Big Data-Enabled Decision Support Using a Hybrid Design Science and Machine Learning Model

Abdul Hamid K[1], Mansur AS[2], Sugiharto[3], Sitti Subaedah[4], and Andi Bahar[5]

{ abdulhamidk1958@gmail.com[1], asmansur@unimed.ac.id[2], sugih.unimed@gmail.com[3], sitti87@unimed.ac.id[4], baharandi@unimed.ac.id[5]}

Department of Educational Technology, Graduate School, Universitas Negeri Medan, Indonesia[1].
Department of Computer Science, Universitas Negeri Medan, Indonesia[2].
Department of Geography Education, Universitas Negeri Medan, Indonesia[3].
Department of Community Education, Universitas Negeri Medan, Indonesia[4].
Department of Automotive Engineering Education, Universitas Negeri Medan, Indonesia[5].

Abstract. This study evaluated the effectiveness and feasibility of an AI-based e-learning system using quantitative measures and expert validation aligned with the ADDIE evaluation phase. System performance was assessed through percentage improvement analysis, user satisfaction indices derived from questionnaires, and expert validation conducted via Focus Group Discussion (FGD). Percentage scores were interpreted using a five-level classification scale. The results show notable improvements in learning efficiency (58%), feedback response time (65%), interactivity (54%), and student motivation (61%). User satisfaction reached 89%, indicating strong acceptance and usability. Expert evaluation further confirmed system readiness, with feasibility scoring 92% and pedagogical relevance 85%. These findings demonstrate that integrating machine learning–based adaptive recommendations, automated feedback, learning analytics, and NLP-based chatbot support can effectively enhance learning performance, engagement, and instructional responsiveness. Overall, the proposed AI-based e-learning system is a feasible and pedagogically relevant solution for supporting adaptive and data-driven learning in graduate education.

Keywords: Big Data; Decision Support System; Machine Learning; Higher Education.

1 Introduction

Higher-education institutions increasingly operate as complex digital organizations where academic services, learning management, human-resource processes, and quality assurance activities generate continuously expanding data [1-2]. While universities have adopted information systems for registration, course delivery, and staff administration, data are often fragmented across heterogeneous platforms, limiting the ability of leaders to make timely and evidence-based decisions [3]. Graduate programs, in particular, require rapid monitoring of student progress, research productivity, workload distribution, and service quality [3-4].

Decision Support Systems (DSS) have been adopted to improve governance by providing integrated reporting and analytic capabilities [5]. However, traditional DSS approaches in

AISTEEL 2025, December 05, Medan, Indonesia

DOI 10.4108/eai.5-12-2025.2363851

higher education typically rely on static reporting dashboards that do not leverage predictive analytics or automated pattern discovery [6]. Big Data architectures and machine learning offer opportunities to transform university management by enabling scalable integration of multi-source data and extracting actionable insights, such as early warnings for academic risk, service bottlenecks, and emerging workload issues [7-8].

This research focuses on the Graduate Program at Universitas Negeri Medan (UNIMED), where academic information systems, HR systems, and e-learning platforms operate as core sources of administrative and learning data. The main challenge is enabling institutional decision-making that is not only descriptive, but also predictive and prescriptive [9]. We proposed a hybrid framework combining Design Science Research (DSR) for systematic artifact construction with machine learning models for predictive analytics and clustering-based pattern discovery.

The contributions of this paper are threefold: (1) a Big Data-enabled DSS architecture integrating academic, HR, and e-learning datasets through a robust ETL and data-warehouse layer; (2) a hybrid DSR+ML approach that combines unsupervised clustering for pattern discovery with supervised predictive modeling to forecast student academic risk and administrative workload; and (3) an empirical evaluation demonstrating measurable operational efficiency gains, predictive model performance, and improved decision-making quality and usability.

2 Related Work

Big Data in higher education has been widely associated with learning analytics, institutional research, and performance monitoring, particularly to support evidence-based governance and continuous improvement [10]. Universities generate multi-source data streams from student information systems, learning management systems, digital libraries, and administrative platforms, creating opportunities for large-scale academic analytics and operational intelligence [10-11]. As a result, institutions increasingly adopt data warehouses, data lakes, and analytics layers to consolidate heterogeneous systems and enable decision-oriented reporting across faculties and administrative units [12]. Such architectures improve data consistency, reporting speed, and institutional transparency, especially when linked to strategic key performance indicators (KPIs) such as graduation rates, retention, research output, and service quality [12].

