# Data Quality Issues and Their Impact on the Accuracy of Analytical Models

**Ansu Mathai Samuel**

Senior Business Analyst at GoDaddy, Tempe, Arizona, USA

**Abstract:** ***This study examines data quality issues and evaluates their impact on the accuracy of analytical models in the context of the digital economy. Particular attention is given to defining data quality metrics (accuracy, completeness, consistency, timeliness, and relevance), analyzing the antecedents that contribute to quality degradation, and assessing the effect of these factors on the predictive performance of analytical systems. A literature review revealed the mechanisms through which data defects negatively affect the accuracy of analytical model forecasts. Additionally, integrated strategies for improving data quality were explored, including automated data cleansing, the use of modern data storage solutions, stream processing, and organizational oversight. The findings of this study have practical implications for optimizing data - driven decision - making processes and serve as a foundation for further empirical research in this field. The insights presented in this article will be of interest to researchers in machine learning, data analysis, and statistics, as they highlight critical relationships between the quality of input data and the accuracy of predictive analytical models. Furthermore, the results will be valuable for professionals in strategic management, information technology, and business analytics, providing theoretically grounded approaches for optimizing big data processing algorithms to enhance decision - making efficiency.***

**Keywords:** data quality, analytical models, forecast accuracy, big data, artificial intelligence, machine learning, automated data cleansing

## 1. Introduction

In modern society, data quality has become a critical factor in the effective operation of analytical systems and managerial decision - making. Deficiencies in data quality can lead to erroneous analytical conclusions, reduced competitiveness, and even strategic miscalculations in organizational management.

The literature on data quality issues demonstrates a variety of approaches to assessing its impact on the accuracy of analytical models. In a theoretical and methodological context, Wang J. et al. [1] provide a comprehensive review of data quality measurements, their antecedents, and their influence on analytical accuracy, establishing a foundation for further research. In parallel, Cho S., Weng C., Kahn M. G., and Natarajan K. [4], through a multi - stage empirical study, focus on specific measurements of data quality generated by wearable devices, emphasizing the need to adapt methodologies to new data types. Ethical aspects related to data quality are explored by Firmani D., Tanca L., and Torlone R. [5], who analyze moral dilemmas and transparency principles in data usage, while Dakkak A., Zhang H., Mattos D. I., Bosch J., and Olsson H. H. [6] delve into challenges associated with continuous data collection, highlighting the link between data quality metrics and the complexities of managing this process.

The practical application of analytical methods in business environments is demonstrated in the studies of Hossain Q. et al. [2] and Noman A. H. M. et al. [3]. Hossain Q. et al. [2] propose the integration of big data into management systems to enhance business intelligence efficiency, whereas Noman A. H. M. et al. [3] focus on improving operational quality through modern analytical tools, underscoring the importance of accurate input data in achieving reliable model outcomes.

At the same time, several studies covering macroeconomic and entrepreneurial aspects, such as those by EbabuEngidaw A. [7], Nayak M., Nayak P. M., Joshi H. G. [8], Zhuo Z., Muhammad B., Khan S. [9], and Zouari G., Abdelhedi M. [10], indirectly highlight the significance of data quality in assessing industry performance, entrepreneurial success, economic growth, and customer satisfaction.

Despite the existence of diverse methodological approaches, certain contradictions remain in the literature. On the one hand, studies focus on detailed characterizations of data quality metrics and nuances of data collection; on the other hand, applied research often overlooks the impact of input data variability on the final accuracy of analytical models. Furthermore, issues related to integrating ethical norms into data quality assessment and balancing quantitative and qualitative aspects remain insufficiently addressed, indicating the need for further research in this area.

The objective of this study is to analyze data quality issues and their impact on the accuracy of analytical models.

The scientific novelty of this work lies in integrating classical data quality assessment concepts with contemporary challenges arising in the era of big data and AI - driven analytics. For the first time, theoretical approaches are combined with empirical data on the influence of automated analytical processes, enabling the proposal of new methodological recommendations to improve the accuracy of analytical models.

