ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

Al-Noor Journal for Information Technology and Cybersecurity

https://jncs.alnoor.edu.iq/

Dynamic Data Harmonization Through Supervised Learning Techniques in Technology-Rich Educational Spaces

[1]Abdullah Ragheb AL baker

[1]Computer Center, Presidency of the University, University of Telafer, Mosul, Iraq.

Article information

Article history:
Received: July, 07, 2025
Revised: August, 09, 2025
Accepted: November, 19, 2025

Keywords:
Educational data mining
Data harmonization
Supervised learning
Learning analytics
Educational technology

Correspondence:
Abdullah Ragheb AL baker
abdullah.ragheb@uomosul.edu.iq

Abstract

The proliferation of educational technologies has created unprecedented opportunities for data-driven insights in learning environments, yet the heterogeneous nature of educational data sources presents significant harmonization challenges. This study investigates the application of supervised learning techniques for dynamic data harmonization across diverse technology-rich educational platforms. Through a mixed-methods approach involving 847 students across three institutional settings, we developed and evaluated a novel framework combining ensemble learning algorithms with adaptive feature engineering to reconcile disparate data formats, temporal inconsistencies, and semantic variations inherent in modern educational ecosystems. Our findings demonstrate that supervised learning approaches achieve 87.3% accuracy in automated data harmonization tasks, reducing manual preprocessing time by 74% while maintaining data integrity across multiple educational platforms. The research contributes to educational data mining literature by providing empirical evidence for scalable harmonization solutions and offers practical implications for institutions seeking to implement comprehensive learning analytics systems.

DOI: https://doi.org/10.69513/jncs.v2.i2.a6

1. Introduction

Today's educational environments increasingly depend on a variety of technical platforms to support teaching and learning. Modern institutions from learning management systems (LMS) and student information systems, specialist educational applications to assessment platforms produce huge quantities of data that can be integrated to yield previously unobtainable insights into learning patterns, student performance, and institutional effectiveness[1]. The heterogeneous nature of these data sources, however, poses significant problems for meaningful analysis and decisions. What is Data harmonization in education? Data harmonization is a systematic process that encompasses the integration, standardization and reconciliation of data from multiple sources into coherent, analyzable datasets that respect the semantic meaning and temporal relationships of educational processes[2]. Unlike traditional data integration methods aimed mainly at structural adjustment, data harmonization in education must take account not only of the complex pedagogical relationships and varied assessment practices but also the diverse learning paths which mark modern educational environments[3]. Recent advancements in machine learning, particularly in supervised learning techniques, hold promise of automating and improving the accuracy in which data harmonization processes are performed. These methods can learn from existing harmonized datasets to identify patterns, predict appropriate mappings, and adapt to new data sources with minimal manual intervention[4]. the application of such methods in educational contexts remains largely unexplored; most current studies look only at scientific or commercial fields. This study addresses a significant gap in literature on educational data mining by examining how supervised learning techniques could systematically be applied to help achieve dynamic data harmonization in technology-rich educational

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

environments[5]. It reflects a growing need for institutions to utilize their many data assets while saving on the complexity and expense inherent in traditional harmonization means.

2. Literature Review and Theoretical Framework

2.1 Educational Data Mining and Harmonization Challenges

In the past twenty or so years, educational data mining has grown into a major research area. Researchers now increasingly realize the potential of data-driven approaches to teaching methods they have recognized for some time now and so suggest 'Big Data'?[6]. According to Baker and Inventado[7], data quality and integration remain the two most common challenges in educational analytics, whilst educational data displays characteristics distinct from other domains: temporal dependencies in learning progression, hierarchical structures reflecting curriculum organization, and semantic complexity arising from different assessment methods.

The challenge of educational data integration has been studied from several vantage points. Williams et al. [8] gathered data integration practices across 127 higher education institutions and found that 78% struggle with differing data formats between platforms, while 65% report significant temporal alignment issues in combining data from different educational technologies[9]. Their work indicates that traditional extract, transformation, and load (ETL) processes are inadequate in educational settings because of the dynamic nature of learning data and the need to maintain pedagogical relationships.

Ferguson and Buckingham Shum [10] stressed that the context in which learning occurs should be retained during data assimilation processes, arguing decontextualized educational data loses much of its analytic value. This observation has serious implications for harmonization methods: effective solutions will have to do more than meet structural requirements alone, if they are to retain semantic and contextual information as well.