Nevertheless, many higher-education Big Data initiatives remain limited to descriptive analytics and dashboard-based monitoring, which may provide visibility but offer limited predictive or prescriptive value for leadership decision-making [13]. Practical challenges frequently include inconsistent data definitions across units, missing or duplicated records, weak master-data management, and limited interoperability between legacy systems and newer digital platforms [14]. In addition, data governance issues such as privacy protection, access control, and ethical use of student data remain critical barriers to sustainable analytics adoption in universities [15]. These limitations highlight the need for structured methods that ensure technical validity while aligning analytics solutions with organizational requirements and academic regulations [16].

Design Science Research (DSR) provides a methodological foundation for building and evaluating information systems artifacts, emphasizing relevance, rigor, iterative refinement, and validated utility in real-world contexts [17]. DSR is particularly suitable for higher-education digital transformation because it enables systematic development of solutions aligned with stakeholder needs, organizational constraints, and measurable quality standards, rather than focusing only on theoretical models. In decision-support contexts, DSR helps translate institutional challenges into artifact requirements, supporting iterative prototyping and evaluation through performance metrics, stakeholder feedback, and deployment readiness assessments [17]. Furthermore, DSR supports reproducibility by encouraging explicit documentation of design principles, design decisions, and evaluation outcomes, which is essential for scalable adoption across institutions [18].

Machine learning has demonstrated strong potential for predictive analytics in higher education, including dropout prediction, academic performance forecasting, course recommendation, and workload estimation for both teaching and administrative services [19]. Predictive models can support early warning systems by identifying risk trajectories before academic failure occurs, allowing universities to implement targeted interventions such as tutoring, counseling, or personalized learning support [20]. In parallel, operational ML models can forecast resource demand and administrative workload patterns, enabling more efficient staff allocation and service planning during peak periods such as enrollment, course registration, and thesis examinations [21-22].

Unsupervised learning methods such as K-Means and DBSCAN are useful for discovering latent groups within student cohorts or institutional processes, supporting segmentation, anomaly detection, and targeted policy actions [21]. For example, clustering can differentiate student engagement profiles (high-performing, at-risk, disengaged, irregular participation), while density-based techniques can detect outliers such as abnormal inactivity in LMS logs or unusually high administrative processing times. Combining Big Data architectures with machine learning techniques is increasingly considered a practical and scalable strategy for data-driven governance, enabling institutions to move beyond static reporting toward predictive decision support and continuous operational optimization [23].

3 Research Method

This study follows a hybrid Design Science Research process integrated with machine learning-based predictive modeling. The DSR component structures the problem identification, requirement elicitation, artifact design, implementation, and evaluation. The machine learning component enables predictive and descriptive analytics within the artifact, transforming integrated institutional data into actionable decision support. Figure 1 illustrates how institutional data from three core systems (academic, HR, and e-learning) are integrated through an ETL process and stored in a centralized data warehouse. Machine learning services (K-Means, DBSCAN, and predictive models) analyze the integrated data to generate insights such as student-risk prediction and workload forecasting. The decision support system (DSS) delivers dashboards, KPI monitoring, and alerts to support administrators and leaders. Evaluation results (performance, model accuracy, and usability) feed back into the iterative DSR cycle to continuously refine and improve the system.

