# Harnessing AI's Potential: Transforming Metadata Management with Machine Learning for Enhanced Data Access and Control

**Swathi Chundru**
*Motivity labs INC*
*QA Automation Engineer*
*Texas, USA*

[1]*Received: 24 January 2024; Accepted: 12 February 2024; Published: 26 February 2024*

## ABSTRACT

The application of artificial intelligence (AI) is expanding across numerous industries, such as banking, transportation, and healthcare. Artificial intelligence relies on large-scale dataset analysis, which requires a constant supply of high-quality data. Nevertheless, there are disadvantages to using data to solve AI problems. This book addresses the challenges of using data comprehensively and critically.

Interpretability, prejudice and fairness, privacy and security, volume of data, quality and explainability, ethical considerations, and technical expertise are some of the issues that AI must address. This essay examines these problems in-depth and recommends how companies and organisations should respond.

By understanding and addressing these challenges, organisations may use AI to make smarter decisions and gain a competitive edge in the digital age. It makes sense that the scientific research community should reevaluate how we approach data strategy for AI, given that this review paper gives and analyses various data issues for AI during the previous ten years. It will also tremendously aid in the development of new and original thoughts.

**Keywords:** *AI; Data Privacy; Data Security; Metadata Management; Machine Learning.*

## INTRODUCTION

Data management must be met for the digital transformation of businesses, governments, and society. A critical component of data management in a corporate setting is master data management. Master data pertains to essential business objects of a company. The goals of master data management include assessment, planning, and control of master data to enable its effective and efficient utilisation. [12]. Taking care of metadata is essential to maintaining master data [12, 9]. Frequently, metadata is known as content-related data [2]. According to Hüner et al., metadata [9] is organised data referencing other data. Metadata either highlights technological features on the characteristics of applying content data [2] or [23].

While managing metadata is a difficult, time-consuming, and thorough process, the issues associated with it are expanding as digitisation increases [12]. Support for metadata management will consequently be required by numerous organisations. Automation, or at least automated support, may be one way to address this problem. Within the field of artificial intelligence, machine learning pertains to the capacity of a machine or software to acquire the ability to do particular tasks. For the analysis of vast amounts of data, machine learning is an effective tool [15]. It seems sense to investigate how well machine learning can assist with metadata management. In this study, we investigate the potential benefits of using machine learning techniques to either fully or partially automate specific metadata management chores.

The application of artificial intelligence in master data management, known as Augmented MDM [11] by the Gartner Group, has sparked significant interest. In this context, our main objective is to address a crucial question: can metadata management be enhanced using machine learning methods?

[1] *How to cite the article:* Chundru S (February, 2024); Harnessing AI's Potential: Transforming Metadata Management with Machine Learning for Enhanced Data Access and Control; *International Journal of Advances in Engineering Research*, Feb 2024, Vol 27, Issue 2, 39-49

Only a limited number of previous studies have evaluated the field of automating metadata procedures. Ganesan et al. [5] have investigated automatic entity identification and relationship. Murthy [5,14] have explored automated data extraction from a collection of master data items has findings of this conventional study were obtained more than a decade ago. As a result, these subject merits a closer examination and a more thorough analysis.

This research aims to determine the potential support that particular machine-learning methods may provide for activities related to metadata management. We seek to respond to two enquiries:

• RQ1: What are the essential requirements for improving metadata management?

• RQ2: Which machine learning techniques enable the application of these specifications?

## THE ROLE OF METADATA IN MODERN DATA ECOSYSTEMS

Metadata, often described as "data about data," serves as a crucial element in data management by providing context, meaning, and structure to raw data. It enables data discovery, integration, and governance, ensuring the data is correct and accessible and usable within different applications. In complex data landscapes, effective metadata management enhances data quality, compliance, and interoperability.

### 1. Data Discovery and Accessibility

Unlike metadata, which is a key enabler of data discovery and accessibility. Information that describes other data assets is known as metadata that helps users search, identify and retrieve proper data. In big & complex data scenarios like data lakes & data warehouses, the volume of data can make it hard to locate a specific data set.

- **Descriptive Metadata:** It provides information about the content, context, and structure of data to help users understand what the data is.
- **Administrative Metadata:** Provides information about the data, such as the source, owners, and lifecycle, helping with data access and governance.
- **Technical Metadata:** Provides information about the technical properties of the data, including file type, format, location of storage, and also processing requirements which help in ensuring and the effective access and usage of data.

