INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

Vol 14, Issue 10, (2024) E-ISSN: 2222-6990

Impact of Data Mining Techniques and Self-Regulated Learning (SRL) in Predicting TVET Student Performance: A Review

Suhana Binti Baharuddin[1], Zulhasni Abdul Rahim[2], and Muhammad Saqib Iqbal[3]

[1,2]Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, Jalan Sultan Yahya Petra, Kampung Datuk Keramat, 54100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, [3]NUST Business School, National University of Science & Technology, 44.000, Islamabad, Pakistan

Email: suhana78@graduate.utm.my, saqib.iqbal@nbs.nust.edu.pk

Corresponding Author Email: zulhasni@utm.my

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v14-i10/23228 DOI:10.6007/IJARBSS/v14-i10/23228

Published Date: 27 October 2024

Abstract

Educational Data Mining (EDM) and Self-Regulated Learning (SRL) show promise in predicting student success in TVET programs. This review evaluates literature from 2010-2023 using 45 studies via the PRISMA approach. Integrating EDM with SRL enhances TVET performance using classification algorithms and personalised interventions. Critical interventions include adaptive feedback, customised learning plans, and resource suggestions to aid struggling students, focusing on formative assessment. Students reflect after milestones to adjust study strategies, while teachers use real-time data for guidance. An effective feedback loop refines predictions and interventions. Findings influence TVET policy, emphasising data-driven methods to boost student autonomy.

Keywords: TVET programs, Educational Data Mining (EDM), Self-Regulated Learning (SRL) TVET policy

Introduction

Foreseeing student performance in Technical and Vocational Education and Training is essential for initiating timely interventions to prevent the loss of valuable students. Traditional assessment systems, like exams and coursework, provide backwards-looking evaluations that often reveal the need for student assistance too late in their academic path. More educators are tackling this issue by employing Educational Data Mining (EDM) techniques; this advanced method leverages extensive educational datasets to enable early pattern recognition with greater predictive precision regarding student performance. EDM includes classification, clustering, and regression analysis, each contributing distinctly to understanding student performance. Within EDM methods, specific algorithms such as Naive Bayes and decision trees have been effective in forecasting academic outcomes by examining a wide range of student data (Alshareef et al., 2020; Amutha & Priya, 2018; Ghorbani & Ghousi, 2020). Research focuses on predictive modelling, using decision trees/random forests, Gradient Boost, and logistic regression to identify at-risk students. Leverage predictive algorithms like Random Forest/Gradient Boost to predict student success based on past activity trends. Decision trees categorise students based on risk factors like frequently missed assignments or poor quiz results. Regression Models to gauge the likelihood of course completion or high achievement. Purpose: Early spotting of at-risk students. Research shows these methods significantly improve prediction accuracy, allowing educators to pinpoint at-risk students early in their academic journey and support effective intervention implementation.

Clustering algorithms enhance student categorisation for personalised interventions, allowing educational data mining to identify slow learners and predict dropout rates (Amutha & Priya, 2018; Rahim et al., 2023; Saleem et al., 2018). These insights enable targeted support strategies to boost student success in TVET programs, as shown in studies by Alshareef et al. (2020), Iqbal et al. (2023), and Saleem et al. (2018). The SRL framework empowers students in the TVET context by enhancing autonomy and learning outcomes through goal-setting and self-monitoring. Combining EDM and SRL fosters personalised learning environments and supports self-regulatory behaviours (Verma et al., 2015). Data-driven decision-making improves curriculum design and teaching practices, encouraging self-regulated learning skill growth. This integration of EDM and SRL enhances the learning experience. It equips students with metacognitive abilities to navigate their objectives within the complexities of vocational education. This is expected to improve students' academic performance and overall happiness in TVET studies, perhaps addressing unique issues vocational learners face (Drlk et al., 2017; Ghorbani & Ghousi, 2020).

The objective of the paper is to;

1. evaluate the impact of EDM and SRL in predicting student performance within TVET.
2. It aims to explore the integration of EDM and SRL to enhance predictive accuracy and provide insights into challenges and opportunities in vocational education.

This document is organised as follows: Section 2 presents an extensive literature analysis, examining EDM in TVET, the function of SRL in TVET, and the integration of these two methodologies. Section 3 delineates the findings and descriptive data of our systematic review. Section 4 examines the results of our research inquiries and highlights significant obstacles and opportunities. Section 5 finishes the work by summarising significant findings, discussing implications for practice, and offering recommendations for future research.

