목 중 널 값이 없는 데이터 항목의 비율을 측정합니다.
완전성 (Completeness)	Com-ML-4	Record completeness (레코드 완전성)	데이터 레코드의 전체 수 대비 비어 있지 않은(non-empty) 데이터 레코드의 비율을 측정합니다.
완전성 (Completeness)	Com-ML-5	Label completeness (라벨 완전성)	데이터셋 내에서 라벨이 누락되거나 불완전하게 라벨링된 샘플의 비율을 측정합니다.
일관성	Con-	Data record consistency (데	데이터셋 내 중복된 데이터 레코드

(Consistency)	ML-1	이터 레코드 일관성)	의 비율을 측정합니다.
일관성 (Consistency)	Con-ML-2	Data label consistency (데이터 라벨 일관성)	유사한 데이터 항목에 동일한 라벨이 할당된 정도를 측정합니다.
일관성 (Consistency)	Con-ML-3	Data format consistency (데이터 포맷 일관성)	데이터 항목들이 데이터 포맷 일관성 요구사항을 충족하는 정도를 측정합니다.
일관성 (Consistency)	Con-ML-4	Semantic consistency (의미 일관성)	데이터 항목들이 의미 일관성 요구사항을 충족하는 정도를 측정합니다.
신뢰성 (Credibility)	Cre-ML-1	Values credibility (값 신뢰성)	데이터 값의 신뢰성을 측정합니다.
신뢰성 (Credibility)	Cre-ML-2	Source credibility (출처 신뢰성)	데이터 출처의 신뢰성을 측정합니다.
신뢰성 (Credibility)	Cre-ML-3	Data dictionary credibility (데이터 사전 신뢰성)	데이터 사전의 신뢰성을 측정합니다.
신뢰성 (Credibility)	Cre-ML-4	Data model credibility (데이터 모델 신뢰성)	데이터 모델의 신뢰성을 측정합니다.
최신성 (Currentness)	Cur-ML-1	Feature currentness (특징 최신성)	특징(feature)에 대해 허용 가능한 날짜 범위 내에 있는 데이터 항목의 비율을 측정합니다.
최신성 (Currentness)	Cur-ML-2	Record currentness (레코드 최신성)	데이터 레코드 내 모든 데이터 항목이 요구되는 연령 범위 내에 속하는 레코드의 비율을 측정합니다.

2. 내재적 및 시스템 의존적 데이터 품질 특성 (Inherent and

System-dependent Characteristics)

소분류 (Sub-classification)	QM ID	QM 항목 (Name) (한/영 병기)	QM 한 줄 설명/개념
접근성 (Accessibility)	Acs-ML-1	User accessibility (사용자 접근성)	ISO/IEC 25024:2015, Table 6.1에 정의된 대로 사용자 접근성을 측정합니다.
접근성 (Accessibility)	Acs-ML-2	Data format accessibility (데이터 포맷 접근성)	ISO/IEC 25024:2015, Table 6.2에 정의된 대로 데이터 포맷 접근성을 측정합니다.
접근성 (Accessibility)	Acs-ML-3	Data accessibility (데이터 접근성)	데이터셋 내 접근 가능한 레코드의 비율을 측정합니다.
규정 준수 (Compliance)	Cmp-ML-1	Data item compliance (데이터 항목 규정 준수)	데이터 항목이 법규, 표준, 규칙 등 규정 준수 요구사항을 충족하는 정도를 측정합니다.
효율성 (Efficiency)	Eff-ML-1	Data format efficiency (데이터 포맷 효율성)	ISO/IEC 25024:2015, Table 9.2에 정의된 대로 데이터 포맷의 효율성을 측정합니다.
효율성 (Efficiency)	Eff-ML-2	Data processing efficiency (데이터 처리 효율성)	ISO/IEC 25024:2015, Table 9.2에 정의된 대로 데이터 처리 효율성을 측정합니다.
효율성 (Efficiency)	Eff-ML-3	Risk of wasted space (공간 낭비 위험)	ISO/IEC 25024:2015, Table 9.2에 정의된 대로 공간 낭비 위험을 측정합니다.
정밀성 (Precision)	Pre-ML-1	Precision of data values (데이터 값 정밀성)	ISO/IEC 25024:2015, Table 10.1에 정의된 대로 데이터 값의 정확도를 측정합니다.
		Traceability of	ISO/IEC 25024:2015,