The research hypothesis suggests that systematically addressing data quality issues through the implementation of integrated technological and organizational solutions (including AI and machine learning methods) enhances the accuracy of analytical models and, consequently, improves the quality of decision - making.

As a methodological foundation, a comparative analysis of publicly available studies was employed.

**1) Theoretical basis of data quality**

Data quality is understood as the degree of suitability of information for use in specific tasks, as well as its correspondence to the actual state of affairs. Figure 1 below outlines the elements that enable the evaluation of data quality.

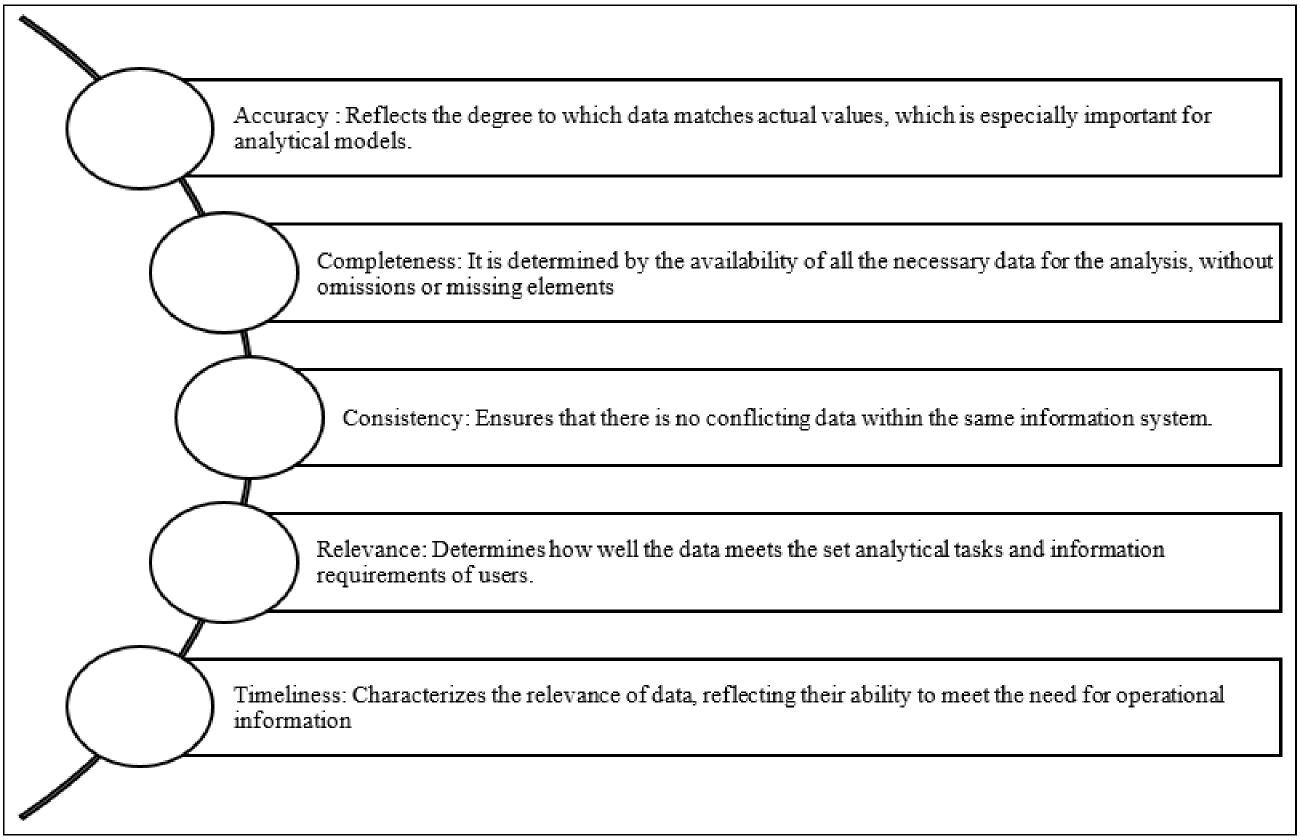


**Figure 1:** Elements that allow us to evaluate the quality of data [1, 2].

Data quality issues can arise due to a variety of reasons, including both technical and organizational factors. Time constraints and information overload affect users' ability to correctly interpret and utilize data [1]. Additionally, users' experience and competencies in working with information systems also determine their perception of data quality [2]. Organizational aspects, such as managerial responsibility and the standardization of data collection and processing procedures, play a crucial role in forming a reliable information base [9, 10].

