Prioritization of ISO/IEC 25012 Data Quality Dimensions Using the AHP Technique

Cláudio Keiji Iwata
Centro Estadual de Educação Tecnológica Paula Souza
São Paulo, Brazil
claudio.iwata@cpspos.sp.gov.br

Márcia Ito
Centro Estadual de Educação Tecnológica Paula Souza
São Paulo, Brazil
marcia.ito@cpspos.sp.gov.br

ABSTRACT

This article presents a method for prioritizing data quality dimensions based on the ISO/IEC 25012 standard, using the Analytic Hierarchy Process (AHP) technique. The method assists in assessing data quality for training predictive models, considering the relevance of dimensions according to the application context. The research combines qualitative analysis, through the collection of expert judgments, and quantitative analysis to calculate relative weights between the selected dimensions. To assess the method's reproducibility and applicability, a simulation was conducted for a predictive model in hospital bed management. The simulation results indicate the feasibility of guiding data quality improvement actions and strengthening information governance for use in predictive modeling.

KEYWORDS

Data Quality, Predictive Models, Healthcare, AHP, AHP-OS, Hospital Management, ISO/IEC 25012, Data Governance.

1 Introduction

The increasing availability of data from the Web and mobile sources has driven research in data analytics, particularly in predictive modeling, which supports decision-making in various domains—especially healthcare [18].

Data quality is critical for ensuring the reliability of clinical data and the outcomes derived from it. Inaccuracies, inconsistencies, or delays in clinical data can significantly compromise the performance of predictive models and the decisions they inform [18]. The ISO/IEC 25012 standard provides a conceptual model to evaluate data quality across various contexts, including healthcare, by defining a set of dimensions such as accuracy, completeness, timeliness, and consistency [6].

While ISO/IEC 25012 outlines important principles for evaluating data quality, it does not offer a clear framework for selecting and prioritizing its dimensions in specific application domains. To bridge this gap, it was introduced a method grounded in the Analytic Hierarchy Process (AHP), implemented using the AHP-OS platform, which converts expert assessments into a structured ranking of relevant quality dimensions.

The proposed method supports the selection of data quality dimensions most appropriate for predictive modeling tasks. It demonstrate its application through a simulation involving hospital bed management, a domain in which data quality plays a critical role in supporting effective decision-making and resource distribution.

- A structured, replicable method for prioritizing data quality dimensions based on ISO/IEC 25012;
- The partial application of AHP technique using AHP-OS to facilitate expert-driven decision-making; and
- The practical validation of the method in a real-world healthcare scenario.

The remainder of this paper is organized as follows: Section 2 presents the theoretical background; Section 3 describes the methodology; Section 4 details the proposed method and simulation; Section 5 discusses the results; and Section 6 concludes with final remarks and future research directions.

2 Theoretical Background

2.1 Data Quality

When data contains issues like missing values, repeated entries, or system-related inconsistencies, the results of any analysis can become less reliable or even misleading. In health-related applications, this often leads to a lack of trust in patient records, confusion in clinical decisions, and reduced performance in machine learning models.

The ISO/IEC 25012 standard outlines 15 different dimensions to assess data quality, divided into two groups: inherent and system-dependent. The first group relates to the natural characteristics of the data itself, such as how accurate or complete it is. The second focuses on how the data is handled or supported by systems, including aspects like accessibility and availability.

Accuracy refers to how correctly data represent the actual value of an attribute, subdivided into syntactic and semantic accuracy. *Completeness* measures whether all expected attributes and instances are present. *Consistency* ensures that data are free from contradictions within or across entities. *Credibility* reflects users' trust in data authenticity. *Currentness* evaluates whether data are timely and relevant.

System-dependent dimensions include *accessibility*, indicating user access capabilities; *compliance*, adherence to standards; *confidentiality*, protection from unauthorized access; *efficiency*, processing performance; *traceability*, ability to track changes; *understandability*, ease of interpretation; *availability*, reliable retrieval; *portability*, ease of migration; and *recoverability*, resilience after failure.

These dimensions are summarized in Table 1.

In AI-based decision systems, data quality is essential to mitigate risks. As cited by Lackoff [7], input errors could introduce biases and jeopardize model performance. Using standardized criteria like ISO/IEC 25012 it's possible to prioritize the most relevant aspects, aligning the evaluation with the dataset's analytical context and ensuring with a consistent method.

Table 1: Data Quality Dimensions Adapted from ISO/IEC 25012 [6].

Dimension	Inherent	System Dependent
Accuracy	X	
Completeness	X	
Consistency	X	
Credibility	X	
Currentness	X	
Accessibility	X	X
Compliance	X	X
Confidentiality	X	X
Efficiency	X	X
Precision	X	X
Traceability	X	X
Understandability	X	X
Availability		X
Portability		X
Recoverability		X

2.2 Analytic Hierarchy Process

The AHP is a multicriteria decision-making technique that combines both qualitative and quantitative elements through pairwise comparisons [12]. It is suitable for complex situations where a standardized prioritization framework is required.

The AHP structures decision problems into hierarchical levels, beginning with the overall objective and followed by criteria, subcriteria (if applicable), and alternatives. This organization enhances clarity and ensures transparency throughout the decision-making process.

The method involves three key steps:

(1) structuring the problem into a hierarchy;
(2) conducting pairwise comparisons among elements at each level using a 1-to-9 scale;
(3) evaluating the consistency of judgments.

The consistency ratio (CR) is an index that evaluates the coherence of judgments in pairwise comparisons,; values above the threshold of 0.10 indicate the need for revision to ensure decision reliability.

AHP is a method that converts subjective expert judgments into numerical values that support objective prioritization. AHP is a multicriteria technique that supports evaluations involving qualitative criteria through a standardized comparison scale, which allows consistent decision-making even without empirical data [4].