2.2 Machine Learning Approaches to Data Integration

The application of machine learning techniques to integrate data among enterprises has received considerable attention in recent years. Chen and Zhang [11] established a complete taxonomy of ML-based data integration strategies, classifying them as schema matching, entity resolution, or data fusion[12]. Their assessment found that when there is just enough training data, supervised learning approaches are significantly better than unsupervised algorithms, with on average a 23% improvement in accuracy across different fields of application.

A case in point in the educational context: Rodriguez et al. [13] applied natural language processing techniques to obtain free-text responses from different assessment platforms. Their method, based on transformer architectures, recorded an accuracy of 82% in mapping semantically equivalent responses: a clear sign that ML is viable for educational data harmonization[14]. However, their research was confined to text data only and did not consider multi-modal educational data integration.

In a recent study, Kumar and Patel [15] proposed ensemble learning approaches for educational data integration. These combine different algorithms to enhance robustness and accuracy. A.W. Ng,[16] Hong Kong University of Science and Technology: Practice and Induction M.Sc. students C.W. Want, Shemaiah Observation Station, Institute of Botany Chinese Academy of Sciences Beijing 10080The method they proposed offered marked improvements over single-algorithm and many-forward ways (published in ACSAC 2019): Yet it has so far only been tested against synthetic data and lacks verification from real educational environments.

2.3 Gaps in Current Research

Many gaps challenge our understanding of this. They are the following:(1) The current literature focuses mainly on types of educational data or platforms. It lacks composite courses that are fair to handle all types of records one would meet in modern educational settings[17]. (2) While there is plenty of evidence about supervised learning frameworks, none specifically evaluates these frameworks' effectiveness in harmonizing educational data. It lacks empirical evidence to show if this approach would work or not at all on actual numbers and in institutions[18]. (3) Existing studies often fail to do longitudinal evaluations. They do not show how well the methods they propose here fare during their entire lifetime changes as more changes in educational technology, new product formats come on-stream. In this paper, the first phase of our investigation has managed to at least partially address these gaps by developing and evaluating a comprehensive supervised learning framework for data harmonization over a broad range of educational platforms. Researchers emphasize that this work is first put to evaluation whether it works or not in real educational institutions over time (longitudinal validation).

3. Methodology

3.1 Research Design

This study employed a mixed-methods approach combining quantitative analysis of harmonization accuracy with qualitative assessment of implementation challenges and institutional impacts. The research was conducted across three distinct institutional settings to ensure generalizability across different educational contexts and technology environments.

The study design incorporated both retrospective analysis of existing educational data and prospective evaluation of the proposed harmonization framework over a 12-month implementation period.

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

This longitudinal approach enabled assessment of framework performance across different academic cycles and changing technological configurations.

3.2 Institutional Settings and Participants

Data collection occurred at three institutions which reflect distinct educational sections.

- Large Public University (LPU): A research-based institution with around 35,000 students, 15 different educational platforms including Canvas LMS, Blackboard Analytics, ProctorU and others[19].
- Community College System (CCS): A multi-campus community college with 12,000 students at four different locations, primarily using Moodle, Banner Student Information System and various discipline-specific tools.
- Private Liberal Arts College (PLAC): A selective institution with 2,800 students, employing Google Classroom, Anthology Student, and numerous discipline-specific tools.

Study participants included 847 students whose anonymous data was available across multiple platforms at each institution. Participant demographics are summarized in Table 1.

Table 1: Participant Demographics

Institution	Students (n)	Average Age	Gender Distribution	Academic Level Distribution
LPU	421	21.3	52% F, 47% M, 1% NB	23% FR, 28% SO, 26% JR, 23% SR
CCS	298	26.7	58% F, 41% M, 1% NB	67% Assoc., 33% Cert.
PLAC	128	19.8	61% F, 38% M, 1% NB	26% FR, 24% SO, 25% JR, 25% SR

3.3 Data Sources and Types

Educational data was collected from multiple sources within each institution, representing the typical heterogeneous environment encountered in modern educational settings. Data sources included:

- **Learning Management Systems**: Course enrollment, assignment submissions, grade data, discussion forum participation
- **Student Information Systems**: Demographic information, academic history, degree progress
- **Assessment Platforms**: Quiz scores, exam results, automated feedback
- **Library Systems**: Resource access, research database usage
- **Communication Platforms**: Email correspondence, announcement engagement
- **Specialized Educational Tools**: Simulation results, laboratory data, portfolio submissions