Three main approaches are used in the study of data quality:

- Intuitive approach: Based on researchers' subjective experience, allowing them to select data quality metrics depending on the specifics of a given task.
- Theoretical approach: Views information systems as data production processes, where quality is determined by objective criteria derived from theoretical models.
- Empirical approach: Focuses on the perspective of end users, for whom data quality is measured by its usability and applicability in specific business processes [1, 2].

Each of these approaches has its advantages and limitations, necessitating a comprehensive analysis to determine effective methodologies for managing data quality in modern conditions.

Analysis shows that data quality is a multidimensional category where standard metrics (accuracy, completeness, consistency, timeliness, and relevance) are supplemented by new aspects related to big data and AI technologies. The antecedents of data quality issues encompass both technical and organizational factors, requiring an integrated approach to management. Comparing methodological approaches (intuitive, theoretical, and empirical) provides a foundation for developing integrated data quality assessment and management systems that enhance the accuracy of analytical models.

**2) The impact of data quality on the accuracy of analytical models**

Any deviations in data can lead to biased, overfitted, or inefficient models, negatively affecting their predictive capability. Thus, data quality directly influences analytical outcomes by determining the accuracy, stability, and interpretability of models.

Analytical models based on machine learning and statistical methods are highly sensitive to errors in raw data. Even minor inaccuracies in data can result in error accumulation during model training, ultimately reducing the accuracy of predictions [1]. Data incompleteness creates situations where the model lacks critical information necessary for accurate forecasting, often leading to biased estimates and reduced reliability. Inconsistencies in data between different sources or even within the same database cause models to train on contradictory information, exacerbating prediction inaccuracies [2, 5]. Timeliness also plays a crucial role, as outdated data may lead to decisions based on obsolete trends, while data relevance determines whether the information


Paper ID: SR25701092036 DOI: https://dx.doi.org/10.21275/SR25701092036

aligns with the specific requirements of the analytical task [3, 6].

Data quality defects have a direct impact on the accuracy of analytical models. Experts with advanced data skills utilize Data Quality Information (DQI) to adjust predictions, demonstrating that models trained on high - quality data exhibit superior accuracy [1, 2]. The presence of artifacts such as missing values and duplicate entries reduces model prediction accuracy, particularly in big data analytics, where large volumes of information complicate real - time data cleansing [4]. Integrating additional data validation mechanisms, such as automated quality control systems, can improve model accuracy by 15–20%, providing substantial benefits for practical applications [1].

Artificial intelligence and machine learning technologies not only detect but also correct data errors. The use of algorithms for automated data cleaning and validation, such as missing value imputation, anomaly detection, and data reconciliation methods, helps mitigate the negative impact of defects on training datasets [2]. AI - driven solutions can adaptively adjust models in real time, which is particularly relevant in environments where data is continuously changing, as seen in large - scale information systems [9, 10]. Therefore, integrating automated data quality control methods becomes essential for improving analytical model accuracy, effectively compensating for data quality issues.

Table 1 below presents the impact of data quality issues on the accuracy of analytical models.

