A Comparative Analysis of Data Mining Models for Student Performance Prediction in School Education

Zahira Noor Quraishi[1], Dr. Atul Dattarya Newase[2]

[1]Research Scholar, [2]Research Supervisor
[1,2]Dr. A. P. J. Abdul Kalam University, Indore, India
zahira16sep@gmail.com , dr.atulnewase@gmail.com

Presented at International Conference on Trends and Innovations in Management, Engineering, Sciences and Humanities (ICTIMESH-24), London, 24-27 June 2024.

Published in IJIRMPS (E-ISSN: 2349-7300), ICTIMESH-24

Abstract
Student performance prediction has become a central focus of Educational Data Mining (EDM) due to its potential to support early intervention and improve learning outcomes. With the increasing availability of educational data, various data mining and machine learning models have been applied to analyze academic, behavioral, and demographic factors influencing student achievement. This study presents a comparative analysis of commonly used data mining models for student performance prediction in school education, with particular emphasis on primary and secondary levels. Drawing upon recent empirical studies and systematic reviews, the research examines traditional machine learning algorithms such as Decision Trees, Random Forest, Naïve Bayes, Regression models, and Neural Networks, alongside pre-processing techniques including feature selection and class imbalance handling. The analysis highlights that ensemble-based models, particularly Random Forest, consistently demonstrate strong predictive performance across diverse datasets, while simpler models offer better interpretability. However, most existing studies rely on higher education or single-institution datasets, limiting their applicability to school-level contexts. By synthesizing methodological trends, performance outcomes, and limitations, this study identifies critical gaps and emphasizes the need for general, explainable, and context-aware models for school education. The findings aim to guide researchers and practitioners in selecting suitable data mining models for effective student performance prediction.

Keywords: Educational Data Mining, Student Performance Prediction, Machine Learning, Primary Education, Secondary Education

1. Introduction

Education systems across the world are increasingly adopting digital technologies for teaching, learning, and administration. As a result, large volumes of educational data are continuously generated from sources such as academic records, attendance logs, learning management systems, assessments, and surveys. Effectively analyzing this data can provide valuable insights into student learning behavior and academic progress [1]. Educational Data Mining (EDM) has emerged as a specialized research field that applies data mining and machine learning techniques to educational datasets with the objective of improving decision-making and learning outcomes [2].

One of the most important applications of EDM is student performance prediction. Predicting academic performance at an early stage enables educators to identify at-risk students, design personalized learning strategies, and implement timely interventions. Over the past decade, numerous data mining models have been developed to forecast student success, dropout risk, and achievement levels. These models typically use academic history, behavioral indicators, demographic attributes, and engagement data as predictive features [3].

Existing literature shows that a wide range of machine learning algorithms has been employed for performance prediction, including Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines, Linear and Logistic Regression, Artificial Neural Networks, and more recently, deep learning models. Comparative studies indicate that ensemble methods such as Random Forest often outperform single classifiers in terms of accuracy and robustness, while regression-based and probabilistic models provide better interpretability. Feature selection techniques, such as Recursive Feature Elimination and correlation-based filtering, are frequently used to enhance model efficiency and reduce noise. Additionally, handling class imbalance through resampling techniques has been shown to significantly improve predictive performance [4].

Despite these advancements, a critical review of the literature reveals that the majority of EDM research focuses on higher education environments. School education, particularly primary and secondary levels remains comparatively under-explored, even though early educational stages play a crucial role in shaping long-term academic trajectories. Many school-level studies are limited in scale, rely on survey-based data, or use older analytical methods, which restricts their applicability in modern data-driven education systems [5].

Another important challenge identified in literature is the lack of model generalizability. Many studies develop prediction models using data from a single institution, region, or country, which limits their transferability to different educational contexts. Furthermore, while advanced models such as neural networks and deep learning offer high accuracy, they often function as “black boxes,” reducing transparency and trust among educators and policymakers. This highlights the importance of comparative analysis to balance predictive performance with interpretability and practical usability [6].

Given these challenges, there is a strong need for systematic comparison of data mining models applied to student performance prediction in school education. Such an analysis can help identify models that not only perform well but are also interpretable, scalable, and suitable for real-world deployment in primary and secondary schools. This study aims to address this need by synthesizing findings from recent research, comparing commonly used data mining techniques, and highlighting methodological trends and gaps. The outcomes of this work are intended to support researchers, educators, and policymakers in selecting effective analytical models for improving student performance in school education systems.

