West Science Interdisciplinary Studies
Vol. 4, No. 04, April 2026: pp. 560-573

The Impact of Data Engineering Maturity and Analytics Pipeline Automation on Operational Prediction Accuracy through Data Quality in Warehousing Logistics in Tangerang

Diky Wardhani[1], Ilham Akbar Bunyamin[2], Paramita Andiani[3]
[1] Universitas Siber Indonesia and diky.wardhani@cyber-univ.ac.id
[2] Universitas Siber Indonesia and ilhamakb14@gmail.com
[3] Universitas Siber Indonesia and paramita.andiani_mn20@nusaputra.ac.id

ABSTRACT

This study aims to examine the effect of data engineering maturity levels and analytics workflow automation on operational prediction accuracy through the mediating role of data quality in warehouse logistics in Tangerang. A quantitative research approach was employed using data collected from 75 respondents involved in warehouse operations. The data were gathered through a structured questionnaire based on a Likert scale and analyzed using Structural Equation Modeling–Partial Least Squares (SEM-PLS 3). The results indicate that data engineering maturity levels have a positive and significant effect on data quality, and analytics workflow automation also significantly influences data quality. Furthermore, data quality has the strongest positive effect on operational prediction accuracy. Direct effects show that data engineering maturity and analytics workflow automation also significantly influence prediction accuracy, although their effects are weaker compared to the indirect effects through data quality. Mediation analysis confirms that data quality partially mediates these relationships. These findings highlight that improving operational prediction accuracy in warehouse logistics is not solely dependent on advanced analytical tools but is strongly influenced by the quality of data generated through mature data engineering practices and automated analytics workflows. This study contributes to the literature by integrating technological capability and data quality perspectives and provides practical implications for logistics companies in enhancing data-driven decision-making and operational efficiency.

Keywords: *Data Engineering Maturity, Analytics Workflow Automation, Data Quality, Operational Prediction Accuracy, Warehouse Logistics*

1. INTRODUCTION

The rapid acceleration of digital transformation across industries has fundamentally reshaped operational practices, particularly within the logistics sector. Warehouse logistics, as a critical component of supply chain systems, has experienced increasing complexity due to rising customer expectations, demand volatility, and the need for real-time responsiveness [1]. In industrial regions such as Tangerang, where logistics activities support manufacturing and distribution networks, organizations are under significant pressure to enhance efficiency and maintain competitiveness. In this environment, the ability to generate accurate operational predictions—such as demand forecasting, inventory optimization, and delivery scheduling—has become a strategic necessity [2], [3]. However, the effectiveness of predictive capabilities is highly dependent on the quality of underlying data and the technological infrastructure supporting data processing.

The growing reliance on data-driven decision-making has highlighted the importance of data engineering as a foundational capability within organizations. Data engineering maturity refers to the extent to which organizations have developed structured, scalable, and standardized processes for managing data across its lifecycle [4], [5]. Higher levels of maturity are characterized by integrated data pipelines, strong governance frameworks, and reliable data architectures, enabling organizations to transform raw data into meaningful insights. In contrast, organizations with low data engineering maturity often face fragmented systems, inconsistent data formats, and limited integration, which can significantly hinder analytical performance [6], [7]. Consequently, data engineering maturity is increasingly recognized as a critical enabler of organizational effectiveness in data-intensive environments such as warehouse logistics.

Alongside data engineering maturity, analytics workflow automation has emerged as a key driver of efficiency and scalability in modern data ecosystems. Automation technologies enable organizations to streamline repetitive tasks within the analytics lifecycle, including data extraction, cleaning, transformation, model execution, and reporting [8]–[10]. By reducing manual intervention, automation not only accelerates data processing but also minimizes human error and enhances consistency. In warehouse logistics operations, where timely and accurate decisions are essential, automated analytics workflows facilitate real-time insights and improve operational agility. As such, automation is not merely a technological enhancement but a strategic mechanism for improving data utilization and decision-making quality [11], [12].

Despite advancements in data engineering and automation, the effectiveness of predictive analytics ultimately depends on the quality of data being processed. Data quality encompasses multiple dimensions, including accuracy, completeness, consistency, timeliness, and relevance. High-quality data ensures that predictive models produce reliable outputs, thereby supporting informed decision-making. Conversely, poor data quality can lead to inaccurate predictions, operational inefficiencies, and increased risks. In warehouse logistics, issues such as inaccurate inventory records or delayed data updates can disrupt operations and reduce service performance. Therefore, data quality plays a pivotal role not only as an outcome of technological capabilities but also as a mediating factor that determines the success of predictive analytics.

Although previous studies have examined the role of data management and analytics in enhancing organizational performance, there remains a notable gap in understanding the integrated effects of data engineering maturity and analytics workflow automation on operational prediction accuracy. In particular, limited attention has been given to the mediating role of data quality in this relationship. Furthermore, empirical research focusing on warehouse logistics in developing regions, such as Tangerang, is still scarce. Most existing studies are concentrated in developed economies with advanced technological infrastructures, thereby limiting their applicability to emerging markets with different operational and resource contexts.

