Framework of Master Data Management in Banking Using Consolidation and Jaro Winkler Algorithm

Ipang Sasono[1*], Mustar Aman[2]
[1,2]Faculty of Computer Science, Universitas Insan Pembangunan Indonesia, Indonesia
[1]ipangsasono@gmail.com, [2]mustarstmik@gmail.com
***Corresponding Author**

Article Info

Article history:

Submission June 23, 2025
Revised August 10, 2025
Accepted August 28, 2025
Published September 16, 2025

Keywords:

Master Data Management
Consolidation Approach
Jaro-Winkler Algorithm
Data Integration
Banking Systems

ABSTRACT

Master Data Management (MDM) is a crucial framework for ensuring the consistency, accuracy, and reliability of key data entities within banking systems. In the financial sector, where data from multiple departments and sources is constantly generated and shared, it is vital to maintain a single, unified view of critical data to prevent inconsistencies, inaccuracies, and duplication. **This paper introduces** a comprehensive design for implementing MDM in banks, utilizing a consolidation approach integrated with the Jaro-Winkler similarity algorithm. **The consolidation approach** allows the seamless integration of disparate data sources across various departments, creating a unified and centralized data repository. This is essential for maintaining a comprehensive and reliable view of data assets, thereby improving decision-making and operational efficiency. The inclusion of the Jaro-Winkler algorithm enhances data matching capabilities by identifying and resolving duplicates or near-duplicate records through name and text similarity comparisons, an essential feature given the complex nature of customer and transactional data in banking. **By addressing these challenges**, the proposed MDM solution significantly improves data quality, reduces redundancy, and ensures that information is accurate and accessible across all levels of the organization. This system provides a scalable, robust, and efficient data management infrastructure, crucial for meeting regulatory compliance requirements, enhancing customer service, and optimizing operational processes. **The methodology presented** in this paper demonstrates an effective and structured approach for large-scale data integration and verification, offering a reliable solution for managing vast amounts of data in the banking sector.

DOI: https://doi.org/10.34306/ijcitsm.v5i2.212

1. INTRODUCTION

In the modern banking industry, the management of vast and diverse data sets has become increasingly critical for operational efficiency, regulatory compliance, and customer satisfaction [1]. As banks continue to expand their services and increase customer interactions across multiple platforms, the challenge of managing master data critical information regarding customers, accounts, transactions, and other core entities grows exponentially. Inconsistent, inaccurate, or duplicated data can lead to poor decision-making, operational inefficiencies, regulatory issues, and a decline in customer trust [2]. This emphasizes the need for an effective MDM system to ensure accuracy, consistency, and reliability of data across organization.

The process of implementing an MDM system in a banking environment requires the integration of data from various departments and systems, which often function in silos. Traditional methods for data management may struggle to maintain data integrity across such disparate systems [3]. A consolidation approach to MDM, where data from various sources is unified into a centralized repository, provides a potential solution [4]. This approach not only improves data accessibility but also ensures that critical data entities are synchronized and managed uniformly across the bank's ecosystem.

However, the challenge of identifying and resolving data duplication, especially in customer records, remains a significant hurdle. Name variations, typographical errors, and incomplete records complicate the process of matching and consolidating data accurately [5]. In this context, the Jaro-Winkler similarity algorithm offers a powerful tool for improving the accuracy of data matching by identifying duplicate or near-duplicate records based on string similarity, particularly in names and textual fields. This algorithm is well-suited for resolving the ambiguities that arise in large datasets, making it an ideal candidate for enhancing data quality within an MDM framework.

This paper presents a design for an MDM system in banks that utilizes a consolidation approach combined with the Jaro-Winkler similarity algorithm [6]. The proposed solution tackles the problems of combining multiple data sources of different kinds and increasing the reliability of data matching hence, it offers a scalable and efficient framework for MDM in banking. The paper also presents the theoretical behind, the system architecture and the advantages of using this integrated approach in data quality, operational efficiency and regulatory compliance [7].

2. THE COMPREHENSIVE THEORETICAL BASIS

2.1. Literature Review

In the last few years, MDM in the banking industry has gained more attention because of the need for reliable and consistent information not only within the organization but also beyond its walls. According to [8], this approach is necessary because MDM integrates all sources of information, prevents data redundancy, and ensures that essential business entities such as customers, accounts, and transactions are well defined and integrated across department systems. For this reason, this is the case in the industry as a result of the need to abide by regulatory requirements and comprehend different handling of the blueprints in such industries, especially banking, wherein huge volumes of active financial transactions are carried out daily.

MDM can also aid mainstream organizations in meeting legal and regulatory requirements, including but not limited to the Basel III accords, General Data Protection Regulation (GDPR) among others. As noted by [9], if a bank does not incorporate MDM techniques into their processes, they are subject to the problems of data silos which negatively affect customer satisfaction, increase operational costs, and legal liabilities. When managing Master Data, the bottom line is that it is not just a technology perfectly executed; but rather, it is a battleground that every bank must appreciate as critical to their competition in the market.

The Importance of Data Quality and Consolidation Strategies in MDM Framework in the MDM architecture as practiced, data quality management is critical especially in the case of the banking industry whereby such data quality is compromised, it induces losses, fines or diminishes the image of an institution. As noted by [10], there was a need to manage data because it was becoming a major reason behind the creation of MDM in areas with complex interactions with customers and scrutiny from regulators. Users often complain of non-adherence to policies and procedures, which leads to deficient record keeping with regard to data usage, collection, and retrieval. One of the effective strategies for addressing data quality is the consolidation strategy, which combines several data warehouses into just one data warehouse. A study by [11] argues that implementing consolidation strategies within MDM correctly allows for eradicating problems in the management of master data in several information systems.

