Data Validation and Standardization in Supply Chain Management Using Artificial Intelligence

1. Introduction

In the increasingly complex and globally interconnected landscape of supply chain management (SCM), the accuracy, consistency, and integrity of data are more critical than ever. Effective decision-making in areas such as inventory control, supplier management, logistics, and order fulfilment hinges on the availability of clean, standardized, and timely data. However, the reality is that supply chains often suffer from fragmented and inconsistent data due to disparate systems, manual inputs, and diverse stakeholder formats.

With the advent of Artificial Intelligence (AI), particularly machine learning (ML), organizations are beginning to address these long-standing issues in novel ways. AI-driven tools can automate data validation and standardization processes, ensuring that information used across supply chain systems is accurate, complete, and formatted in a consistent manner. This paper explores how AI can be implemented to improve data quality in SCM, reviews existing literature, evaluates real-world applications, and proposes a structured framework for integrating AI into supply chain data governance.

2. Problem Statement

Traditional supply chain systems are built upon multiple data sources that often do not communicate seamlessly with one another. Manual entry, outdated legacy systems, and non-standardized formats contribute to data discrepancies. This misalignment can result in significant operational inefficiencies including stockouts, overstocking, delayed shipments, invoice mismatches, and supplier miscommunication.

The key challenge lies not just in cleaning data but in creating an intelligent, scalable system capable of continuously validating and standardizing data inputs. Manual data audits are time-consuming and error-prone, and rule-based automation is often too rigid to handle the dynamic nature of supply chain data. Therefore, there is a need for adaptable and intelligent systems that can continuously learn and improve—an ideal use case for machine learning.

3. Objectives

The primary objectives of this research are:

- To explore the role of AI and machine learning in automating data validation and standardization across SCM systems.
- To evaluate the impact of AI on improving data quality and decision-making in supply chains.
- To identify industry practices and tools currently in use and assess their effectiveness.
- To propose an integrated framework for implementing AI-driven data governance in supply chain operations.

4. Literature Review

4.1. Automated Data Validation Tools

- A study by Omar et al. (2021) highlights the increasing interest in automated data validation frameworks within machine learning pipelines. These systems are designed to identify data anomalies, ensure schema consistency, and detect drifts in data quality. The authors propose best practices and a maturity model for adopting such tools, emphasizing their role in improving reliability in industrial ML applications, including SCM.

4.2. Graph-Based and Federated Models

Zhang et al. (2025) present a novel approach to enhancing supply chain visibility using Federated Learning and Graph Convolutional Networks (GCNs). By enabling decentralized training and modeling complex relationships among supply chain nodes, these models preserve data privacy while improving transparency. Such models are particularly useful for validating inter-organizational data without requiring centralized data pooling.

4.3. Systematic Review on ML in SCM

Chong et al. (2024) conducted a comprehensive review of machine learning applications in SCM, identifying key areas such as demand forecasting, procurement, and logistics. One of the major obstacles highlighted is the poor quality of underlying data and the absence of

standardized formats. The review underscores the necessity for AI systems that not only perform analytics but also prepare and standardize the data feeding those analyses.

5. AI Techniques for Data Validation and Standardization

5.1. Supervised and Unsupervised Learning

Supervised ML can be used to train classification models that detect whether a data record is likely valid based on historical labeling.

Unsupervised ML, such as clustering or anomaly detection, helps in identifying unusual data patterns, flagging potential errors, and suggesting corrections.

5.2. Natural Language Processing (NLP)

NLP plays a significant role in standardizing free-text fields such as product descriptions, supplier comments, and order notes. Named Entity Recognition (NER) and semantic analysis allow AI models to extract structured data from unstructured inputs, improving consistency and searchability.

5.3. Data Cleansing Pipelines

AI-based pipelines automate data profiling, missing value imputation, duplicate detection, and standardization. These pipelines can be integrated with ERP systems and updated continuously, reducing reliance on static validation rules.

6. Real-World Case Studies

6.1. Healthcare Supply Chains

Hospitals like Mayo Clinic and Cleveland Clinic have integrated AI into their supply chain systems to manage inventory and streamline order processing. AI models forecast inventory needs based on historical usage, predict shortages, and automate procurement, leading to more resilient healthcare logistics.

6.2. Retail Forecasting and Stock Optimization

Retailers such as Coles in Australia are leveraging machine learning platforms to forecast product demand. By incorporating data from external factors such as holidays and weather, these systems significantly improve ordering accuracy and reduce waste.

6.3. Invoice and Document Processing

AI tools are being used to validate invoices against purchase orders, extract key fields from scanned documents, and identify mismatches automatically. This reduces human error and accelerates accounts payable cycles.

7. Proposed Framework

We propose a four-layered AI-powered data governance model for SCM:

Layer	Description
1. Data Ingestion	Collects data from internal systems, third-party logistics (3PL), IoT sensors, and suppliers.
2. Data Preprocessing	Standardizes formats (e.g., date, currency), removes duplicates, and identifies missing values.
3. Machine Learning Models	Apply anomaly detection, NLP, and clustering algorithms for validation and enrichment.
4.Feedback and Learning	Incorporates user feedback to retrain models and improve over time, enabling adaptive governance.

This framework ensures that data flowing through various SCM nodes is validated, corrected, and standardized in real-time.

8. Benefits

1. Implementing AI for data validation and standardization offers multiple advantages:
2. Improved Accuracy: AI reduces manual errors and ensures high-quality data inputs.
3. Scalability: Automated systems can process vast volumes of data with minimal human intervention.
4. Operational Efficiency: Timely and accurate data enables just-in-time inventory, better supplier coordination, and streamlined workflows.
5. Enhanced Decision-Making: Reliable data provides a stronger foundation for predictive analytics and strategic planning.

9. Challenges and Considerations

Despite the potential, several challenges remain:

- **System Integration**: AI tools must work with legacy ERP and SCM systems.
- **Data Privacy**: Especially in decentralized supply chains, compliance with data sharing regulations (e.g., GDPR) is crucial.
- **Model Bias and Drift**: Without continuous monitoring, AI models may learn incorrect patterns or become obsolete.

Organizations must adopt a continuous monitoring and governance strategy to mitigate these risks.

10. Conclusion

Artificial Intelligence presents a powerful solution to the persistent challenges of data validation and standardization in supply chain management. By deploying intelligent models that learn and adapt, businesses can ensure their supply chain data is accurate, consistent, and actionable. As global supply chains become more complex and data-driven, AI will be indispensable in ensuring that data integrity keeps pace with operational demands.

References

1. Omar, M., et al. (2021). *Adoption of Automated Data Validation Tools in Industrial ML Projects*. arXiv:2103.04095.
2. Zhang, Y., et al. (2025). *A Federated Learning and GCN Approach to Supply Chain Visibility*. arXiv:2503.07231.
3. Chong, A., et al. (2024). *Machine Learning in Supply Chain Management: A Systematic Review*. IJPR. DOI: 10.1080/00207543.2025.2466062.
4. Business Insider. (2025). *AI Optimization in Hospital Inventory*.
5. The Australian. (2025). *Coles Adopts AI for Liquor Demand Forecasting*.
6. Intellias. (2025). *Machine Learning for Supply Chain Automation*.
7. ISO 8000 Standard. *Data Quality and Enterprise Master Data*. Wikipedia.