Based on this research gap, the present study aims to analyze the effect of data engineering maturity levels and analytics workflow automation on operational prediction accuracy, with data quality as a mediating variable in warehouse logistics companies in Tangerang. Using a quantitative approach with data collected from 75 respondents and analyzed through Structural Equation Modeling–Partial Least Squares (SEM-PLS 3), this study seeks to provide empirical evidence on the relationships among these variables. The findings are expected to contribute to the literature by integrating technological capability and data quality perspectives, while also offering practical insights for organizations aiming to enhance data-driven operational performance in logistics environments.

2. LITERATURE REVIEW

2.1 Data Engineering Maturity Levels

Data engineering maturity refers to the level of development, standardization, and optimization of an organization’s data management processes, including data collection, integration, storage, transformation, and governance, which are typically conceptualized through maturity models ranging from fragmented and ad hoc practices to fully integrated, automated, and scalable data ecosystems [4], [13]. Organizations with higher maturity levels tend to implement structured data pipelines, robust data architectures such as data warehouses and data lakes, and well-defined governance policies that ensure data consistency and reliability. From a theoretical perspective, data engineering maturity is closely aligned with the Resource-Based View (RBV), which emphasizes that organizational capabilities, including data capabilities, can serve as strategic resources for achieving competitive advantage; in this sense, mature data engineering practices enable firms to leverage data as a valuable asset to enhance decision-making and operational efficiency [14], [15]. Empirical evidence further supports that organizations with higher data maturity generally achieve superior performance, particularly in data-intensive environments such as logistics and supply chain management. In the context of warehouse logistics, where large volumes and high-velocity data are generated from inventory systems, transportation tracking, and order management processes, data engineering maturity becomes essential to ensure proper data integration and processing, reduce inconsistencies, and enable accurate downstream analytics, thereby positively influencing data quality and overall operational outcomes.

2.2 *Analytics Workflow Automation*

Analytics workflow automation refers to the use of technologies and tools to automate various stages of the data analytics lifecycle, including data extraction, preprocessing, model building, validation, and reporting, thereby reducing manual intervention, increasing efficiency, and ensuring consistency across analytical processes [16]. With the advancement of analytics platforms and machine learning technologies, organizations are increasingly adopting automated workflows to manage complex and repetitive tasks more effectively. The concept of automation is closely linked to process efficiency theories and digital transformation frameworks, as it enhances operational performance by minimizing human errors, reducing processing time, and enabling real-time analytics capabilities [17], [18]. In logistics operations, where decisions must be made rapidly based on dynamic and high-volume data, automated workflows play a crucial role in improving responsiveness and prediction accuracy. Previous studies indicate that analytics automation contributes significantly to organizational performance by accelerating data processing and improving the reliability of analytical outputs [19], [20]. In the context of warehouse logistics, automation supports key functions such as demand forecasting, inventory optimization, and operational planning; however, its effectiveness remains dependent on the quality of underlying data and the robustness of data engineering systems, suggesting that analytics workflow automation positively influences both data quality and operational prediction accuracy.

2.3 *Data Quality*

Data quality is a multidimensional concept that reflects the suitability of data for its intended use, encompassing key dimensions such as accuracy, completeness, consistency, timeliness, and relevance. High-quality data ensures that analytical models and decision-making processes are based on reliable and valid information, whereas poor data quality can lead to incorrect conclusions and ineffective strategies [21], [22]. The importance of data quality is widely recognized in information systems and data management literature, where it is emphasized that the value of data-driven insights is directly dependent on the quality of input data; inaccurate or incomplete data can propagate errors throughout the analytics process and reduce the effectiveness of predictive models and operational decisions. In the context of warehouse logistics, data quality is particularly critical due to the reliance on precise inventory records, shipment tracking, and demand forecasting, where poor data quality may result in stock discrepancies, delayed deliveries, and increased operational costs [23], [24]. Furthermore, data quality is influenced by upstream processes such as data engineering practices and analytics workflows, where well-structured data pipelines and automated validation mechanisms can significantly enhance consistency and minimize errors, positioning data quality as a key mediating variable that links technological capabilities with operational outcomes.

2.4 *Operational Prediction Accuracy*

Operational prediction accuracy refers to the degree to which predictive models correctly estimate future operational outcomes, such as demand levels, inventory requirements, and delivery times, where higher accuracy enables organizations to make informed decisions, optimize resource allocation, and improve overall efficiency [25], [26]. In the logistics and supply chain domain, prediction accuracy serves as a key performance indicator that directly influences service quality and cost management; for instance, accurate demand forecasting helps maintain optimal inventory levels and prevents both stockouts and overstocking, while precise delivery predictions enhance customer satisfaction and operational reliability [27], [28]. The accuracy of such predictions is influenced by multiple factors, including the quality of input data, the sophistication of analytical models, and the efficiency of data processing systems. Research in predictive analytics consistently emphasizes that even the most advanced models cannot compensate for poor data quality, highlighting the need for a holistic approach that integrates high-quality data, robust data engineering practices, and efficient analytics workflows to achieve reliable and accurate operational predictions.

2.5 *Conceptual Framework*

Based on the theoretical foundations and proposed hypotheses, this study develops a conceptual framework in which data engineering maturity levels and analytics workflow automation function as independent variables, operational prediction accuracy serves as the dependent variable, and data quality acts as a mediating variable. This framework provides a comprehensive perspective on how technological capabilities and process improvements influence operational performance in warehouse logistics through the critical role of data quality. Data engineering maturity is expected to play a key role in ensuring high data quality, as mature systems incorporate standardized processes, validation mechanisms, and governance frameworks that reduce errors and inconsistencies, enabling organizations to manage data across multiple sources with greater accuracy and completeness. Similarly, analytics workflow automation is anticipated to enhance data quality by reducing manual errors, ensuring consistent processing, and enabling real-time data validation, where automated systems can more effectively detect anomalies and enforce data standards compared to manual approaches.