2. Literature review

Choi et al. (2023), Programming education is essential but challenging for beginners, and EDM is increasingly used to understand learning behavior and improve outcomes in programming courses. This systematic literature review synthesizes research from the last five years on EDM-based performance prediction in programming education. It examines common data sources and influential features, predictive targets, modeling approaches, pre-processing steps, validation strategies, and evaluation metrics used to assess model quality. The review also discusses limitations and challenges across prediction approaches and proposes directions for future work, aiming to guide researchers toward more robust and meaningful prediction systems in programming learning contexts. **[1]**

ilva Filho et al. (2023), To address a common limitation in EDM—weak causal reasoning—this study combines EDM techniques with theory-driven causal models to better interpret performance interventions. Using large-scale Brazilian assessment data, the authors map unobserved confounders with causal graphs and apply a two-way fixed-effects logistic regression to control for confounding. The model's predictive ability is evaluated and then explored via classification rules and decision trees to generate interpretable insights. Findings emphasize the influence of socio-economic factors and highlight the impact of faculty education policies, including variation across Brazilian states, demonstrating how causal modeling can strengthen the usefulness of EDM findings for decision-makers. **[2]**

Chytas et al. (2023), An interactive system is proposed to assess and improve learning processes using data generated by online university services, analyzed across periods before, during, and after the COVID-19 outbreak at a Greek university. By examining learning paths, online presence, and service participation, the system derives performance insights and predicts future learning progression. The study argues such analytics can help universities refine learning design, adjust online and in-person delivery, and strengthen strategic planning. Overall, it positions institutional service data as a resource for improving quality, supporting students, and enabling more targeted teaching practices. **[3]**

Gök et al. (2023), This study applies data mining to understand factors influencing primary teachers' mathematics teaching anxiety and motivation, using Random Forest for prediction and K-Means clustering to define profiles. Survey data from 485 Turkish teachers included demographic variables alongside standardized anxiety and motivation scales, with outcomes transformed into low/high categories. Across both models, "grade level taught" had the highest predictive importance, followed by "length of service." The work demonstrates how EDM methods can reveal educator profiles and key predictors, potentially informing targeted professional support and interventions for mathematics teaching. **[4]**

Wongvorachan et al. (2023), Educational data mining (EDM) enables data-driven applications such as early warning systems and academic performance prediction, yet class imbalance remains a major challenge. Many predictive models assume balanced class distributions, which is rarely the case in educational datasets. This study compares resampling techniques across different imbalance ratios using the High School Longitudinal Study of 2009 dataset. Random oversampling, random under-sampling, and a hybrid SMOTE-NC plus under-sampling approach were evaluated with a Random Forest classifier. Results indicate random oversampling performs best for moderately imbalanced data, while hybrid resampling is more effective for extremely imbalanced datasets, offering practical guidance for EDM applications. **[5]**

Yang et al. (2023), This research explores the integration of data mining techniques with teaching quality evaluation in vocational education. Addressing limitations in existing evaluation systems, the study designs a diversified teaching quality indicator framework using Analytic Hierarchy Process (AHP) for weighting indicators. Drawing from domestic and international practices, the proposed system improves fairness and efficiency in evaluation. By implementing a process-oriented evaluation model supported by data mining, the study enhances objectivity and rationality in assessing vocational teachers' teaching quality, providing a methodological reference for higher vocational institutions. **[6]**

Guleria et al. (2023), Machine learning and Explainable AI are combined in this study to propose an intelligent framework for student career counseling. Acting as a decision-support expert system, the framework analyzes academic and employability attributes to guide students toward suitable career paths. Both white-box and black-box ML models were trained on educational datasets. Among the evaluated algorithms, Naïve Bayes achieved the best performance, with recall and F-measure scores above 90%. The framework demonstrates how interpretable and predictive AI models can support informed career decisions and improve employability outcomes. **[7]**

Abdelmagid et al. (2023), Using the Orange data mining platform, this study investigates learning patterns and predicts academic performance among university students using Blackboard LMS data. K-Means clustering revealed three performance-based student groups, while multiple predictive algorithms were tested. Results show that activities and periodic assessments strongly predict performance in most courses. Linear Regression emerged as the most effective prediction model, whereas SVM contributed the least. The findings emphasize the role of learning analytics in identifying meaningful performance indicators and tailoring instructional support in e-learning environments. **[8]**