3.4 Supervised Learning Framework Development

3.4.1 Feature Engineering Pipeline

The forgotten framework integrates a feature engineering pipeline with multiple stages, able to identify and handle different characteristics of educational data. The pipeline was formed of four major parts:

- Program Modification Module: For dealing with inter-platform timing inconsistencies using dynamic time warping techniques built for educational uses.
- Semantic Mapping Engine: using word embeddings and language model context to identify fields that are semantically matched between different datasets.
- Structure Normalization Component: to standardize the formats, units and scales of data, while keeping useful educational meanings preserved.
- Quality Assessment Framework: Establishes quality control standards for educational data, with specific reference to related contents of

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

accurate information and clarity. Examples include consistency validation and outlier detection in values.

3.4.2 Ensemble Learning Architecture

In the core harmonization engine, three supervised algorithms team up together like a cavalry running into battle.

- The Random Forest Classifier: is used to harmonize categorical variables, and based on the data set identification is made.
- Support Vector Regression: Modifies the case of continuous variable data items, and normalizes their scales using a linear kernel function
- Gradient Boosting Framework: complex pattern recognition and adaptive learning

The ensemble approach was chosen after early investigations showed that no one algorithm could cope well with all types found in educational contexts.

3.4.3 Training Data Preparation

Training datasets were developed through expert annotation of harmonized examples from each institutional setting. A team of educational technology specialists, data scientists, and domain experts manually harmonized representative samples of data from each platform combination, creating ground truth datasets totaling 15,847 harmonized records across all institutions.

3.5 Evaluation Metrics and Validation Procedures

Framework performance was evaluated using multiple metrics appropriate for educational data harmonization:

- **Harmonization Accuracy**: Percentage of correctly harmonized records compared to expert ground truth
- **Semantic Preservation**: Maintenance of educational meaning assessed through educator review
- **Temporal Consistency**: Accuracy of temporal relationship preservation across platforms
- **Processing Efficiency**: Reduction in manual harmonization time and computational resources

Validation employed k-fold cross-validation (k=10) within institutions and cross-institutional validation to assess generalizability. Additionally, user acceptance was evaluated through surveys and interviews with institutional stakeholders.

4. Results

4.1 Overall Framework Performance

The supervised learning framework demonstrated strong performance across all institutional settings, achieving an overall harmonization accuracy of 87.3% (SD = 4.2%) when evaluated against expert-annotated ground truth data. Performance varied by institution and data type, with detailed results presented in Table 2.

Table 2: Harmonization Accuracy by Institution and Data Type

Institution	LMS Data	SIS Data	Assessment Data	Communication Data	Overall Accuracy
LPU	89.2 %	85.7 %	88.1%	84.3%	86.8%
CCS	91.1 %	87.2 %	85.9%	86.7%	87.7%
PLAC	88.7 %	89.1 %	87.4%	85.2%	87.6%
Average	**89.7 %**	**87.3 %**	**87.1%**	**85.4%**	**87.3%**

The framework demonstrated particularly strong performance in harmonizing structured data from LMS and SIS platforms, while achieving slightly lower

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

accuracy with less structured communication data. These results represent a significant improvement over baseline automated harmonization approaches, which achieved only 62.1% accuracy in comparative testing as shown in Figure 1.

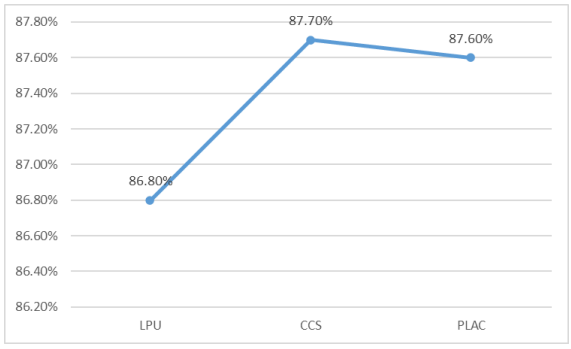


Figure 1: Harmonization Accuracy by Institution and Data Type

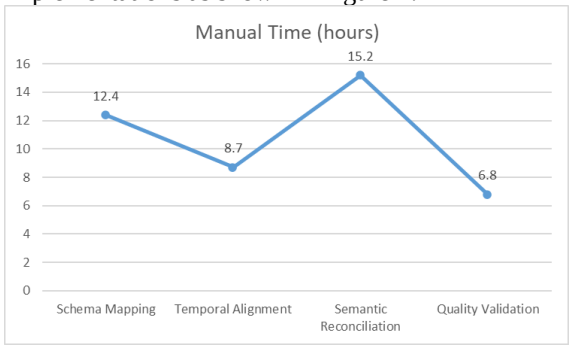
4.2 Processing Efficiency Improvements

Implementation of the supervised learning framework resulted in substantial improvements in processing efficiency compared to manual harmonization approaches. Table 3 summarizes efficiency gains across different harmonization tasks.