AHP-OS is a platform that automates the AHP process by generating comparison matrices, calculating priority weights, and performing consistency checks [3].

2.3 Digital Health

Digital health integrates information and communication technologies to improve healthcare delivery, enhancing efficiency, accessibility, and service quality [15]. Solutions such as telemedicine, mobile health apps, and electronic health records have expanded the volume and complexity of clinical data, reinforcing the need for robust data quality practices.

The effectiveness of predictive models in healthcare relies heavily on the quality of training data. Inadequate data—such as duplicates, missing values, or inconsistencies—can degrade model performance and compromise decision-making [2]. Key data quality dimensions like accuracy, completeness, consistency, and relevance are essential for ensuring reliable analytical outcomes in digital health environments.

In Big Data environments, good data quality helps organizations work better and make clearer decisions. Studies show that even data from indirect sources can improve user satisfaction and make intelligent systems more trustworthy [16].

Thus, implementing structured data quality assessment aligned with domain-specific needs is critical for maximizing the value of digital health initiatives and supporting reliable AI-driven healthcare solutions.

3 Research Methodology

Data quality should be evaluated considering coherence and correspondence with the application domain, a principle called **ontological foundation** [17]. This foundation is necessary to ensure that the selected **quality dimensions** reflect the characteristics of the context in which the data will be used, guaranteeing **semantic validity** and **information integrity**.

Considering that each context presents its own characteristics and requirements, the selection of appropriate dimensions is ensured through the coordinated action of a **project manager**, responsible for defining specifications and managing activities, and a **business specialist**, tasked with analyzing the specific needs of each context. This collaboration allows adapting the **prioritization of quality dimensions** to the project's objectives, contributing to **decision-making in data quality management**.

From this concept, the **quality dimensions are hierarchized and prioritized** to focus on those that present greater relevance to the project domain, facilitating **data quality management**.

After defining the **data quality dimensions**, the application of **AHP technique** was structured to obtain the **relative importance weights** among the selected criteria. The proposed method consists of transforming **subjective judgments** into **numerical values** through **pairwise comparisons** between criteria, using a **scale of importance** to rank the elements. The technique was operationalized through the use of the **AHP-OS** tool, which enables practical execution of these comparisons, automatic calculation of **eigenvectors** (relative weights), and verification of the **consistency index** of judgments to ensure response coherence [3]. The weights obtained from **AHP technique application** were used to build a prioritization matrix among the selected **data quality dimensions**. This matrix functions as the method's result, establishing the order of relative importance of the dimensions and serving as a reference to support future data quality evaluation processes, enabling targeted qualification efforts in accordance with specific **analytical project objectives**.

For the method's application, an **experiment** was designed with the participation of specialists from the **health**, **data**, and **information technology** areas. For this purpose, a hypothetical system was conceived for the simulation with the design of a conceptual **database** structure. The objective was to contextualize the method's application. This structure was developed from attributes commonly found in **hospital systems**, such as **patient_id, admission_date, length_of_stay**, and **primary_diagnosis**. This simulation was used solely to assess the method's applicability and does not involve real or confidential data; therefore, ethical committee approval was not required.

Finally, the method's application was conducted in a representative scenario of the health domain, related to **hospital bed management**. This application aims to validate the "model's' utility as a tool to support the prioritization of **quality criteria** in **databases** that underpin decision-making. The method was developed and formalized through a **flowchart** representing the sequential, modular, and replicable model of method application. This structure facilitates the **understanding**, **adaptation**, and **repetition** of the process in different **analytical contexts**.

4 Proposed Method

The proposed method supports the prioritization of data quality dimensions from the ISO/IEC 25012 standard for predictive models, aligning with the specific needs of different domains and analytical contexts. It is structured in sequential and modular steps, from project scoping to the generation of data governance recommendations.

This approach addresses a gap in both literature and practice regarding the operationalization of ISO/IEC 25012, which presents 15 conceptual data quality dimensions—grouped into inherent, system-dependent, and relational categories [6]—but does not provide guidance on selecting or ranking them based on specific application needs.

Each context has its distinct requirements; it is important to analyze project goals and constraints before deciding to prioritize of quality dimensions. This ensures focus on dimensions that truly impact data usage, improving the efficiency of quality improvement efforts. This analysis is important to avoid irrelevant quality dimensions may be included in the study, reducing effectiveness and decision reliability of the model.

To ensure contextual alignment, the method involves both a project manager and a business analyst. Together, they define technical, operational, and strategic requirements and validate the selected dimensions, promoting a prioritization process oriented toward practical results.

The methodology involves both a project manager and a business analyst, responsible for ensuring the proper execution of all phases and for validating the selected data quality dimensions for the prioritization process.

The method also allows modular adjustments for different domains, supporting the inclusion of new dimensions or evaluation criteria when needed. This flexibility enables the consideration of additional factors such as data lifecycle, decision impact, and resource limitations for quality improvements.

AHP was chosen as the supporting technique, as it converts subjective expert judgments into quantitative weights for decision making [4, 12, 14]. The AHP-OS tool [3] facilitates the process by ensuring consistency and traceability in results.

The method structures the prioritization process and adapts it to project-specific requirements, helping direct quality improvement efforts and supporting the development of reliable and effective predictive models.

Prioritization Flowchart of Data Quality Dimensions. The process of prioritizing data quality dimensions for predictive models was organized into four stages:

(1) Presentation of the overall flow via a visual flowchart;
(2) Definition of the project scope;
(3) Partial application of AHP technique to rank the dimensions;
(4) Preparation of a report with recommendations.