### 2. Data Quality and Governance

Metadata plays a key role in ensuring data quality and enabling data governance initiatives. Metadata is designed to capture information about the origin, transformation, and utilization of the data — ensuring that the data is accurate, consistent, and reliable.

- **Data Lineage:** Metadata captures the lineage of data, tracking its origin and movement across different systems and processes, allowing for transparency into the transformations and usage of data. By doing so, errors can be captured and corrected, maintaining the integrity of the data.
- **Compliance and Auditing:** Metadata aids in compliance with regulatory requirements by providing a record of data handling practices, access controls, and usage policies. It can also help in auditing by acting as a record of data activities.

### 3. Data Integration and Interoperability

This is particularly the case in modern data ecosystems where data is both distributed across numerous systems/platforms & formats. In achieving seamless interoperability across diverse systems, metadata drives the integration and harmonization of this data.

- **Standardization:** By providing a consistent format and structure to data, metadata aids in the standardization of data collection, allowing for easier integration of data from different sources. This is key to a unified data view across multi-cloud or hybrid environments.

- **Data Mapping:** Metadata helps to map data elements across different systems so that they can be correctly interpreted and used in integrated applications.

**4. Data Analytics and Machine Learning**

Metadata allows Data to drive insights but helps analytics and ML use the data more effectively. It helps in locating the appropriate data sets, performs feature engineering and increases model accuracy.

- **Contextual Metadata:** This metadata offers contextual information about data (e.g., the circumstances surrounding the collection of the data) that is critical for accurate analysis and modelling.
- **Feature Metadata:** This feature metadata describes the features used in the ML models and helps data scientists pick the right features and understand the relationships between different data attributes.

**5. Data Security and Privacy**

After all, a significant part of modern data ecosystems includes the use of various data security and privacy services as well, and these services are very dependent on metadata. Metadata with rich access, use, and sensitivity information can help organizations safeguard sensitive information and comply with data privacy regulations.

- **Access Control Metadata:** Defines who can access specific data sets and what actions they can perform, ensuring that data is protected from unauthorized access.
- **Sensitivity Metadata:** Identifies data elements that contain sensitive or personal information, enabling organizations to implement appropriate data protection measures, such as encryption or anonymization.

**6. Data Collaboration and Sharing**

Metadata helps teams and organizations to achieve a common understanding of the data that supports better collaboration and sharing of data. It helps maintain the compatibility of external and internal systems to make sure that any data can be effectively shared and used in cooperative projects such as data science projects or cross-division analytics.

- **Collaborative Metadata:** Captures information about data usage and insights generated by different users, enabling knowledge sharing and collaboration.
- **Data Catalogues:** Metadata-driven data catalogues provide a centralized repository of data assets, allowing users to discover and share data across the organization.

**7. Enhancing the Value of Data Assets**

Metadata adds value to data assets through the enabling of its usability, understandability, and actability. Metadata found in data-driven organizations is an asset that not only enables decision-making processes but also drives innovation and competitive advantage.

- **Data Enrichment:** Metadata adds value to raw data by providing additional context, annotations, and classifications, making it more meaningful and useful for various applications.
- **Data Monetization:** In certain instances, metadata can itself be monetized, especially when offering unique insights or enabling value-added services like data analytical or tailored recommendations.

## CHALLENGES AND CONSIDERATIONS IN AI-DRIVEN METADATA MANAGEMENT

The potential of metadata management using AI (artificial intelligence) and ML (machine learning) seems massive; they can significantly improve efficiency, accuracy, and scalability. But also, there are some challenges and considerations that organisations should focus on to provide successful implementations of AI-based metadata management systems.

**1. Data Privacy and Security**

AI-enabled metadata management needs access to large-scale data sets to train model and generate precise metadata. This may also pose major issues concerning data privacy and security.

- **Data Access and Control are another Risks of using AI:** AI systems require vast amounts of data to learn and operate effectively, potentially leading to the exposure of sensitive or personal information. It is critical that only authorized AI systems have access to sensitive data.
- **Compliance with Regulations:** Organizations need to comply with different data privacy regulations (such as GDPR, CCPA) that sets out stringent rules for data collection, processing, and sharing. Compliance with these regulations is a significant consideration for AI-driven metadata management.
- **Secure Data Storage and Processing:** AI systems should store metadata safely to avoid access, leaks and breaches. This encompasses encrypting information when it is stored and being transferred as well as deploying strong access controls.

