TOJSR
The Oxford Journal of Science and Research

ISSN: 2581 - 9380

TOJSR. 2025, 4 (1)

KNOCKING THE INFLUENCE OF LEARNING ANALYTICS DASHBOARD IN EDUCATION 5.0

Mrs. Priya S.

Department of Computer Science and Applications, The Oxford College of Science, Bangalore, Karnataka, India

*Email:priyatm31@gmail.com

ABSTRACT: LAD stands for Learning Analytics Dashboard: a tool used in educational settings to visualize and analyze data related to student learning and performance. These dashboards provide real-time insights to educators and learners, helping them make informed decisions to improve learning outcomes and personalize the educational experience. In the evolving landscape of Education 5.0, characterized by technological advancements and personalized learning, Learning Analytics Dashboards (LADs) serve as crucial tools for shaping and enhancing educational experiences. This research investigates the impact of LADs on student engagement, academic performance, and decision-making within the Education 5.0 framework. By examining how these dashboards support learning outcomes through the integration of real-time data visualization, predictive analytics, and personalized feedback, the study highlights their role in empowering both educators and students. Additionally, the paper addresses potential challenges associated with LAD adoption, such as privacy concerns, data misinterpretation, and digital inequality. Ultimately, the findings emphasize LADs' transformative potential in redefining educational experiences and setting a new standard for global classroom objectives. This work aims to analyze a set of the most recent versions of LAD frameworks, focusing on their data collection, preprocessing, data interpretation methodologies, advantages, and limitations.

KEY WORDS*:* Adaptive Learning Systems, Data-Driven Education, Educational Data Mining, Learning Analytics Dashboards, Machine Learning (ML), Predictive Analytics, Student Performance Tracking, Artificial Intelligence (AI)

INTRODUCTION

As tech is taking over a fast pace in Education 5.0, the artificial intelligence, machine learning and data analytics are being integrated to innovate how to deliver & experience the education. The Learning

Analytics Dashboard (LAD) is a common, powerful tool being applied to collect rich data in many educational settings[1]. LADs provide real-time insights into student performance, engagement, and learning patterns, enabling educators to make data-driven decisions that enhance the overall educational experience[2]. As Education 5.0 emphasizes personalized learning and the seamless integration of technology into education, LADs play a crucial role in supporting these goals by offering a window into the learning process that was previously unimaginable[3].

Learning Analytics Dashboards are beyond merely data visualization tools, in spite they are dynamic systems that analyze, interpret, and present educational data in ways that are actionable and accessible to educators and students alike[4]. By leveraging AI and ML, LADs can identify trends, predict outcomes, and even suggest interventions to improve student learning. This capability aligns perfectly with the goals of Education 5.0,

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

which seeks to create a more personalized, adaptive, and student-centered learning environment[5]. With LADs, educators can tailor instruction to meet the unique needs of each student, monitor progress in real-time, and adjust teaching strategies accordingly. This not only enhances learning outcomes but also fosters a more engaging and responsive educational experience[6].

Despite their benefits, the implementation of LADs in Education 5.0 comes with its own set of challenges such as data privacy, the potential for data misinterpretation, and the digital divide must be carefully considered to ensure that the benefits of these tools are fully realized. Despite these challenges, the potential of LADs to revolutionize education is immense. As schools and institutions continue to embrace Education 5.0, the role of LADs in shaping the future of learning becomes increasingly important, offering a pathway to more effective, equitable, and innovative educational practices. This work is indented to delve within the existing implementations of AI and ML towards the construction of an ideal LAD framework.

2. Existing methods

A set of most recent endeavors towards application of AI and ML in LADs are chosen here to study the methodologies used, advantages and the limitations in detail. These works are selected based on the performance and their preamble period.