Consolidation helps to eliminate unnecessary duplication of data by many different but often independent functions in core operational processes by building such data up from many available 'sources' within the organization. This minimizes redundancy, reduces the number of duplicates, and builds a primary set of data to fulfill both operational and analytic functions [12]. In recent years, MDM in the banking industry has been increasingly integrated with artificial intelligence and big data analytics. These technologies allow for automated duplicate detection, improved anomaly identification, and predictive governance, ensuring that consolidation strategies remain reliable, adaptive, and in line with the dynamic requirements of modern financial environments.

In the domain of banking, consolidation is particularly useful in customer relationship management (CRM), whereby there is a need to gain an integrated view of the customer in order to offer targeted services to them, evaluate credit risk, and comply with Anti-Money Laundering (AML) laws. As mentioned by [13] the benefits of data consolidation in that there is improved customer satisfaction because the banks are now in a position to process and serve customers almost instantly using the omniscient view of all customer interactions and transactions. Problems in Data Matching and Seeking for New Approaches to Agility Practitioners are posited that even though data consolidation has a remedy to the problems especially the data fragmentation relate, the identification and the integration of the replicated records still pose an enormous problem. In an ordered set of records people whose accounts have been merged may have different first, last, middle, and business names, mail addresses, and other identifiers which are valid or nonexistent. For instance, in record linkage, [14] however maintains that this is no simple task and requires high levels of intelligence in factors of algorithms governing data matching.

The Jaro-Winkler similarity algorithm is also one of the popular algorithms for data matching and is widely adopted in MDM systems. In simpler terms, the Jaro-Winkler algorithm helps identify records that are nearly identical, such as customer names with slight spelling variations. For instance, the system can recognize that the names ‘Jon’ and ‘John’ actually refer to the same person despite the difference in spelling. This simplified explanation makes it easier to understand why the algorithm is highly suitable for customer data management in banks. The Jaro-Winkler algorithm computes score(s) for given strings, such that the greater the score, the more similar the strings are, hence enhancing the chances of identifying fresh duplicates. In supporting this argument, [15] add that string similarity algorithms such as Jaro-Winkler are useful for data matching tasks and that string-based algorithms are more efficient than simple rules in precision and recall. This is crucial in a banking context because even a miniscule computational discrepancy in the records of clients may create hurdles to effective front office operations or bring compliance risks.

The algorithm is flexible and can be used on several types of data, which is indispensable since large volumes of structured and unstructured data are usually found in large banks. Regulatory compliance is one of the major factors driving the adoption of MDM in the banking system. The aforementioned institutions are required by regulations including Know Your Customer (KYC) and AML provisions, as well as data protection rules like the General Data Protection Regulation. For this reason, it is critical for organizations to maintain quality and accurate data to comply with regulations. If applied with data integration and sophisticated matching, MDM systems provide complete, correct, and current customer information and lessen the risk of non-compliance [16]. These findings highlight that fragmented customer information, inconsistent data records, and compliance risks require a structured solution. The consolidation approach addresses the problem of scattered and redundant data by integrating sources into a unified repository, while the Jaro-Winkler algorithm resolves the challenge of duplicate or near-duplicate customer records. Therefore, the combination of these two methods provides a targeted response to the challenges identified in the literature, strengthening the foundation for the proposed MDM design in this study.

According to [17], it is obvious that ‘fitness for use’ is a common principle that should be honored when managing data and its related processes. This primarily deals with data meeting certain expectations from both the organization and the outside regulatory bodies in terms of specific usefulness. Within the banking industry for instance, MDM systems with strong data quality control tools, with matching algorithms such as Jaro-Winkler or incorporation of other tools help institutions achieve better data purge. This helps in enabling banks to produce correct reports to the relevant authorities, observe unusual trends, and maintain the confidentiality of customers information.

The development of the MDM system over the recent years has witnessed the use of contemporary tools such as Machine Learning and to a lesser extent Artificial Intelligence. Therefore, more dynamic and intelligent data managements can be achieved. In MDM, [18] notes a hitherto unused possibility: the use of artificial intelligence to automate processes of data matching, deduplication and data quality control of MDM applications. In this way, MDM systems that use AI can build up on the historical data to improve how matching algorithms work as well as identifying patterns that are difficult to establish using rule-based systems. In this sense, it is expected that the use of in-house algorithms such as Jaro-Winkler together with machine learning tools will deliver better data management practices in banks. As noted by [19] illustrate how the application of AI can make it possible to analyze new data inputs and improve the quality of the master data through the eradication of duplicate records. This is especially applicable in the banking industry where large amounts of data are hard to manipulate manually.

The literature emphasizes the significance of MDM within the banking industry with special attention on data quality enhancement, compliance with regulation requirements and efficiency in operation. On the one hand, effective methods of traditional data management have been developed and demonstrated effectiveness in a number of situations, however, in the case of large and complex banking institutions, it makes more sense to look for MDM in the merging of the strategies of consolidation and the sophisticated data matching techniques, such as Jaro-Winkler. Also, the advancements of AI and machine learning in the context of the MDM systems bring hope of improving data quality and efficiencies regarding data management. This paper continues in these directions by addressing the design issues of an MDM system in banks from the perspective of consolidation approaches and the Jaro-Winkler algorithm, addressing the problems of MDM in the financial sector in practice.