Eldeen et al. (2023), This research evaluates the effectiveness of a flipped classroom approach integrated with the Cognitive Theory of Multimedia Learning and Bloom's revised taxonomy in vocational education. Educational data mining was applied to Moodle log data, including pre-class activities, attendance, and assignments. Correlation analysis revealed that students using the flipped approach outperformed peers in traditional settings, with after-class assignments being a strong predictor of final results. The study demonstrates how EDM supports evidence-based evaluation of pedagogical strategies and enhances decision-making in blended learning environments. **[9]**

Cardona et al. (2023), Predicting student retention is critical for improving graduation rates in higher education. This systematic review examines machine learning approaches used to forecast dropout, attrition, and completion risk. The study analyzes methodologies, predictive factors, and performance contexts across selected literature. Key findings highlight commonly used algorithms, influential academic and behavioral factors, and methodological challenges such as data quality and generalizability. By synthesizing existing research, the review provides guidance for developing more robust retention prediction models and supports intentional advising strategies to enhance student persistence. **[10]**

Parhizkar et al. (2023), This study focuses on building generalizable student performance prediction models across geographical contexts. Using questionnaire data from domestic and international students, several machine learning classifiers and feature selection techniques were applied. The research emphasizes model transferability rather than single-context accuracy. Results show that Random Forest and CNN achieved the best balance between accuracy and F-score, demonstrating strong generalizability. The findings highlight the importance of diverse datasets in EDM and contribute to developing scalable models suitable for global educational applications. **[11]**

Elsamani et al. (2023), Beyond education, data mining techniques are applied in this study to explore the relationship between employee well-being and innovativeness. Using citation network analysis and semantic similarity across decades of literature, a multilevel conceptual framework was developed encompassing individual, organizational, and market levels. The framework identifies key constructs and interactions influencing innovation outcomes. As the first study to integrate data mining with well-being and innovation research, it provides theoretical and practical guidance for organizations seeking to foster innovative capacity through holistic workforce strategies. **[12]**

Sánchez et al. (2023), This study applies the CRISP-DM methodology to analyze large-scale data from an online education center in Chile spanning nearly two decades. Using the Knowledge Discovery in Databases process, key variables influencing student success in e-learning programs were identified, including age, gender, educational level, and locality. By uncovering hidden patterns in 18,610 records, the research demonstrates how EDM can support strategic decision-making, enhance student success understanding, and improve the sustainability of online education programs. **[13]**

Al-Alawi et al. (2023), This systematic literature review examines the use of Educational Data Mining in predicting academic performance of prospective students to improve admission processes. Analyzing studies published between 2018 and 2022, the review identifies 13 EDM techniques and 35 predictive factors categorized into socio-demographic, secondary education, and admission data. Findings reveal how machine learning models support better student selection and reduce dropout risk. The study also presents recommendations and future research directions to strengthen data-driven admissions in higher education. **[14]**

Chen et al. (2023), This comparative study evaluates multiple machine learning methods for student performance prediction using diverse educational datasets. Unlike prior research focusing on single data types, the study examines binary and multi-class tasks across three task-oriented datasets. Seven optimized algorithms were assessed using multiple evaluation metrics and visual analyses. Results show Random Forest achieved the most consistent and superior performance, while Decision Tree and Artificial Neural Network models also demonstrated strong potential. The findings provide empirical guidance for selecting robust models in varied educational prediction scenarios. **[15]**

Singh et al. (2016), Predicting student performance is an important concern in primary education, especially for enabling timely support and improving next-year results. This study analyzes rural and urban primary school students in Betul district, Madhya Pradesh (India), using a survey-cum-experimental approach to build a dataset from both primary and secondary sources. The objective is to identify factors linked to prior exam performance and select a suitable data mining algorithm to predict students' grades. Hypothesis testing indicates that school type does not significantly influence performance, while school area (rural/urban) and students' previous results (along with related background factors such as occupation) play a major role in predicting grades. **[16]**