2.1. Examining university teachers' self-regulation in using a learning analytics dashboard for online collaboration (SRLLAD)

In 2023, Lingyun Huang et.al., presented SRLLAD[8] work to probe about how self-regulated learning (SRL) impacts teachers' use of Learning Analytics Dashboards (LADs) for data-driven instruction. It found that teachers initially engaged in cognitive regulation but increasingly adopted metacognitive regulation over time. Effective self-regulators, identified through a case study with ten participants using LADs for asynchronous online problem-based learning, displayed more frequent monitoring and evaluation, resulting in a more dynamic regulatory approach. Teachers with strong SRL were better at diagnosing issues in student learning and collaboration, highlighting the critical role of SRL in optimizing the use of LADs for educational purposes.
Two recruitment methods were used for the study. Initially, information about the study was distributed via email to members of the problem-based learning (PBL) special interest group (SIG) within the American Educational Research Association (AERA), targeting university professors and teachers involved in PBL research or implementation. Additionally, doctoral students from American universities with PBL knowledge or experience were also invited. As a result, ten participants were recruited from public universities in East Asia and North America. This group included four university faculty members, two university teaching fellows, one research fellow, and three doctoral students. Their experience with PBL varied: one had over ten years of experience, four had between 2–5 years, three had less than one year, and two did not report their experience. Their online teaching experience ranged from under one year to over ten years, with several required details. The SRLLAD work introduces a new LAD named as HOWARD.

HOWARD is a Learning Analytics Dashboard(LAD) featuring four visualizations: Conversation Explorer, Social Network View (SNA), Task Progress View, and Activity View. The Conversation Explorer tracks students' participation and group interactions over time, providing insights into conversation development and highlighting frequently used words. The SNA visualization uses color-coded nodes and lines to represent the volume and flow of student interactions, showing how students engage with each other and

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

identifying dominant discussion patterns. The Task Progress View uses grids to show task completion status, with completed tasks highlighted in green. The Activity View, presented as a line chart, tracks the frequency of chat posts or total word counts over time, helping to gauge student engagement and flag potential issues with participation.
These visualizations collectively illustrate student learning and interaction patterns. HOWARD was used to analyze discussion data from five student groups: one group showed strong engagement and knowledge-building, another had one student dominating discussions, a third group exhibited limited interaction with parallel contributions, and a fourth group experienced social loafing, where some members reduced their participation. The study was conducted in an online setting using ZOOM for communication between researchers and participants, including audio and video recording of sessions. Prior to the study, a researcher sent each participant an invitation letter, study introduction, HOWARD tutorial video, and a pre-study survey. During the implementation phase, two researchers conducted virtual meetings with each participant, reintroducing the study's purpose, addressing any questions about the HOWARD platform, and guiding the participant through the process. Participants navigated the platform while verbalizing their thoughts, with their screen and voice recorded.

The LAD task was divided into two sessions. In the first 30 minutes (Session 1), participants logged into HOWARD to explore student discussions for Part 1. They were given a 5-minute break to prevent fatigue before continuing with Part 2 (Session 2), where they reviewed students' collaboration on the dashboard. Participants were reminded when a session was about to end. After completing the sessions, a post-study survey was emailed to participants, who were requested to return it promptly. SRLLAD work employed two process mining algorithms to analyze teachers' self-regulated learning processes. The conformance checker compared the alignment between a Petri Net model of SRL and actual data, while the fuzzy miner visualized event clusters and their relationships. This mixed-method approach combined transcript coding, process mining, and content analysis to address research questions through both quantitative and qualitative insights.

Innovative methodology, Insights into SRL, and implications for practice are the observed advantages of SRLLAD work, whereas, Context dependency, inherited properties of think-aloud protocols, and Higher time consumption are the noted limitations of SRLLAD work.

2.2. EX-LAD: an Explainable Learning Analytics Dashboard in Higher Education (EX-LAD)

Tesnim Khelifi et.al., introduced (EX-LAD) [9] work to clearly and intuitively display data on student performance, engagement, and perseverance. The dashboard aims to make this information accessible to both teachers and students, even those without advanced data analysis skills. It helps teachers monitor student progress, identify at-risk learners, and offer targeted support, while also allowing students to track their own learning, pinpoint strengths and weaknesses, and make informed decisions to enhance their academic performance. Featuring visualizations of key learning aspects, the dashboard's effectiveness was demonstrated through a case study conducted with real data from ESIEE-IT, an engineering school in France, during the 2021-2022 academic year, showcasing its impact and benefits in an educational setting. The authors conducted a case study using real data from the Learning Management System (LMS) of ESIEE-IT, an IT school in France. ESIEE-IT offers various computer science programs, including artificial intelligence, cybersecurity, and

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

information systems, catering to bachelor, engineering, and master's students. The study involved 128 students enrolled in a Python programming course, comprising 22 in Master Green, 48 in an engineering program, 29 in BTS, and 29 in a Master's program in Big Data. The sample included 117 male and 11 female students. Data was collected during the 2021-2022 academic year, with measures taken to anonymize the data to comply with ethical standards.