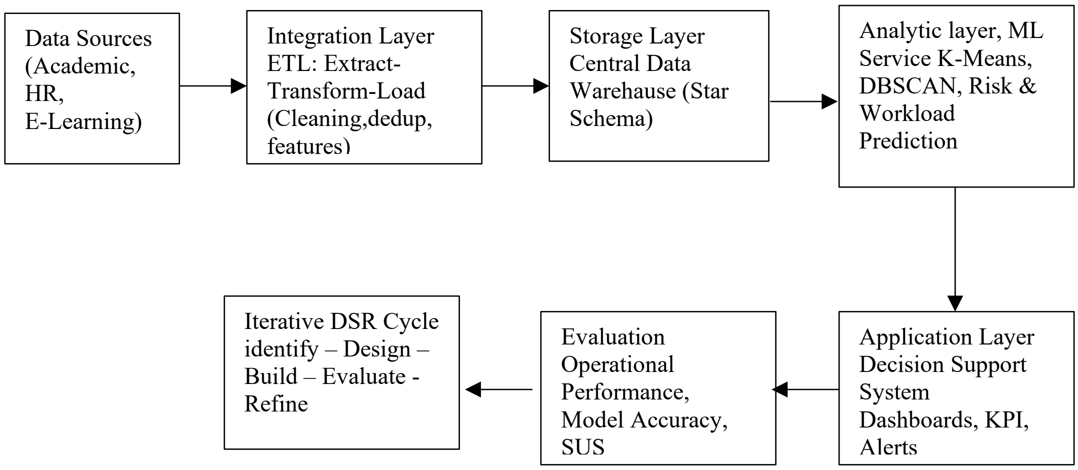

Fig. 1. Work flow of study

The research context is the Graduate Program of UNIMED, involving three major data sources: (1) the academic system (student profiles, enrollment, grades, course history), (2) the HR system (staff workload indicators, administrative assignments, service logs), and (3) the e-learning system (course activity logs, participation indicators, assessment records). These datasets are integrated into a centralized data warehouse to support analysis and decision-making.

3.1 Hybrid DSR + ML Framework

The proposed framework consists of five layers: data sources, integration (ETL), storage (data warehouse), analytics (machine learning services), and application (DSS interface). DSR guides the iterative development of these layers, ensuring that the artifact meets functional requirements such as data freshness, interoperability, and reporting accuracy, as well as non-functional requirements such as performance, usability, and scalability.

3.2 Data Integration and Warehouse Design

Data was extracted from the three core systems through scheduled jobs and API/database connectors. Transformation included schema alignment, data cleaning, deduplication, missing-value handling, and creation of analytic features such as GPA trend, course completion rate, attendance proxy from LMS activity, and administrative workload proxies from HR logs. Loading was executed into a star-schema warehouse design with fact tables capturing learning events and administrative service events, and dimension tables representing students, courses, semesters, staff, and organizational units.

3.3 Machine Learning Models

Unsupervised learning was applied for latent pattern discovery. K-Means clustering was used to group students based on multi-dimensional indicators (academic performance, engagement, and progression), while DBSCAN was applied to detect outliers representing anomalous student behaviors or atypical workload patterns. In addition, supervised predictive modeling was used to forecast academic risk (binary classification) and administrative workload (regression and classification variants depending on scenario).

Model training used stratified k-fold cross-validation. Feature scaling was applied where necessary, and model performance was reported using ROC-AUC and F1-score for classification. For clustering, internal validation was performed using silhouette coefficient for K-Means, and neighborhood-based sensitivity analysis for DBSCAN hyperparameters.

3.4 System Architecture and Implementation

The DSS artifact was implemented as a modular architecture. Data ingestion services performed extraction from the academic, HR, and e-learning systems. The ETL pipeline executed validation and transformation rules before storing data in the warehouse. An analytics layer exposed machine learning services through APIs to support real-time scoring and periodic batch analysis. The application layer provided dashboards for leaders and administrative staff, including KPI monitoring, student-risk alerts, workload heatmaps, and drill-down reports.

Security and privacy controls were incorporated through role-based access control and aggregation policies. Decision-support outputs were designed to be interpretable, showing contributing indicators (e.g., declining GPA, reduced LMS activity, missing assessments) for student-risk predictions. This supports transparency and fosters trust among academic stakeholders.

4 Evaluation and Results

This study evaluated the proposed Big Data-enabled Decision Support System (DSS) using three complementary dimensions: (1) operational performance, (2) predictive model performance and (3) system usability. The evaluation compared baseline institutional processes-characterized by fragmented systems and manual reconciliation—with the integrated DSS deployment, which enables centralized data warehousing and machine learning–based analytics. The overall results are summarized in Table 1, while the comparative improvements and performance indicators are visualized in Figure 2 and Figure 3.