**2. Bias and Fairness in AI Models**

Analysis of data using AI and ML models in metadata management might be biased, which means that these models might lead to unfair or inaccurate outcomes. In some domains, this may have dire consequences, particularly in cases where metadata plays a role in important decisions.

- **Leverageable Viewing/Scrutiny:** What are the different use cases for which data is viewed and scrutinized containing personal information in metadata? The biases of the training data can, for instance, find their way into the metadata generated by the models as the performance of the underlying models is reflected in the quality of the metadata.
- **Algorithmic Bias:** Even if the training data is free from bias (which is often not the case), specific algorithms might process information in a biased manner. A fundamental step in the process of ensuring AI algorithms operate fairly is the careful selection and testing of the models on which they are built.
- **Mitigation in AI:** Organizations need to develop methodologies for identifying, tracking, and reducing bias in AI-built metadata processing. This can involve employing diverse training datasets, implementing fairness constraints during model training, and routinely auditing AI systems for unfair results.

**3. Scalability and Performance**

Although AI possesses the potential to improve the scalability of metadata management, the use of AI-based systems in enterprise data environments has many differences.

- **Resource Requirements:** AI-oriented systems tend to need significant computational power, especially to train and run ML models. It is also important to ensure that these systems scale as data volumes grow.
- **Redefining metrics upon changes to data in the environment:** In many dynamic data environments – like Active-Active advertisement firms having multiple locations to serve customers (for example, usage of microservice to segregate data of various businesses) – our team might have to refactor logic and generate metadata instantaneously. Such systems based on AI needs to be able to process data at high throughput without sacrificing accuracy or performance.
- **Ingesting and Querying of Massive Data Sets:** As such, AI-driven metadata management systems must support large and heterogeneous data sets, spanning structured, unstructured, and semi-structured data. One of the biggest challenges is handling these data sets efficiently without degrading the performance.

**4. Data Quality and Integrity**

The power of AI-driven metadata management is only as effective as the quality and integrity of the data that drives the process itself.

- **Garbage In, Garbage Out (GIGO):** Bad data creates bad metadata, leading to poor data discovery, poor data governance, and poor data analytics. So, data quality is a must for dependable AI-powered metadata management**.**
- **Inconsistency in Data:** Differences in data formats, structures, and standards can be a challenge for generating and integrating metadata. AI systems need to be prepared to deal with and resolve these discrepancies.
- **Maintaining Data Integrity:** AI-driven systems must ensure that the metadata accurately reflects the underlying data, without introducing errors or distortions during the processing stages.

**5. Ethical Considerations**

Introducing AI in metadata management poses ethical concerns regarding data ownership and transparency, as well as implications for existing human roles historically involved in metadata management activities and workflows.

- **Transparency and Explainability:** When implemented, AI-driven metadata management systems operate as a "black box," making it difficult for users to gain an appropriate level of visibility into decision-making. Model transparency and explainability can provide assurances about accountability in the decision-making process of AI technology.

- **Impact on Employment:** The introduction of AI could potentially automate specific roles in metadata management, leading to reduced human efforts. This requires organizations to address the disruption to jobs, especially factories and manufacturing jobs, and to find opportunities to retrain and reskill these workers to work alongside AI.

- **Ethical Use of AI:** Organizations should adhere to ethical standards in the handling of AI in metadata to prevent practices that can potentially result to manipulation, discrimination, or exploitation.

**6. Integration with Existing Systems**

The implementation of AI-driven metadata management often involves integration with existing data management systems and workflows, which can be complex.

- **Compatibility with Legacy Systems:** Organizations often employ legacy systems for metadata management. Existing systems may require extensive changes or upgrades to integrate solutions powered by AI.

- **Interoperability:** The integration of metadata across various platforms and formats is essential to maximize the potential of AI systems.

- **Change Management:** Deploying AI promptings may necessitate changes to organizational processes, roles, and culture. Properly managing these changes is critical to a successful transition and adoption.