추적성 (Traceability)	Tra-ML-1	data values (데이터 값 추적성)	Table 11.1에 정의된 대로 데이터 값의 추적 가능성을 측정합니다.
추적성 (Traceability)	Tra-ML-2	User access traceability (사용자 접근 추적성)	ISO/IEC 25024:2015, Table 11.2에 정의된 대로 사용자 접근의 감사 추적 가능성을 측정합니다.
추적성 (Traceability)	Tra-ML-3	Data values traceability (데이터 값 추적성)	ISO/IEC 25024:2015, Table 11.2에 정의된 대로 데이터 값의 추적 가능성을 측정합니다.
이해 가능성 (Understandability)	Und-ML-1	Symbols understandability (기호 이해 가능성)	ISO/IEC 25024:2015, Table 12.1에 정의된 대로 데이터의 기호 이해 가능성을 측정합니다.
이해 가능성 (Understandability)	Und-ML-2	Semantic understandability (의미론적 이해 가능성)	ISO/IEC 25024:2015, Table 12.1에 정의된 대로 데이터의 의미론적 이해 가능성을 측정합니다.
이해 가능성 (Understandability)	Und-ML-3	Data values understandability (데이터 값 이해 가능성)	ISO/IEC 25024:2015, Table 12.1에 정의된 대로 데이터 값의 이해 가능성을 측정합니다.
이해 가능성 (Understandability)	Und-ML-4	Data representation understandability (데이터 표현 이해 가능성)	ISO/IEC 25024:2015, Table 12.2에 정의된 대로 데이터 표현의 이해 가능성을 측정합니다.

3. 시스템 의존적 데이터 품질 특성 (System-dependent Data Quality Characteristics)

--	--	--	--

소분류 (Sub-classification)	QM ID	QM 항목 (Name) (한/영 병기)	QM 한 줄 설명/개념
가용성 (Availability)	Ava-ML-1	Data availability ratio (데이터 가용성 비율)	ISO/IEC 25024:2015, Table 13에 정의된 대로 데이터 가용성 비율을 측정합니다.
이식성 (Portability)	Por-ML-1	Data portability ratio (데이터 이식성 비율)	ISO/IEC 25024:2015, Table 14에 정의된 대로 데이터 이식성 비율을 측정합니다.
이식성 (Portability)	Por-ML-2	Prospective data portability (예상 데이터 이식성)	ISO/IEC 25024:2015, Table 14에 정의된 대로 예상 데이터 이식성을 측정합니다.
복구 가능성 (Recoverability)	Rec-ML-1	Data recoverability ratio (데이터 복구 가능성 비율)	ISO/IEC 25024:2015, Table 15에 정의된 대로 데이터 복구 가능성 비율을 측정합니다.
복구 가능성 (Recoverability)	Rec-ML-2	Feature recoverability ratio (특징 복구 가능성 비율)	단계적으로 전송된 데이터셋 특징이 복구 가능한 정도를 측정합니다.

4. 추가 데이터 품질 특성 (Additional Data Quality Characteristics)

이 특성들은 분석(Analytics) 및 기계 학습(ML)의 맥락에서 데이터셋의 **충실도(Fidelity)**를 평가하는 데 중요합니다.

소분류 (Sub-classification)	QM ID	QM 항목 (Name) (한/영 병기)	QM 한 줄 설명/개념
감사 가능성 (Auditability)	Aud-ML-1	Audited records (감사된 레코드)	데이터셋 내 감사(audit)를 거친 레코드의 비율을 측정합니다.
	Aud-		데이터셋 내 감사에 활용