Sehaj Singh et al. (2021), Secondary and senior secondary education (Classes IX–XII) in India forms a crucial bridge between school and college, making strong educational infrastructure essential for holistic and interactive learning. This paper examines the growth of Madhya Pradesh's secondary education sector by mapping infrastructure development trends. While the state has shown progress through initiatives such as the Education Guarantee Scheme (EGS) and Muft Cycle Yojana, the study argues that further efforts are required to meet SDG 4 (Quality Education) and its targets. Using a blend of primary and secondary data, the research analyzes micro-level progress and proposes interventions to help policymakers design inclusive, equity-focused education policies so that no learner is left behind. **[17]**

Table 1. Systematic literature review

Ref	Author (First author et al.)	Year	Title	Methods	Result	Advantage	Limitation	Data Used
[1]	Choi et al.	2023	A systematic literature review on performance prediction in learning programming using educational data mining	Systematic literature review (last 5 years); analysis of data sources, features, models, validation, metrics	Identified common predictors, models, preprocessing steps, and challenges in programming education	Comprehensive synthesis guiding robust EDM-based programming prediction	Focused only on programming education; no empirical validation	Prior EDM studies on programming courses
[2]	Silva Filho et al.	2023	Leveraging causal	Causal graphs; two-	Socio-economic factors and faculty	Integrates causal	Complex modeling;	Large-scale

			reasoning in educational data mining: an analysis of Brazilian secondary education	way fixed-effects logistic regression; decision trees for interpretation	education policies significantly influence performance	reasoning with EDM, improving interpretability for policy decisions	requires large, high-quality datasets	Brazilian secondary education assessment data
[3]	Chytas et al.	2023	Educational data mining in the academic setting: employing the data produced by blended learning	Learning analytics; performance prediction; comparative temporal analysis (pre/during/post COVID-19)	Online service participation and learning paths predict future learning progression	Demonstrates institutional service data value for strategic planning	Context limited to a single Greek university	Online university service and blended learning data
[4]	Gök et al.	2023	Investigation of variables affecting primary school teachers' anxiety and motivation in mathematics	Random Forest prediction; K-Means clustering	Grade level taught and teaching experience are key predictors	Identifies teacher profiles to inform targeted interventions	Focuses on teachers, not direct student outcomes	Survey data from 485 Turkish primary teachers
[5]	Wongvorachan et al.	2023	A comparison of undersampling, oversampling, and SMOTE methods in EDM	Random oversampling, undersampling, SMOTE-NC + undersampling with Random Forest	Oversampling best for moderate imbalance; hybrid best for extreme imbalance	Practical guidance for handling class imbalance in EDM	Evaluated on one longitudinal dataset	High School Longitudinal Study of 2009 (USA)
[6]	Yang et al.	2023	Application of AHP algorithm based on data mining in vocational education evaluation	Analytic Hierarchy Process (AHP); data mining-based evaluation model	Improved fairness and objectivity in teaching quality evaluation	Structured, process-oriented evaluation framework	Focused on vocational education only	Teaching quality evaluation indicators
[7]	Guleria et al.	2023	Explainable AI and machine learning for EDM-inspired career counseling	Naïve Bayes, interpretable & black-box ML; Explainable AI	Naïve Bayes achieved >90% recall and F-measure	Combines high accuracy with explainability	Career counseling focus, not academic prediction	Academic and employability datasets
[8]	Abdelmagid et al.	2023	Utilizing	K-Means	Learning	Demonstra	Results	Blackboa