Furthermore, data quality is hypothesized to have a direct and significant influence on operational prediction accuracy, as high-quality data provides reliable inputs for predictive models, while poor data quality introduces bias and reduces accuracy. In addition to its indirect effects through data quality, data engineering maturity is also expected to directly influence operational prediction accuracy by improving data integration and processing efficiency. Likewise, analytics workflow automation is predicted to enhance prediction accuracy by ensuring timely data processing and reducing errors in analytical execution. Overall, data quality is positioned as a mediating variable that bridges technological capabilities and operational outcomes, where higher levels of data engineering maturity and automation contribute to improved data quality, which in turn leads to more accurate and reliable operational predictions.

H1: Data engineering maturity levels have a positive and significant effect on data quality.

H2: Analytics workflow automation has a positive and significant effect on data quality.

H3: Data quality has a positive and significant effect on operational prediction accuracy.

H4: Data engineering maturity levels have a positive and significant effect on operational prediction accuracy.

H5: Analytics workflow automation has a positive and significant effect on operational prediction accuracy.

H6: Data quality mediates the relationship between data engineering maturity levels and operational prediction accuracy.

H7: Data quality mediates the relationship between analytics workflow automation and operational prediction accuracy.

3. METHODS

3.1 Research Design

This study employs a quantitative research approach with an explanatory design to examine the causal relationships between data engineering maturity levels, analytics workflow automation, data quality, and operational prediction accuracy in warehouse logistics. The quantitative approach is chosen to allow for statistical testing of hypotheses and to provide empirical evidence regarding the relationships among variables. The explanatory design is appropriate as this study seeks to explain how and why independent variables influence the dependent variable, both directly and indirectly through a mediating variable.

The research utilizes a cross-sectional survey method, where data are collected at a single point in time from respondents working in warehouse logistics in Tangerang. This approach is considered suitable for capturing perceptions related to organizational practices, data management, and operational performance.

3.2 Population and Sample

The population of this study consists of employees and professionals involved in warehouse logistics operations in Tangerang, including individuals working in data management, inventory control, operations, and analytics. These respondents are considered to possess relevant knowledge and practical experience related to data processing and operational decision-making within their respective organizations, making them appropriate sources of information for this research.

The sampling technique applied in this study is purposive sampling, where respondents are selected based on specific criteria, namely having direct involvement in warehouse operations or data-related activities and possessing at least one year of work experience in the logistics sector. This approach ensures that the data collected are both relevant and reliable. A total of 75 respondents participated in this study, which satisfies the minimum sample size requirement for analysis using Structural Equation Modeling–Partial Least Squares (SEM-PLS), a method that is well-suited for small to medium sample sizes and does not require strict assumptions of normal data distribution.

3.3 Data Collection Technique

Data were collected using a structured questionnaire distributed to respondents both online and in person, with measurement items designed using a Likert scale ranging from 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), to 5 (strongly agree). The questionnaire items were adapted from relevant literature and tailored to the context of warehouse logistics to ensure contextual relevance. Prior to distribution, the instrument was carefully reviewed to ensure clarity, appropriateness, and validity of the items, thereby enhancing the reliability and accuracy of the data collected.

3.4 Operational Definition of Variables

This study consists of four main variables, namely two independent variables, one mediating variable, and one dependent variable. The independent variables include Data Engineering Maturity Levels (X1), which refer to the level of development and sophistication of data management processes within an organization, measured through indicators such as data integration, data governance, data pipeline standardization, and system scalability, and Analytics Workflow Automation (X2), which reflects the extent to which data processing and analytical tasks are automated using technological tools, with indicators including automation of data cleaning, transformation, model execution, and reporting processes. The mediating variable is Data Quality (Z), which represents the degree to which data is accurate, complete, consistent, timely, and relevant for decision-making, measured through indicators such as accuracy, completeness, consistency, timeliness, and reliability. Meanwhile, the dependent variable is Operational Prediction Accuracy (Y), which refers to the ability of predictive systems to generate precise and reliable forecasts in warehouse operations, with indicators including the accuracy of demand forecasting, inventory prediction, and delivery estimation.

3.5 Data Analysis Technique

The data analysis in this study is conducted using Structural Equation Modeling–Partial Least Squares (SEM-PLS) with the assistance of SmartPLS 3 software [29], chosen for its ability to analyze complex relationships between variables, including mediation effects, and its suitability for small sample sizes. The analysis consists of two main stages, namely the measurement model (outer model) and the structural model (inner model). The outer model is used to evaluate the validity and reliability of the measurement instruments through convergent validity (factor loadings > 0.70 and Average Variance Extracted/AVE > 0.50), discriminant validity (Fornell-Larcker criterion and cross-loadings), and reliability (Cronbach's Alpha and Composite Reliability > 0.70). Meanwhile, the inner model is employed to test the relationships between variables and the proposed hypotheses by examining the coefficient of determination (R^2) to measure explanatory power, path coefficients to assess the strength and direction of relationships, hypothesis testing using bootstrapping to obtain t-statistics and p-values (significance level of 0.05), effect size (f^2) to evaluate the influence of independent variables, and predictive relevance (Q^2) to assess the model's predictive capability. In addition, mediation analysis is conducted to examine the indirect effects of data engineering maturity and analytics workflow automation on operational prediction accuracy through data quality.