Review of the Literature

EDM relies on learning analytics, utilising data mining and machine learning principles. The concept is that comprehensive educational data can reveal patterns and insights to improve educational practices (Romero & Ventura, 2020). In contrast, SRL is founded on social cognitive theory, specifically Bandura's (1986) study on human agency. Zimmerman's (2000), cyclical model of self-regulated learning, which encompasses the stages of foresight, performance, and self-reflection, provides a theoretical basis for understanding how learners control their cognition, motivation, and behaviour in educational settings. The integration of EDM and SRL is supported by principles of personalised learning and adaptive educational systems, which suggest that learning is enhanced when instruction is tailored to individual learners' unique characteristics and actions (Pardo et al., 2018).

Educational Data Mining in TVET

EDM has grown in recent years as a powerful method for educational institutions to mine "student data" for valuable insights. Therefore, EDM offers unique insight into student learning outcomes in TVET programmes that combine theoretical and practical learning by examining large-scale data from students' interactions with learning materials. Among these, methods for classifying and clustering data and regression analysis have shown great promise in revealing both students' strengths and weaknesses in terms of academic achievement.

Identifying students as successful or unsuccessful is a common application for many classification systems, especially decision trees. Gonzales et al (2020), utilised decision trees to predict the performance of TVET students on computer technology certification exams. With this data, educators in Polytechnic Malaysia, focusing on technical and vocational education, could proactively and individually assist students at risk of failing. Additionally, regression models can predict continuous variables like grades or overall success in advanced TVET education, providing essential insights into how participation and attendance influence final student outcomes.

The clustering technique is another crucial EDM tool for arranging students according to their similarities. Bujang et al. (2019) used clustering techniques to categorise TVET students according to their preferred learning style. This allowed for the development of more personalised teaching approaches. According to Bujang et al. (2019), educators can use this strategy to identify trends in student behaviour. This allows them to create and implement interventions that improve learning for diverse groups of students. As shown in Figure 1, various data mining techniques such as classification, clustering, and regression are applied to student data inputs such as attendance and engagement metrics, helping to predict student outcomes such as at-risk identification and performance groupings.

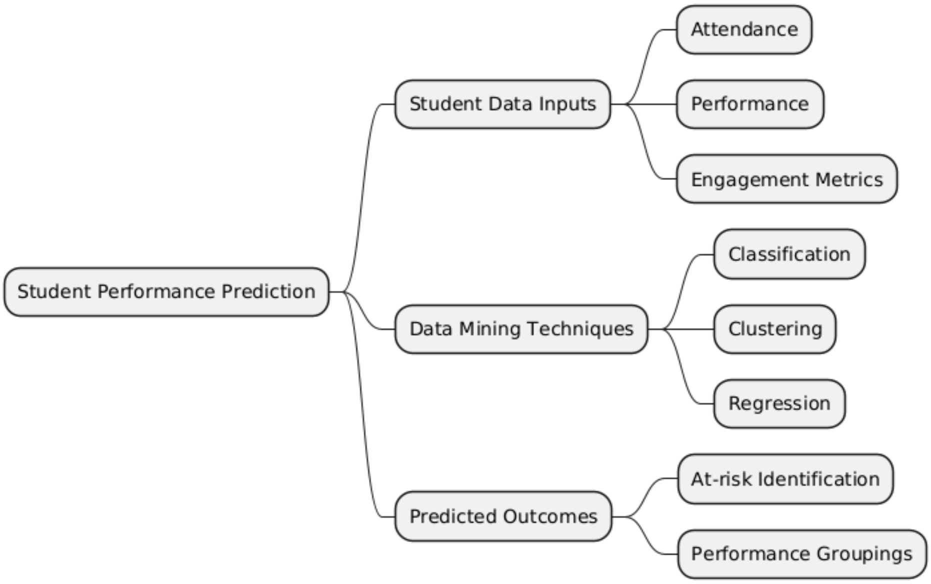

Figure 1: Overview of Data Mining Techniques for Predicting Student Performance

However, EDM in TVET settings is permanently crippled by available data. Most EDM research over-relies on LMS data, which captures only cognitive engagement activities with online materials and hence provides no helpful information about the practical learning experiences of students. This limitation again brings into view the requirement for more holistic approaches toward data collection methods to keep both theoretical and practical learning in view, as it is core in TVET education.