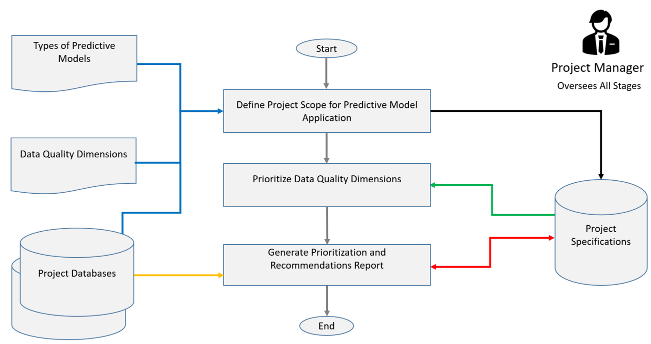

Figure 1: Flowchart for Prioritization of Data Quality Dimensions

As shown in Figure 1, the method begins with assigning a project manager responsible for overseeing all steps. The project scope is then defined through the collection of key information such as the predictive model type, relevant data quality dimensions, and specification parameters.

This information is stored in a structured database, which serves as a repository for the project's core elements. Simultaneously, different predictive models are assessed to determine the most suitable one based on data characteristics and project goals, following recommended practices for sensitive data contexts [18].

Prioritization considers the ontological alignment between data and the application domain, ensuring relevance to the analytical context, as outlined in the theoretical framework.

Once the scope and critical variables are established, the quality dimensions are organized into a hierarchy using routines based on ISO/IEC 25012 criteria. These dimensions are recorded in the specification database and serve as the foundation for later stages of evaluation and treatment.

The project receives information from various databases originating from diverse sources. This data must be transformed into a standardized format to facilitate analysis by the predictive model [2]. Standardization involves multiple steps, such as formatting, structural verification, and other data harmonization procedures.

After the scope definition phase is completed, the data quality dimensions prioritization phase begins. Hierarchization serves as an input for generating the final recommendation report, which utilizes the project's data as foundational inputs [8].

Project Scope Definition. The scope definition stage, shown in Figure 2, marks the initial phase of the data quality dimension prioritization method. The process begins with formalizing the objective of the predictive model, with active participation from the business analyst, who helps define responsibilities and align goals. This objective is recorded in the project specifications database and serves as the basis for later decisions.

Next, the appropriate type of predictive model is selected based on alignment with project goals, data characteristics, and technical requirements. Information about the model is stored in the same specification database.

Specialists then define the set of data quality dimensions to be used for evaluating the relevant databases. This selection is based on project needs, model requirements, and the ISO/IEC 25012 standard [6], ensuring the reliability of the data. The selected data sources are assessed for compatibility with project goals, data availability, and structure. These characteristics are documented in the project database.

Business specialists review the specifications to confirm that the selected definitions align with model requirements and validate all critical components.

In the subsequent stage, data quality specialists, database experts, and business experts are selected to participate in the AHP survey and establish the prioritization of data quality dimensions. By leveraging techniques from diverse areas of expertise, ensuring that multiple perspectives are considered. According to [5], the integration of technical and business expertise enhances the analysis of data maturity by taking into account both data quality and its strategic value. This multi-faceted approach contributes to the generation of consistent and context-specific priorities for the targeted application, thereby supporting improvements in the performance of predictive systems [7, 18].

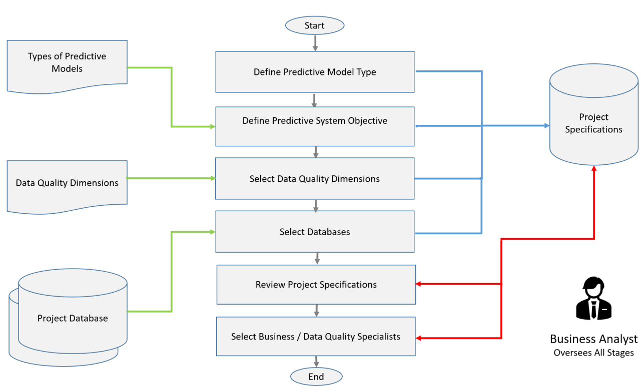

Figure 2: Define Project Scope to Apply Predictive Model

Hierarchizing Data Quality Dimensions. After defining the project scope, the next step is to prioritize the selected data quality dimensions, as shown in Figure 3. This phase is to organize the dimensions through the partial application of AHP technique, ensuring consistency and structure in the evaluation.

The project database serves as input to create the AHP survey using the AHP-OS tool [3]. The survey generated by the tool allows experts to perform pairwise comparisons in accordance with Saaty's method [12]. The survey is completed by the selected specialists, who input their responses directly into the system. The responses are stored automatically within the system, ensuring traceability and data integrity [18].

Once responses are collected, AHP-OS processes the data, generating priority vectors and consistency indices to assess the reliability of judgments [9]. The tool calculates the maximum eigenvalue ($\lambda_{\max}$) of each matrix, from which the Consistency Index (CI) is computed: $\mathrm{CI} = (\lambda_{\max} - n)/(n-1)$, with n as the number of criteria. The Consistency Ratio (CR) is derived by dividing the CI by the Random Index (RI), where $RI = 0.58$ for $n = 3$, 0.90 for $n = 4$, and 1.12 for $n = 5$ [12].

Judgments are accepted when $CR < 0.10$ [13]. If a participant exceeds this threshold, their matrix is revised; if it remains inconsistent, their input is excluded to ensure the validity of results.

Although the calculations of CI and CR are performed automatically by the AHP-OS tool, their computations are described in this paper for educational purposes, to provide an understanding of the mathematical theory underlying the assessment of these consistency parameters. In practice, the research relied on the automated processing capabilities of the tool, ensuring accuracy and standardization.