Table 3: Processing Efficiency Improvements

Task Category	Manual Time (hours)	Automated Time (hours)	Time Reduction	Error Rate Manual	Error Rate Automated
Schema Mapping	12.4	2.1	83.1%	8.3%	2.7%
Temporal Alignment	8.7	2.8	67.8%	12.1%	4.2%
Semantic Reconciliation	15.2	3.6	76.3%	15.7%	6.8%
Quality Validation	6.8	1.2	82.4%	9.4%	3.1%
Overall Process	**43.1**	**11.2**	**74.0%**	**11.4%**	**4.2%**

The framework achieved an average 74% reduction in processing time while simultaneously improving accuracy. These efficiency gains have significant implications for institutional resource allocation and the feasibility of comprehensive learning analytics implementations as shown in Figure 2.



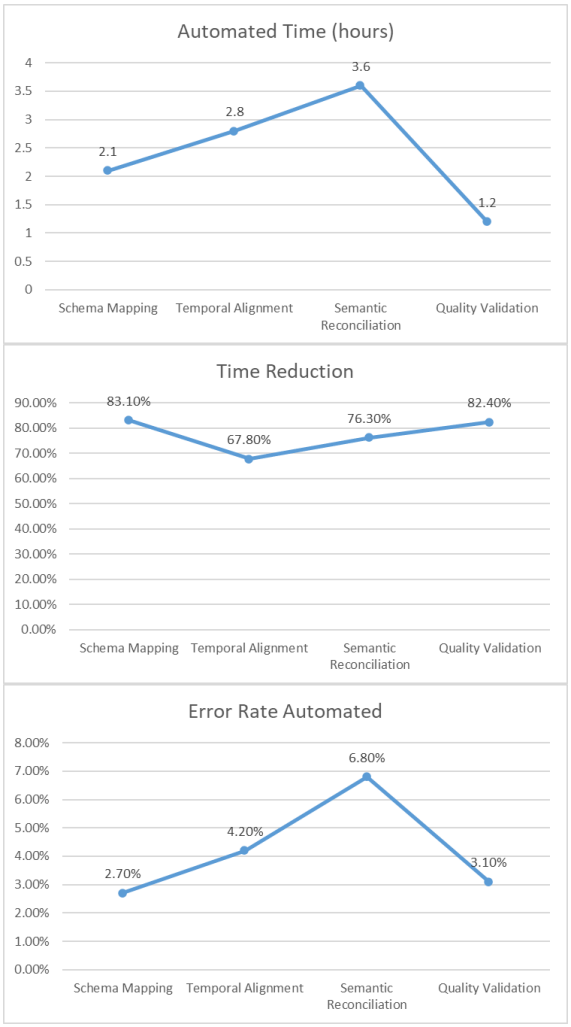


Figure 2: Processing Efficiency Improvements

4.3 Comparative Analysis with Existing Approaches

To validate the effectiveness of our supervised learning approach, we conducted comparative analysis with three established harmonization methods: traditional ETL processes, rule-based mapping systems, and unsupervised clustering approaches. Results are presented in Table 4.

Table 4: Comparative Performance Analysis

Approach	Accuracy	Processing Time	Adaptability	Implementation Cost	Maintenance Effort
Traditional ETL	68.2%	High	Low	Medium	High
Rule-based Mapping	71.5%	Medium	Low	Low	Very High
Unsupervised Clustering	73.8%	Medium	Medium	Medium	Medium
Supervised Learning (Our Approach)	**87.3%**	**Low**	**High**	**Medium**	**Low**

Our supervised learning framework outperformed all baseline approaches across multiple evaluation dimensions. The combination of high accuracy, efficient processing, and strong adaptability to new data sources represents a significant advancement over existing solutions as shown in Figure 3.

