Global Journal of Multidisciplinary Research and Development (GJMRD)
Volume.6, Issue. 3, May-June, 2025. pp.23-27
https://gjmrd.com/

Optimizing Data Management Strategies for Enhanced Performance of Artificial Intelligence in the Banking Sector

PAWAN KUMAR,
Ai Researcher
USA.

Abstract

Artificial Intelligence (AI) is transforming the banking sector by automating services, enhancing risk management, improving customer experiences, and detecting fraud. However, the efficacy of AI models is fundamentally reliant on data quality and management strategies. This paper explores optimized data management practices to enhance the performance of AI systems in banking. By reviewing pre-2023 literature and integrating structured methodologies, the study analyzes how robust data governance, integration, cleansing, and ethical handling of data elevate AI capabilities. Furthermore, practical frameworks and visual models are presented to outline best practices for implementation. This research provides insights for financial institutions aiming to balance innovation with operational reliability and compliance.

Keywords: Artificial Intelligence, Data Management, Banking Sector, Data Governance, Machine Learning, Financial Technology, Risk Management, Data Quality, Big Data, Digital Banking.

How to Cite: Kumar, P. (2025). Optimizing data management strategies for enhanced performance of artificial intelligence in the banking sector. *Global Journal of Multidisciplinary Research and Development (GJMRD)*, *6*(3), 23–.

1. Introduction

The rapid digitization of financial services has amplified the need for intelligent systems capable of processing vast, complex datasets in real time. Artificial Intelligence (AI) is now deeply embedded in the architecture of modern banking, powering everything from predictive credit scoring to anti-money laundering (AML) solutions. Yet, the backbone of any successful AI implementation lies in how effectively data is managed.

Despite significant investments in AI, many banks still struggle with legacy systems, fragmented data silos, and inconsistent data governance protocols. Poor data quality can introduce significant model errors, while inefficient storage systems can bottleneck real-time processing. Therefore, the success of AI in banking is directly correlated to the efficiency and robustness of underlying data management strategies. This paper proposes a comprehensive approach to optimizing data handling and storage architectures, emphasizing both technological and procedural aspects to achieve superior AI performance.

2. Literature Review

The interplay between AI and data management has attracted scholarly attention across finance, computer science, and management disciplines. Prior to 2023, studies consistently underscored the critical role of high-quality, well-curated datasets in facilitating effective AI deployment in banking environments.

For instance, Sun et al. (2020) emphasized that data heterogeneity and inconsistency were among the leading barriers to AI adoption in traditional banks. Similarly, Chen et al. (2019) outlined how poor data preprocessing negatively impacts machine learning model accuracy. In a related study, Sharma and Krishna (2021) demonstrated that decentralized and unstructured data repositories reduce AI efficacy in credit scoring models. Furthermore, Deloitte (2020) found that financial institutions that implemented centralized data governance frameworks reported up to 45% higher AI-driven ROI compared to those without.

Several researchers also focused on data security and regulatory compliance as vital components of successful data management. According to Wang and Li (2022), ethical AI in finance is impossible without clear data lineage and transparent audit trails. In their comparative study of European and Asian banks, Ahmed et al. (2018) identified GDPR compliance mechanisms as a key enabler for responsible AI.

The literature thus consistently advocates for an integrated, compliant, and scalable approach to data management in the banking sector, laying a solid foundation for the present study.

3. Data Management Challenges in AI-Driven Banking Systems

3.1 Legacy Systems and Data Silos

Legacy infrastructures still dominate banking institutions, particularly in developing economies. These systems often operate in isolated environments, causing data fragmentation. This makes it difficult for AI algorithms to obtain real-time, contextual insights.

Additionally, integrating newer AI solutions with older core banking systems poses compatibility challenges. Without robust ETL (Extract, Transform, Load) processes, discrepancies in data formats and schema create further processing errors. Addressing these technical gaps requires both architectural and cultural transformation within organizations.

3.2 Regulatory and Security Constraints

Banks operate under stringent regulatory frameworks such as GDPR, CCPA, and Basel III, which mandate secure and auditable data handling. These requirements often conflict with the agile data practices needed to train dynamic AI models.

Furthermore, anonymization and encryption—while essential for data protection—can degrade the quality of AI inputs. Balancing privacy with performance is a key concern that underscores the need for smarter, policy-aware data pipelines in AI systems.

4. Optimized Data Management Strategies

4.1 Centralized Data Governance Models

Implementing centralized data governance is vital for standardizing formats, establishing ownership, and enforcing quality control across departments. A single source of truth not only reduces inconsistencies but also provides traceable data lineage, aiding in AI model audits.

Banks should adopt enterprise data catalogs and metadata management platforms to facilitate real-time discovery and classification of data. Technologies like Apache Atlas or Informatica Metadata Manager can offer a unified view of data assets, thus improving AI training accuracy.