**Table 1:** The impact of key data quality issues on the accuracy of analytical models [1, 2, 7, 8].

| Data quality issue | Mechanism of impact | Effect on model accuracy |
|---|---|---|
| Inaccuracy | Measurement errors lead to biased training datasets, increasing the likelihood of systematic errors. | Reduced prediction accuracy, increased error variance. |
| Incompleteness | Missing critical information results in incomplete training data, impairing the model's ability to generalize patterns. | Overfitting on incomplete data, increased prediction bias, decreased reliability. |
| Inconsistency | Conflicting records and duplicate entries create artifacts that mislead learning algorithms. | Increased data noise, reduced model interpretability, higher prediction errors. |
| Timeliness | Using outdated data fails to reflect current trends and changes in analyzed processes. | Reduced model adaptability, higher risk of misinterpreting current data, leading to erroneous decisions. |
| Relevance | Irrelevant data introduces noise, making it harder to identify meaningful patterns. | Decreased model ability to focus on key features, leading to poor prediction quality and result interpretation. |

Thus, data quality plays a critical role in shaping accurate analytical models. Systematic errors caused by deficiencies in raw data can significantly reduce model predictive performance. Even minor deviations in data quality metrics can negatively impact analytical outcomes. In modern environments, integrating artificial intelligence methods for automated data cleansing and correction offers an effective way to minimize these issues, ultimately improving analytical model accuracy. A comprehensive approach to data quality management is becoming a fundamental requirement for achieving high efficiency in analytics and making well - informed managerial decisions.

## 2. Strategies for improving data quality and enhancing analytical accuracy

Enhancing data quality has become a fundamental requirement for improving the accuracy of analytical models and the effectiveness of decision - making. One key approach is the implementation of automated data control and cleansing systems that leverage artificial intelligence and machine learning methods. Advanced algorithms, such as anomaly detection, missing value imputation, and data reconciliation techniques, help reduce errors in raw data and improve its reliability [4, 5]. The application of such solutions enables real - time data quality monitoring and rapid responses to changes in the information environment, which is particularly relevant in the context of big data analytics [7, 9].

From a technological perspective, the use of modern data storage and processing systems, such as NoSQL databases and data lakes, is essential for ensuring the flexible integration of heterogeneous information sources [2, 6]. The integration of stream processing technologies, such as Apache Kafka and Apache Flink, enhances data timeliness and enables preliminary data processing, reducing the risk of using outdated information in analytical models [1, 8].

Beyond technical solutions, organizational approaches to data quality management also play a crucial role. The implementation of Total Data Quality Management (TDQM) systems and the development of data quality standards help establish a unified corporate culture focused on high standards of information processing. Training specialists in new analytical tools and developing competencies in data science enhance users' ability to detect and correct errors while optimizing data interpretation processes [1, 2].

Regular data quality audits and the application of metrics for assessing data conditions are essential components of an organizational strategy. These measures help identify systematic deficiencies and enable timely adjustments in data collection, storage, and processing processes. The introduction of internal data quality control systems promotes transparency in information flows and reduces the risk of decision - making based on erroneous data.

The integration of hybrid models that combine traditional statistical processing methods with modern AI technologies represents a promising direction for further improving the accuracy of analytical models. Such models allow for adaptive forecast adjustments based on the dynamics of input data quality [2]. Additionally, developing feedback mechanisms between users and data quality systems will enable rapid model adjustments based on expert evaluations [9, 10].

Table 2 below illustrates how a combination of technological and organizational strategies enhances data quality, ultimately leading to improved analytical accuracy.

**Table 2:** The main strategies for improving data quality and their impact on analytical accuracy [1, 2, 4, 6].

| Strategy | Description | Expected effect |
|---|---|---|
| Automated data cleansing | Application of machine learning algorithms for anomaly detection, missing value imputation, and data reconciliation. | Reduction of errors, improved reliability of training datasets, and enhanced analytical model accuracy. |
| Use of modern data storage | Implementation of NoSQL databases and data lakes for integrating heterogeneous sources and ensuring scalable infrastructure. | Increased storage flexibility, improved data timeliness and relevance, and enhanced processing efficiency. |
| Stream data processing | Utilization of Apache Kafka and Apache Flink for real - time data processing, enabling immediate responses to data changes. | Minimized use of outdated data, improved data timeliness and processing efficiency. |
| Organizational data quality control | Implementation of data quality management systems (e. g., TDQM), regular audits, and specialist training for continuous process improvement. | Establishment of a quality - driven culture, reduction of systematic errors, and increased transparency in information flows. |
| Hybrid analytical models | Integration of traditional statistical methods with modern AI technologies for adaptive forecast corrections based on data quality. | Enhanced adaptability and accuracy of analytical models through dynamic adjustments based on real - time data quality assessments. |

Thus, the adoption of modern technological solutions and organizational measures is essential for improving data quality and, consequently, analytical accuracy. The implementation of automated data cleansing and processing systems, the use of flexible storage infrastructures, and systematic organizational control help mitigate the impact of data defects on predictive models.