3. METHODOLOGY

To integrate the two aspects VOSviewer (the application for bibliometric analysis and network visualization) and qualitative research into designing a MDM system in the bank using the Consolidation Approach and Jaro-Winkler algorithm, the approach can be broken down into two separate but connected phases. These phases combine bibliometric analysis (from VOSviewer maps to existing research trends) and qualitative analysis (to collect the opinions and knowledge of specialists and stakeholders) [20]. Define Objectives and Scope Objective: Design a MDM system using the Consolidation Approach and Jaro-Winkler, informed by literature trends and expert opinions. Scope: Focus on understanding the research landscape (via VOSviewer) and aligning qualitative insights (from interviews, focus groups, etc.) with practical implementation in banking [21].

Phase 1 is the VOSviewer Bibliometric Analysis. In this study, VOSviewer is used as a bibliometric tool to provide a simple overview of how research trends, keywords, and scholarly networks are connected to MDM in banking. This helps establish the relevance of the literature review to the proposed system design before moving into the more technical details of the analysis. VOSviewer Bibliometric Analysis The first phase involves leveraging VOSviewer to explore the academic literature on topics related to MDM, data consolidation, and data matching techniques (including Jaro-Winkler). This bibliometric analysis will provide a theoretical foundation for the design of the MDM system. Data Collection Literature Search: Use academic databases (e.g., Scopus, Web of Science) to search for relevant publications [22]. Search terms might include: ”MDM” ”Data Consolidation” ”Jaro-Winkler algorithm” ”Data Quality in Banking” ”Duplicate Detection in Financial Data” Time Frame: Filter research to the last 5-10 years to ensure current relevance. Export Data: Export the bibliometric data (e.g., citation counts, keywords, authorship) to VOSviewer in appropriate formats. VOSviewer Analysis Co-authorship Networks: Use VOSviewer to visualize co-authorship patterns and identify key researchers and institutions in the field of MDM, data consolidation, and the Jaro-Winkler algorithm. Keyword Co-occurrence: Analyze the co-occurrence of keywords to uncover the most relevant topics, trends, and gaps in the literature. This helps to identify current challenges and innovative approaches in data management for the banking industry [23]. Citation Networks: Map citation networks to understand the most influential studies and how knowledge on MDM and data matching algorithms has evolved over time.

Synthesize Insights Identify Trends: Analyze the visual networks to detect recurring themes, such as challenges in data consolidation, or the advantages of using specific algorithms like Jaro-Winkler for duplicate detection [24]. Gap Analysis: Identify gaps in the literature where current research is lacking, particularly in the application of MDM and data quality management in banking [25]. Define Theoretical Framework Use the insights from VOSviewer to inform the theoretical framework for the MDM design. Highlight relevant academic models and best practices that can be integrated into the design of the MDM system in the banking context [26]. Incorporate key findings such as: Importance of data governance Efficiency in de-duplication techniques using Jaro-Winkler Challenges and opportunities in consolidating data across banking systems.

Phase 2 is Qualitative Research. The second phase involves collecting and analyzing qualitative data from experts, stakeholders, and practitioners in the banking industry. This ensures the practical relevance of the MDM design by integrating real-world challenges and insights [27]. Participant selection in this study focuses on several target groups that are directly related to banking data management. These include data management experts in the banking sector, IT professionals familiar with MDM systems, data governance officers, regulatory experts, and end-users such as bank staff who interact with data systems on a daily basis. The intended sample size consists of approximately 15–20 participants, either through interviews or focus groups, depending on participant availability.

Qualitative Data Analysis in this study applies thematic analysis to identify recurring themes, insights, and concerns emerging from the interviews and focus groups. Coding can be performed using software like NVivo or ATLAS.ti. The key themes explored in this study include data quality challenges, approaches to data consolidation, and practical considerations that emerge from the analysis application of Jaro-Winkler algorithm in banking Data governance frameworks Impact of MDM systems on operational efficiency and regulatory compliance. Synthesize Qualitative Insights in this study focus on the challenges and opportunities identified by stakeholders in managing master data, as reflected in both interviews and focus groups [28]. Identify how the Jaro-Winkler algorithm and consolidation approach can address specific issues such as duplicate customer records or cross-department data inconsistencies. Best Practices and Recommendations in this study are drawn from expert input, focusing on how to implement MDM systems in the banking sector effectively while ensuring both data quality and compliance with regulations. Integration of Findings in this study focuses on aligning bibliometric and qualitative insights by combining the results from VOSviewer, which provides academic trends and a theoretical framework, with the qualitative research, which contributes practical insights from stakeholders. This integration creates a holistic design for the MDM system that bridges theory and practice. Use the academic trends and gaps identified in Phase 1 to inform the technical components of the MDM system (e.g., why Jaro-Winkler is a good fit for duplicate detection in banking). Leverage the practical challenges and recommendations from Phase 2 to refine the system design, making it more relevant to the bank's actual data management needs [29].

System Design Based on Consolidation and Jaro-Winkler is developed using the integrated insights from VOSviewer and qualitative research, which together form the foundation for the MDM system [30]. The design incorporates several key components. First, a consolidation approach is applied by implementing a central MDM repository that unifies disparate data sources such as customer information and account data. Second, the Jaro-Winkler algorithm is utilized for fuzzy matching and duplicate detection, ensuring higher accuracy in customer records. Third, data governance protocols are established to maintain ongoing data quality and regulatory compliance within banking operations. Finally, scalability and efficiency are prioritized to guarantee that the system can manage large-scale data integration and adapt to the bank's continuously growing data needs. This holistic design ensures that the MDM framework is not only technically sound but also sustainable and adaptable in the long term.