The Python programming course is delivered in a hybrid format, with 80% of the instruction occurring online and 20% in-person. Online lessons are accessed through the school's LMS, Blackboard Learn, while in-person sessions require students to be physically present to engage with teachers and ask questions. The Blackboard course consists of various sequences, each including: (a) video lectures, (b) supplementary notes in multiple formats, (c) instructional documents and exercise corrections, and (d) quizzes with 5 to 10 questions and a final test of 20 questions. Student interactions with Blackboard Learn, such as clicks, time spent, and platform access, are recorded in the Snowflake data warehouse. The next section will detail the steps of the dashboard developed for this study. As the first step, the authors gathered digital learning traces from the SnowFlake data warehouse, encompassing 128 instances and 106 features related to students. The dataset is organized into several sections: personal data including names, email addresses, and courses of study; platform access metrics such as connection and time spent; academic performance indicators including grades, ranks, submissions, and average scores; and engagement metrics such as interaction-oriented investment and access-oriented investment. Additionally, the dataset includes a feature for difficulty, categorizing students into four profiles based on the issues they face.

In the preprocessing stage, the raw data was prepared for analysis and visualization by cleaning and organizing it. The data was collected from various tables and combined into a single dataset, incorrect and mislabeled entries were addressed. Incomplete and duplicate records were removed to avoid inaccuracies and misleading results. NaN (Not a Numeric) and NaT (Not a Time) values were replaced with "0" to ensure compatibility with numerical calculations. Additionally, the data was anonymized to comply with General Data Protection Regulation (GDPR) requirements, with identifying information such as email addresses and names removed. The EX-LAD dashboard features various visualizations, including raw data, bar charts, line graphs, and scatter diagrams, to present indicators clearly and offer insights into student performance, engagement, and profile evolution. Designed with user-friendliness in mind, the visualizations are easy to interpret, require no data science knowledge, and include text descriptions and color coding to enhance clarity and usability for all stakeholders.
Clarity, Transparency, Ethical considerations, User friendliness and Support are the primary advantages achieved by EX-LAD work. Limited engagement indicators, Individual performance indication deficiency, and increased computational complexity are the identified limitations of EX-LAD work.

2.3. Developing a multimodal classroom engagement analysis dashboard for higher-education students (DMCEA)

Alpay Sabuncuoglu et.al., proposed DMCEA[10] work in 2023, to explore the development of learning analytics dashboards in response to the growing accessibility of online learning tools in K-12 and higher education. It details the design and analysis process of a multimodal engagement dashboard, focusing on students' physical and cognitive involvement in learning. The project involved creating a machine learning model using deep learning features related to

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

face and pose and designing a dashboard to track engagement and identify learning patterns. User studies with undergraduate and graduate students provided feedback on the dashboard's design. The paper contributes by presenting a student-centric, open-source dashboard, establishing a baseline architecture for engagement analysis, and offering design insights for future LAD development. The research aims to support novice teacher education, student self-evaluation, and engagement assessment in diverse classroom settings.

The Classroom Engagement Dataset consists of 1280x720 resolution JPEG frames capturing group views, with individual faces cropped to 320x320 resolution using dlib's feature extractor in DMCEA work. The dataset includes self-evaluation engagement scores, audio recordings in WAV format, and transcripts in SRT subtitle format. For the first learning session, 6969 frames were initially collected, with 5381 frames remaining after removing low-confidence ones for training the engagement level classifier. The second session provided 4972 frames, and after filtering out low-confidence frames, 4855 frames were used for classifier training. DMCEA work employed OpenFace's Face Feature Extractor, generating 1562 features per vector, including AU Intensities, 3D Eye Landmarks, 3D Face Landmarks, Gaze Directions, and Head Pose, with each frame providing a feature vector at a rate of one frame per second. These features, such as eye gaze and head position, are valuable for learning analytics. OpenPose was used to extract pose features from group videos, leveraging its ability to encode global context and part affinity fields to achieve accurate results. The extracted pose features for each participant included twenty-five key points in 3D locations, resulting in a total of seventy-five points

