ISO/IEC 5259-2:2024-11 (E)

Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 2: Data quality measures

Contents

Page

Foreword v

Introduction vi

1 Scope 1

2 Normative references 1

3 Terms and definitions 1

4 Symbols and abbreviated terms 5

5 Data quality components and data quality models for analytics and machine learning 5
5.1 Data quality components in data life cycle 5
5.2 Data quality model 6

6 Data quality characteristics and quality measures 8
6.1 General 8
6.2 Inherent data quality characteristics 9
6.2.1 Accuracy 9
6.2.2 Completeness 10
6.2.3 Consistency 12
6.2.4 Credibility 13
6.2.5 Currentness 14
6.3 Inherent and system-dependent data quality characteristics 15
6.3.1 Accessibility 15
6.3.2 Compliance 15
6.3.3 Efficiency 16
6.3.4 Precision 16
6.3.5 Traceability 17
6.3.6 Understandability 17
6.4 System-dependent data quality characteristics 18
6.4.1 Availability 18
6.4.2 Portability 18
6.4.3 Recoverability 19
6.5 Additional data quality characteristics 19
6.5.1 Auditability 19
6.5.2 Balance 20
6.5.3 Diversity 22
6.5.4 Effectiveness 23
6.5.5 Identifiability 24
6.5.6 Relevance 25
6.5.7 Representativeness 25
6.5.8 Similarity 26
6.5.9 Timeliness 27

7 Implementing a data quality model and data quality measures for an analytics or ML task 28

8 Data quality reporting 28
8.1 Data quality reporting framework 28

8.2 Data quality measure information 29
8.3 Guidance to organizations 29

Annex A (informative) Design and document of a measurement function 30

Annex B (informative) UML model of data quality measure framework 32

Annex C (informative) Overview of data quality characteristics 33

Annex D (informative) Alternative groups of data quality characteristics 35

Annex E (informative) Comparison between data quality characteristics of ISO/IEC 25012 and ISO/IEC 5259-2 36

Bibliography 37