	al.		EDM techniques for detecting patterns and predicting academic performance	clustering; Linear Regression; SVM using Orange	activities and assessments strongly predict performance	tes effectiveness of LMS-based analytics	vary across courses; LMS-dependent	rd LMS data
[9]	Eldeen et al.	2023	Analysis of students' performance in flipped classroom-based vocational education	Correlation analysis; EDM on Moodle logs	Flipped classroom students outperformed traditional peers	Supports evidence-based evaluation of pedagogy	Limited to vocational education	Moodle logs, attendance, assignments
[10]	Cardona et al.	2023	Data mining and machine learning retention models in higher education	Systematic review of ML-based retention models	Identified key predictors and methodological challenges	Guides development of robust retention systems	Higher education focus only	Prior retention and dropout studies
[11]	Parhizkar et al.	2023	Student performance prediction evaluating geographical generalizability	Random Forest, CNN, feature selection	RF and CNN showed best accuracy–F-score balance	Emphasizes model transferability	Questionnaire-based data may limit depth	Domestic and international student surveys
[12]	Elsamani et al.	2023	Employee well-being and innovativeness using data mining	Citation network analysis; semantic similarity	Developed multilevel conceptual innovation framework	Novel application of DM beyond education	Not focused on student learning	Literature datasets across decades
[13]	Sánchez et al.	2023	Sustainable e-learning by data mining	CRISP-DM; KDD process	Age, gender, locality influence student success	Large-scale, long-term analysis	Limited to one institution	18,610 online education records (Chile)
[14]	Al-Alawi et al.	2023	EDM utilization to support admission processes	Systematic literature review (2018–2022)	Identified 13 EDM techniques and 35 predictors	Supports data-driven admissions	Admission-focused, not performance after entry	Prior admission-related EDM studies
[15]	Chen et al.	2023	Comparative study on student performance prediction	Random Forest, DT, ANN; multi-task datasets	Random Forest most consistent performer	Empirical comparison across datasets	Still largely traditional ML	Multiple educational datasets
[16]	Singh et al.	2016	Using data	Survey-cum-	Prior exam results	Early	Limited	Rural

			mining to predict primary school student performance	experimental ; hypothesis testing; DM models	and background factors predict grades	Indian study on primary education EDM	scale; older methods	and urban primary school data (M.P., India)
[17]	Sehaj Singh et al.	2021	Development trajectory of educational infrastructure in M.P.	Descriptive analysis using primary & secondary data	Identified infrastructure growth and policy gaps	Context-specific insights for policymak ers	Not predictive or EDM-based	Educatio n infrastruc ture data of M.P.

3. Research Gap

A clear research gap from the above studies is that robust, generalizable student performance prediction for primary and secondary schooling—especially in local/regional contexts—remains limited, despite strong progress in higher education and specialized domains. Several works are systematic reviews that consolidate methods and predictors but do not translate into deployable, school-level models (e.g., programming education prediction and admissions/retention reviews) [1], [10], [14], while many empirical studies rely on single-institution LMS datasets (Blackboard/Moodle) that may not represent school environments or broader populations [8], [9], [13]. Where methods advance beyond prediction to address interpretability, causal reasoning, and bias, this is mostly demonstrated in secondary education at scale in Brazil rather than being extended to diverse school systems and regions [2]. Important methodological issues—such as class imbalance handling and cross-geographical generalizability—are studied, but they are not yet integrated into a unified framework tailored for school-level interventions [5], [11], [15]. In the Indian/Madhya Pradesh context, existing work is either older and limited in scale/modeling depth for primary student prediction [16] or focuses on infrastructure trends rather than EDM-driven performance analytics, leaving a gap in building up-to-date, explainable, and policy-relevant EDM models for improving primary and secondary student outcomes in M.P. [17].

4. Systematic Result Analysis

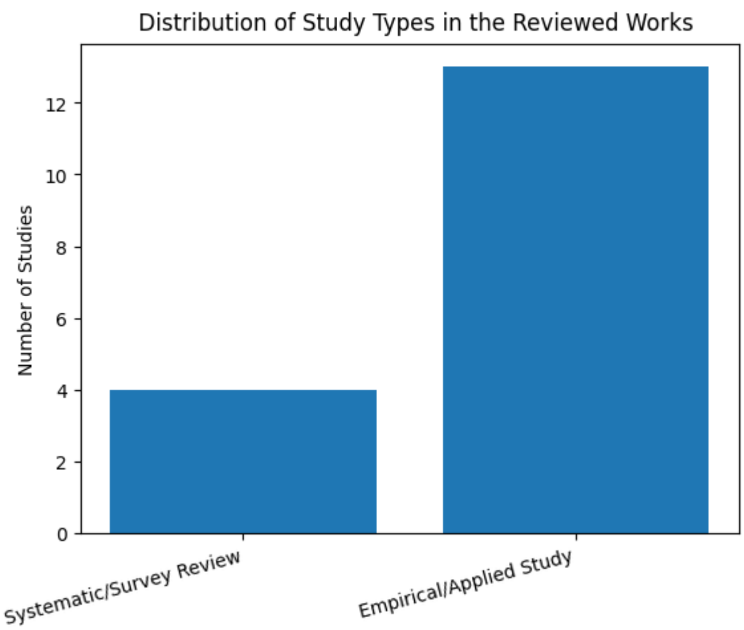

Figure 1: Distribution of Study Types in the Reviewed Works

This figure 1 compares the number of **systematic/survey review studies** versus **empirical/applied studies** in the selected literature. It shows that most works are empirical/applied, while a smaller portion focuses on literature synthesis.