Different machine learning models were tested in DMCEA work to provide baseline scores on the engagement dashboard. The evaluation involved logistic regression (LR), k-Nearest-Neighbor (kNN), Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, random forest (RF), and boosted trees (BT) for classifying five levels of engagement, using OpenFace and OpenPose for feature extraction. Scikit-learn was used for training the models, and performance was assessed using the weighted F1-score, which balances precision and recall, particularly useful for imbalanced data. The average F1-scores for the models were: LR at 0.516, kNN at 0.684, RBF-SVM at 0.354, RF at 0.445, and BT at 0.606. The kNN Classifier and Regressor were chosen as the baseline model for the study, providing a simple and effective method for understanding engagement patterns, with the data pipeline detailed in the open-source repository.

The impact of individual features and their combinations on classifiers was examined in DMCEA work using InterpretML, an open-source Python library with LIME for explaining model behavior. Due to limitations in InterpretML for multi-class classification, interpretable rules were produced for binary predictions by running algorithms on a binarized version of the data. The significant rules affecting model decisions were shared on the dashboard. This process aimed to enhance the interpretability of the model's decisions.

Customizable dashboard, Data pipeline accessibility, and improved data insights are the stated advantages of DMCEA work, whereas, vulnerable transparency, lack of numeric based visualizations, and Sensitivity dependence on individual performance are the observed limitations of DMCEA work.

2.4. Learning analytics and the Universal design for learning (UDL): A clustering approach

UDL work[11] is introduced by Marvin Roski in 2024 to make learning accessible for all

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

students. However, research on how students interact with UDL-guided elements in digital environments is limited. This study analyzes the usage patterns of 384 9th and 10th graders on the I3Learn platform, which uses UDL principles to teach chemistry. By examining detailed log files, including time spent on videos, texts, tasks with or without assistance, and self-assessments, the study employed Exploratory Factor Analysis (EFA) to identify key usage behaviors. The results revealed four main factors influencing student engagement, leading to six distinct clusters of usage patterns, which can guide the development of personalized learning support and inform educators through an analytics dashboard.

UDL uses a methodological approach involves four key steps such as gathering data, preparing the data, performing exploratory factor analysis, and applying k-means clustering. Data collection took place on the I3Learn learning platform, involving 580 learners from 27 classes who generated nearly 500,000 log files during the first half of 2022. These logs, detailing learners' time-accurate behavior, were supplemented by pre- and post-test data on conceptual knowledge about ions and their bonding, assessed using an adapted Bonding Representations Inventory with a reliability score of 0.87. Additional data collected included information on chemical self-concept, socio-economic background, reading and cognitive abilities, language spoken at home, gender, school type, and grade. The data collection process was approved by the Ministry of Science and Culture, Lower Saxony, Germany.

To address the research questions, the duration of usage for UDL-guided elements—texts, videos, tasks with or without assistance, and self-assessments—was extracted from log files. Despite additional UDL-guided elements being integrated into I3Learn, only these three were used in the analysis due to their measurability and the decision-making process they require. From the dataset, 23,072 log files were extracted, with data points replaced by zero for non-usage, reflecting actual time spent on these elements. Usage duration was calculated from timestamps of learner actions, with a 30-minute cut-off for videos and texts and a 5-minute cut-off for self-assessments and tasks to manage interruptions and ensure adequate engagement. A minimum usage time of 30 seconds was set for videos and texts to ensure meaningful interaction, while there was no minimum threshold for self-assessments and tasks. Specific tasks, such as "Ratio Formula 1" and "ion grid," were not included in the “tasks” feature category if they required help or were only available with assistance, although associated self-assessments were classified as UDL-guided elements. This approach ensured accurate representation and analysis of learner engagement with the UDL-guided elements.