Based on valid responses (i.e., $CR \leq 0.10$), normalized techniques are used to determine the relative weights of each dimension. These weights define the final ranking of the data quality dimensions, which is recorded in the project database. This classification supports future decisions on data quality assessments and guides improvement strategies [2, 7].

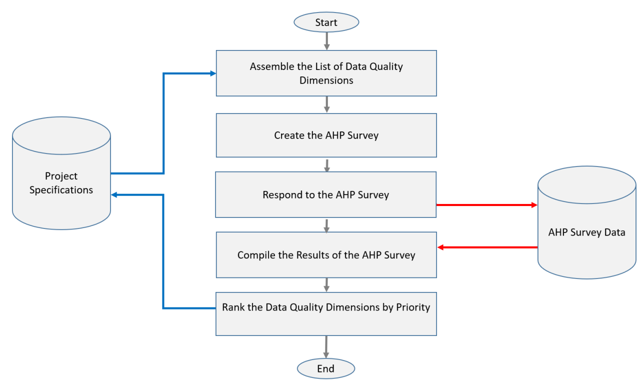

Figure 3: Rank Data Quality Dimensions

Report of Prioritization and Recommendations. After completing the hierarchy of data quality dimensions, the next step is to retrieve the prioritized list from the project database, as shown in Figure 4. This ordering is based on the weights from AHP technique and considers predefined project parameters, such as the selected predictive model.

The prioritized list is then used to guide further stages of analysis and database structuring, ensuring the model is built according to quality criteria and project goals [18].

With the dimensions ranked, the final database is defined. This includes selecting data sources, choosing the variables that will compose the dataset, and identifying key relationships. These steps follow the project specifications, ensuring alignment between requirements and data structure [2].

Once the classification and prioritization of data quality dimensions are completed, the data are recorded in the database. The database is then used to identify which data sources will be utilized to create the foundational dataset for the predictive model. These steps adhere to the project specifications, ensuring alignment between the requirements and the data structure [2].

A consolidated report is generated at the conclusion of the process. The results of the prioritization and the analysis of the database structure are systematically utilized for system improvement and also serve as a basis for future adjustments to the predictive model.

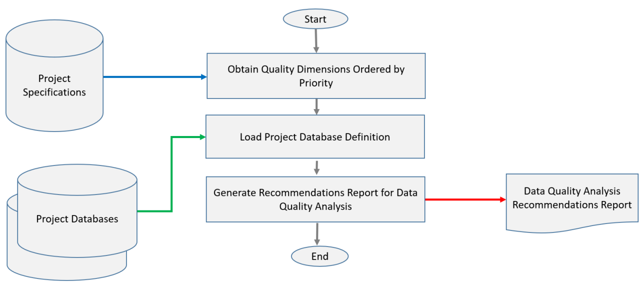

Figure 4: Generate Prioritization Reports and Recommendations

5 Simulation

This section presents the application of the proposed method in the context of hospital bed management. The procedure followed the steps illustrated in Figure 1.

Hospital bed management is a strategic area of healthcare administration, influencing resource allocation and care quality. The reliability of predictive models used in this context depends on the quality of data. Prioritizing specific data quality dimensions supports better system operation and decision-making, especially in complex hospital environments.

Accurate bed allocation relies on complete, consistent, and up-to-date data. Inaccurate information can lead to operational inefficiencies, longer patient wait times, and increased pressure on clinical teams [11]. The partial application of AHP technique enables the prioritization of data quality dimensions, helping managers focus on the most critical aspects for decision.

Indicators such as bed occupancy require reliable data that accurately reflect hospital dynamics. Improving data quality in this context supports efficient resource use and enhances patient safety.

The method was applied to the Intensive Care Unit (ICU) admission regulation process, which requires high-quality data due to its critical nature. The Federal Council of Medicine (CFM) Resolution No. 2.156/2016 provides five criteria used to prioritize ICU admissions:

(1) Diagnosis and patient need;
(2) Available medical services;
(3) Prioritization by clinical condition;
(4) Bed availability;
(5) Potential therapeutic benefit and prognosis.

Effective bed management balances capacity, admission demand, and clinical criteria, ensuring care continuity, efficient resource use, and overall institutional performance [2]. According to CFM guidelines, ICU admission should consider severity, recovery potential, and expected treatment outcomes, while discharge decisions should be based on stabilization or clinical irreversibility [1].

Reliable clinical and administrative data from hospital systems are essential for traceability, information integrity, and effective decisions. Poor-quality data—whether incomplete, inconsistent, or outdated—can impair care coordination and negatively affect performance metrics [10, 18].

Defining structured criteria for data quality assessment is a key strategy to improve data recording and analysis. In environments where predictive models support clinical or managerial decisions, selecting appropriate data quality dimensions is vital to ensure the usefulness and reliability of the analysis [2].

This simulation aimed to define the problem context and identify critical elements for building the predictive model, including the database, key variables, data sources, and record characteristics. Based on this, the most relevant data quality dimensions for the context were selected.

Project Scope Definition. This stage defines the scope of applying the proposed method to hospital bed management. The predictive model aims to support ICU bed prioritization and allocation, based on clinical criteria from CFM Resolution No. 2,156/2016, as detailed in Table 2. The definition involved specialists in hospital management and data analysis, focusing on evidence-based decisions aligned with public healthcare dynamics.

Table 2: Prioritization of Admission to the Intensive Care Unit [1]

Priority	Rule
§ 1	Life-support interventions with high recovery probability, no therapeutic limitations.
§ 2	Intensive monitoring with high risk of intervention, no therapeutic limitations.
§ 3	Life-support needed, low recovery chance or therapeutic limitations.
§ 4	Intensive monitoring, high risk, but with therapeutic limitations.
§ 5	Terminal patients, no recovery possibility. Admission only in exceptional cases.