Table 1. Summary of Evaluation Results for the Proposed Big Data–Enabled DSS Framework

Metric	Result	Category
Data-retrieval latency	−61%	Operational efficiency
Report generation time	−48%	Operational efficiency
Manual reconciliation effort	−38%	Operational efficiency
Decision-cycle time	−30%	Decision-making quality
KPI alignment	+22 points	Decision-making quality
ROC-AUC	0.86	Predictive model performance
F1-score	0.76	Predictive model performance
SUS score	81.3	Usability

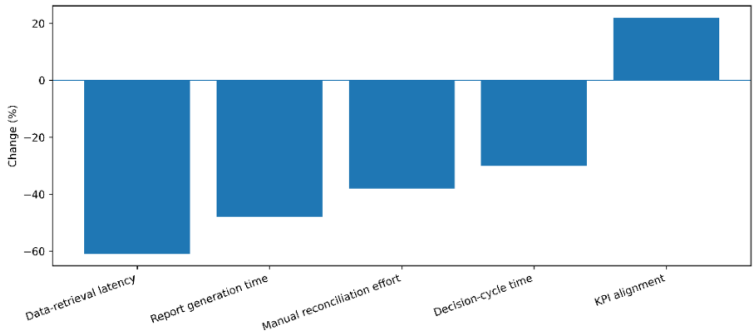

Fig. 2. Operational Efficiency and Decision-Making Improvements

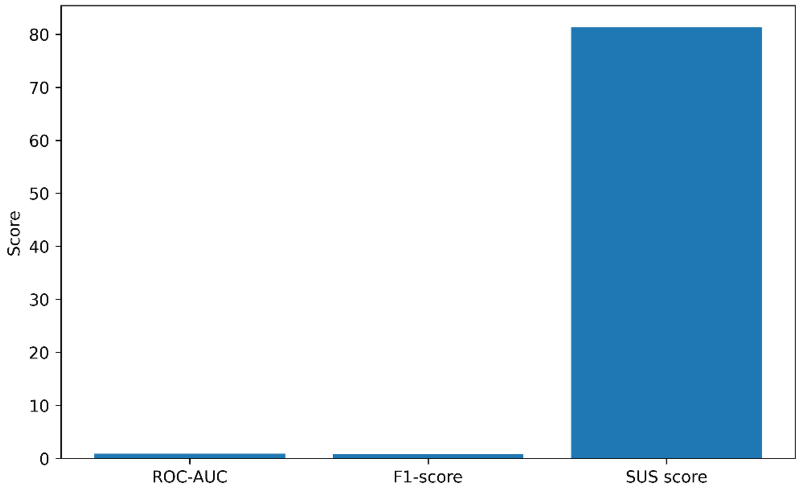

Fig. 3. Predictive Model Performance and System Usability

4.1 Operational Performance Evaluation

Operational performance was assessed by examining how efficiently institutional data can be retrieved and transformed into management reports. Under baseline conditions, academic and administrative staff typically relied on separate data sources and manual verification workflows to reconcile inconsistencies between systems. Following implementation of the integrated DSS, standardized ETL processing and centralized data storage enabled faster access and reporting.

As reported in Table 1, the proposed approach resulted in substantial efficiency gains. Data-retrieval latency decreased by 61%, indicating a significant improvement in system responsiveness after centralized integration. Similarly, report generation time decreased by 48%, demonstrating that automated consolidation and preprocessed data structures accelerate institutional reporting. Furthermore, manual reconciliation effort decreased by 38%, confirming that the DSS reduces repetitive administrative work related to cross-system validation. These operational improvements are illustrated in Figure 2, which highlights the comparative reductions in key efficiency metrics under the proposed deployment.

4.2 Predictive Model Performance Evaluation

To support proactive institutional decision-making, the proposed DSS incorporates machine learning models that forecast student academic risk and administrative workload trends. Predictive model performance was evaluated using ROC-AUC and F1-score metrics, which indicate discrimination ability and classification balance.

The results show that the predictive models achieved ROC-AUC = 0.86 and F1-score = 0.76 (Table 1). These values demonstrate strong predictive capability and suggest that the models are suitable for identifying early risk signals and supporting preventive interventions. The predictive model outcomes are further summarized in Figure 3, which presents the achieved ROC-AUC and F1-score alongside usability results.