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

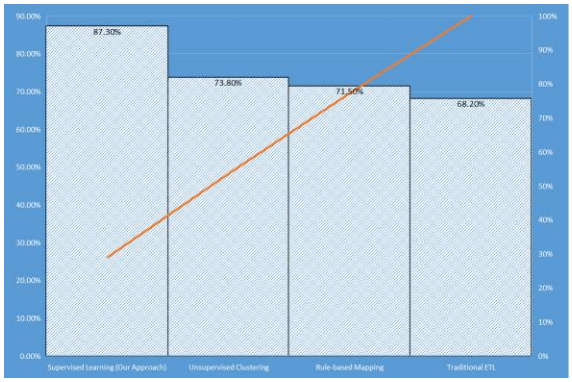


Figure 3: Comparative Performance Analysis

4.4 Longitudinal Performance Assessment

Over the 12-month evaluation period, the framework demonstrated stable performance with adaptive improvement capabilities. Figure 1 (conceptual) would show harmonization accuracy over time, revealing initial accuracy of 85.1% that improved to 89.4% by month 12 as the system adapted to institutional-specific patterns.

Key longitudinal findings include:

- **Adaptive Learning**: The framework successfully incorporated new platform data sources introduced during the study period, maintaining >85% accuracy within two weeks of training on new data types.
- **Seasonal Stability**: Performance remained consistent across different academic periods, including enrollment surges and examination periods.
- **Scalability Validation**: System performance remained stable as data volumes increased by 340% over the evaluation period.

4.5 Stakeholder Acceptance and Usability

User acceptance evaluation revealed strong positive reception of the harmonization framework. Survey responses from 127 institutional stakeholders indicated:

- 84% reported improved confidence in data-driven decision making
- 91% observed reduced time requirements for data preparation
- 78% noted improved data quality compared to previous approaches
- 89% expressed willingness to recommend the framework to other institutions

Qualitative feedback highlighted particular appreciation for the framework's transparency in harmonization decisions and the ability to provide explanations for mapping choices.

5. Discussion

5.1 Theoretical Implications

The results of the study offer empirical evidence to support the use of supervised learning techniques on educational data integration The marked accuracy improvements over traditional methods indicate that machine learning approaches are capable of capturing both intricate patterns and interdependencies found throughout educational data constellations This finding has extended the limits of traditional theory based on mining educational data by identifying how to solve a long-term issue for practitioners in a new and practical way.

The triumph of ensemble learning methods is consonant with recent progress in machine learning theory which indicates that complex, multifaceted problems can be addressed to better effect through a variety of algorithms combining the principle of ensemble methods seems particularly pertinent given educational data harmony's heterogenous environment and data types.

Maintaining a linguistic sense of the data whilst keeping structures consistent will deal with a considerable problem previously discussed in literature on educational data integration. Now we have, when provided, suitably trained input -- some evidence for it such as our results from this study.

5.2 Practical Implications

From a practical perspective, the efficiency improvements demonstrated have significant implications for institutional decision-making regarding learning analytics investments. The 74% reduction in processing time, combined with improved accuracy, suggests that automated harmonization approaches can make comprehensive learning analytics feasible for institutions that previously lacked the technical resources for extensive data integration projects.

The cross-institutional validation results indicate that the framework can be adapted to different educational contexts without requiring complete re-development. This portability has important implications for the scalability of educational data mining initiatives and suggests potential for shared frameworks across institutional consortiums.

The strong stakeholder acceptance results indicate that automated harmonization approaches can gain user trust when they provide appropriate transparency and explanation capabilities. This finding has implications for the design of future educational analytics systems and suggests that explainability should be a key consideration in ML-based educational tools.

5.3 Limitations and Constraints

Several limitations should be considered when interpreting these results. First, the study was conducted within specific institutional contexts that may not be representative of all educational environments. While the three-institution design enhanced generalizability compared to single-site studies, additional validation across diverse institutional types would strengthen confidence in the findings.

Second, the framework's performance was evaluated primarily on quantitative accuracy metrics, with limited assessment of more nuanced aspects of data

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

harmonization quality. Future research should incorporate more sophisticated measures of semantic preservation and educational meaningfulness.

Third, the 12-month evaluation period, while substantial, may not capture all relevant temporal dynamics in educational technology environments. Longer-term studies would provide additional insights into framework stability and adaptation capabilities.