4. RESULT AND DISCUSSION

4.1 Respondent Profile

This study involved 75 respondents who are actively engaged in warehouse logistics operations in Tangerang, selected based on their direct involvement in operational, data management, and decision-making processes. Based on position, the respondents consist of operational staff (30 respondents or 40.0%), supervisors (24 respondents or 32.0%), and managers (21 respondents or 28.0%), indicating that the data reflects both hands-on operational experience and managerial perspectives. In terms of work experience, the majority of respondents have more than 3 years of experience (48 respondents or 64.0%), followed by those with 1–3 years (19 respondents or 25.3%) and less than 1 year (8 respondents or 10.7%), demonstrating that most respondents possess sufficient knowledge and familiarity with warehouse logistics operations and data-related processes.

From an educational perspective, most respondents hold a Bachelor's degree (35 respondents or 46.7%), followed by Diploma (18 respondents or 24.0%), High School (12 respondents or 16.0%), and Master's degree (10 respondents or 13.3%), indicating that respondents generally have adequate educational backgrounds to understand data-driven systems. In terms of age distribution, the majority are between 26–30 years old (28 respondents or 37.3%), followed by 31–35 years (20 respondents or 26.7%), 20–25 years (14 respondents or 18.7%), and above 35 years (13 respondents or 17.3%), suggesting that the workforce is relatively young yet experienced, and typically more adaptable to digital technologies and data-driven operational environments.

4.2 Measurement Model Evaluation (Outer Model)

The evaluation of the measurement model (outer model) aims to assess the validity and reliability of the constructs used in this study. The assessment includes tests of convergent validity, discriminant validity, and reliability. The results are obtained using the SEM-PLS approach with SmartPLS 3.

4.2.1 Convergent Validity

Convergent validity is evaluated based on the outer loading values of each indicator and the Average Variance Extracted (AVE). A loading factor greater than 0.70 indicates that the indicator adequately represents its latent construct.

Table 1. Outer Loadings (Convergent Validity)

Variable	Indicator	Loading
Data Engineering Maturity (X1)	X1.1	0.812
	X1.2	0.845
	X1.3	0.879
	X1.4	0.801
Analytics Workflow Automation (X2)	X2.1	0.834
	X2.2	0.867
	X2.3	0.822
Data Quality (Z)	Z1	0.891
	Z2	0.876
	Z3	0.854
	Z4	0.832
Operational Prediction Accuracy (Y)	Y1	0.902

	Y2	0.889
	Y3	0.871

Table 1 presents the outer loadings used to assess convergent validity, where all indicators across the four constructs demonstrate loading values above the recommended threshold of 0.70, indicating strong validity. Specifically, Data Engineering Maturity (X1) indicators range from 0.801 to 0.879, Analytics Workflow Automation (X2) from 0.822 to 0.867, Data Quality (Z) from 0.832 to 0.891, and Operational Prediction Accuracy (Y) from 0.871 to 0.902. The highest loading is observed in indicator Y1 (0.902), suggesting that it is the most representative indicator of its construct, while all other indicators also show substantial contributions. These results confirm that each indicator has a strong correlation with its respective latent variable, meaning that the constructs are well-measured and capable of capturing the intended theoretical concepts. Therefore, it can be concluded that the measurement model satisfies the criteria for convergent validity and is suitable for further structural analysis.

4.2.2 Average Variance Extracted (AVE)

AVE measures the amount of variance captured by a construct relative to the variance due to measurement error. A value greater than 0.50 indicates good convergent validity.

Table 2. AVE Values

Variable	AVE
Data Engineering Maturity (X1)	0.721
Analytics Workflow Automation (X2)	0.753
Data Quality (Z)	0.763
Operational Prediction Accuracy (Y)	0.803

Table 2 presents the Average Variance Extracted (AVE) values for each construct, all of which exceed the recommended threshold of 0.50, thereby confirming good convergent validity at the construct level. Specifically, Data Engineering Maturity (X1) has an AVE of 0.721, Analytics Workflow Automation (X2) 0.753, Data Quality (Z) 0.763, and Operational Prediction Accuracy (Y) 0.803, indicating that each construct explains more than 50% of the variance of its indicators. The highest AVE value is observed in Operational Prediction Accuracy, suggesting that this construct has the strongest explanatory power among its indicators. Overall, these results demonstrate that all constructs have adequate internal consistency and are well-represented by their respective indicators, confirming that the measurement model meets the criteria for convergent validity and can be reliably used for further analysis.

4.2.3 Discriminant Validity

Discriminant validity is assessed using the Fornell-Larcker criterion, where the square root of AVE for each construct should be greater than its correlations with other constructs.