4.3 Decision-Making Quality Assessment

In addition to operational improvements and predictive performance, the evaluation assessed decision-making outcomes. The objective was to determine whether the DSS improves governance effectiveness by strengthening alignment to institutional KPIs and reducing decision-cycle delays.

The findings indicate that KPIs alignment improved by 22 percentage points, reflecting more consistent use of institutional performance indicators in decision processes (Table 1). In parallel, the average decision-cycle time decreased by 30%, showing faster institutional response due to improved access to integrated information and analytics support. These decision-making improvements are also visible in Figure 2, where reductions in decision-cycle time and improvements in KPI alignment are highlighted.

4.4 Usability Evaluation

System usability was measured using the System Usability Scale (SUS) to assess user acceptance and practical feasibility. Usability evaluation is essential because a DSS delivers long-term benefits only when it can be adopted consistently by institutional stakeholders.

The DSS achieved a SUS score of 81.3, categorized as Excellent (Table 1). This result indicates that users perceived the system as intuitive and effective for routine decision-support tasks. The SUS score is presented in Figure 3, confirming that strong technical performance is accompanied by high usability.

4.5 Summary of Findings

Overall, the evaluation provides empirical evidence that the hybrid Big Data + DSR + ML framework improves institutional operations and decision-making. Operational efficiency was enhanced through reduced latency, faster reporting, and lower manual reconciliation effort (Figure 2). Predictive modeling achieved strong accuracy for academic risk and workload forecasting (Figure 3), while decision quality improved through better KPI alignment and shorter decision cycles. The system’s usability was also high, indicating strong potential for real-world adoption.

5 Discussion

The findings demonstrate that a Big Data–enabled decision support system, built on a hybrid Design Science Research (DSR) and machine learning (ML) framework, can significantly strengthen institutional decision-making in higher-education management. The integration of a scalable Big Data architecture with machine learning analytics was shown to improve the accuracy, timeliness, and comprehensiveness of information available to academic leaders, thereby increasing the rationality and responsiveness of administrative decisions.

The implementation of a structured ETL (Extract–Transform–Load) pipeline, together with a centralized data warehouse, was found to reduce operational burdens across the institution by eliminating fragmented reporting workflows and consolidating disparate data sources into a single, coherent repository. Standardized definitions of key performance indicators (KPIs)-such as enrollment rates, academic risk indicators, and teaching-load metrics-were established and applied uniformly across faculties and administrative units. As a result, data inconsistencies and time-consuming manual reconciliations were minimized, enabling administrators to access timely, reliable, and comparable information with minimal manual intervention [24].

Furthermore, unsupervised learning techniques contributed substantial value in uncovering latent patterns and structures within complex institutional datasets. Clustering algorithms were applied to multidimensional records of student engagement, attendance, and academic performance, and they successfully generated actionable segments of student profiles. These segments highlighted distinct groups such as "high-risk" learners, "consistently underperforming" cohorts, and "at-risk but rapidly improving" students. Similar clustering was also employed to analyze patterns of administrative workload, revealing regions of overburdened staff and periods of chronic resource strain. Institutional leaders were thus able to prioritize interventions, allocate support staff more effectively, and target mentoring or remedial programs to the most vulnerable student groups at an early stage of the academic cycle.

In addition, predictive analytics were used to extend decision-making from reactive reporting to proactive governance. Forecasting models trained on historical student and administrative data were capable of estimating the probability of academic risk, predicting peaks in administrative workload, and identifying potential bottlenecks before they occurred. These forecasts were integrated into dashboards and reporting tools, enabling leaders to anticipate resource requirements, schedule interventions (such as academic mentoring, tutoring, or workload redistribution), and adjust staffing and service capacity during high-demand periods. The reliance on delayed, retrospectively compiled reports was therefore reduced, and decision-making became more anticipatory and evidence-driven.

Critically, the hybrid DSR + ML approach ensured that the system was not only technically sound but also organizationally grounded and practically usable. The design artifacts were iteratively evaluated in collaboration with key stakeholders, including academic managers, data officers, and IT staff, and refinements were made based on feedback obtained during prototype testing and pilot deployments. This iterative process helped align system features with real-world decision-making routines, facilitated user acceptance, and supported the design of an extensible architecture that could accommodate new data sources and additional analytical models in future implementations.