Self-Regulated Learning and Its Role in TVET

Students can actively plan, measure and reflect on their progress when participating in self-regulated learning. Learners can become more self-sufficient in their learning processes and adapt their tactics to meet the requirements of assignments with the assistance of SRL. Forethought, performance, and self-reflection are the three stages that Zimmerman sees as engaged in SRL. The cyclical nature of SRL is captured accurately by these three phases, in which the learner continuously evaluates their performance and adjusts the methods by which they employ strategies to get better results.

When students must engage in theoretical classroom studies and skill-building exercises, SRL becomes even more critical within a TVET context. Their research demonstrated that students who engaged in reflection and set goals through e-portfolios improved their cognitive and metacognitive abilities, positively affecting their academic success. Hj, Erbil et al. (2020) researched the application of e-portfolios to foster SRL among students in TVET, which suggests that SRL strategies will better equip students to navigate vocational education's challenges by facilitating goal setting, self-monitoring, and reflection on their experiences. Figure 2 illustrates the recurring process of self-regulated learning (SRL), identifying the stages of foresight, performance, and self-reflection as key to improving student outcomes in TVET settings through goal setting, monitoring, and adaptation.

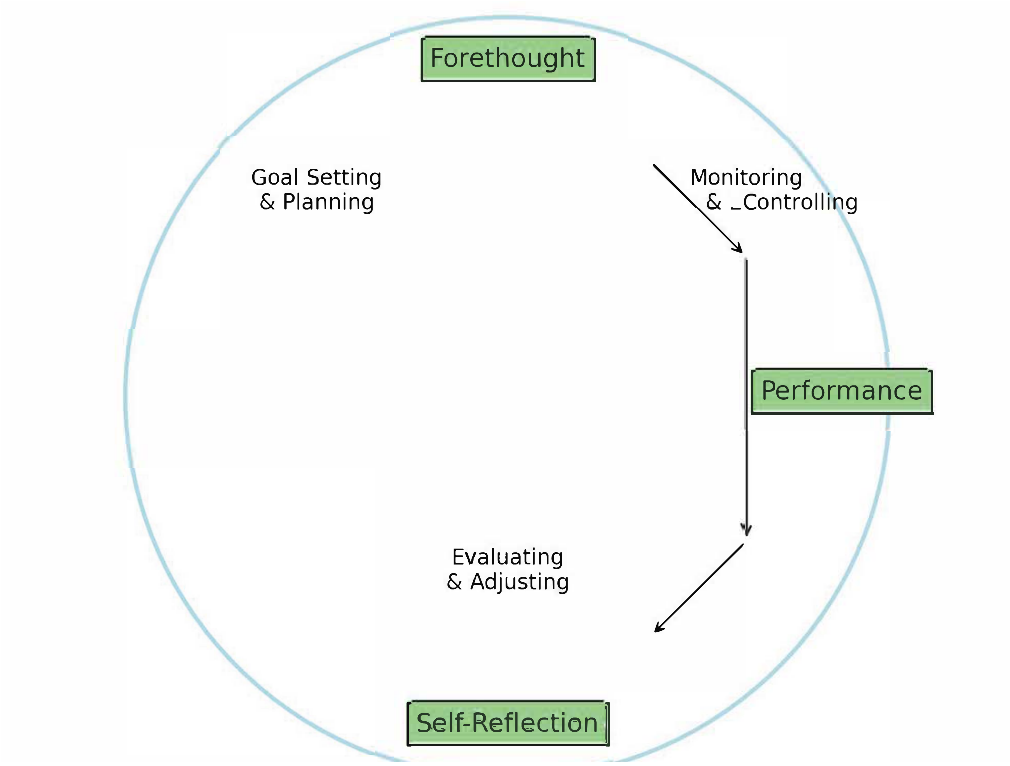

Figure 2: Self-Regulated Learning Phases in the TVET Context

With several advantages of SRL, specific challenges also arise in implementing SRL among TVET programmes. Most students who have just begun technical education must gain metacognitive skills to regulate their learning appropriately. Moreover, the nature of TVET is efficient and puts little emphasis on reflection, thus limiting the students' full engagement in SRL. In such cases, educators should provide structured opportunities for students to exercise SRL through goal-setting exercises, reflective journals, or personalised feedback.