## 3. Conclusion

The literature review identified key dimensions of data quality—accuracy, completeness, consistency, timeliness, and relevance—and highlighted how deficiencies in these areas reduce the effectiveness of analytical systems. To enhance data quality and improve the accuracy of analytical models, a comprehensive set of strategies is proposed, encompassing both technological and organizational measures.

Among the technological solutions, automated data cleansing using artificial intelligence and machine learning algorithms, the adoption of modern data storage systems such as NoSQL and data lakes, and the implementation of real - time data processing are particularly noteworthy. These approaches enable timely responses to changes in data, minimizing the risk of outdated or inconsistent information affecting analytical outcomes.

Organizational measures, including the implementation of data quality management systems, regular audits, and continuous professional development of specialists, contribute to establishing a corporate culture focused on data quality and transparency in information flows.

The integration of advanced technological solutions with structured organizational approaches significantly enhances the accuracy of analytical models, which directly impacts the efficiency of business processes and strategic management. However, despite these advancements, challenges remain in optimizing data quality control methods and developing hybrid analytical models capable of adaptively adjusting forecasts in response to the dynamic evolution of information environments. These aspects present promising directions for future research and further refinement of analytical practices in the digital economy.

## References

[1] Wang J. et al. Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality //Journal of the Knowledge Economy. – 2024. – Vol.15 (1). – pp.1159 - 1178.
[2] Hossain Q. et al. Integration of Big Data Analytics in Management Information Systems for Business Intelligence //Saudi J Bus Manag Stud. – 2024. – Vol.9 (9). – pp.192 - 203.
[3] Noman A. H. M. et al. Enhancing Operations Quality Improvement through Advanced Data Analytics //Journal of Computer Science Engineering and Software Testing. – 2024. – Vol.10 (1). – pp.1 - 14.
[4] Cho S., Weng C., Kahn M. G., Natarajan K. Identifying data quality dimensions for persongenerated wearable device data: Multi - method study. JMIR mHealth and uHealth. – 2021. – Vol.9 (12). - pp. – 1 - 9.
[5] Firmani D., Tanca L., Torlone R. Ethical dimensions for data quality. Journal of Data and Information Quality (JDIQ). – 2019. – Vol.12 (1). – pp.1–5.
[6] Dakkak A., Zhang H., Mattos D. I., Bosch J., Olsson H. H. Towards continuous data collection from in - service products: Exploring the relation between data dimensions and collection challenges. In 2021 28th Asia - Pacifc Software Engineering Conference (APSEC). IEEE. – 2021. – pp. - 243–252.
[7] EbabuEngidaw A. (2021). The efect of external factors on industry performance: The case of Lalibela City micro and small enterprises, Ethiopia. Journal of Innovation and Entrepreneurship. – 2021. – Vol.10 (1). – pp.1–14.
[8] Nayak M., Nayak P. M., Joshi H. G. Determinants influencing the entrepreneurial success of MSMEs in emerging economies: a study of Indian women entrepreneurs //Cogent Economics & Finance. – 2025. – Vol.13 (1). – pp.2472585.

[9] Zhuo Z., Muhammad B., Khan S. Underlying the relationship between governance and economic growth in developed countries. Journal of the Knowledge Economy. – 2021. – Vol.12 (3). – pp.1314–1330.
[10] Zouari G., Abdelhedi M. Customer satisfaction in the digital era: Evidence from Islamic banking. Journal of Innovation and Entrepreneurship. – 2021. – Vol.10 (1). - pp 1–18.