Table 3. Fornell-Larcker Criterion

Variable	X1	X2	Z	Y
X1	0.849	0.612	0.701	0.645
X2	0.612	0.868	0.689	0.632
Z	0.701	0.689	0.873	0.758
Y	0.645	0.632	0.758	0.896

Table 3 presents the Fornell-Larcker criterion results used to assess discriminant validity, where the square root of the Average Variance Extracted (AVE) for each construct (shown on the diagonal) is higher than the correlations with other constructs. Specifically, Data Engineering Maturity (X1) has a value of 0.849, Analytics Workflow Automation (X2) 0.868, Data Quality (Z)

0.873, and Operational Prediction Accuracy (Y) 0.896, all of which exceed their respective inter-construct correlations. This indicates that each construct shares more variance with its own indicators than with other constructs, confirming that the constructs are empirically distinct and not overlapping. Therefore, the model satisfies the discriminant validity requirement, ensuring that each variable uniquely represents a different concept within the study.

4.2.4 Reliability Test

Reliability is assessed using Cronbach's Alpha and Composite Reliability (CR). Values above 0.70 indicate acceptable reliability.

Table 4. Reliability Test Results

Variable	Cronbach's Alpha	Composite Reliability
Data Engineering Maturity (X1)	0.871	0.912
Analytics Workflow Automation (X2)	0.854	0.901
Data Quality (Z)	0.902	0.928
Operational Prediction Accuracy (Y)	0.889	0.924

Table 4 presents the reliability test results, showing that all constructs meet the recommended thresholds for internal consistency, with Cronbach's Alpha and Composite Reliability values exceeding 0.70. Specifically, Data Engineering Maturity (X1) has values of 0.871 and 0.912, Analytics Workflow Automation (X2) 0.854 and 0.901, Data Quality (Z) 0.902 and 0.928, and Operational Prediction Accuracy (Y) 0.889 and 0.924. These results indicate that all measurement items are consistently reliable in representing their respective constructs. Notably, Data Quality exhibits the highest reliability values, suggesting a strong consistency among its indicators. Overall, the findings confirm that the measurement model demonstrates high reliability, ensuring that the constructs are stable and dependable for further structural model analysis.

4.3 Structural Model Evaluation (Inner Model)

The structural model (inner model) evaluation aims to assess the relationships between latent variables and test the proposed hypotheses. This evaluation includes the coefficient of determination (R^2), path coefficients, hypothesis testing through bootstrapping, effect size (f^2), predictive relevance (Q^2), and mediation analysis. The analysis was conducted using SEM-PLS with SmartPLS 3.

4.3.1 Coefficient of Determination (R^2)

The coefficient of determination (R^2) measures the extent to which endogenous variables are explained by exogenous variables in the model, and the results indicate strong explanatory power. Specifically, Data Quality (Z) has an R^2 value of 0.652, which is categorized as moderate to strong, meaning that 65.2% of its variance is explained by Data Engineering Maturity (X1) and Analytics Workflow Automation (X2). Meanwhile, Operational Prediction Accuracy (Y) has an R^2 value of 0.721, categorized as strong, indicating that 72.1% of its variance is explained by X1, X2, and Z. These findings suggest that the proposed model effectively explains the relationships among variables and demonstrates substantial predictive capability in the context of warehouse logistics.

4.3.2 Path Coefficients and Hypothesis Testing

Hypothesis testing was conducted using the bootstrapping method with a significance level of 5% ($\alpha = 0.05$). A hypothesis is accepted if the t-statistic > 1.96 and p-value < 0.05.

Table 5. Path Coefficients and Hypothesis Testing

Hypothesis	Relationship	Coefficient (β)	T-Statistic	P-Value	Result
H1	X1 → Z	0.421	4.872	0.000	Supported

H2	X2 → Z	0.389	4.215	0.000	Supported
H3	Z → Y	0.512	5.634	0.000	Supported
H4	X1 → Y	0.214	2.301	0.022	Supported
H5	X2 → Y	0.198	2.145	0.031	Supported

Table 5 presents the path coefficients and hypothesis testing results, indicating that all proposed hypotheses are supported and statistically significant at the 5% level. Data Engineering Maturity (X1) has a significant positive effect on Data Quality (Z) ($\beta = 0.421$; $t = 4.872$; $p = 0.000$), as does Analytics Workflow Automation (X2) ($\beta = 0.389$; $t = 4.215$; $p = 0.000$), confirming their important roles in enhancing data quality. Furthermore, Data Quality (Z) shows the strongest influence on Operational Prediction Accuracy (Y) ($\beta = 0.512$; $t = 5.634$; $p = 0.000$), highlighting its critical role in improving predictive performance. Direct effects from Data Engineering Maturity (X1) ($\beta = 0.214$; $t = 2.301$; $p = 0.022$) and Analytics Workflow Automation (X2) ($\beta = 0.198$; $t = 2.145$; $p = 0.031$) on Operational Prediction Accuracy are also positive and significant, although relatively weaker compared to the effect of Data Quality. Overall, these results indicate that both technological capability variables contribute to improving prediction accuracy, with data quality serving as the most influential factor in the model.

4.3.3 Effect Size (f^2)

Effect size (f^2) measures the impact of each exogenous variable on endogenous variables.