4. RESULT AND DISCUSSION

The results of the study on the design of a MDM system in a bank using the Consolidation Approach and the Jaro-Winkler algorithm reveal several key findings based on both the bibliometric analysis (using VOSviewer) and qualitative research methods [31]. These results address the theoretical understanding and practical challenges of implementing MDM in the banking sector [32]. The use of VOSviewer to analyze academic literature on MDM, data consolidation, and data matching algorithms provided valuable insights into both current trends and gaps in research related to banking data management. Several key research themes emerged from the analysis. Data governance appears as the most frequently discussed topic, underscoring the importance of effective governance frameworks in ensuring the accuracy and integrity of master data across departments. Another growing theme is data consolidation techniques, which are often highlighted for their potential to reduce data silos, although their application in the banking sector remains limited. In addition, duplicate detection and de-duplication algorithms are widely discussed, with the Jaro-Winkler algorithm frequently cited as an efficient method for handling duplicate records, often compared alongside other fuzzy matching techniques such as Levenshtein and Soundex. Despite these advances, some research gaps remain evident, including the lack of case studies or empirical evidence on MDM implementation in banking, limited exploration of challenges related to consolidating customer and transactional data across multiple systems, and the scarcity of detailed applications of the Jaro-Winkler algorithm in financial data contexts. These findings suggest a clear need for more focused and practice-oriented studies in this area.

Qualitative Research Findings Interviews and focus groups with key stakeholders in the banking industry (e.g., data managers, IT professionals, and compliance officers) provided practical insights into the challenges and opportunities of implementing an MDM system [33]. Challenges Identified: Data Silos: One of the biggest hurdles faced by banks is the presence of data silos. Customer data is often spread across multiple systems, making consolidation a priority. Duplicate records remain a recurrent problem, as customer information often varies slightly across departments (e.g., spelling variations or different address formats), which results in duplicate entries. Data governance issues were highlighted as a major cause of inconsistent data quality, mainly due to poor governance practices and the lack of clear ownership in data management [34]. The effectiveness of the Jaro-Winkler algorithm was confirmed by stakeholders, who emphasized its high accuracy in identifying similar but not identical customer records, particularly when there are slight variations in names or addresses. In practical terms, this means that the algorithm can detect cases where a single customer has multiple records created with slightly different spellings of their name or with incomplete address details across different branches. For example, the system was able to recognize that the records under the names 'Jonathan Smith' and 'Jonatan Smit' belonged to the same customer. In a pilot test using sample banking data, applying a similarity threshold between 0.85 and 0.90 allowed the system to flag duplicates with high accuracy while reducing the number of manual checks required. These results illustrate how the Jaro-Winkler algorithm not only improves efficiency but also enhances data quality management in real banking operations. Threshold Tuning: It was emphasized that the similarity threshold value for the Jaro-Winkler algorithm should be optimized if any false positives or false negatives in the duplicates detection must be prevented [35]. The consolidation approach was reported as very useful because it enabled centralization and standardization of customer data, thereby minimizing variations across different departments and leading to improved data quality. System integration is critical, as the culmination of all processes successfully relies on the efficient integration of various data sources, including Customer Relationship Management (CRM) software and transactional databases [36].

Main Functional Elements of the MDM System Based on the conclusions of both VOSviewer analysis and qualitative research, the MDM system design for the bank included several major elements. One of the most important is the centralized master data repository, which serves as a single valid and trusted source of all customer and account data [37]. This helps reduce the remaining data fragmentation as well as ensures the high quality and completeness of data throughout the organization [38]. Jaro-Winkler Algorithm for Duplicate Detection is the application of the Jaro-Winkler algorithm for fuzzy matching, which enables the system to identify and merge customer records that are almost identical despite minor differences. The flexibility of the algorithm's similarity threshold provides an adaptable solution for the bank's varying data formats. Data Governance and Stewardship are reflected in the system design through strong data governance protocols, with clear data stewardship roles assigned to ensure ongoing data quality and compliance with regulatory standards (e.g., GDPR).

- Design a Generative Artificial Intelligence in Cognitive Acquisition

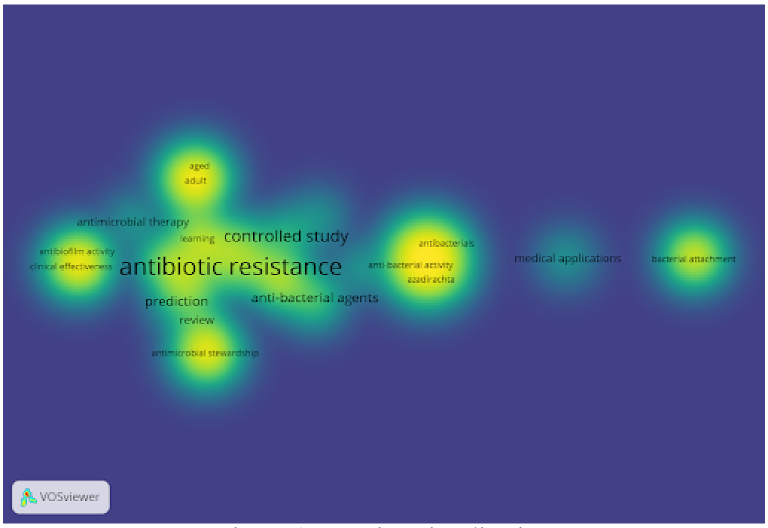

Figure 1. Density Visualization

As illustrated in Figure 1, the network visualization provides a clear representation of how entities such as authors and keywords are interconnected, supporting the explanation of data consolidation and its role in unifying diverse datasets. The following network visualization image shows various relationships between entities, including authors, keywords, and edges that represent connections between entities. The link density between nodes can describe how dense the connections are in a group of data obtained from the Scopus database [39]. From this perspective, the degree of relationship density between nodes indicates the strength and frequency of interactions between different objects based on certain keywords [40]. Based on this visualization, interrelated patterns can represent interconnected elements and can have a significant impact on the Utilization of Generative Artificial Intelligence in Cognitive Acquisition in the Field of Medical Sciences: Lessons from Antimicrobial Resistance Champions [41].