Based on the ICU admission rules, it was defined the most suitable predictive model to ensure greater adherence. The most appropriate model was identified as a decision support system that integrates clinical and administrative data from hospital information systems. Among the predictive model options, such as linear regression, decision trees, and Gradient Boosting, the eXtreme Gradient Boosting (XGBoost) model was selected due to its capability to

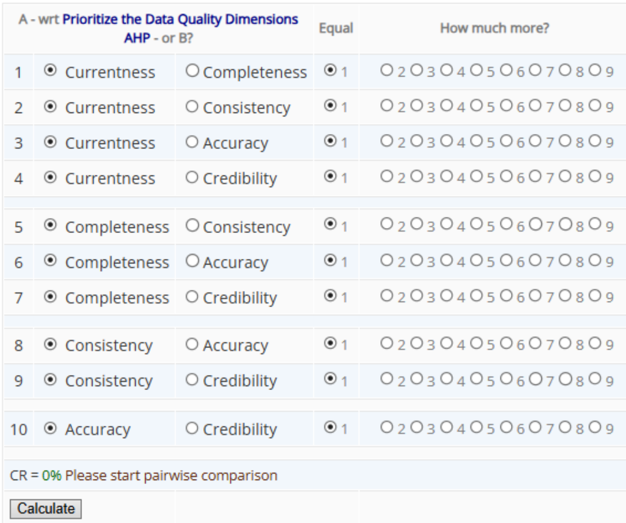

Figure 5: Weight Selection Screen for each Group of Data Quality

handle complex interactions among variables, its tolerance to missing data, and its ability to produce interpretable results. Decisions are based on these characteristics, which contribute to transparent decision-making in the triage of patients for critical care.

Next, the most relevant data quality dimensions were selected—those classified as *inherent* by ISO/IEC 25012: accuracy, completeness, consistency, credibility, and timeliness. Dimensions related to system dependencies were excluded, based on project scope criteria.

Data sources were defined, including hospital and mortality information, utilizing clinical data to meet the criteria established by CFM, such as diagnosis, severity score, admission status, and length of stay. Compatibility between the originate data and the objectives of the predictive system must be aligned to ensure coherence.

A hospital regulation specialist reviewed the technical and functional specifications—covering variables, sources, and prioritization logic—to ensure alignment with real-world constraints and healthcare system guidelines.

Experts in hospital management, data quality, health IT, and machine learning were then selected to support prioritization and evaluation. Their combined expertise ensured coherence with project goals and methodological consistency.

Finally, a business analyst validated the project's definitions and data, confirming their alignment with model requirements and ensuring the reliability of information used during model development.

Hierarchizing Data Quality Dimensions. After defining the scope and selecting relevant data quality dimensions for hospital bed management, the next step was to establish their relative importance using AHP technique via the AHP-OS platform.

The AHP-OS interface allows users to input pairwise comparisons, generating real-time results, weight charts, and reports. Each specialist accessed the system individually to provide judgments in a controlled environment. Figure 5 shows the weight input screen,

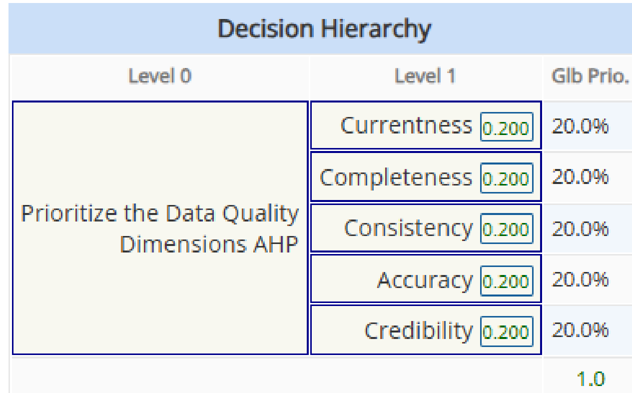

Figure 6: Screen with Details of Data Quality Dimensions Source

while Figure 6 displays the dimension detail view and the initial global consistency index.

The specialist selection criteria included higher education, experience in healthcare or data science, and technical and/or business expertise to perform comparisons of data quality dimensions. The selected participants possessed diverse profiles, including machine learning and database technicians, as well as healthcare professionals, which enhanced the variety of opinions and perspectives regarding data quality.

All had at least an undergraduate degree, along with postgraduate education in areas like software engineering, data science, or healthcare. Additional criteria included experience in data cleaning, familiarity with ISO/IEC 25012, and understanding of the AHP method.

AHP technique was employed and demonstrated suitability for evaluating expert judgments, particularly in situations where objective data were not fully concrete [12, 17]. A global consistency index was calculated to assess the level of agreement among responses. Alignment is crucial to ensure reliability, as inconsistent judgments can significantly impact the prioritization results [12]. The resulting weights were recorded in the project's database for subsequent use in the next phase of the method.

Participation occurred remotely, with judgments entered directly in the AHP-OS tool. The sample included 20 professionals, with 65% from the private sector and 35% from other domains (public sector, freelancers, etc.). Regarding AHP experience, 55% reported being experienced or highly experienced.

Following data collection, a consensus analysis was conducted. Participants with outlier judgments were asked to review their inputs. Responses not aligned with the group were excluded to improve result consistency, as recommended in multi-expert AHP studies [12].

Generate Report of Prioritization and Recommendations. Following the completion of the prioritization phase of the data quality dimensions, information stored in the project database was extracted. This task aimed to organize the dimensions according to the priority order assigned by the specialists. The ranking considered results obtained through the partial use of AHP technique, as well as predefined parameters, including the type of predictive model to be adopted in the context of bed management.