**7. Continuous Learning and Adaptation**

Training and evolving AI metadata management solutions must have the capability to learn and adapt to the ever-evolving data environment and user needs.

- **Model Maintenance:** AI models should be retrained regularly to keep them accurate and relevant, so the metadata will also be up to date. It needs continuous monitoring, data capturing, and assessment.

- **Handling Evolving Data:** As data ecosystems change, AI based systems should be able to adapt to new data types, formats and standards. A big thing to be mindful of is keeping these systems flexible and adaptable.

- **User Feedback:** Incorporating user feedback into AI-driven metadata management systems can help improve accuracy and relevance over time. However, collecting and integrating this feedback in a meaningful way can be challenging.

## METHODOLOGY

Figure 1 presents our methodology for the study. Otto and Hüner's proposed functional architecture for master data management [17] serves as the foundation for the first stage. The term "master data management" refers to the overall system. Metadata management is one of the many sub architectures that make it up. We also conducted a literature review to elicit functional requirements for tools supporting metadata management. We employed the methodology Fettke [4], and Webster and Watson [26] suggested to analyse the data. To find pertinent material, we used Google Scholar, Web of Science, the AIS eLibrary, IEEE Xplore, and Springer Link. Using the search terms "metadata management" OR "Metadata management," we examined the keywords and abstracts of each publication. As shown in Fig. 1, we considered 29 articles.

Articles about managing metadata within the framework of master data management were included. Based on the review's conclusions, we developed a set of needs using a requirements elicitation procedure in accordance with ISO 29148 [10]. Thirty-seven needs were found, and they were categorised into seven groups.

During the master data management roundtable at Heilbronn University of Applied Sciences, we convened with fifteen participants to examine each need's practical significance and rank them in order of importance. In eight German companies, the participants manage master data strategically or operationally. We also interviewed four experts. Each specialist works in digital sales, process management, or research and development, and they are all in charge of master data management.

There are six open-ended questions and 25 closed-ended questions in our questionnaire.

For the analysis of responses to the closed-ended questions, we employed the Mayring-proposed qualitative content analysis [13], a method known for its robustness. The results are presented in Table 1 [26], a concept matrix based on Webster and Watson.