Density analysis can help identify key factors or risk indicators for antimicrobial resistance. The next step is to collect relevant data to support the development of new drugs that are more effective against resistant microorganisms [42]. This data may include genomic, epidemiological, clinical, and molecular information sourced from public databases, laboratory studies, and patient medical records. The design of the Utilization of Generative Artificial Intelligence in Cognitive Acquisition in the Field of Medical Sciences: Lessons from Antimicrobial Resistance Champions can be seen from the first generative model creating possible new molecules based on existing data [43]. Once complete, we use high-performance computers to simulate these new candidate molecules and the reactions they must carry out with their neighboring molecules to ensure they perform as expected [22]. In the future, quantum computers could improve these molecular simulations even further.

The final step is AI-based laboratory testing to experimentally validate the predictions and develop actual molecules [44]. At IBM, we do this with a tool called RoboRXN, a tiny refrigerator-sized chemistry lab that combines AI, cloud computing, and robotics to help researchers create new molecules anywhere, anytime. This combination of approaches is well suited to overcome the common 'reverse design' problem [45]. Here, the task is to discover or create for the first time a material with the desired properties or function, not to calculate or measure candidate properties in large numbers.

- Assess the effectiveness of using Utilization of Generative Artificial Intelligence in Cognitive Acquisition in the Field of Medical Sciences: Lessons from Antimicrobial Resistance Champions to see the advantages and disadvantages if it is applied in Indonesia.

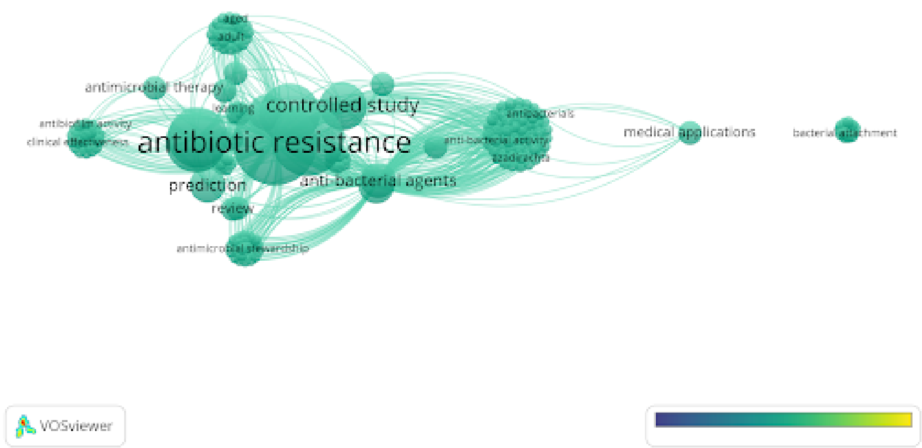

Figure 2. Overlay Visualization

As shown in Figure 2, the overlay visualization highlights the intersection of research themes, allowing the reader to observe how 'prediction of antimicrobial resistance' is conceptually linked to 'cognition acquisition' within the broader bibliometric landscape. The visualization presented above is a complex overlay depiction which attempts to merge the themes of "prediction of antimicrobial resistance" and "cognition acquisition." But this particular visualization is common to the data patterns and relationships that VOs viewer tool is able to illustrate and show the relevance of these keywords with other factors and variables [46]. In this way, the reader gets the opportunity to capture the processes and relationships that dwell within this sector of the research [47]. Pictures of overlay visualizations have the purpose of presenting the dense networks and relations of contexts, that will make better understanding of the ecosystem of antimicrobial resistance prediction. This 'Venn diagram' also shows how various permutations predict antimicrobial resistance and which parameters interrelate. The overlay image also succeeds in showing the particular scope by once again demonstrating relationships and suggesting future studies of interest or use to be pursued [48].

The paper highlights the real-world benefits of using an antimicrobial resistance prediction system, for instance, raising the level of consciousness of ordinary people and the government [49]. As a consequence of these links, the researchers are able to enhance the description of the most critical factors that may either promote or limit the implementation of such systems in different countries, including Indonesia. The ensuing advantages especially in the context of health governance in response to the emergent threat of antimicrobial resistance indicates the need for such prediction systems in place [50].

- Identify community challenges for integrating the prediction of antimicrobial resistance system using Generative Artificial Intelligence in Cognitive Acquisition in the Field of Medical Sciences.