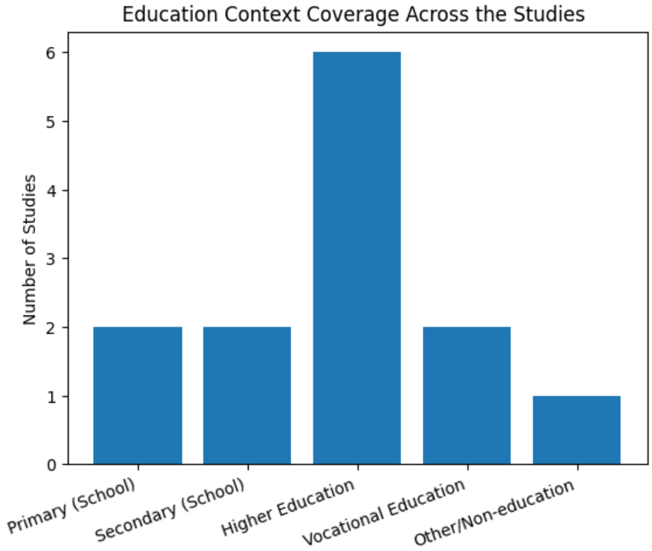

Figure 2: Education Context Coverage Across the Studies

This figure 2 summarizes which education contexts are most studied. It indicates that the reviewed literature is **heavily concentrated in higher education**, while **primary and secondary school contexts** are comparatively fewer, reflecting a gap in school-level EDM research.

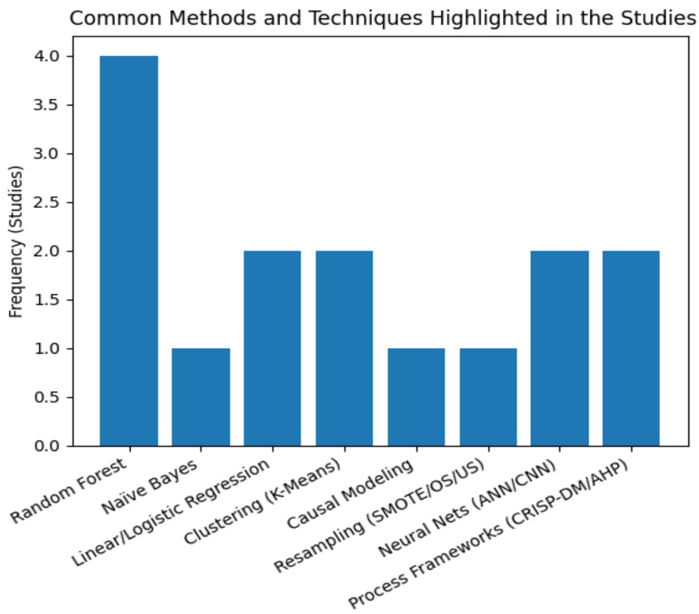

Figure 3: Common Methods and Techniques Highlighted in the Studies

This figure 3 shows the most frequently reported methods. Random Forest appears most often, followed by other techniques such as regression, clustering, neural networks (ANN/CNN), and process frameworks (CRISP-DM/AHP). Resampling and causal modeling appear less frequently, suggesting scope for broader adoption.

5. Conclusion

This study presented a comparative analysis of data mining models used for student performance prediction in school education, drawing insights from recent empirical research and systematic reviews. The analysis indicates that traditional machine learning models, particularly Random Forest and other ensemble-based approaches, consistently demonstrate strong predictive performance across diverse educational datasets. Simpler models such as regression and Naïve Bayes offer advantages in terms of interpretability, making them suitable for educational decision-making. However, the review also highlights significant gaps, including the dominance of higher education-focused studies, limited school-level datasets, and insufficient attention to model generalizability and explainability. Addressing these limitations is essential for effective implementation of EDM in primary and secondary education. Strengthening data quality, incorporating contextual factors, and balancing accuracy with transparency can significantly enhance the usefulness of performance prediction models. Future work should focus on developing explainable, generalizable, and region-specific EDM frameworks validated on large-scale primary and secondary school datasets.