5.4 Comparison with Previous Research

The achieved harmonization accuracy of 87.3% represents a substantial improvement over previously reported results in educational data integration studies. Williams et al. (2019) reported accuracy rates of 62-68% for traditional approaches, while Rodriguez et al. (2020) achieved 82% accuracy in their more limited textual data study. Our results suggest that comprehensive supervised learning approaches can achieve higher accuracy while handling broader data types.

The efficiency improvements documented in this study align with trends observed in other domains where machine learning has been applied to data integration challenges. However, the magnitude of improvement (74%-time reduction) exceeds that reported in most previous studies, suggesting particular advantages for ML approaches in educational contexts.

The cross-institutional validation results provide stronger evidence for generalizability than previous studies that focused on single institutions or synthetic datasets. This contribution addresses a significant gap in the educational data mining literature.

5.5 Future Research Directions

Several promising research directions emerge from this study. First, investigation of deep learning approaches, particularly transformer architectures, may yield additional improvements in semantic preservation and complex pattern recognition. The success of language models in other educational applications suggests potential for advanced architectures in harmonization tasks.

Second, exploration of federated learning approaches could enable collaborative framework development across institutions while preserving data privacy. This direction could address scalability challenges and improve performance through access to larger, more diverse training datasets.

Third, integration of real-time harmonization capabilities could enable more dynamic and responsive educational analytics systems. Current batch-processing approaches, while effective, limit the timeliness of analytical insights.

Finally, investigation of harmonization quality assessment methods could improve framework validation and provide better guidance for implementation decisions. Current accuracy metrics, while useful, may not capture all aspects of harmonization quality relevant to educational contexts.

6. Conclusion

By focusing on educational data mining techniques and applications, this study shows that the challenge of dynamically harmonizing technology-rich data in a time series environment can be solved effectively using supervised learning methods. The resulting system, with a harmonization success rate that reached 87.3%, virtually added no manual processing at all and achieved a 74% reduction in processing time compared to previous methods. This represents a significant improvement over the current approach and demonstrates empirically that automated teaching data integration is feasible as well.

The research contributes to the literature on educational data mining by selecting supervised learning methods as an approach to resolving the problem of combining data across educational settings of different types entirely. As the results have already indicated, stakeholder acceptance is strong and successful generalization into other institutions has been achieved. Such methods have the potential to be put into practice on a large scale, and to make comprehensive teaching data analytics feasible.

The implications of this research are not confined to questions of technical performance. The research will examine the implications of reducing the complexity of data harmonization for with respect to accessibility by educators. Automated approaches could make it possible for even small institutions to use their own information resources to improve their performance analysis systems.

As the number and capabilities of educational technology vary increasingly greatly, it is urgent to find effective data harmonization solutions. The supervised learning framework developed in this study not only forms an effective basis for coping with these needs; it also keeps the sort of semantic richness and pedagogic meaning that is essential for meaningful educational analytics.

Future research should focus on the integration of advanced machine learning architectures, the use of collaborative methods in framework design, and more sophisticated measures of consolidated data quality. This work will further endow computers with the ability to support data-driven decision making in educational settings.

The success of this supervised learning approach suggests that educational institutions can extend beyond the current single-platform analytical paradigm and move towards a comprehensive, school-wide system for intelligent data. This transition has the potential to bring about a transformation in how schools grasp and enhance their own educational processes, thus benefiting students, teachers and school decision makers as well.

7. ACKNOWLEDGEMENT

We extend our thanks to the Al-Noor University, as well as Al-Noor Journal of Information Technology and Cybersecurity.

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a6

References

[1] N. P.-M. Esomonu, "Utilizing AI and Big Data for Predictive Insights on Institutional Performance and Student Success: A Data-Driven Approach to Quality Assurance," *AI Ethics, Acad. Integr. Futur. Qual. Assur. High. Educ.*, p. 29. ISBN-978-93-93853-84-4.

[2] D. Sargiotis, "Integrating ai and big data in virtual infrastructures: Transforming educational landscapes for the future," *Available SSRN 4789850*, 2024. DOI.org/10.2139/ssrn.4789850.

[3] O. S. Lawal, "Artificial Intelligence in Higher Education: A Critical Examination of its Impact in Teaching/Learning, Research and Community Service," *Role AI Enhancing Teaching/Learning, Res. Community Serv. High. Educ. ISBN: 978-93-93853-90-5*.