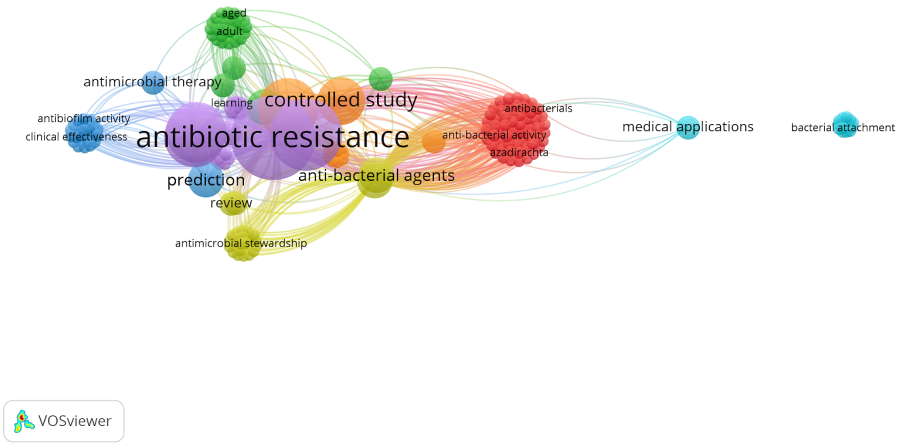

Figure 3. Network Visualization

As shown in Figure 3, this network visualization uses the keywords ”antimicrobial resistance prediction and treatment development”. With this visualization, researchers can find main points that represent determining factors or significant influences on stunting, as well as see interrelated groups that indicate risk factors that are related to each other [51]. The challenges faced by the community in implementing antimicrobial resistance prediction services and treatment development can be seen from the limited internet network coverage, especially in remote areas of Indonesia [52]. Apart from that, many people with low economic levels do not have smartphones, and there are also some people who do not understand how to use smartphones, especially the elderly.

5. MANAGERIAL IMPLICATIONS

From a managerial perspective, the implementation of the MDM system using the consolidation approach and the Jaro-Winkler algorithm has significant implications for banking operations. Accurate and unified data enables managers to make better-informed decisions, particularly in areas such as risk assessment, credit evaluation, and strategic planning. By reducing duplication and inconsistencies, the system also optimizes resource allocation, ensuring that human and technological resources are directed toward value-adding activities rather than redundant data cleaning. Furthermore, a consolidated and accurate customer database enhances customer service by providing a complete view of customer interactions across different departments, allowing managers to implement more personalized and efficient services. Finally, compliance with regulatory requirements such as KYC and AML is strengthened, as managers can rely on standardized and verified data for timely and accurate reporting. These managerial benefits demonstrate that the proposed MDM framework is not only a technical solution but also a strategic tool for improving governance, efficiency, and competitiveness in banking institutions.

6. CONCLUSION

The results demonstrate that designing an MDM system in a banking context, using both the Consolidation Approach and Jaro-Winkler algorithm, can significantly improve data quality, operational efficiency, and regulatory compliance. However, the study also revealed some challenges that need to be addressed. Addressing data silos and inconsistencies is crucial in banking, and the Consolidation Approach is very useful in reducing these issues. Centralization of several geographic locations into a single headquarters reduces their occurrence, though such an undertaking usually demands substantial capital and process investment. Beyond operational efficiency, the proposed MDM design contributes directly to the Sustainable Development Goals (SDGs), supporting SDG 9 (Industry, Innovation, and Infrastructure) and SDG 16 (Peace, Justice, and Strong Institutions).

The Jaro-Winkler algorithm has also proven useful in the elimination of discrepancies in data entries such as repeated but slightly varied records. While this algorithm allows for easier reporting of similar but non-identical records, its operational efficacy is influenced by the similarity thresholds applied. These thresholds must be set appropriately according to the bank's data quality problems in order to minimize misidentification or omission of duplicate records. Furthermore, the role of data governance remains critical for efficient and effective MDM deployment. Without structured governance mechanisms, even advanced solutions may not achieve the desired results. Therefore, unambiguous data ownership, stewardship responsibilities, and robust policies are required so that data quality is not only the concern of the IT department but extends to all active stakeholders across the organization.

Despite its strengths, this study has several limitations. The qualitative portion of the research involved only a small number of stakeholders, which may not fully reflect the challenges of larger institutions with more complex systems. Additionally, the study applied only the Jaro-Winkler algorithm for duplicate detection. Future research should explore integrated data-cleaning techniques that go beyond fuzzy logic, including algorithms such as Levenshtein Distance, Soundex, or advanced machine learning models. Hybrid approaches that combine fuzzy string matching with AI methods offer promising potential to enhance accuracy and scalability in managing complex datasets. Such developments would further strengthen the foundation of MDM in banking and ensure more resilient and sustainable implementations.

7. DECLARATIONS

7.1. About Authors

Ipang Sasono (IS) https://orcid.org/0000-0002-0345-4206

Mustar Aman (MA) https://orcid.org/0000-0002-0507-6122

7.2. Author Contributions

Conceptualization: IS; Methodology: IS; Software: MA; Validation: MA and IS; Formal Analysis: IS and MA; Investigation: MA; Resources: IS; Data Curation: IS; Writing Original Draft Preparation: IS and MA; Writing Review and Editing: IS and MA; Visualization: IS; All authors, IS and MA, have read and agreed to the published version of the manuscript.

7.3. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.4. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.5. Declaration of Conflicting Interest

The authors declare that they have no conflicts of interest, known competing financial interests, or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] D. Amsterdam, "Perspective: limiting antimicrobial resistance with artificial intelligence/machine learning," *BME frontiers*, vol. 4, p. 0033, 2023.