Exploratory Factor Analysis (EFA) was employed in UDL work to uncover the underlying structure of student usage data, simplifying complexity for subsequent cluster analysis and enhancing interpretability. Factors, representing latent variables, correlate with observed features like usage time, while factor loadings measure these correlations. The Bartlett’s sphericity test and Kaiser-Meyer-Olkin (KMO) coefficient assess data suitability for EFA, with a significant Bartlett’s test and a KMO value above 0.6 indicating adequacy. EFA calculates eigenvalues for each variable and determines the number of factors using a Scree plot, followed by factor rotation with the “Varimax” method for clearer interpretation. A cut-off value of 0.4 for factor loadings was used to ensure meaningful factors and minimize noise, with the analysis performed using the Python module factor analyzer. K-means clustering was utilized to explore usage behaviors of UDL-guided elements, with the number of clusters determined by metrics such as the silhouette score, Bayesian Information

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

Criterion (BIC), and Akaike Information Criterion (AIC). The algorithm assigns data points to centroids based on similarity using Euclidean distance, updating centroids until assignments stabilize. Features for clustering were derived from four factors identified in EFA, with each learner's weighted sum score used as a feature for k-means. K-means was chosen for its simplicity, efficiency, and interpretability, with data scaling and clustering performed using the Python library scikit-learn.

Usage pattern identification, Educator insight improvements, and Self-Regulated Learning support are the observed advantage of UDL work, at the same time, Lack of outcome measurement, Local clustering bias dependency, and proneness to inaccurate data are the drawbacks of UDL work.

2.5. Cluster-based performance of student dropout prediction as a solution for large scale models in a moodle LMS (CPSDPS)

In 2023, Louis-Vincent Poellhuber et.al., introduced CPSDPS[12] work to address challenges in student engagement, success, and retention, learning analytics are employed to develop predictive models and dashboards. However, managing large volumes of data is complex and requires significant expertise. This paper explores improving student dropout prediction across a large set of courses by clustering courses based on design and similarity, then applying machine learning algorithms to each cluster. The proposed methodology provides a foundational framework that can be adapted and refined for future research. The research team pinpointed eight tables with relevant data related to behavioral engagement. These tables include details about the course structure, evaluation criteria and weights, user roles within the course, and records of each user's actions, as documented in the log file.

The model architecture begins with two datasets: the student behavior dataset, which captures students' actions in a course, and the clustering dataset, which outlines the course's composition. Both datasets are derived from raw Moodle files. The clustering dataset is used to select and evaluate a clustering algorithm, with the optimal one generating course clusters. Various classification algorithms are trained and tested on each course cluster using cross-validation, with the best algorithm chosen based on binary cross-entropy (log loss). The final step involves testing the selected algorithm for each cluster by splitting the dataset into training and testing sets and manually evaluating the results. This process ensures that the models are robust and accurately reflect the specific characteristics of each course cluster. Additionally, performance metrics are analyzed to assess the effectiveness of the chosen algorithms in predicting student outcomes within each cluster.

In CPSDPS work, the clustering dataset was pre-processed by counting the number of different modules in each course, including forums, quizzes, resources, URLs, and assignments, and by including additional variables such as the course year, number of students, and dropout proportion. Multiple clustering algorithms were trained on this dataset using Scikit-Learn, including agglomerative clustering, OPTICS, k-means, mean shift, and Gaussian mixture. Algorithms requiring a predetermined number of clusters were evaluated with default hyperparameters, except for the number of clusters. Performance was assessed using the silhouette score, the Calinski-Harabasz score (variance ratio criterion), and the cluster length standard deviation to ensure clusters were neither too large nor too small. The selected clustering model's predictions defined the clusters used in the subsequent classification model.

The CPSDPS classification model selection involved combining datasets from each course within clusters to create clustered student behavior datasets. Several

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

classification algorithms—logistic regression, random forest, gradient boosting, extreme gradient boosting, and adaptive boosting—were trained and tested using default hyperparameters. Due to potential small or sparse clusters, tenfold cross-validation was employed to enhance model selection, with a total of 210 models trained and tested across various clusters. Model performance was assessed using log loss, or cross-entropy, and a final set of 42 models was retained, one for each cluster. This final step involved re-training the best models identified during the selection phase. Framework versatility, Effective clustering, Extensive improvements are the advantages are the CPSDPS methodology. Insufficient accuracy, Uncertain impact, Data tracking issues and Threshold fine-tuning dependency are the limitations of CPSDPS work