감사 가능성 (Auditability)	ML-2	Auditable records (감사 가능한 레코드)	가능한 레코드의 비율을 측정합니다.
균형 (Balance)	Bal-ML-1	Brightness balance (밝기 균형)	이미지 샘플의 평균 밝기 대비 이미지 샘플의 밝기 차이가 최대인 값의 역수를 측정합니다.
균형 (Balance)	Bal-ML-2	Resolution balance (해상도 균형)	이미지 샘플의 평균 해상도 대비 이미지 샘플의 해상도 차이가 최대인 값의 역수를 측정합니다.
균형 (Balance)	Bal-ML-3	Balance of images between categories (범주 간 이미지 균형)	데이터셋의 평균 범주 크기(샘플 수) 대비 최대 범주 크기 차이의 역수를 측정합니다.
균형 (Balance)	Bal-ML-4	Bounding box height to width ratio balance (바운딩 박스 종횡비 균형)	데이터셋 내 바운딩 박스 종횡비 평균 대비 최대 차이의 역수를 측정합니다.
균형 (Balance)	Bal-ML-5	Category bounding box area balance (범주 바운딩 박스 영역 균형)	데이터셋 내 모든 샘플의 평균 바운딩 박스 영역 대비 범주별 평균 영역의 최대 차이의 역수를 측정합니다.
균형 (Balance)	Bal-ML-6	Sample bounding box area balance (샘플 바운딩 박스 영역 균형)	데이터셋 내 모든 샘플의 평균 바운딩 박스 영역 대비 샘플별 바운딩 박스 영역의 최대 차이의 역수를 측정합니다.
균형 (Balance)	Bal-ML-7	Label proportion balance (라벨 비율 균형)	특정 라벨 값을 가진 두 범주 간 데이터 항목 비율의 차이를 측정합니다.
	Bal-	Label distribution	라벨 분포와 균일 (uniform) 라벨 분포 사이

균형 (Balance)	ML-8	balance (라벨 분포 균형)	의 발산 정도를 측정합니다.
다양성 (Diversity)	Div-ML-1	Label richness (라벨 풍부도)	데이터셋 내 고유한 (distinct) 라벨의 비율을 측정합니다.
다양성 (Diversity)	Div-ML-2	Relative label abundance (상대적 라벨 풍부도)	데이터셋 내 특정 라벨을 가진 개별 데이터(항목, 레코드, 프레임)의 비율을 측정합니다.
다양성 (Diversity)	Div-ML-3	Category size diversity (범주 크기 다양성)	품질 요구사항에 정의된 임계값보다 범주화된 데이터 항목 수가 적은 범주의 비율을 측정합니다.
유효성 (Effectiveness)	Eft-ML-1	Feature effectiveness (특징 유효성)	데이터셋 내 허용 가능한 특징(acceptable feature)을 가진 샘플의 비율을 측정합니다.
유효성 (Effectiveness)	Eft-ML-2	Category size effectiveness (범주 크기 유효성)	범주화된 샘플 수가 임계값보다 낮은 범주의 비율을 측정합니다.
유효성 (Effectiveness)	Eft-ML-3	Label effectiveness (라벨 유효성)	데이터셋 내 허용 가능한 라벨을 가진 샘플의 비율을 측정합니다.
식별 가능성 (Identifiability)	Idn-ML-1	Identifiability ratio (식별 가능성 비율)	데이터셋 내 식별 가능성 (PII)에 사용될 수 있는 데이터 레코드의 비율을 측정합니다.
적합성 (Relevance)	Rel-ML-1	Feature relevance (특징 적합성)	주어진 맥락(context)에 적합한 데이터셋 내 특징(feature)의 비율을 측정합니다.
	Rel-		주어진 맥락(context)에

적합성 (Relevance)	ML-2	Record relevance (레코드 적합성)	적합한 데이터셋 내 레코드의 비율을 측정합니다.
대표성 (Representativeness)	Rep-ML-1	Representativeness ratio (대표성 비율)	목표 모집단(Target Population)의 관련 속성 대비 데이터셋에서 발견된 관련 속성의 비율을 측정합니다.
유사성 (Similarity)	Sim-ML-1	Sample similarity (샘플 유사성)	클러스터링 알고리즘 결과로 도출된 클러스터 수를 활용하여 데이터셋 내 유사 샘플의 비율을 측정합니다.
유사성 (Similarity)	Sim-ML-2	Samples tightness (샘플 밀집도)	정규화된 데이터셋의 최대 고유값과 최소 고유값 간의 차이를 측정합니다 (데이터셋의 밀집도를 기하학적으로 측정).
유사성 (Similarity)	Sim-ML-3	Samples independency (샘플 독립성)	PCA(주성분 분석) 방법을 사용하여 데이터셋의 차원 축소 가능성(샘플 독립성)을 측정합니다.
적시성 (Timeliness)	Tml-ML-1	Timeliness of data items (데이터 항목 적시성)	적시성 요구사항을 충족하는 데이터 항목의 비율을 측정합니다.