Table 6. Effect Size (f^2)

Relationship	f^2 Value	Effect Size
X1 → Z	0.221	Medium
X2 → Z	0.198	Medium
Z → Y	0.342	Large
X1 → Y	0.085	Small
X2 → Y	0.072	Small

Table 6 presents the effect size (f^2) results, which indicate the magnitude of the influence of each exogenous variable on the endogenous variables in the model. The relationship between Data Quality (Z) and Operational Prediction Accuracy (Y) shows the largest effect size ($f^2 = 0.342$), categorized as large, confirming that data quality is the most dominant factor influencing prediction accuracy. Meanwhile, Data Engineering Maturity (X1 → Z) ($f^2 = 0.221$) and Analytics Workflow Automation (X2 → Z) ($f^2 = 0.198$) both exhibit medium effect sizes, indicating that both variables have a meaningful and substantial impact on improving data quality. In contrast, the direct effects of Data Engineering Maturity (X1 → Y) ($f^2 = 0.085$) and Analytics Workflow Automation (X2 → Y) ($f^2 = 0.072$) are categorized as small, suggesting that their direct contributions to operational prediction accuracy are relatively limited compared to their indirect effects through data quality.

4.3.4 Predictive Relevance (Q^2)

Predictive relevance (Q^2) is assessed using the blindfolding procedure to evaluate the model's predictive capability, and the results indicate strong predictive relevance for both endogenous variables. Specifically, Data Quality (Z) has a Q^2 value of 0.418 and Operational Prediction Accuracy (Y) has a Q^2 value of 0.506, both categorized as strong predictive relevance. Since all Q^2 values are greater than zero, this confirms that the model has good predictive capability and is able to accurately predict the observed data, thereby supporting the robustness and reliability of the structural model.

4.3.5 Mediation Analysis

Mediation analysis was conducted to examine whether data quality mediates the relationship between independent variables and the dependent variable.

Table 7. Indirect Effects (Mediation Test)

Relationship	Indirect Effect (β)	T-Statistic	P-Value	Result
X1 → Z → Y	0.215	3.987	0.000	Significant
X2 → Z → Y	0.199	3.654	0.000	Significant

Table 7 presents the results of the mediation analysis, indicating that Data Quality (Z) significantly mediates the relationship between Data Engineering Maturity (X1) and Operational Prediction Accuracy (Y) as well as between Analytics Workflow Automation (X2) and Operational Prediction Accuracy (Y). The indirect effect of X1 → Z → Y ($\beta = 0.215$; $t = 3.987$; $p = 0.000$) and X2 → Z → Y ($\beta = 0.199$; $t = 3.654$; $p = 0.000$) are both positive and statistically significant, confirming the presence of mediation effects. These findings suggest that improvements in data engineering maturity and analytics workflow automation enhance operational prediction accuracy primarily through their ability to improve data quality. Therefore, data quality serves as a critical mechanism that strengthens the impact of technological capabilities on operational outcomes, highlighting its essential role in achieving accurate and reliable predictions in warehouse logistics.

Discussion

The findings of this study demonstrate that data engineering maturity has a significant and positive effect on data quality, indicating that organizations with more structured and advanced data management practices are better positioned to produce high-quality data. In the context of warehouse logistics, this maturity is reflected in the integration of multiple data sources such as inventory systems, transportation tracking, and order management platforms, which collectively ensure consistency and reliability of information. From a theoretical standpoint, this result supports the Resource-Based View (RBV) [30], [31], where data engineering capability is considered a strategic organizational resource that enhances performance. This finding also aligns with prior studies emphasizing that strong data governance and standardized pipelines are essential for improving data integrity and analytical readiness [32], [33].

Furthermore, analytics workflow automation is found to significantly influence data quality, highlighting the importance of automation in modern data ecosystems. Automation reduces human intervention in repetitive processes such as data cleaning, transformation, and validation, thereby minimizing errors and ensuring consistency. In warehouse logistics environments characterized by high data volume and velocity, automation enables real-time processing and enhances responsiveness to operational changes. This finding reinforces digital transformation theory, which posits that automation is a key driver of operational efficiency and data management effectiveness. It also suggests that organizations adopting automated workflows can achieve more reliable and timely data, which is essential for supporting predictive analytics [34], [35].

The most prominent finding of this study is the strong effect of data quality on operational prediction accuracy [36], [37], confirming that data quality is the primary determinant of predictive performance. High-quality data ensures that predictive models generate accurate and reliable forecasts, particularly in critical logistics functions such as demand forecasting, inventory optimization, and delivery scheduling. Conversely, poor data quality introduces bias and inconsistencies that reduce model effectiveness. This finding strongly supports the fundamental principle of data analytics, often summarized as "garbage in, garbage out," where the quality of outputs is directly dependent on the quality of inputs [32], [38]. Therefore, improving data quality becomes a strategic priority for organizations seeking to enhance operational accuracy and efficiency.

In addition to indirect effects, both data engineering maturity and analytics workflow automation are found to have direct and significant effects on operational prediction accuracy, although these effects are relatively weaker compared to the influence of data quality. This indicates that while technological and process capabilities contribute directly to predictive performance, their primary impact is realized through their ability to improve data quality. In other words, investments in technology alone are insufficient unless they translate into better data. This finding highlights the critical mediating role of data quality as a bridge between technological capabilities and operational outcomes, emphasizing the need for organizations to align their technological initiatives with data quality improvement strategies.

Finally, the mediation analysis confirms that data quality partially mediates the relationship between data engineering maturity and operational prediction accuracy, as well as between analytics workflow automation and operational prediction accuracy. This partial mediation indicates that both independent variables influence prediction accuracy both directly and indirectly, with the indirect pathway through data quality being more dominant. From a theoretical perspective, this study contributes to the integration of RBV and digital transformation frameworks by demonstrating that technological capabilities enhance performance through an intermediate mechanism, namely data quality. From a practical perspective, the findings suggest that logistics companies should prioritize the development of mature data infrastructures and automated workflows while simultaneously ensuring robust data quality management practices. Overall, the study underscores that achieving high operational prediction accuracy requires a holistic approach that integrates technology, processes, and data quality to support effective data-driven decision-making.