2.6. Content-Focused Formative Feedback Combining Achievement, Qualitative and Learning Analytics Data (CFFCAQ)

Cecilia Martinez et.al. established CFFCAQ [13] work in 2023, to explore how content-focused feedback impacts student achievement in higher education, this study presents an empirical case of learning analytics (LA)-based feedback. The model integrates quantitative achievement indicators, such as pretest results, practice exercise participation, and exam grades, with qualitative data from in-depth interviews with students of varying performance levels to identify learning challenges. Feedback, tailored to student performance, is provided every two weeks, resulting in statistically significant improvements in final grades and increased participation in problem-solving among those who received feedback compared to those who opted out. The contributions of this study to LA-based formative feedback include: (a) a model that combines quantitative and qualitative data to address and understand student achievement challenges, (b) templates for designing effective pedagogical and research-based formative feedback, (c) positive outcomes evidenced by both quantitative and qualitative data, and (d) a detailed account of the practical implementation process.

Design-based research is a suitable methodology for designing and analyzing educational interventions within specific contexts. It aligns with action research principles, emphasizing reflection on problems while also exploring new theoretical frameworks and variable relationships. In this approach, a team of professors at TU/e in the Netherlands assessed student performance and teaching practices in their Electromagnetics II course to develop and test an intervention. The process involved continuously analyzing the educational problem, crafting a theoretically grounded solution, implementing the intervention, and documenting the results. The LMS data includes student attempts to solve exercises, navigation history on digital resources, and personal data, which the team could not access due to privacy regulations. Analyzing correlations between student activities and assessments was essential for developing a prognosis system to identify at-risk students and offer targeted feedback. The ethical review board at TU/e approved the use of student data, ensuring compliance with GDPR through a data protection impact assessment and anonymizing data to prevent individual identification. Students gave explicit consent for the use of their data, with options to limit its use or permit it for scientific research, and could withdraw their participation at any time. Data collection and analysis were carried out in phases following a design-based research approach.

The first phase of the study aimed to understand learning challenges through learning analytics (LA) and interviews, using data from 675 students who had enrolled in the EMII course over three years. Quantitative data included pretest results, practice exercise participation, midterm and final exam grades, and overall LMS participation. The pretest utilized items from the Conceptual Survey of Electricity and

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

Magnetism (CSEM) to assess prior knowledge. Analysis revealed that the number of exercises presented during practice hours had the highest correlation with final exam grades, followed by pretest results. Qualitative data were gathered through six confidential interviews with students and a focus group interview in 2023, capturing diverse perspectives from students with varying performance levels. Data coding and analysis were performed using the Saturate App, which facilitated the identification of patterns and themes.

In the second phase of CFFCAQ, data were collected from 366 students enrolled in the EMII course during 2022 and 2023, with 55% and 56% of students from each year respectively consenting to have their LMS data analyzed for feedback. A quasi-experimental design was used to compare final grades and SLT participation between students who received feedback and those who did not, with SLT participation being a significant achievement indicator. Additionally, a survey with 11 responses and a focus group with six students were conducted to assess experiences with the APSPMS project.

Data source integration facility, Adaptiveness, Timeliness, Positive performance impact are the stated advantages of CFFCAQ method. Data privacy constrains, Self-Selection bias, and Constrained Feedback personalization are noted as the limitations of CFFCAQ work.

2.7. Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT (BPLAMXAI)

Teo Susnjak et.al., introduced BPLAMXAI[14] work in 2023 for the purpose of enhancing the effectiveness of Learning Analytics, recent research has increasingly utilized machine learning to predict at-risk students, aiming to initiate timely interventions and improve retention and completion rates. However, most studies have focused primarily on predictive accuracy without adequately addressing the interpretability of these models or the communication of their predictions to stakeholders. To address this gap, the field of eXplainable AI has emerged, offering tools for transparent predictive analytics and tailored prescriptive advice. This study proposes a novel framework that integrates transparent machine learning with prescriptive analytics and leverages advanced large language models, such as ChatGPT, for generating personalized feedback. The framework is demonstrated using a real-world dataset of approximately 7000 learners from 2018 to 2022, showcasing its application in identifying learners at risk of non-completion. Additionally, the study includes two case studies illustrating how predictive modeling can be combined with prescriptive analytics to produce clear, actionable feedback for at-risk students.