Based on this prioritization, the structure of the database to be employed in the project was defined. This definition included the selection of institutional data sources, the choice of variables composing the dataset, and the identification of relevant relationships among these variables. The definition process was guided by the scope of the predictive model and the model to be applied in bed allocation, following guidelines established by Resolution CFM No. 2.156/2016 [1].

From the database structuring and dimension prioritization, a technical report was prepared containing recommendations related to data quality assessment. The report indicated procedures for analyzing the prioritized dimensions, including verification of data integrity, identification of gaps, and analysis of internal consistency of the variables.

This document aimed to guide the execution of the evaluation and treatment stages of the database, according to the needs of the predictive model. The recommendations were based on the weights assigned to the quality dimensions and on the relationships established between these dimensions and the operational data used in the clinical decision-making process.

Based on the analyses, the final recommendation report was developed, presenting the results of AHP technique application and the evaluation of the database structure. This report provides an expert opinion to support the implementation of corrective measures, the development of predictive systems, and the governance of data quality.

Information quality governance encompasses practices and standards that ensure data integrity, reliability, consistency, and security. The report supports decision-making processes. The prioritized data quality dimensions (based on the ISO/IEC 25012 standard and AHP technique) credibility, consistency, accuracy, timeliness, and completeness—are employed to evaluate and enhance hospital databases utilized in bed management.

Discussion on the Application of the Method. After completing the prioritization of data quality dimensions, the project database was consulted to extract and organize the dimensions by importance. The ranking was based on the partial application of AHP technique and on project-specific parameters, such as the chosen predictive model for hospital bed management.

This prioritization guided the definition of the project's database structure, including the selection of institutional data sources, relevant variables, and relationships among them. The structuring process aligned with the scope and requirements defined by CFM Resolution No. 2.156/2016 [1].

A technical report was developed with recommendations for assessing the prioritized dimensions. It outlines procedures for evaluating data integrity, identifying missing data, and verifying the consistency of the data, with a focus on the requirements of the predictive model.

Finally, the recommendations were based on the dimension weights and their relevance to operational data used in clinical decisions. This document supported the evaluation and treatment stages of the dataset, offering practical guidance for improving data quality and data governance.

The prioritized dimensions of credibility, consistency, accuracy, timeliness, and completeness followed ISO/IEC 25012 and the AHP results. These guided the assessment and enhancement of hospital data quality for bed management.

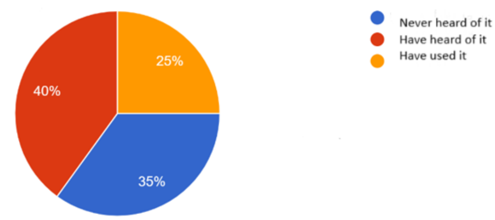

Figure 7: Familiarity with AHP

6 Results

All 20 participants signed the Free and Informed Consent Form (TCLE), ensuring ethical compliance with CNS Resolution No. 196/96. The TCLE confirmed participants' understanding of the study's goals and authorized the use of secondary data.

Participants were categorized into two main areas: healthcare, and data science/information technology. Around 40% work with data quality, governance, or database management; 30% specialize in modeling, prediction, or integration; and 20% are from healthcare, higher education, or applied research. Although fewer were from healthcare directly, most had practical experience with data quality in clinical or administrative settings.

About 55% reported being "experienced" or "very experienced" with the research topic, showing adequate familiarity with the subject, even among those outside the healthcare domain. This supports the suitability of using AHP, which relies on structured expert judgments [12].

Regarding employment sectors, 65% work in the private sector, mainly in data-related fields. The remaining 35% are distributed across public, freelance, or other sectors. Understanding the participants' work contexts is relevant, as it can influence how they perceive and prioritize decision criteria in multicriteria analyses [3].

Most participants (70%) had over 10 years of professional experience; 25% had 6–10 years, and only one had less than five. This suggests strong contextual knowledge, which may enhance judgment quality when applying methods like AHP [12].

Data were collected using a structured questionnaire via the AHP-OS platform [3]. Figure 7 shows participant familiarity with AHP technique. Eight participants (40%) had basic knowledge; five (25%) had prior experience using the method; and seven (35%) had never heard of it. For the nine participants interviewed in person (45%), an explanation was provided.

Despite varying levels of familiarity, the structured AHP-OS interface allowed all participants to engage with the method. The

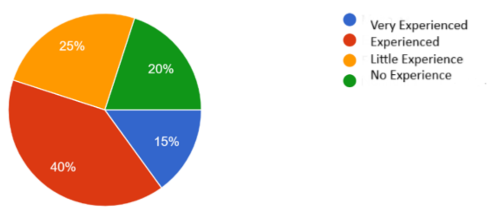

Figure 8: Experience with Research Topic

technique requires understanding of pairwise comparisons, weight calculation, and consistency checking [3, 12].

Participants' experience with the study topic was also assessed. Figure 8 shows that 40% identified as "experienced," 15% as "very experienced," 25% as having "little experience," and 20% as having none. All participants had knowledge in databases, machine learning, or health, but not all had direct experience in hospital bed management. This variety enabled the analysis of AHP's applicability across different expertise levels and its impact on judgment consistency.

Experience level is important in AHP-based evaluations, as the method relies on the decision maker's ability to interpret and rank criteria. Prior knowledge enhances result robustness, especially when dealing with subjective domains like data quality [3, 12].

The results obtained from the prioritization of data quality dimensions using AHP technique demonstrated consistent and structured findings. To ensure methodological rigor, first it was analyzed the consistency ratio (CR) of all pairwise comparison matrices. Following Saaty's recommendation [13], responses with CR values above 0.10 were considered inconsistent and excluded.

Data collection was carried out through the AHP-OS tool with 20 professionals from the healthcare and data science sectors. After evaluating individual CR values, five responses were excluded. Among the remaining participants, those with divergent judgments were contacted to review their comparisons. Four additional responses were excluded due to continued disagreement with group consensus.