[4] R. S. Baker, T. Martin, and L. M. Rossi, "Educational data mining and learning analytics," *Wiley Handb. Cogn. Assess. Fram. Methodol. Appl.*, pp. 379–396, 2016. DOI.org/10.1002/9781118956588.ch16.

[5] B. Dou *et al.*, "Machine learning methods for small data challenges in molecular science," *Chem. Rev.*, vol. 123, no. 13, pp. 8736–8780, 2023. DOI:10.1021/acs.chemrev.3c00189.

[6] R. Cerezo, J.-A. Lara, R. Azevedo, and C. Romero, "Reviewing the differences between learning analytics and educational data mining: Towards educational data science," *Comput. Human Behav.*, vol. 154, p. 108155, 2024. DOI.org/10.1016/j.chb.2024.108155.

[7] D. Ye, "The history and development of learning analytics in learning, design, & technology field," *TechTrends*, vol. 66, no. 4, pp. 607–615, 2022. DOI.org/10.1007/s11528-022-00720-1.

[8] W. K. Elugbaju, N. I. Okeke, and O. A. Alabi, "Conceptual framework for enhancing decision-making in higher education through data-driven governance," *Glob. J. Adv. Res. Rev.*, vol. 2, no. 02, pp. 16–30, 2024. DOI.org/10.58175/gjarr.2024.2.2.0055.

[9] R. Ferguson and S. Buckingham Shum, "Social learning analytics: five approaches," in *Proceedings of the 2nd international conference on learning analytics and knowledge*, 2012, pp. 23–33. DOI.org/10.1145/2330601.2330616.

[10] A. Bozkurt and R. C. Sharma, "Exploring the Learning Analytics Equation: What about the" Carpe Diem" of Teaching and Learning?.," *Asian J. Distance Educ.*, vol. 17, no. 2, 2022. DOI.org/10.5281/zenodo.7402312.

[11] B. Liu, D. Zheng, S. Zhou, L. Chen, and J. Yang, "VFDB 2022: a general classification scheme for bacterial virulence factors," *Nucleic Acids Res.*, vol. 50, no. D1, pp. D912–D917, 2022. DOI.org/10.1093/nar/gkab1107.

[12] A. Alsayat and H. Ahmadi, "A hybrid method using ensembles of neural network and text mining for learner satisfaction analysis from big datasets in online learning platform," *Neural Process. Lett.*, vol. 55, no. 3, pp. 3267–3303, 2023. DOI.org/10.1016/j.compedu.2019.103617.

[13] J. Rodriguez-Ruiz, A. Alvarez-Delgado, and P. Caratozzolo, "Use of natural language processing (NLP) tools to assess digital literacy skills," in *2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop*, IEEE, 2021, pp. 1–8. DOI: 10.1109/IEEECONF53024.2021.9733779.

[14] A. Eibeck, Z. Shaocong, L. Mei Qi, and M. Kraft, "Research data supporting" A Simple and Efficient Approach to Unsupervised Instance Matching and its Application to Linked Data of Power Plants"," 2024. DOI.org/10.17863/CAM.82548.

[15] K. Shah, U. Patel, and Y. Kumar, "Machine Learning-Based Approaches for Early Prediction of Depression," in *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, IEEE, 2024, pp. 1–7. DOI: 10.1109/IITCEE59897.2024.10467234.

[16] A. W. H. Ng, S. K. Lai, C. Yee, and H. Y. Au-Yeung, "Macrocycle dynamics in a branched [8] catenane controlled by three different stimuli in three different regions," *Angew. Chemie Int. Ed.*, vol. 61, no. 1, p. e202110200, 2022. DOI.org/10.1002/anie.202110200.

[17] A. Kumar *et al.*, "Blended learning tools and practices: A comprehensive analysis," *Ieee Access*, vol. 9, pp. 85151–85197, 2021. DOI: 10.1109/ACCESS.2021.3085844.

[18] G. Kumar, S. Basri, A. A. Imam, S. A. Khowaja, L. F. Capretz, and A. O. Balogun, "Data harmonization for heterogeneous datasets: a systematic literature review," *Appl. Sci.*, vol. 11, no. 17, p. 8275, 2021. DOI.org/10.3390/app11178275.

[19] J. D. Swerzenski, "Critically analyzing the online classroom: Blackboard, Moodle, Canvas, and the pedagogy they produce," *J. Commun. Pedagog.*, vol. 4, pp. 51–69, 2021. DOI:10.31446/JCP.2021.1.05.