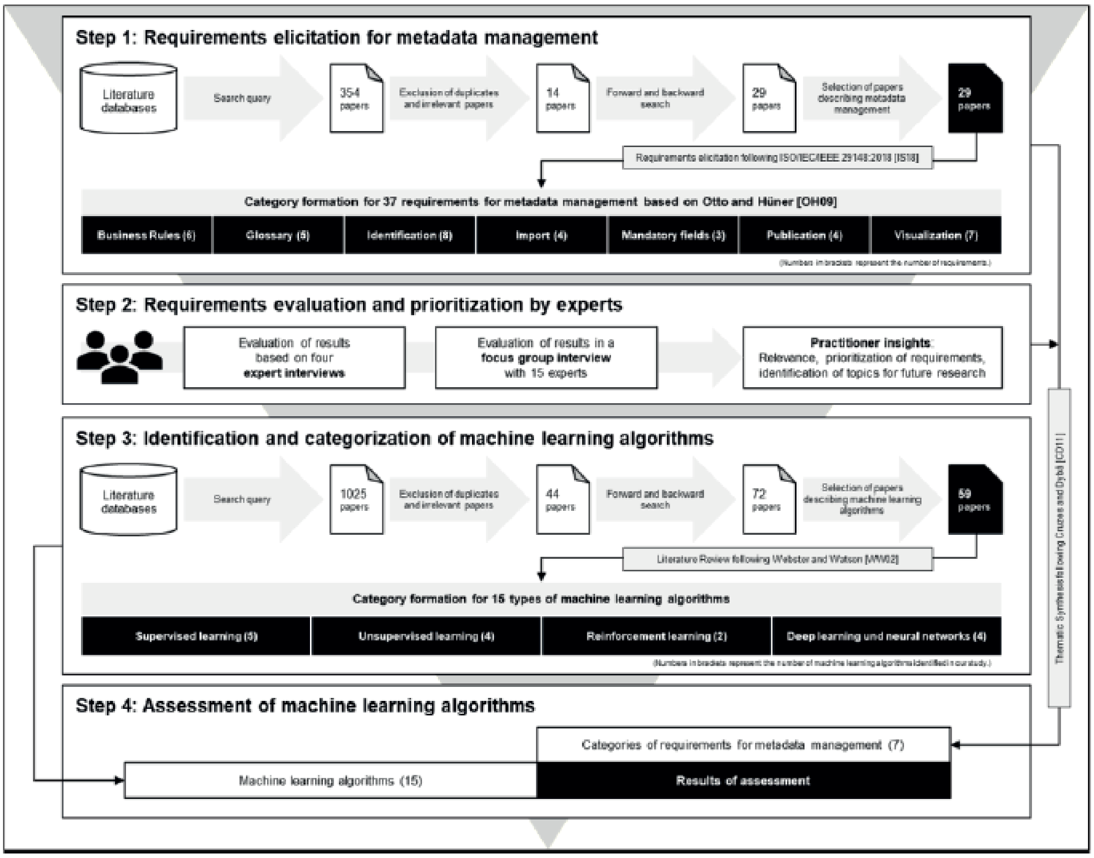


Fig. 1: Research approach

In the third step, we conducted a second literature study to classify and identify machine learning methods. We followed the same course of action as in the first evaluation, outlined in Step 1. The search terms that we used were [“machine learning“ OR „machinery Lernen“] AND [„master data management“ OR „metadata management“ OR “Stammdatenmanagement” OR “Metadata management” OR “algorithms” OR “Algorithmen”].

Following our analysis of abstracts and keywords, we systematically considered 59 articles released between 2005 and 2020 that covered machine learning algorithms and machine learning. We then took a systematic approach to categorize the fifteen different kinds of machine learning algorithms into four groups: reinforcement learning, unsupervised learning, supervised learning, and neural networks. This systematic categorization ensures that our findings are well-organized and easy to understand.

Using these algorithms, we identified their features and functions, possible applications, and advantages and disadvantages.

The fourth phase involves evaluating particular machine learning algorithms. Using the 37 criteria determined in step 1, we evaluated the algorithms using an argumentative-deductive methodology. By combining sub-requirements into requirements, we constructed a model of higher-order themes from our data using the methodology described by Cruzes and Dybå [3]. We applied the four-eye principle put forward by Peffers et al. [19]. Our results are shown in section 3.4.

## FINDINGS

### Necessities for Management of Metadata

Guided by the principles of ISO 29184 [10], our research process was significantly influenced by the insights gained from an extensive literature study. This study led us to the requirements elicitation procedure, which resulted in the identification of 37 needs. These needs were then categorized into seven groups, six of which were originally proposed by Otto and Hüner [17].

We included the category identification of metadata since the literature views this as a crucial metadata management component. We briefly review the seven types of metadata management duties in the following section (see Figure 1).

• The goal of documenting business rules is to capture staff members' unique expertise, process-related standards, and partially codified rules. Process automation requires well-documented business rules. Metadata definitions can be used to help build these rules. They may refer to a collection of this data or specific master data pieces.

• The process of finding, recording, and preserving master data is a challenging task in data management. It requires the creation of a lexicon or dictionary. However, with the use of metadata, we can confidently develop and update a lexicon, ensuring the preservation of master data and demonstrating our data management skills.

It facilitates communication and documentation of business processes that call for master data [12].

• Identification entails adding metadata to master data that already exists. Then, this metadata needs to be updated regularly. Gathering all definitions of master data creates an organization's foundation for metadata management [12].

A data dictionary or catalogue may have these definitions [8].

• Importing metadata facilitates unifying disparately formatted metadata into a logical framework. As Loshin [12] explains, one component of integration operations is the import of metadata from heterogeneous systems.

• When master data is modified or added to a database, it's crucial that mandatory fields have valid values. Controlling required fields is not just a task but a responsibility that guarantees that master data is of sufficient quality. Fields marked as required or optional are defined by metadata, underscoring the importance of your role in maintaining data quality.

Making metadata accessible in operational information systems is known as metadata publication. The service facilitates a logical application of master data. When integrating information systems, glossaries and data dictionaries support consistent use of terminology and formats [12].

The purpose of visualisation is to present master data in a graphical format. This process uses metadata to create visual representations of data objects, such as Entity-Relationship or UML-Diagrams.

