LITERATURE REVIEW

Learning analytics refers to the measurement, collection, analysis, and reporting of data about learners and their contexts, with the aim of understanding and optimizing learning and the learning environments. It has emerged as a powerful tool to inform decision-making and improve educational practices. With learning analytics, a data driven approach is being used to understand and support learners' progress, engagement, and achievement. Learning analytics can help educators to identify patterns, predict student performance, personalize learning experiences, and provide timely interventions. Baker and Siemens (2014) emphasized the potential of learning analytics in enabling evidence-based decision making for educators and administrators. These data-driven approaches are important when predicting student success (Siemens, 2013), in policy making and developing online learning platforms (Waheed *et al*., 2020; Bilal *et al*., 2022).

To have a comprehensive understanding of the dynamics of student engagement, several research studies have explored the role of learning analytics and the various factors that influence student interactions within virtual learning environments (VLEs). Student engagement in virtual learning environments is a critical factor in promoting effective learning outcomes. They offer flexibility, accessibility, and a range of multimedia resources. VLEs enable learners to engage in self-paced learning, collaborate with peers, and access resources from anywhere at any time.

This literature review aims to explore the intersection and synthesize key findings from recent studies that have investigated the relationship between learning analytics, student engagement, and VLEs.

Engagement with the VLE and student performance as studied by researchers

Different variables and factors have been studied by researchers to explore their interactions with the VLE and how they affect student performance. Siemens and Gašević (2012), in their study titled "Exploring Patterns of Student Interactions with a Virtual Learning Environment," analysed the log data of students' interactions with a VLE to identify patterns and behaviours indicative of student engagement and success. Their study revealed that certain types of interactions, such as frequent forum participation and regular access to learning materials, were associated with better learning outcomes.

Khalil and Ebner (2014) analysed various variables such as forum posts, access to course resources, and online activity to investigate student data from blended courses that utilized the Moodle learning management system (LMS). The study identified significant predictors of student success. Al-Balushi and Al-Badi (2016) examined factors such as content relevance and accessibility to investigate students' usage patterns and behaviours in a blended learning environment.

Furthermore, Ma *et al*. (2019) investigated the relationship between reflective thinking and student engagement in online learning environments. Their study emphasized the significance of promoting reflective thinking as a means to enhance student engagement within VLEs. Romero *et al*. (2019) developed predictive models using learning analytics to forecast students' academic performance and dropout risk in VLEs by exploring various factors, such as online behaviour and social network analysis. Gašević *et al*. (2020) and Verbert *et al*. (2021) investigated the impact of learning analytics dashboard on students' interactions and learning outcomes. The study revealed that the use of the user-friendly and actionable dashboards positively influenced student engagement, goal setting, and self-monitoring. Hernández-Leo *et al*. (2020) focused on the early detection of academic challenges by analysing students' sentiment through sentiment analysis of their interactions in VLEs. Their study demonstrated the potential of sentiment analysis in predicting students' difficulties and providing timely support.

Predictive analysis using the Open University Learning Analytics Dataset

Different types of machine learning models have been employed by researchers to predict student performance in the Learning Analytics field using the Open University Learning Analytics dataset. Many of these models are predominantly classification and regression algorithms. Decision Trees and related algorithms most commonly appear in literature (Hussain *et al*., 2018; Tomasevic, Gvozdenovic and Vranes, 2020; Bilal *et al*., 2022). These types of models have high interpretability and allow the features which contribute to the classification to be viewed. The common method implemented when developing machine learning models is to train and test multiple models and chose the model by examining precision, accuracy, and recall (Rizvi, Rienties and Khoja, 2019; Bilal *et al*., 2022).

Predictive analytics plays a crucial role in identifying at-risk students and providing targeted support. Predictive models can forecast potential challenges and suggest appropriate interventions by analyzing patterns in learner data, Predictive models have uses when developing student-centred learning environments, including predicting students as risk of failing in order to provide tailored support systems (Rizvi, Rienties and Khoja, 2019). Ramesh *et al*. (2018) conducted a study using the Open University Learning Analytics Dataset and developed predictive models to identify students who were likely to drop out. Their findings demonstrated the potential of predictive analytics in improving retention rates in virtual learning environments.

Aljohani, Fayoumi and Hassan (2019) used the clickstream data in the OULAD to predict student performance. Predictions were made on the possibility of a student passing/failing as early as 10 weeks of their interaction with the VLE based on their weekly interaction with the learning environment. They reported the superiority of the long short-term memory (LSTM) model over the other models tested with 90% accuracy, 93.46% precision and 75.79% recall. Their findings emphasized the effectiveness of deep learning models, such as deep long short-term memory (LSTM) and artificial neural networks (ANN), in early prediction of student performance. However, their study did not take into consideration the number of attempts., i.e., the performance of students who repeated their courses, and the influence of activity type on performance, which is one of the objectives of our study.

Drousiotis *et al*. (2021) used demographical and behavioural data to predict the final results of students and to identify students at risk of dropping out at the beginning of the course. They identified the first assessment mark as a strong predictor of student performance using the BART (Bayesian Additive Regression Trees) model, which outperformed the use of random forest and decision trees.