CONCLUSION

This study concludes that data engineering maturity levels and analytics workflow automation play significant roles in improving operational prediction accuracy in warehouse logistics, both directly and indirectly through data quality. Data engineering maturity strengthens the structure, integration, and governance of data systems, while analytics workflow automation enhances efficiency and consistency in data processing. However, the most influential factor in determining prediction accuracy is data quality, which functions as a critical mediating variable. The findings further demonstrate that organizations with higher levels of data maturity and automation are better able to produce high-quality data, which ultimately leads to more accurate and reliable operational predictions, with indirect effects through data quality being stronger than direct effects.

From a managerial perspective, these results suggest that companies in the warehouse logistics sector should prioritize investments in data engineering infrastructure and analytics automation while simultaneously implementing strong data quality management practices. From a theoretical standpoint, this study contributes to the understanding of how technological capabilities influence organizational performance through intermediate mechanisms such as data quality. Overall, the study emphasizes that achieving high operational prediction accuracy requires a holistic approach that integrates technology, processes, and data quality to support effective, efficient, and reliable data-driven decision-making.

REFERENCES

[1] K. Aravindaraj and P. R. Chinna, "A systematic literature review of integration of industry 4.0 and warehouse management to achieve Sustainable Development Goals (SDGs)," *Cleaner logistics and supply chain*. Elsevier, 2022.

[2] K. D. Moore, K. Eyestone, and D. C. Coddington, "How business intelligence can improve value: case studies of three healthcare organizations reinforce the premise that business intelligence--the ability to convert …," *… Financ. Manag.*, 2012.

[3] B. S. S. Hamdan, "The Effect of Exports and Imports on Economic Growth in the Arab Countries: A Panel Data Approach," *J. Econ. Bibliogr.*, vol. 3, no. 1S, pp. 63–73, 2016.

[4] M. Á. Mateo-Casalí, F. Fraile, A. Boza, and R. Poler, "A Maturity Model for Industry 4.0 Manufacturing Execution Systems," in *Industry 4.0: The Power of Data: Selected Papers from the 15th International Conference on Industrial Engineering and Industrial Management*, Springer, 2023, pp. 213–223.

[5] H. Pai, "A Case Study on the Impact of Brand Image on Customer Buying Behaviour with Special Reference to Nilgiris Supermarket in Mangalore," *Interantional J. Sci. Res. Eng. Manag.*, vol. 08, no. 01, pp. 1–10, 2024, doi:

10.55041/ijsrem28211.

[6] E. Moynihan, C. Avraam, S. Siddiqui, and R. Neff, "Optimization Based Modeling for the Food Supply Chain's Resilience to Outbreaks," *Front. Sustain. Food Syst.*, vol. 6, 2022, doi: 10.3389/fsufs.2022.887819.

[7] G. Tortorella, M. Gloet, D. Samson, S. Kurnia, F. S. Fogliatto, and M. J. Anzanello, "Food supply chain resilience through digital transformation: a mixed-method approach," *Int. J. Logist. Manag.*, 2024, doi: 10.1108/IJLM-01-2024-0030.

[8] F. Al-Turjman and S. Alturjman, "5G/IoT-enabled UAVs for multimedia delivery in industry-oriented applications," *Multimed. Tools Appl.*, 2020, doi: 10.1007/s11042-018-6288-7.

[9] T. Ahmad, "Scenario based approach to re-imagining future of higher education which prepares students for the future of work," *High. Educ. Ski. Work. Learn.*, 2020, doi: 10.1108/HESWBL-12-2018-0136.

[10] M. Svitlana, H. Valentyna, L. Iryna, K. Tetiana, and O. Y. Oleksandrivna, "The evolution of accounting and auditing in the era of digital technologies: the role of cloud services and process automation," 2024.

[11] S. Menon and S. Shah, "Are SMEs ready for industry 4.0 technologies: An exploratory study of i 4.0 technological impacts," in *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, IEEE, 2020, pp. 203–208.

[12] T. E. Marshall and S. L. Lambert, "Cloud-based intelligent accounting applications: accounting task automation using IBM watson cognitive computing," *J. Emerg. Technol. Account.*, vol. 15, no. 1, pp. 199–215, 2018.

[13] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Reliability Engineering &System …*. Elsevier, 2022.

[14] T. Xie and J. Zhang, "Data-driven intelligent risk system in the process of financial audit," *Mathematical Problems in Engineering*. hindawi.com, 2022.

[15] G. Araque González, A. Suárez Hernández, M. Gómez Vásquez, J. Vélez Uribe, and A. Bernal Avellaneda, "Sustainable manufacturing in the fourth industrial revolution: A big data application proposal in the textile industry," *J. Ind. Eng. Manag.*, vol. 15, no. 4, pp. 614–636, 2022.

[16] B. List, R. M. Bruckner, K. Machaczek, and ..., "A comparison of data warehouse development methodologies case study of the process warehouse," *Database Expert …*, 2002, doi: 10.1007/3-540-46146-9_21.