References

1. Choi, Wan-Chong, Chan-Tong Lam, and António José Mendes. "A systematic literature review on performance prediction in learning programming using educational data mining." In *2023 IEEE Frontiers in Education Conference (FIE)*, pp. 1-9. IEEE, 2023.
2. Silva Filho, Rogério Luiz Cardoso, Kellyton Brito, and Paulo Jorge Leitão Adeodato. "Leveraging causal reasoning in educational data mining: an analysis of Brazilian secondary education." *Applied Sciences* 13, no. 8 (2023): 5198.
3. Chytas, Konstantinos, Anastasios Tsolakidis, Evangelia Triperina, and Christos Skourlas. "Educational data mining in the academic setting: employing the data produced by blended learning to ameliorate the learning process." Data Technologies and Applications 57, no. 3 (2023): 366-384.
4. Gök, B., Akkuş, E.B., Kavak, G. and KASAP, P.Y., 2023. Investigation of the Variables Affecting Primary School Teachers' State of Anxiety and Motivation in Mathematics Teaching through Data Mining Methods. Current Psychology, 42(31), pp.27678-27693.
5. Wongvorachan, Tarid, Surina He, and Okan Bulut. "A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining." Information 14, no. 1 (2023): 54.
6. Yang, Zhongying. "Research on the Application of AHP Algorithm Based on Data Mining in Vocational Education Teaching Evaluation System." In *3rd International Conference on Internet, Education and Information Technology (IEIT 2023)*, pp. 856-862. Atlantis Press, 2023
7. Guleria, Pratiyush, and Manu Sood. "Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling." *Education and Information Technologies* 28, no. 1 (2023): 1081-1116.
8. Abdelmagid, A. S., and A. I. M. Qahmash. "Utilizing the Educational Data Mining Techniques" Orange Technology" for Detecting Patterns and Predicting Academic Performance of University Students." *Inf. Sci. Lett* 12, no. 3 (2023): 1415-1431.
9. Eldeen, Yousra O. Kamal, Mohamed AbdelRaheem, and Taysir Hassan A. Soliman. "Analysis of students performance in flipped classroom-based vocational education using data mining tools." In *2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 387-394. IEEE, 2023.
10. Cardona, Tatiana, Elizabeth A. Cudney, Roger Hoerl, and Jennifer Snyder. "Data mining and machine learning retention models in higher education." *Journal of College Student Retention: Research, Theory & Practice* 25, no. 1 (2023): 51-75.
11. Parhizkar, Amirmohammad, Golnaz Tejeddin, and Toktam Khatibi. "Student performance prediction using datamining classification algorithms: Evaluating generalizability of models from geographical aspect." *Education and Information Technologies* 28, no. 11 (2023): 14167-14185.

12. Elsamani, Yousif, Cristian Mejia, and Yuya Kajikawa. "Employee well-being and innovativeness: A multi-level conceptual framework based on citation network analysis and data mining techniques." *Plos one* 18, no. 1 (2023): e0280005.
13. Sánchez, Aurora, Cristian Vidal-Silva, Gabriela Mancilla, Miguel Tupac-Yupanqui, and José M. Rubio. "Sustainable e-Learning by data mining—Successful results in a Chilean University." *Sustainability* 15, no. 2 (2023): 895.
14. Al-Alawi, Adel Ismail, Maryam Ali Abu Alfateh, and Amal Mohammed Alrayes. "Educational Data Mining Utilization to Support the Admission Process in Higher Education Institutions: A Systematic Literature Review." In *2023 International Conference On Cyber Management And Engineering (CyMaEn)*, pp. 332-339. IEEE, 2023.
15. Chen, Yawen, and Linbo Zhai. "A comparative study on student performance prediction using machine learning." *Education and Information Technologies* 28, no. 9 (2023): 12039-12057.
16. Singh, Manmohan, Harish Nagar, and Anjali Sant. "Using data mining to predict primary school student performance." IJARIIE 2, no. 1 (2016): 43-46.
17. Sehaj Singh, Utkarsha Bansal, "Development Trajectory Of Educational Infrastructure in Madhya Pradesh At Secondary and Senior Secondary Education-Levels", IJRAR December 2021, Volume 8, Issue 4, 2021, pp. 240-246.