Bilal *et al*. (2022) developed techniques to predict final semester performance of students from pre-admission features and first-semester performance using data from students studying Doctor of Veterinary Medicine (DVM). The students were split into high and low performing. Splitting the students into two classes appears to be commonly used in the field of learning analytics (Bilal *et al*., 2022; Waheed *et al*., 2020). The authors also explore the predominant features affecting final semester performance. They used five supervised machine learning algorithms: Decision Tree, Random Forest, Support Vector Machines, K – Nearest Neighbours and Logistic Regression, and used precision, accuracy and recall to evaluate these models. From their Decision Tree Model, they defined attributes and factors which affected the performance of the students in their data. The findings indicate that the subjects studied, and the marks attained, prior to the DVM course, had an impact on the students' performance. It was noted that the demographic features did not play a role in the prediction of student performance.

However, other studies have shown that demographic features affect learning outcomes and should be included in models in the field of Educational Data Mining and Learning Analytics (Rizvi, Rienties and Khoja, 2019). Rizvi, Rienties and Khoja (2019) examined the contribution of the demographic features to the students' final result. This was done by measuring the variable importance from Decision Tree models. Region and IMD band were found to play a role in the prediction of students learning outcomes whereas features such as gender and disability were not significant. The previous education level was seen to be a contributing feature towards student performance in assessments early in the course but became less of a prominent feature for later performance. Therefore, it is uncertain from the current evidence whether demographic features have significance in predicting student performance. Given the current uncertainty around this topic, it warrants further investigation.

A study by Waheed *et al*. (2020) used Artificial Neural Networks to predict students that are 'at-risk' of failing or withdrawing. The authors found demographics and the student's interaction with the VLE to be important features. They mentioned that further study is required to assess the activity types and their importance on student performance.

Hussain *et al*. (2018) compared the engagement of different activity types and how this relates to student performance. The authors investigated factors affecting student motivation and the main features which contribute to higher student engagement. They found the activity types (forumng, oucontent and subpage) to have higher engagement and that students with higher engagement performed better than students with low engagement. However, there appears to be discrepancies in the way that engagement is classified in this paper. High engagement was classified as a combination of the variables: total number of clicks; score on assessment and final results, and low engagement as the number of clicks. This study used the machine learning classification models: Decision Trees; J48; Classification and Regression Tree (CART); JRIP Decision Rules; Gradient Boosting Trees; Naïve Bayes Classifier. The J48 model had the highest accuracy and recall. They also found that clicks on the activities oucontent, forumng, subpage, and homepage are most important features when predicting engagement. This study used the Open University Learning Analytics dataset and looked at a single social sciences course. Future research could evaluate the engagement of activity types in different courses and across the student cohort. This would provide more insight into the engagement of the different activity types and how this relates to student performance.

In summary, the utilization of the OULAD and related citations in this literature review demonstrates the value of leveraging large-scale educational datasets for research purposes. The findings from previous research demonstrate the potential of learning analytics to provide valuable insights into student interactions, predictive modelling, visualization, and early detection of academic challenges, ultimately leading to enhanced learning outcomes.

References

Aljohani, N.R., Fayoumi, A. and Hassan, S.-U. (2019) Predicting at-risk students using clickstream data in the virtual learning environment. Sustainability [online]. 11(24), 7238. http://dx.doi.org/10.3390/su11247238 [Accessed 10 April 2023].

Drousiotis, E., Shi, L. and Maskell, S. (2021) Early Predictor for Student Success Based on Behavioural and Demographical Indicators. In: A. I. Cristea and C. Troussas (Eds.): ITS 2021, LNCS 12677, pp. 161–172. https://doi.org/10.1007/978-3-030-80421-3_19 [Accessed 10 April 2023].

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (pp. 253-274). Cambridge University Press.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. Computers & Education, 104, 18–33.

Ramesh, A.; Goldwasser, D.; Huang, B.; Daume III, H.; and Getoor, L. 2014. Learning latent engagement patterns of students in online courses. In Twenty-Eighth AAAI Conference on Artificial Intelligence.

Kuzilek, J. *et al*. Open University Learning Analytics dataset. Sci. Data 4:170171 doi: 10.1038/sdata.2017.171 (2017)

Bilal, M., Omar, M., Anwar, W., Bokhari, R.H. and Choi, G.S. (2022) The role of demographic and academic features in a student performance prediction. Scientific Reports [online]. 12 (1), p. 12508. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9307570/ [Accessed 6 July 2023].

Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R. (2018) Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. Computational Intelligence and Neuroscience [online]. 2018, p. 6347186. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6189675/ [Accessed 6 July 2023].

Rizvi, S., Rienties, B. and Khoja, S.A. (2019) The role of demographics in online learning; A decision tree based approach. Computers & Education [online]. 137, pp. 32–47. Available from: https://www.sciencedirect.com/science/article/pii/S0360131519300818 [Accessed 5 July 2023].

Tomasevic, N., Gvozdenovic, N. and Vranes, S. (2020) An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & Education [online]. 143, p. 103676. Available from: https://www.sciencedirect.com/science/article/pii/S0360131519302295 [Accessed 6 July 2023].

Waheed, H., Hassan, S.-U., Aljohani, N.R., Hardman, J., Alelyani, S. and Nawaz, R. (2020) Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior [online]. 104, p. 106189. Available from: https://www.sciencedirect.com/science/article/pii/S0747563219304017 [Accessed 4 July 2023].