The final analysis included 11 consistent participants. The aggregated results showed a group consensus index of 75.1% and relative homogeneity of 86.6%, both classified as high by the AHP-OS platform. The global CR was 0.0010, confirming the internal validity of the judgments.

Table 3: Global Weights of Data Quality Dimensions

Dimension	Weight (%)
Credibility	31.1
Consistency	23.3
Accuracy	20.8
Timeliness	13.7
Completeness	11.2

Participant	Currentness	Completeness	Consistency	Accuracy	Credibility	CR_{max}
Group result	13.7%	11.2%	23.3%	31.1%	20.8%	0.1%
Participant 1	25.2%	16.9%	21.8%	10.9%	25.2%	1.7%
Participant 2	13.6%	16.4%	19.5%	16.6%	33.9%	8.8%
Participant 3	8.2%	4.8%	13.0%	34.6%	39.5%	6.4%
Participant 4	35.5%	7.0%	10.9%	22.2%	24.4%	3.0%
Participant 5	8.0%	27.8%	19.4%	36.5%	8.4%	1.5%
Participant 6	11.8%	7.5%	25.2%	47.7%	7.8%	6.9%
Participant 7	18.7%	9.8%	35.7%	17.9%	17.9%	7.8%
Participant 8	11.5%	9.8%	21.1%	36.5%	21.1%	9.3%
Participant 9	5.7%	5.0%	30.4%	30.4%	28.4%	0.4%
Participant 10	4.5%	17.8%	26.7%	22.4%	28.5%	1.9%
Participant 11	18.0%	6.5%	14.6%	56.4%	4.4%	5.5%

Figure 9: Group Consensus of Respondents

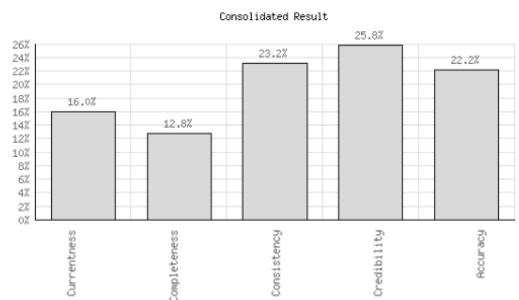

Figure 10: Graph of Global Priorities Consolidation

Credibility received the highest weight (31.1%), followed by consistency (23.3%) and accuracy (20.8%). These findings reflect the professionals' view that trust in data sources and internal coherence are more critical than data completeness in the ICU bed management context. This aligns with previous research [7, 18], which highlights how unreliable data undermine predictive accuracy and clinical trust.

Completeness ranked lowest (11.2%), likely due to the understanding that useful models can still be developed using incomplete datasets, provided the available information is accurate and logically coherent. In hospital systems, prioritizing data veracity and coherence may be more impactful than ensuring every field is filled.

To support implementation, a technical report was produced, applying the prioritized dimensions to assess relevant fields in the hospital bed management database. Below are examples of how each dimension was operationalized:

Completeness. Critical fields such as `bed_number`, `bed_type`, and `sofa_score` were reviewed for missing data. Automated validation was suggested to block incomplete entries.

Consistency. Fields including `occupancy_status`, `transfer_date`, and `destination_department` were checked for logical conflicts. Cross-field validations were recommended.

Credibility. Indicators like `ward_occupancy_rate` and `data_source` were analyzed for anomalies. Alert mechanisms and regular audits were proposed.

Accuracy. Discrepancies in `physical_bed_number`, `active_bed_number`, and `length_of_stay` suggested tracking errors. Periodic verification was advised.

Timeliness. Fields such as `status_update_datetime` were assessed for delay. Monitoring and data refresh routines were proposed.

The report provided practical guidance for improving hospital data quality and for supporting model reliability. By converting expert judgment into quantifiable priorities, AHP technique helped align technical requirements with clinical expectations.

These results reinforce that structured prioritization enhances data governance practices, especially in critical areas like ICU management. Although credibility and consistency emerged as top priorities, the multidimensional nature of data quality suggests that ongoing assessment and adaptation are essential for effective use in predictive systems.

7 Conclusion, Limitations, and Future Work

This study presented a method for prioritizing data quality dimensions using the ISO/IEC 25012 standard, with partial application of the Analytic Hierarchy Process (AHP), supported by the AHP-OS platform. Designed as a sequence of structured steps, the method was validated through a simulation focused on hospital bed management.

The approach proved feasible and effective, enabling the identification and ranking of the most relevant dimensions for the healthcare context analyzed. By translating expert judgment into quantitative data, AHP technique facilitated prioritization in a traceable, automated, and replicable way. The AHP-OS tool supported this process by standardizing inputs and producing consistent outputs.

The main contribution of this work lies in offering a replicable methodology for dimension prioritization grounded in international standards and multicriteria decision-making. In addition to its successful application in healthcare, the method shows potential for adaptation to other domains that rely on structured data, such as logistics, finance, or education. Its reliance on expert input and flexible structure makes it broadly applicable across sectors.

Scientifically, the method supports operationalizing the ISO/IEC 25012 standard in real-world scenarios, integrating theoretical principles with practical prioritization needs. Practically, it provides a decision-making framework for professionals seeking to improve the quality of databases used in predictive analytics projects, ensuring that qualification efforts focus on the dimensions that matter most.

Despite its strengths, the study has limitations. One key limitation is its reliance on expert judgment. As the prioritization is subjective, the final weights are influenced by the profile and diversity of the respondent group. Homogeneous or inexperienced groups could produce biased results, limiting generalization.

Another constraint is the method's domain-specific application. Although it is designed for broader use, this implementation was limited to hospital bed management. Generalizing to other fields will require contextual adaptation, including the selection of relevant dimensions and experts.