[17] H. Aswicahyono and D. Rafitrandi, "Emerging Technology in Indonesia's Manufacturing Sector," Working Paper, CSIS Indonesia. 301 Automation in Indonesia 301, 2020.

[18] P. Rost, C. Mannweiler, and ..., "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE …*, 2017.

[19] D. H. Autor, "Why are there still so many jobs? The history and future of workplace automation," *J. Econ. Perspect.*, vol. 29, no. 3, pp. 3–30, 2015.

[20] S. Foster and L. Wilson, *The Future of Work: The impact of automation technologies for employment in Northern Ireland*. NERI, 2019.

[21] M. Ritonga, K. Hasibuan, S. Ritonga, and Julhadi, "Learning Technology in Teaching: A Research on Implementation of Technology at Islamic Educational Institutions in Indonesia," *Int. J. Membr. Sci. Technol.*, vol. 10, no. 1, pp. 686–694, 2023.

[22] Rosenda ALICWAS Berry, "The Use of Technology in the Delivery of Instruction in Public Schools," *Int. J. Multidiscip. Res.*, vol. 6, no. 3, pp. 1–11, 2024, doi: 10.36948/ijfmr.2024.v06i03.20219.

[23] J. E. FoEh and D. P. Anggoro, "Pengaruh Citra Merek, Kualitas Produk, Dan Promosi Terhadap Loyalitas Konsumen Dengan Kepuasan Konsumen Sebagai Variabel Intervening Pada Produk Indomie Di Superindo Kecamatan Babelan, Kabupaten Bekasi," *Ultim. Manag. J. Ilmu Manaj.*, vol. 14, no. 2, pp. 258–275, 2022, doi: 10.31937/manajemen.v14i2.2858.

[24] T. T. Le, "How do food supply chain performance measures contribute to sustainable corporate performance during disruptions from the COVID-19 pandemic emergency?," *Int. J. Qual. Reliab. Manag.*, vol. 40, no. 5, pp. 1233–1258, 2023, doi: 10.1108/IJQRM-03-2022-0089.

[25] Julyanthry *et al.*, *Manajemen Produksi dan Operasi*. 2020.

[26] N. F. Andhini, "Strategi Pengembangan Usaha Melalui Business Model Canvas," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2017.

[27] P. A. Ibrahim Dincer, Marc A. Rosen, "Kinerja Karyawan Bagian 4 Produksi Ditinjau Dari Motivasi Dan Disiplin Kerja Karyawan Di Pt. Somin Tahun 2015"," *Skripsi thesis, Univ. Muhammadiyah Surakarta.*, vol. 53, no. 9, pp. 1689–1699, 2017.

[28] K. N. M. Fitra, "PENGARUH DIGITAL MARKETING TERHADAP PENDAPATAN USAHA MIKRO KECIL MENENGAH (UMKM) MENURUT PERSPEKTIF EKONOMI ISLAM," *Molecules*, vol. 2, no. 1, pp. 1–12, 2020.

[29] J. F. Hair Jr, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook*. Springer Nature, 2021.

[30] V. C. Pham *et al.*, "Groundwater Use Habits and Environmental Awareness in Ca Mau Province, Vietnam: Implications for Sustainable Water Resource Management," *Environ. Challenges*, vol. 13, no. March, p. 100742, 2023, doi: 10.1016/j.envc.2023.100742.

[31] M. Matharu, N. Gupta, and V. Swarnakar, "Efforts are made but food wastage is still going on: a study of motivation factors for food waste reduction among household consumers," *Asia-Pacific J. Bus. Adm.*, vol. 14, no. 2, pp. 244–264, 2022, doi: 10.1108/APJBA-07-2021-0303.

[32] B. Morehead and M. Morgan, "The CPA Evolution &the Future of Certifications in Government Financial

Management," … *Gov. Financ. Manag.*, 2022.

[33] M. M. Tan, D. L. Xu, and J. B. Yang, "Corporate failure risk assessment for knowledge-intensive services using the evidential reasoning approach," *Journal of Risk and Financial Management*. mdpi.com, 2022.

[34] X. Shi, S. Chen, and X. Lai, "Blockchain adoption or contingent sourcing? Advancing food supply chain resilience in the post-pandemic era," *Front. Eng. Manag.*, vol. 10, no. 1, pp. 107–120, 2023, doi: 10.1007/s42524-022-0232-2.

[35] P. Filzmoser and K. Nordhausen, "Robust linear regression for high-dimensional data: An overview," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 13, no. 4, pp. 1–18, 2021, doi: 10.1002/wics.1524.

[36] G. R. Nagiah and N. Mohd Suki, "Linking environmental sustainability, social sustainability, corporate reputation and the business performance of energy companies: insights from an emerging market," *Int. J. Energy Sect. Manag.*, 2024, doi: 10.1108/IJESM-06-2023-0003.

[37] R. Sachan, "Assessment of Effective Teaching Using by TOPSIS method," *J. Innov. Teach. Learn.*, vol. 2, no. 2, pp. 11–20, 2023, doi: 10.46632/jitl/2/2/2.

[38] D. Pandya, G. Kumar, and S. Singh, "Aligning sustainability goals of industrial operations and marketing in Industry 4.0 environment for MSMEs in an emerging economy," *J. Bus. Ind. Mark.*, no. August, 2023, doi: 10.1108/JBIM-04-2022-0183.