### Evaluated and Prioritized Requirements

As previously indicated, we spoke with experts in the field to assess and rank the outcomes of our requirement elicitation process. We reviewed and discussed the findings of the interviews with a more extensive group of specialists during the ensuing focus group sessions. We asked them to rank the seven metadata management responsibilities mentioned above. The experts prioritised the following tasks: master data visualisation, metadata identification, and business rule documentation. The establishment of mandatory fields and metadata import were given priority number two, and the creation of a glossary and metadata publication was given priority number three. These criteria offer helpful direction for choosing valuable projects that will assist machine learning-based metadata management.

### Selected Machine Learning Algorithms

In step 3, we distinguished and grouped four different kinds of machine learning algorithms. In the next step, we evaluate their potential to enhance metadata management.

The capacity of algorithms to assign a given set of inputs to a set of outputs is known as supervised learning. We also assessed Sindhu et al.'s [21] four subcategories of supervised learning: "decision trees," "support vector machines," "naïve Bayes," and "linear regression." We also include "semi-supervised learning."

*Unsupervised learning aims to identify patterns in a given data set [6]. Nock and Nielsen [16] describe four types of algorithms as important examples: "harmonic k-means clustering," "fuzzy c-means," "gaussian-expectation-maximization," and "K-means clustering."*

*Reinforcement learning teaches algorithms using a set of "rewards" and/or "punishments" [20]. Weber cites "q-learning," "temporal difference learning," and "generative adversarial networks" [24] as examples.*

*The foundation of neural networks is the idea that the electrochemical activity of networks of neurons accounts for the majority of mental activity [20]. The calculation uses "artificial neurons" placed in a neural network. Three layers of artificial neurons make up the basic architecture: an input layer, a hidden layer, and an output layer [7]. Neural networks come in a variety of forms [20].* We decided to address ‘auto-encoders’, ‘convolutional neural networks’ and ‘recurrent neural networks.

**Selected Findings of Assessments**

We evaluated the algorithms' level of compliance with the requirements in the fourth step of our research procedure. The assessment's results are displayed in Table 1. We can only display the findings at the level of the prerequisite categories due to space constraints. The authors may provide the original presentation with all 37 requirements upon request. We use Harvey-Balls to show the findings. A white ball indicates that an algorithm does not meet the requirements, while a whole black ball indicates a strong match between the needs and the functioning of an algorithm.

Four different kinds of algorithms were determined to be unfit for any use.

The main goal of "linear regression" is to forecast future states.

This does not meet any of the prerequisites. The same holds for "q-learning" and "temporal difference learning," two Reinforcement Learning Algorithms. These algorithms are primarily appropriate for issues that Markov chains can help solve. "Generative adversarial networks" are especially well-suited for producing data sets, such as creating or altering photographs. According to the expert’s, identifying information is one of the most crucial functional needs. According to our assessment, the algorithm with the best chance of supporting metadata identification is "support vector machines" (SVM). We, therefore, concentrate on this issue to demonstrate how machine learning met, hods can a, id in metadata management.

Among the most widely used supervised learning algorithms are SVM. Algorithms are trained to classify and draw boundaries through a given data collection [20].

Tab. 1: Assessment of machine learning algorithms with metadata management requirements

| **Machine learning algorithms** | **Requirements** Business rules | Glossary | Identification | Import | Mandatory fields | Publication | Visualization |
|---|---|---|---|---|---|---|---|
| **Supervised learning** | | | | | | | |
| Support vector machines | ◔ | ● | ● | ◕ | ◑ | ● | ● |
| Naïve bayes | ◕ | ● | ◕ | ◕ | ◑ | ● | ◕ |
| Decision trees | ◑ | ◕ | ◑ | ◑ | ◑ | ◑ | ◑ |
| Semi-supervised learning | ◑ | ◕ | ◕ | ◑ | ◕ | ◕ | ● |
| Linear regression | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **Unsupervised learning** | | | | | | | |
| K-means-clustering | ◑ | ◕ | ◑ | ◑ | ◔ | ◕ | ● |
| Fuzzy-c-means | ◑ | ◕ | ◑ | ◑ | ◔ | ◕ | ● |
| Gaussian-EM | ◑ | ◕ | ◕ | ◑ | ◔ | ◕ | ● |
| Harmonic-k-means-clustering | ◑ | ◕ | ◑ | ◑ | ◔ | ◕ | ● |
| **Reinforcement learning** | | | | | | | |
| Q-learning | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Temporal difference learning | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Generative adversarial networks | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **Neural networks** | | | | | | | |
| Auto encoders | ◑ | ● | ◑ | ◕ | ● | ◕ | ● |
| Convolutional neural networks | ◔ | ◑ | ◕ | ◔ | ◔ | ◑ | ◑ |
| Recurrent neural networks | ◑ | ◕ | ◕ | ◔ | ◑ | ◕ | ◑ |