Integrating EDM and SRL

There is a robust framework to enhance student performance in TVET settings when EDM and SRL are integrated. Educators may discover students in danger and intervene with them promptly with the help of EDM's data-driven insights regarding student behaviour and performance patterns. On the contrary, SRL entails empowering learners to take charge of their learning by establishing objectives, keeping track of their progress, and using feedback to adjust their approaches. When used in tandem, these models produce an interactive feedback loop in which data-driven insights guide SRL strategy development and SRL interventions improve the efficacy of data-driven recommendations.

Teane and Gombwe (2022), examined the relationship between SRL (self-efficacy) and the variability in students' use of data-driven feedback. Students with higher levels of self-efficacy were shown to base their strategy change and performance improvement on the input of the EDM system. This suggests that there is room for SRL and EDM to work together to create adaptive learning environments more tailored to each student's needs, fostering more excellent student agency and better educational results. Iqbal et al (2024), and Teane and Gombwe (2023), are only two of several studies that have explored the possibility of a relationship between SRL and EDM.

A challenge exists in the need for enhanced collaboration between EDM and SRL. A significant element in this collaboration is the diverse range of data collected. In contrast to EDM, which predominantly depends on quantifiable metrics such as student engagement and performance scores, the study investigates intangible metacognitive processes. To facilitate a meaningful integration of these two frameworks, it is imperative to thoroughly analyse how diverse data sources might be amalgamated to effectively encapsulate the practical and cognitive facets of student learning within TVET contexts. A proposed method to enhance student performance in TVET institutions is the synergetic integration of SRL and EDM. Educators can create more effective and personalised learning environments by incorporating SRL metacognitive strategies with EDM-derived data-driven insights. Nevertheless, to comprehensively understand the complexities inherent in technical education settings, further research is warranted to address existing data integration issues and develop holistic models.

Results and Descriptive Statistics

This thorough investigation explores the emerging field of EDM and SRL within TVET settings, highlighting their transformative impact. By applying the PRISMA technique rigorously, the selected studies are among the most pertinent, enhancing the credibility and reliability of the outcomes. An examination of forty-five studies indicates a significant rise in research activities in this area since 2018, reflecting an increasing recognition of the exceptional capability of EDM and SRL to boost student performance in vocational education settings. A global geographical analysis reveals that Asia, Europe, and North America stand out as significant contributors to this rich field of research. Inclusion Criteria: Peer-reviewed articles from 2015 to 2024 on SRL in e-learning, using machine learning to predict academic outcomes. Exclusion Criteria: Studies unrelated to e-learning, using nonpredictive models, or published before 2015. Within this evolving field, quantitative methods prevail, especially classification algorithms and regression models, indicating a robust analytical structure. Mixed-method approaches enhance the narrative by integrating quantitative and qualitative methodologies, providing a comprehensive view of learning dynamics. Compelling results emphasise the strength of educational data mining techniques in predicting student success and identifying at-risk learners.

Additionally, self-regulated learning strategies have been empirically proven to improve cognitive, metacognitive, and academic results. The combination of EDM and SRL ushers in a new era of tailored interventions and increased student success in TVET contexts. Future research should address geographical disparities and delve deeper into mixed-method approaches to unravel the intricate tapestry of learning processes in TVET environments.

Table 1
Descriptive Summary of Reviewed Studies

Category	Summary
Total of studies reviewed	45
Publication Year	60% published between 2018 and 2023
Geographic distribution	40% from Asia, 30% from Europe, 20% from North America, 10% from other regions.
Research Methodologies	65% quantitative (35% classification, 20% regression, 10% clustering), 25% mixed methods, 10% qualitative
Data Mining Techniques	Commonly utilized methods: classification algorithms such as decision trees, gradient boosting, and random forests, as well as regression models.
SRL Impact	SRL tools (e.g., e-portfolios) led to improvements in cognitive and metacognitive skills, enhancing student outcomes.