Despite these positive outcomes, several limitations were identified. First, the performance and reliability of the system were shown to depend heavily on the quality, completeness, and consistency of data across the integrated source systems. Missing records, inconsistent coding schemes, and temporal misalignments across data streams could degrade model accuracy and reduce the trustworthiness of generated insights. Second, predictive models required continuous monitoring, periodic recalibration, and careful governance of feature definitions to prevent performance drift and concept-shift degradation over extended periods.

Future research should therefore expand the current framework by integrating additional institutional data sources, such as research output repositories, third-party survey platforms, and financial management systems. Such an extension would enable a broader range of governance functions, including research performance monitoring, strategic planning, and budgetary decision-making. Further improvement may also be achieved by incorporating explainable AI (XAI) techniques into the analytical pipeline, which could enhance the transparency, interpretability, and stakeholder trust in predictive outcomes. Finally, longitudinal evaluations conducted across multiple academic cycles are recommended to assess the sustained institutional impact of the system, its contribution to policy effectiveness, and its long-term return on investment in data-driven governance.

6 Conclusion

This paper presented a Big Data–enabled decision support framework that combines Design Science Research (DSR) with machine learning–based predictive modeling for graduate-program governance at Universitas Negeri Medan. By integrating academic, human-resource, and e-learning data through an ETL pipeline and a centralized data warehouse, the proposed artifact enabled scalable reporting and analytics for institutional decision-making.

The results indicate that the hybrid DSR + ML approach improved operational efficiency through faster data access, reduced reporting time, and lower manual reconciliation effort. The predictive modeling component provided effective forecasting of academic risk and administrative workload trends, supporting proactive institutional interventions. Decision-making quality also improved through enhanced KPI alignment and shorter decision-cycle time, while usability outcomes confirmed strong user acceptance with an excellent SUS score. Overall, the proposed framework offers a scalable model for data-driven governance and can be adapted by other higher-education institutions seeking integrated and intelligent management systems.

Acknowledgments

This research was conducted as part of the PENELITIAN PENGEMBANGAN IPTEKS PNPB program, funded by LPPM Universitas Negeri Medan (UNIMED) in 2025. The authors would like to thank the Graduate Program of Universitas Negeri Medan for facilitating data access, stakeholder coordination, and the evaluation activities carried out in this study.

References

[1] M. Alenezi, "Digital learning and digital institution in higher education," *Educ. Sci. (Basel).*, vol. 13, no. 1, p. 88, 2023.

[2] P. S. Aithal and A. K. Maiya, "Development of a new conceptual model for improvement of the quality services of higher education institutions in academic, administrative, and research areas," *International Journal of Management, Technology, and Social Sciences (IJMTS)*, vol. 8, no. 4, pp. 260–308, 2023.

[3] K. C. Gonugunta and K. Leo, "Role of data-driven decision making in enhancing higher education performance: A comprehensive analysis of analytics in institutional management," *International Journal of Acta Informatica*, vol. 3, no. 1, pp. 149–159, 2024.

[4] S. Mpofu and D. Chasokela, "Data-informed decision-making: using analytics to drive strategic management in higher education," in *Building Organizational Capacity and Strategic Management in Academia*, IGI Global Scientific Publishing, 2025, pp. 103–138.

[5] R. Hasan and S. Akter, "Information System-Based Decision Support Tools: A Systematic Review Of Strategic Applications In Service-Oriented Enterprises," *Review of Applied Science and Technology*, vol. 1, no. 04, pp. 26–65, 2022.

[6] N. Neiroukh and D. Ça\uglar, "Information Systems Quality and Corporate Sustainability: Unpacking the Interplay of Financial Reporting, Artificial Intelligence, and Green Corporate Governance," *Systems*, vol. 13, no. 7, p. 537, 2025.

[7] Y. Song and N. Wang, "Application of Big Data Analytics in Education Management: Enhancing Teaching Quality and Resource Allocation Efficiency," *International Journal of High Speed Electronics and Systems*, p. 2540637, 2025.