A substantial amount of high-quality data is required for SVM training [25]. SVM is used by Bahmani et al. [1] to clean up data and find duplicates so they can be combined into a single data object. Text classification is identified by Singh et al. [23] as the primary use case for SVM. Their main advantages are the algorithms' excellent accuracy, capacity to prevent overfitting and fit solid for fact generalisation. Its drawbacks are SVM's intricacy, the required training effort, and the fact that performance depends on selecting the correct parameters.

The eight requirements outlined in Figure 1 and Table 1 under the heading "identification of metadata" are covered in the following paragraphs.

We evaluated classification, entity identification, and entity resolution problems mainly in SVM. Classification issues relate a discrete number of outputs to a given set of inputs [Mu12]. Text classification challenges describe objects using specific words or word combinations [23]. Duplicates fall into particular object groups distinguished by their characteristics [1]. Duplicate object conflicts can be resolved by joining the two objects together, which also helps to represent semantic characteristics [9, 12] or by replacing them with a single object [1].

"The algorithm shall list all definitions of master data objects that exist within an organisation in the form of metadata," states the first criterion. This is possible because SVM can identify definitions and data objects using text categorisation. But this calls for a lot of training.

Since it is a duplication detection problem combined with text recognition, the second criterion, "The algorithm shall recognise and list similarities and differences of master data definitions in the form of metadata," may be easily satisfied. However, a significant quantity of training will likely be needed.

"The algorithm shall standardise and harmonise similar definitions of master data objects in the form of metadata," states the third criterion. Since the primary objective here is text recognition, the implementation is doable.

Algorithm training will likely take a while. Experts may need to refine the results. "The algorithm shall name them differently in the metadata according to a uniform scheme" is the fourth criterion, which states that "if similar definitions of master data objects cannot be standardised." The method is likewise relatively straightforward since this is another text recognition problem. However, the naming scheme must already be in place for the results to be valid. "The algorithm shall connect the metadata of data objects if similar definitions cannot be standardised, but data objects are similar," states the fifth condition. Since connecting definitions is a classification issue, this is conceivable. However, the task's difficulty significantly raises the amount of work needed to train the algorithm. Experts may also need to improve the outcomes.

Since this is primarily a comparative text recognition problem, SVM can be used to implement the sixth criterion, which states, "If the existing definitions of master data objects do not match newly determined definitions, the algorithm shall replace them in the metadata." "The algorithm shall indicate that two definitions of master data objects exist at the same hierarchical level, and if they have the same fields, they may be synonyms and establish a link in the referring metadata," states the seventh criterion. Given that this challenge involves text recognition and classification, SVM is a good fit for it.

Similar to earlier criteria, there will likely be a substantial training effort, and the outcomes might require expert improvement.

Machine learning is not required for the eighth criterion, which states, "The algorithm shall enter definitions of master data objects into a repository." As a result, this need has not been considered in our research.

## CONCLUSION

Within the framework of master data management, we have determined 37 requirements for enhancing metadata management, which we have categorised into seven groups. After assessing our results, professionals ranked the requirements. Based on this, we can now respond to RQ 1: Documenting business rules, identifying metadata, and visualising master data are essential for enhancing metadata management.

Additionally, we have classified 15 different kinds of machine-learning algorithms into four groups. We investigated which algorithms can satisfy which of the requirements listed in the first phase of our research.

Here is how we respond to RQ 2: Neural networks, with some limitations, and supervised and unsupervised learning algorithms have the most potential to facilitate metadata management. Nevertheless, reinforcement learning doesn't appear to be a good fit for improving metadata management.