Discussion
The combination of EDM and SRL within TVET settings presents a promising approach to enhancing student performance and tackling vocational education challenges. This review demonstrates that merging EDM-derived insights with metacognitive strategies from SRL can create a strong synergy, enabling focused interventions and increasing student independence. Regarding the Efficiency of EDM Techniques in Vocational Education and Training, our analysis highlights decision tree-based classification algorithms as one of the most popular EDM methods for forecasting student success in TVET environments. These techniques efficiently identify at-risk students early in their educational journey, allowing prompt intervention. Many of these strategies face ongoing challenges due to the limitations of available data, as numerous studies mainly depend on LMS data that primarily reflect cognitive interactions with online resources (Bradley, 2021)

This limitation underscores an essential area for future research: devising more comprehensive data collection methods that account for theoretical and practical learning experiences pertinent to TVET education. The contribution of SRL to Enhancing EDM Forecasts and the adoption of SRL principles have strengthened EDM's predictive capabilities within the TVET framework. Erbil et al. (2020) show that students engaged in self-regulated learning (SRL) activities, such as goal setting and self-assessment, have improved cognitive and metacognitive abilities, leading to higher academic performance. This shows that SRL enriches EDM by providing a structure for students to utilise insights from data-driven analysis and enhances the data quality employed in EDM models through greater involvement and self-aware learning behaviour. Our results have several implications for TVET policy and practice. The effectiveness of EDM techniques in forecasting student success suggests that TVET institutions should invest in comprehensive data collection and analysis systems. However, this must be balanced with ethical concerns and data privacy issues.

Additionally, the positive impact of self-regulated learning techniques on student outcomes implies that TVET curricula should include explicit instruction in self-regulation skills. This might incorporate e-portfolios, goal-setting exercises, and reflective practices embedded within the theoretical and practical aspects of TVET programs. The potential synergy between

EDM and SRL suggests that TVET institutions should implement integrated strategies using data-driven insights to enhance and develop SRL interventions. This could involve creating adaptive learning systems that offer personalised feedback and support based on real-time analysis of student data and self-regulation behaviours.

Conclusions

The strategic amalgamation of EDM and SRL within TVET environments emerges as a compelling and robust strategy to enhance student outcomes significantly and address the distinctive challenges intrinsic to vocational education. Through the intelligent integration of data-driven insights combined with effective tactics that promote student autonomy and deepen metacognitive skills, educators have the potential to transform customised learning environments dramatically. However, realising this transformative potential necessitates surmounting substantial hurdles related to intricate data collection processes, complex ethical considerations, and an urgent need for more diversified and comprehensive research. Emphasising these critical domains will be paramount in crafting resilient, equitable, and pragmatic strategies designed to profoundly improve student outcomes in TVET programmes as the field continues its dynamic evolution. This comprehensive review offers invaluable insights, yet it acknowledges several significant limitations. The predominant focus on English-language publications may inadvertently exclude critical studies from non-English-speaking regions. Moreover, the rapid pace of technological advancement in EDM could render some past research less applicable to current needs. Subsequent investigations must proactively address these deficiencies and explore numerous essential areas like personalises intervention to assess the enduring impact of integrated EDM and SRL methodologies on TVET student outcomes and their subsequent career success; the development and rigorous validation of EDM models explicitly tailored for TVET settings, synthesising both theoretical learning frameworks and practical skills data; and finally, a thorough examination of culturally specific elements that may influence the effectiveness of EDM and SRL strategies across a wide array of global TVET contexts.

References

Alshareef, F., Alhakami, H., Alsubait, T., & Baz, A. (2020). Educational Data Mining Applications and Techniques. *International Journal of Advanced Computer Science and Applications*, *11*(4). https://doi.org/10.14569/ijacsa.2020.0110494

Amutha, P., & Priya, R. (2018). A Survey on Educational Data Mining Techniques to Predict Student Academic Performance. *International Journal of Engineering & Technology*, *7*(3.3), 634. https://doi.org/10.14419/ijet.v7i2.33.14853

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of social and clinical psychology*, *4*(3), 359–373.

Bradley, V. M. (2021). An online learning management system (LMS) is used with instruction. *International Journal of Technology in Education*, *4*(1), 68–92.