[8] M. Y. Shakor and M. I. Khaleel, "Modern Deep Learning Techniques for Big Medical Data Processing in Cloud," *IEEE Access*, 2025.

[9] A. F. A. Hassan, H. Mohamad, N. I. Zulkifle, N. S. Zainudin, and P. K. I. Robi, "A MOBILE LEARNING APPLICATION FOR MALAYSIAN SIGN LANGUAGE," *SUSED2025*, p. 52, 2025.

[10] M. Beerkens, "An evolution of performance data in higher education governance: a path towards a 'big data'era?," *Quality in Higher Education*, vol. 28, no. 1, pp. 29–49, 2022.

[11] Z. He and W. Fang, "Research data management in institutional repositories: an architectural approach using data lakehouses," *Digit. Libr. Perspect.*, vol. 41, no. 1, pp. 145–178, 2025.

[12] P. Tamla, "Towards a Reference Model for an Information Extraction Lifecycle supporting Data Collection, Data Preparation, Information Extraction, Information Organization, Knowledge Organization, Information Access and Retrieval".

[13] A. Sorour and A. S. Atkins, "Big data challenge for monitoring quality in higher education institutions using business intelligence dashboards," *Journal of Electronic Science and Technology*, vol. 22, no. 1, p. 100233, 2024.

[14] L. Haoran and C. Yuxin, "Strategies for Ensuring Accuracy and Reliability of Business Critical Master Data in Complex Enterprise Systems," *Journal of Applied Big Data Analytics, Decision-Making, and Predictive Modelling Systems*, vol. 9, no. 9, pp. 1–11, 2025.

[15] M. M. Ncube and P. Ngulube, "A Systematic Review of Postgraduate Programmes Concerning Ethical Imperatives of Data Privacy in Sustainable Educational Data Analytics.," *Sustainability (2071-1050)*, vol. 16, no. 15, 2024.

[16] A. H. Adepoju, A. Eweje, A. Collins, and O. Hamza, “Developing strategic roadmaps for data-driven organizations: A model for aligning projects with business goals,” *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 4, no. 6, pp. 1128–1140, 2023.

[17] S. Kroop, “Artifact Validity in Design Science Research (DSR): A Comparative Analysis of Three Influential Frameworks BT - Local Solutions for Global Challenges,” S. Chatterjee, J. vom Brocke, and R. Anderson, Eds., Cham: Springer Nature Switzerland, 2025, pp. 199–215.

[18] S. Reining, F. Ahlemann, B. Mueller, and R. Thakurta, “Knowledge Accumulation in Design Science Research: Ways to Foster Scientific Progress,” *SIGMIS Database*, vol. 53, no. 1, pp. 10–24, Jan. 2022, doi: 10.1145/3514097.3514100.

[19] K. Fahd, S. Venkatraman, S. J. Miah, and K. Ahmed, “Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature,” *Educ. Inf. Technol. (Dordr).*, vol. 27, no. 3, pp. 3743–3775, 2022, doi: 10.1007/s10639-021-10741-7.

[20] S. Akter and S. F. Wamba, *Handbook of Big Data Research Methods*. Cheltenham, UK: Edward Elgar Publishing, 2023. doi: 10.4337/9781800888555.

[21] T. Begum, “Chapter 10: Predictive analytics for machine learning and deep learning,” in *Handbook of Big Data Research Methods*, Cheltenham, UK: Edward Elgar Publishing, 2023, pp. 148–164. doi: 10.4337/9781800888555.00014.

[22] M. Badawy, N. Ramadan, and H. A. Hefny, “Healthcare predictive analytics using machine learning and deep learning techniques: a survey,” *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, p. 40, 2023, doi: 10.1186/s43067-023-00108-y.

[23] G. Nookala, “Adaptive Data Governance Frameworks for Data-Driven Digital Transformations,” *Journal of Computational Innovation*, vol. 4, no. 1, pp. 1–20, 2024.

[24] D.-M. Córdova-Esparza *et al.*, “Predicting and Preventing School Dropout with Business Intelligence: Insights from a Systematic Review,” *Information*, vol. 16, no. 4, p. 326, 2025.