4.2 Automation and AI-Augmented Data Cleaning

Automated data wrangling tools, supported by AI itself, can help detect anomalies, remove duplicates, and impute missing values in training datasets. These tools reduce human error and accelerate data preparation timelines.

Incorporating natural language processing (NLP) for unstructured data parsing—such as from emails or customer chats—also adds to the richness of datasets used for training AI systems.

Table 1: Comparison of Manual vs. Automated Data Cleaning

Feature	Manual Cleaning	AI-Augmented Cleaning
Speed	Slow	Fast
Error Rate	High	Low
Scalability	Limited	High
Cost	High (labor-intensive)	Moderate (tool licensing)

5. AI Model Performance Enhancement through Better Data Strategies

5.1 Improved Predictive Accuracy

Optimized data strategies significantly boost the predictive power of AI models used in fraud detection, credit scoring, and customer segmentation. Clean, labeled, and context-rich datasets allow models to uncover subtle patterns and reduce false positives.

Banks deploying AI in customer relationship management (CRM) have reported up to 30% improvement in prediction accuracy after implementing centralized and cleaned datasets, as per McKinsey (2019).

5.2 Real-time Decision-Making Capabilities

Efficient data streaming infrastructures, such as Apache Kafka or AWS Kinesis, can help AI models process transactions in real-time, crucial for fraud detection and algorithmic trading. Low-latency data systems enable instant decision-making, giving banks a competitive edge.

6. Implementation Framework for Banks

6.1 Phased Roadmap for Data Strategy Optimization

A structured roadmap includes:

1. **Assessment of Current Infrastructure**
2. **Data Audit and Cataloging**
3. **Tool and Vendor Selection**
4. **Governance Policy Rollout**
5. **Continuous Monitoring and Refinement**

Banks should follow agile methodologies for iterative deployment of these components to allow feedback-based adjustments.

6.2 Metrics for Success and ROI Measurement

Success should be measured through quantifiable KPIs such as:

- Model accuracy improvement rates
- Time-to-insight reduction
- Regulatory audit scores
- Customer satisfaction index
- Operational cost savings

7. Conclusion

Optimizing data management strategies is not merely a technical requirement but a strategic imperative for banks seeking to harness the full power of AI. Through centralized governance, automated data pipelines, and real-time data flows, banks can ensure compliance while enhancing the operational performance of AI systems. The paper underscores that without clean, contextual, and accessible data, even the most sophisticated AI models will underperform. Future research should explore industry-specific case studies and cross-national benchmarking for broader applicability.

References

1. Sun, Y., Zhang, K., & Wang, F. (2020). *Data quality challenges in AI deployment: Evidence from banking*. Journal of Financial Data Science.
2. Biru, S. (2025). Transforming Investment Banking Middle Office: A Framework for Advanced Security and Data Management. International Journal of Scientific Research

in Computer Science, Engineering and Information Technology, 11(1), 608–616. https://doi.org/10.32628/CSEIT25111268

3. Chen, L., Liu, H., & Zhou, M. (2019). *Preprocessing financial data for machine learning models: Best practices and pitfalls*. IEEE Access.

4. Biru, S. (2025). Intelligent Automation in Banking Operations: Impact Analysis on Renewable Energy Investment Assessment. International Journal of Computer Engineering and Technology (IJCET), 16(1), 673–687. https://doi.org/10.34218/IJCET_16_01_056

5. Sharma, R., & Krishna, A. (2021). *Evaluating data silos in Indian banking AI systems*. International Journal of Information Management.

6. Deloitte. (2020). *AI-driven banking: Harnessing data governance for competitive advantage*. Deloitte Insights.

7. Wang, J., & Li, Y. (2022). *Transparent data lineage for ethical AI in finance*. ACM Transactions on Management Information Systems.

8. Biru, S. (2025). AI-Powered Deduplication in Investment Banking Middle Office. International Journal of Research in Computer Applications and Information Technology (IJRCAIT), 8(1), 1713–1723. https://doi.org/10.34218/IJRCAIT_08_01_125

9. Ahmed, R., Lee, S., & Tomovic, D. (2018). *Comparative regulatory analysis of AI compliance in banking*. Journal of Banking Regulation.

10. McKinsey & Co. (2019). *Smart banking: Leveraging big data and AI*. McKinsey Financial Services Report.

11. KPMG. (2021). *Operationalizing AI in finance through integrated data pipelines*. KPMG AI in Finance Series.

12. Ghosh, P., & Malhotra, S. (2020). *AI readiness and data infrastructure in public sector banks*. Indian Journal of AI & Data Science.

13. Accenture. (2019). *Banking on AI: Why data quality is a strategic asset*. Accenture Research Report.