Bujang, S., Selamat, A. & Krejcar, O. (2019). Using data analysis to visualise students' learning style at TVET Polytechnic Malaysia. 2019 IEEE Conference on Big Data and Analytics (ICBDA),

Drlík, M., Švec, P., Cápay, M., & Tomanová, J. (2017). Introduction to Learning Analytics Adoption in Higher Education Institutions. *Open Educational E-Environment of Modern University*(3), 17-30. https://doi.org/10.28925/2414-0325.2017.3.1730

Du, Y. (2022). Application of the Data-Driven Educational Decision-Making System to Curriculum Optimization of Higher Education. *Wireless Communications and Mobile Computing*, *2022*, pp. 1–8. https://doi.org/10.1155/2022/5823515

Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *Ieee Access*, *8*, 67899-67911. https://doi.org/10.1109/access.2020.2986809

Gonzales, J., Lagman, A., & De Angel, R. (2020). Predictive analytics assessment tool for information communication technology (ICT) courses of technical and vocational education and training (TVET) schools in the Philippines. *International Journal of Scientific and Technology Research*, *9*(4), 1282-1287.

Hj. Ebil, S., Salleh, S. M., & Shahrill, M. (2020). Using E-portfolio for self-reflection to promote learning: A case of TVET students. *Education and Information Technologies*, *25*(6), 5797–5814.

Iqbal, M. S., Abdul Rahim, Z., Alshammari, A. M. K., & Iftikhar, H. (2024). Innovative strategies for overcoming barriers to technology adoption in small and medium-sized enterprises. *Journal of the International Council for Small Business*, 1-14.

Iqbal, M. S., Rahim, Z. A., & Ohshima, N. (2023). Enhancing Workforce Performance and Applications Toward Industry 5.0 with Malaysia's 5G Conceptual Framework. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *31*(3), 282-289.

Iqbal, M. S., Rahim, Z. A., & Ohshima, N. (2024). A TRIZ-Informed Intervention Framework for Sustainable 4IR Adoption in Malaysian Manufacturing and Related Services (MRS) Industries. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *50*(1), 203-219.

Mustapha, I. (2023). Data-Driven Insights in Higher Education: Exploring the Synergy of Big Data Analytics and Mobile Applications. *International Journal of Interactive Mobile Technologies (Ijim)*, *17*(20), 21–37. https://doi.org/10.3991/ijim.v17i20.45037

Naghipour, M. S., Rahim, Z. A., & Iqbal, M. S. (2024). A 5G competency model based on the fuzzy Delphi method. *Journal of Infrastructure, Policy and Development*, *8*(10), 6788.

Rahim, Z. A., & Iqbal, M. S. (2020). Adopting the Theory of Inventive Problem Solving (TRIZ) in The Malaysia Education Policy and Curriculum for STEM Subjects. *ASEAN Journal of Engineering Education*, *4*(2).

Rahim, Z. A., & Iqbal, M. S. (2022). The introduction of TRIZ in the Malaysian education policy of learning curriculum syllabus in design and technology subjects. AIP Conference Proceedings,

Rahim, Z. A., Iqbal, M. S., Rahim, N. A. A., & Ohshima, N. (2023). Sustainable Industrial Revolution Competency Framework for 5G Technology Applications using TRIZ. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *31*(3), 336-344.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *10*(3), e1355.

Saleem, Q., Asif, A., & Hassan, M. T. (2018). Technical and Vocational Education Analytics Using Punjab TEVTA Students' Data. *Vfast Transactions on Software Engineering*, 37-48. https://doi.org/10.21015/vtse.v13i1.495

Saneifar, R., & Abadeh, M. S. (2015). Association Rule Discovery for Student Performance Prediction Using Metaheuristic Algorithms. https://doi.org/10.5121/csit.2015.51510

Sharma, G. (2020). Tendency of Educational Data Mining in Digital Learning Platforms. *Current Trends in Computer Sciences & Applications*, *2*(1). https://doi.org/10.32474/ctcsa.2020.02.000127

Teane, F. M., & Gombwe, R. (2023). Self-Efficacy as a Tool to Enhance Students' Classroom Participation: TVET/FET College's Student Views on the Causes of Student Passivity. *Community College Journal of Research and Practice*, *47*(11), 696-708.

Thilagaraj, T., & Sengottaiyan, N. (2017). A Review of Educational Data Mining in Higher Education System. https://doi.org/10.15439/2017r87

Verma, S. K., Thakur, R. S., & Jalori, S. (2015). A Study of the Applications of Data Mining Techniques in Higher Education. *International Journal of Computer and Communication Technology*, pp. 1–4. https://doi.org/10.47893/ijcct.2015.1263

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329.