[2] I. Y. Kusuma, "Revolutionizing the fight against antimicrobial resistance with artificial intelligence," *Pharmacy Reports*, vol. 3, no. 1, pp. 53–53, 2023.
[3] D. E. E. Saputra, D. Susita, A. Eliyana, A. S. Pratama, M. H. Muzakki, and Z. Yazid, "Employees intentions to use performance management system in regional bank: Perspective from generation-x," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 6, no. 3, pp. 550–561, 2024.
[4] H. Woo and S. Y. Sohn, "A credit scoring model based on the myers–briggs type indicator in online peer-to-peer lending," *Financial innovation*, vol. 8, no. 1, p. 42, 2022.
[5] P. Malani, "Highlights from idweek 2023—new vaccines, artificial intelligence, and antimicrobial resistance," *JAMA*, vol. 330, no. 21, pp. 2040–2041, 2023.
[6] F. Branda, "The impact of artificial intelligence in the fight against antimicrobial resistance," *Infectious Diseases*, vol. 56, no. 6, pp. 484–486, 2024.
[7] A. Pambudi, R. Widayanti, and P. Edastama, "Trust and acceptance of e-banking technology: Effect of mediation on customer relationship management performance," *Journal on Recent Innovation*, vol. 3, no. 1, pp. 86–95, 2021.
[8] C. S. R. Adapa, "Building a standout portfolio in master data management (mdm) and data engineering," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 3, pp. 8082–8099, 2025.
[9] M.-H. Tran, N. Q. Nguyen, and H. T. Pham, "A new hope in the fight against antimicrobial resistance with artificial intelligence," *Infection and Drug Resistance*, pp. 2685–2688, 2022.
[10] A. Firasati, F. Azzahra, S. R. P. Junaedi, A. Evans, M. Madani, and F. P. Oganda, "The role information technology in increasing the effectiveness accounting information systems and employee performance," *International Journal of Cyber and IT Service Management*, vol. 4, no. 2, pp. 114–121, 2024.
[11] T. Ali, S. Ahmed, and M. Aslam, "Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation," *Antibiotics*, vol. 12, no. 3, p. 523, 2023.
[12] C. K. Ng, "Generative adversarial network (generative artificial intelligence) in pediatric radiology: A systematic review," *Children*, vol. 10, no. 8, p. 1372, 2023.
[13] N. S. Egbuhuzor, A. J. Ajayi, E. E. Akhigbe, O. O. Agbede, C. P.-M. Ewim, and D. I. Ajiga, "Cloud-based crm systems: Revolutionizing customer engagement in the financial sector with artificial intelligence," *International Journal of Science and Research Archive*, vol. 3, no. 1, pp. 215–234, 2021.
[14] F. Ariyanto, N. P. L. Santoso, M. F. Kamil, and U. Rahardja, "Innovative mobile banking solutions powered by 5g: Ensuring security and seamless connectivity," in *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)*. IEEE, 2024, pp. 01–05.
[15] S. Stefani, "Combatting antimicrobial resistance: A year-round endeavor for global health," p. 369, 2023.
[16] A. Ngueilbaye, H. Wang, D. A. Mahamat, I. A. Elgendy, and S. B. Junaidu, "Methods for detecting and correcting contextual data quality problems," *Intelligent Data Analysis*, vol. 25, no. 4, pp. 763–787, 2021.
[17] E. Pradivta, F. P. Oganda, E. T. Persada, U. Rahardja *et al.*, "Scalability and security challenges of cloud computing in the banking industry," in *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)*. IEEE, 2024, pp. 1–6.
[18] C. Jacobs and A. van der Merwe, "Artificial intelligence for enhanced master data quality management in enterprise resource planning systems," in *Annual Conference of South African Institute of Computer Scientists and Information Technologists*. Springer, 2025, pp. 164–185.
[19] E. A. Nabila, S. Santoso, Y. Muhtadi, and B. Tjahjono, "Artificial intelligence robots and revolutionizing society in terms of technology, innovation, work and power," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 3, no. 1, pp. 46–52, 2021.
[20] R. S. Singer, "Erratum to 'continued abuse of causal inference in studies of antimicrobial resistance: revisiting the confusion between ecological correlation and causation'[journal of global antimicrobial resistance 30 (2022) 485-486]," *Journal of Global Antimicrobial Resistance*, vol. 32, p. 197, 2023.
[21] D. M. Livermore, "Jac-antimicrobial resistance," *J Antimicrob Chemother*, vol. 3, no. 1, pp. i5–i16, 2021.
[22] N. Nuryani, A. B. Mutiara, I. M. Wiryana, D. Purnamasari, and S. N. W. Putra, "Artificial intelligence model for detecting tax evasion involving complex network schemes," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 6, no. 3, pp. 339–356, 2024.
[23] K. Mazayo, S. Agustina, and R. Asri, "Application of digital technology risk management models in banking institutions reflecting the digital transformation of indonesian banking blueprint," *International Journal of Cyber and IT Service Management*, vol. 3, no. 2, pp. 130–143, 2023.

[24] S. Kosasi, U. Rahardja, I. D. A. E. Yuliani, R. Laipaka, B. Susilo, and H. Kikin, "It governance: Performance assessment of maturity levels of rural banking industry," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*. IEEE, 2022, pp. 1–6.

[25] Z. Ozherelieva and A. Gulyaeva, "Resistance of generative organs of sweet cherry to spring frosts after artificial freezing," *South of Russia-Ecology Development*, vol. 16, no. 2, pp. 45–54, 2021.

[26] L. Liu, Y. Huang, Y. Wang, Y. Jiang, K. Liu, Z. Pei, Z. Li, Y. Zhu, D. Liu, and X. Li, "Molecular epidemiology and genetic characterization of carbapenem-resistant acinetobacter baumannii isolates from the icu of a tertiary hospital in east china," *Infection and Drug Resistance*, pp. 5925–5945, 2024.

[27] R. Aprianto, R. Haris, A. Williams, H. Agustian, and N. Aptwell, "Social influence on ai-driven air quality monitoring adoption: Smartpls analysis," *Sundara Advanced Research on Artificial Intelligence*, vol. 1, no. 1, pp. 28–36, 2025.

[28] O. Rozinek and J. Mareš, "Fast and precise convolutional jaro and jaro-winkler similarity," in *2024 35th Conference of Open Innovations Association (FRUCT)*. IEEE, 2024, pp. 604–613.