A more thorough analysis revealed that SVM might be the most suitable method for supporting metadata identification. Algorithms for unsupervised machine learning are ideal for enhancing metadata visualisation. These algorithms aid in the identification, categorisation, and relationship-making of the components to be visualised. Regretfully, there are better methods for supporting business rule documentation.

## REFERENCES

[1] Bahmani, Z.; Bertossi, L.; Vasiloglou, N.: ERBlox: Combining Matching Dependencies with Machine Learning for Entity Resolution. In International Journal of Approximate Reasoning, 2017, 83; pp. 118–141.

[2] Burnett, K.; Ng, K. B.; Park, S.: A comparison of the two traditions of metadata development. In J. Am. Soc. Info. Sci., 1999, 50; pp. 1209–1217.

[3] Cruzes, D. S.; Dybå, T.: Recommended Steps for Thematic Synthesis in Software Engineering. In (Ed. IEEE Computer Society): ESEM 2011 Proceedings, 2011; pp. 275–284.

[4] Fettke, P.: State-of-the-Art des State-of-the-Art. In WIRTSCHAFTSINFORMATIK, 2006, 48; pp. 257–266.

[5] Ganesan, B. et al.: Link Prediction using Graph Neural Networks for Master Data Management, 2020.

[6] Guérin, E.; Aydin, O.; Mahdavi-Amiri, A.: Artificial Intelligence. In (Eds. Guo, H.; Goodchild, M. F.; Annoni, A.): Manual of Digital Earth. Springer, Singapore, 2020; pp. 357–385.

[7] Haykin, S. S.: Neural networks and learning machines. Pearson, New York, 2009.

[8] Eds. Hildebrand, K. et al.: Daten- und Informationsqualität. Auf dem Weg zur Information Excellence. Springer Vieweg, Wiesbaden, 2021.

[9] Hüner, K. M.; Otto, B.; Österle, H.: Collaborative management of business metadata. In International Journal of Information Management, 2011, 31; pp. 366–373.

[10] ISO/IEC/IEEE: Systems and Software Engineering: Life Cycle Processes Requirements Engineering (ISO/IEC/IEEE 29148), 2018.

[11] Judah, S.; White, A.: Hype Cycle for Data and Analytics Governance and Master Data Management, 2020, 05.04.2021.

[12] Loshin, D.: Master Data Management. Elsevier, Amsterdam, 2008.

[13] Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz Verlag, Basel, 2015.

[14] Murthy, K. et al.: Content-Aware Master Data Management: Proceedings of the 16th COMAD, Nagpur, India, 2010; pp. 206–210.

[15] Murphy, K. P.: Machine learning. A probabilistic perspective. MIT Press, Cambridge, 2012.

[16] Nock, R.; Nielsen, F.: On weighting clustering. In IEEE transactions on pattern analysis and machine intelligence, 2006, 28; pp. 1223–1235.

[17] Otto, B.; Hüner, K. M.: Funktionsarchitektur für unternehmensweites Stammdatenmanagement. Universität St. Gallen, St. Gallen, 2009.

[18] Otto, B.; Österle, H.: Corporate Data Quality. Springer, Berlin, 2016.

[19] Peffers, K. et al.: Design Science Research Evaluation. In (Eds. Peffers, K.; Rothenberger, M.; Kuechler, B.): Design science research in information systems. Springer, Berlin, 2012; pp. 398–410.

[20] Russell, S. J. et al.: Artificial intelligence. A modern approach. Pearson, Boston, 2016.

[21] Sindhu Meena, K.; Suriya, S.: A Survey on Supervised and Unsupervised Learning Techniques. In (Eds. Kumar, L. A.; Jayashree, L. S.; Manimegalai, R.): AISGSC. Springer, Cham, 2020; pp. 627–644.

[22] Singh, A.; Thakur, N.; Sharma, A.: A review of supervised machine learning algorithms: 2016 3rd INDIACom, 2016; pp. 1310–1315.

[23] Tozer, G. V.: Metadata management for information control and business success. Artech House, Boston, 1999.

[24] Ed. Weber, F.: Künstliche Intelligenz für Business Analytics. Springer Fachmedien, Wiesbaden, 2020.

[25] Ed. Wennker, P.: Künstliche Intelligenz in der Praxis. Springer Fachmedien Wiesbaden, Wiesbaden, 2020.

[26] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In MIS Quarterly, 2002, 26; pp. xiii–xxiii.