International Standard

ISO/IEC 5259-2

First edition
2024-11

Artificial intelligence — Data quality for analytics and machine learning (ML) —

Part 2: Data quality measures

Intelligence artificielle — Qualité des données pour les analyses de données et l'apprentissage automatique —

Partie 2: Mesure de la qualité des données

Reference number
ISO/IEC 5259-2:2024(en)

© ISO/IEC 2024



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

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**

© ISO/IEC 2024 – All rights reserved

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

Bibliography **37**

© ISO/IEC 2024 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial Intelligence.*

A list of all parts in the ISO/IEC 5259 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

© ISO/IEC 2024 – All rights reserved

Introduction

Data-supported decision-making brings new challenges to data quality management in data analytics and artificial intelligence (AI) based on machine learning (ML). Issues in data quality, such as incomplete, false or outdated data, can adversely affect analytics and ML processes and outcomes. Data from various sources, including structured data (e.g. relational databases) and unstructured data (e.g. documents, images, audios), can be directly consumed into the data life cycle for analytics and ML model development. Data are transformed in each stage of the data life cycle of analytics and ML. A holistic standardized approach to control, produce and deliver sufficient high-quality data is necessary for data analytics and ML models to be safe, reliable and interoperable. To develop credible data quality management for analytics and ML, intrinsic data quality International Standards, including concepts and use cases, characteristics and measurements, management requirements, and process framework, can be considered.

This document is a part of the ISO/IEC 5259 series. This document builds upon the ISO 8000 series, ISO/IEC 25012 and ISO/IEC 25024. The purpose of this document is to describe a data quality model through the definition of data quality characteristics and data quality measures based on ISO/IEC 25012 and ISO/IEC 25024. Data quality models can be extended or modified according to this document.

© ISO/IEC 2024 – All rights reserved

Artificial intelligence — Data quality for analytics and machine learning (ML) —

Part 2:
Data quality measures

1 Scope

This document specifies a data quality model, data quality measures and guidance on reporting data quality in the context of analytics and machine learning (ML).

This document is applicable to all types of organizations who want to achieve their data quality objectives.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 5259-1, *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples*

ISO/IEC 25024, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality*

ISO/IEC 22989, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 5259-1, ISO/IEC 22989 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
data
re-interpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

Note 1 to entry: Data can be processed by humans or by automatic means.

[SOURCE: ISO/IEC 2382:2015, 2121272]

© ISO/IEC 2024 – All rights reserved