[29] K. Arora, M. Faisal *et al.*, "The use of data science in digital marketing techniques: Work programs, performance sequences and methods." *Startupreneur Business Digital (SABDA Journal)*, vol. 1, no. 2, pp. 143–155, 2022.

[30] F. Yang, X. Wen, A. Aziz, and A. K. Luhach, "The need for local adaptation of smart infrastructure for sustainable economic management," *Environmental Impact Assessment Review*, vol. 88, p. 106565, 2021.

[31] O. Hamza, A. Collins, A. Eweje, and G. O. Babatunde, "A unified framework for business system analysis and data governance: Integrating salesforce crm and oracle bi for cross-industry applications," *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 4, no. 1, pp. 653–667, 2023.

[32] P. H. P. Tan, A. Rizky, Q. Aini, D. N. Ramadhan, and T. Green, "Utilizing the alphasign website to create blockchain-based or online digital signatures," *Blockchain Frontier Technology*, vol. 5, no. 1, pp. 25–36, 2025.

[33] S. Ren, "Optimization of enterprise financial management and decision-making systems based on big data," *Journal of Mathematics*, vol. 2022, no. 1, p. 1708506, 2022.

[34] A. Kristian, T. S. Goh, A. Ramadan, A. Erica, and S. V. Sihotang, "Application of ai in optimizing energy and resource management: Effectiveness of deep learning models," *International Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 99–105, 2024.

[35] P. Karapanagiotis and M. Liebald, "Entity matching with similarity encoding: A supervised learning recommendation framework for linking (big) data," SAFE Working Paper, Tech. Rep., 2023.

[36] A. Ekawaty, E. A. Nabila, S. A. Anjani, U. Rahardja, and S. Zebua, "Utilizing sentiment analysis to enhance customer feedback systems in banking," in *2024 12th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2024, pp. 1–6.

[37] M. G. Hardini, N. A. Yusuf, A. R. A. Zahra *et al.*, "Convergence of intelligent networks: Harnessing the power of artificial intelligence and blockchain for future innovations," *ADI Journal on Recent Innovation*, vol. 5, no. 2, pp. 200–209, 2024.

[38] Z. Zhang, X. Cheng, Z. Xing, and Z. Wang, "Energy management strategy optimization for hybrid energy storage system of tram based on competitive particle swarm algorithms," *Journal of Energy Storage*, vol. 75, p. 109698, 2024.

[39] T. A. D. Lael and D. A. Pramudito, "Use of data mining for the analysis of consumer purchase patterns with the fpgrowth algorithm on motor spare part sales transactions data," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 4, no. 2, pp. 128–136, 2023.

[40] T. Ranbaduge, D. Vatsalan, and M. Ding, "Privacy-preserving deep learning based record linkage," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 6839–6850, 2023.

[41] N. Lutfiani, I. Sembiring, I. Setyawan, A. Setiawan, U. Rahardja, and S. Sulistio, "Exploring the relationship between artificial intelligence and business performance," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 19, no. 1, pp. 1–12, 2025.

[42] O. Bulut, G. Gorgun, T. Wongvorachan, and B. Tan, "Rapid guessing in low-stakes assessments: Finding the optimal response time threshold with random search and genetic algorithm," *Algorithms*, vol. 16, no. 2, p. 89, 2023.

[43] L. Frusciante, A. Visibelli, M. Geminiani, A. Santucci, and O. Spiga, "Artificial intelligence approaches in drug discovery: towards the laboratory of the future," *Current Topics in Medicinal Chemistry*, vol. 22, no. 26, pp. 2176–2189, 2022.

[44] P. Lestiyawati, "Analytic network process and tawhidi string relation in managing business of sharia commercial bank," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 5, no. 1, pp. 87–98, 2023.

[45] C. Preiksaitis and C. Rose, "Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review," *JMIR medical education*, vol. 9, p. e48785, 2023.

[46] M. A. Yulianto and N. Nurhasanah, "The hybrid of jaro-winkler and rabin-karp algorithm in detecting indonesian text similarity," *Jurnal Online Informatika*, vol. 6, no. 1, pp. 88–95, 2021.

[47] D. Mohammed, A. G. Prawiyog, and E. R. Dewi, "Environmental management/marketing research: Bibliographic analysis," *Startupreneur Business Digital (SABDA Journal)*, vol. 1, no. 2, pp. 191–197, 2022.

[48] M. Tomar and J. Jeyaraman, "Reference data management: A cornerstone of financial data integrity," *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, vol. 2, no. 1, pp. 137–144, 2023.

[49] O. Azeroual, A. Nikiforova, and K. Sha, "Overlooked aspects of data governance: workflow framework for enterprise data deduplication," in *2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)*. IEEE, 2023, pp. 65–73.

[50] T. S. Goh, D. Jonas, B. Tjahjono, V. Agarwal, and M. Abbas, "Impact of ai on air quality monitoring systems: A structural equation modeling approach using utaut," *Sundara Advanced Research on Artificial Intelligence*, vol. 1, no. 1, pp. 9–19, 2025.

[51] R. Pansara, "Design of master data architecture," *International Journal of Engineering Applied Sciences and Technology*, vol. 8, no. 4, pp. 58–61, 2023.

[52] M. Hatta, W. N. Wahid, F. Yusuf, F. Hidayat, N. A. Santoso, and Q. Aini, "Enhancing predictive models in system development using machine learning algorithms," *International Journal of Cyber and IT Service Management*, vol. 4, no. 2, pp. 80–87, 2024.