Additionally, only part of the AHP method was applied—namely, the comparison of criteria. The analysis did not extend to alternatives or complex decision trees, thus not fully leveraging AHP's capabilities. Scalability is also a concern: as the number of dimensions increases, the number of comparisons grows rapidly, increasing the risk of inconsistent judgments and cognitive overload for participants.

These limitations, while not compromising the study's validity, highlight areas for improvement and suggest avenues for future research. One direction involves applying the method in other domains, such as logistics or education, assessing adaptability and effectiveness across different operational realities. In such cases, the most critical dimensions and expert profiles would need to be revisited.

Future work could also explore hybrid approaches, combining AHP with other multicriteria techniques such as TOPSIS, to strengthen decision-making in uncertain or multi-alternative scenarios. This integration could enhance robustness and applicability, especially in complex environments.

Expanding the sample size and diversity of expert participants is another recommended avenue. Comparative studies involving professionals from clinical, managerial, and technical backgrounds could offer insights into how perspective influences prioritization outcomes.

Finally, incorporating objective metrics and quantitative validation of the prioritized dimensions could further strengthen the model. By linking data quality indicators to predictive performance, future research may quantify the impact of each dimension on decision-making, enhancing both methodological rigor and practical relevance.

REFERENCES

[1] Brasil. Ministério da Saúde. Conselho Nacional de Saúde. 2012. Resolução nº 466, de 12 de dezembro de 2012. Diário Oficial da União: seção 1, Brasília, DF, p. 59, 13 jun. 2013. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html.

[2] W. Elouataoui, I. El Alaoui, S. El Mendili, and Y. Gahi. 2022. An Advanced Big Data Quality Framework Based on Weighted Metrics. *Big Data and Cognitive Computing* 6, 153 (2022), 1–22. doi:10.3390/bdcc6040153

[3] Klaus Goepel. 2018. Implementation of an Online Software Tool for the Processo de Análise Hierárquica (AHP-OS). *International Journal of the Analytic Hierarchy Process* 10, 3 (2018), 469–485. doi:10.13033/ijahp.v10i3.590

[4] C. González-Pavón, C. V. Palau, J. Manzano Juárez, V. Estruch-Guitart, S. Guillem-Picó, and I. Balbastre-Peralta. 2019. Optimization of Collective Irrigation Network Layout through the Application of the Analytic Hierarchy Process (AHP) Multi-criteria Analysis Method. *Water* (2019).

[5] P. C. Grossler and J. Fleck. 2017. An application of the Analytic Hierarchy Process to the evaluation of companies' data maturity. *Journal of Decision Systems* 26, 4 (2017), 344–355.

[6] ISO/IEC. 2008. ISO/IEC 25012:2008 Software engineering – Software product quality requirements and evaluation (SQuaRE) - Data quality model. Disponível em: https://iso25000.com/index.php/en/iso-25000-standards/iso-25012.

[7] Steven Labkoff et al. 2024. Toward a responsible future: recommendations for AI-enabled clinical decision support. *Journal of the American Medical Informatics Association* 31, 11 (2024), 2730–2739. doi:10.1093/jamia/ocae209

[8] S. E. Labkoff and D. F. Sittig. 2017. Who watches the watchers: working towards safety for EHR knowledge resources. *Applied Clinical Informatics* 8, 2 (2017), 680–685. doi:10.4338/ACI-2017-02-IE-0032

[9] R. Li, L. J. Sun, H. Zhang, C. Yao, and G. Luo. 2017. Research on DC voltage class series with AHP. *The Journal of Engineering* 13 (2017), 1993–1998. doi:10.1049/joe.2017.0678

[10] S. Malacaria et al. 2023. An Application of the Analytic Hierarchy Process to the Evaluation of Companies' Data Maturity. *SN Computer Science* 4 (2023), 696.

doi:10.1007/s42979-023-02065-9

[11] Seyede Maryam Najibi, Farhad Lotfi, Erfan Kharazmi, Payam Farhadi, Payam Shojaei, Peivand Bastani, and Zahra Kavosi. 2021. Identification and Prioritization of Local Indicators of Hospital Bed Allocation in Iran. *Research Square* (2021). doi:10.21203/rs.3.rs-609451/v1

[12] Thomas L. Saaty. 1994. *How to Make a Decision: The Analytic Hierarchy Process.* RWS Publications, Pittsburgh. 101 pages.

[13] T. L. Saaty. 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences* 1, 1 (2008), 83–98.

[14] G. M. Sani, A. M. Abas, N. Yusoff, and M. F. Said. 2023. Site selection for electric vehicle charging stations using GIS with MCDM AHP FAHP and TOPSIS techniques: A Review. *IOP Conference Series: Earth and Environmental Science* 1274 (2023), 012019. doi:10.1088/1755-1315/1274/1/012019

[15] J. A. Silva and M. R. Oliveira. 2020. Saúde digital: desafios e oportunidades na era da transformação digital. *Revista Brasileira de Saúde Digital* 5, 2 (2020), 45–58. doi:10.5220/0012185900003598

[16] Samuel Fosso Wamba, Shahriar Akter, Laura Trinchera, and Marc de Bourmont. 2019. Turning information quality into firm performance in the big data economy. *Management Decision* 57, 8 (2019), 1756–1783. https://ro.uow.edu.au/gsbpapers/536

[17] Y. Wand and R. Y. Wang. 1996. Anchoring data quality dimensions in ontological foundations. *Commun. ACM* 39, 11 (1996), 87–93. doi:10.1145/240455.240456

[18] N. G. Weiskopf et al. 2017. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 5, 1 (2017), 1–19. http://doi.org/10.13063/egems.1280.s1