BPLAMXAI introduces an innovative prescriptive analytics framework designed to enhance Learning Analytics by identifying at-risk students and initiating timely and effective interventions. The framework demonstrates a more comprehensive use of both predictive and prescriptive analytics than previously achieved. It shows how machine learning models can be developed to predict qualification completion outcomes, with transparency at both global and individual levels to address stakeholder needs. Additionally, the study provides case examples of how prescriptive analytics tools can be employed to automatically generate specific, evidence-based feedback, which is then translated into natural language using ChatGPT to offer actionable suggestions. The Prescriptive Learning Analytics Framework (PLAF) of BPLAMXAI is presented, featuring two main components: the predictive and prescriptive phases. It assumes that data identification and acquisition have been completed before the framework is applied. The first step involves cleaning and preprocessing the raw data to

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

prepare it for further analysis. This step includes performing exploratory data analyses and evaluating data reliability, such as checking for and imputing missing values. In addition, PLAF assesses whether the data is suitable for deeper analyses in subsequent phases.

The data for the study was sourced from an Australasian higher education institution, including information from both the Student Management System (SMS) and Moodle, the Virtual Learning Environment (VLE). The dataset included undergraduate students who started their studies between 2018 and 2022 and tracked their outcomes of either completing or abandoning their studies. The dataset was evenly balanced between the two outcome categories, with 52% (3,693) of students completing their studies and 48% (3,415) abandoning them. The target variable for prediction was program or qualification completion. The dataset captured students' progress as snapshots across each academic year, resulting in multiple data points for students enrolled over several years. For instance, a student completing a three-year bachelor's program would have three data points, each labeled as 'completed.' The dataset contained a total of 14,918 data points, with 72% (10,736) representing students who completed their program and 28% (4,182) representing those who did not, resulting in a relatively unbalanced dataset. The predictive problem was both formative, evaluating student outcomes at various points during their studies, and summative, predicting outcomes at the end of the program or semester using course-level models.

Enhanced transparency, Comprehensive approach, and Human interpretable feedback are the major advantages identified in BPLAMXAI method. Implementation complexity, AI model dependency, and Inferior impact on high performance students are the notable limitations of BPLAMXAI work.

2.8. Utilizing random forest algorithm for early detection of academic underperformance in open learning environments (RFEDA)

In 2023, Balabied SAA et.al., introduced RFEDA[15] work to address the challenges faced by Open Learning Environments (OLEs), such as high student failure rates and inadequate support, this study focuses on early prediction of at-risk students. Machine learning algorithms, including decision trees, random forests, and neural networks, are employed to build predictive models from large datasets. These models analyze learner behavior data, such as time spent on tasks and performance on quizzes, to identify patterns and predict future performance. Key factors in evaluating these models include data quality, model accuracy, and predictor relevance. The study utilizes the random forest algorithm on behavioral data from the OULAD dataset, sourced from a Moodle-like educational system at the Open University. This approach aims to deliver timely interventions to support at-risk students and improve their academic success.

In RFEDA work, the OULAD dataset from the Open University is used which includes information on student demographics, course details, and interactions with the virtual learning environment (VLE). It encompasses data from seven modules, with courses offered in February and October, and notes that February semesters are 20 days shorter. The dataset, created from 22 modules, includes demographic details for 32,593 students and records over 10 million entries of clickstream data. The dataset's focus is on student data, detailing interactions with VLE resources, course assessments, and student performance. It also contains specific information on course assessments, including their weight and timing within the module. Missing values in the assessment date column are filled with the mean number of days, while missing values in the Student Assessment data, being relatively few, are deleted. In the Student information data,

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

missing values in the imd_band column, a categorical variable, are filled with the mode. Additionally, columns with nearly 80% missing values in the VLE data, specifically week from and week to, are removed in the RFEDA preprocessing state.

RFEDA data merging involves combining various tables to analyze interactions and correlations within the dataset. The student VLE table is merged with the VLE table to understand student interactions with the virtual learning environment. The Student Registration table is merged with the courses table to explore the relationship between course registrations and durations. Additionally, the Student Info table is merged with the Student Registration table, and the Assessments table is merged with the Student Assessment table to examine correlations between performance and assessments. Finally, VLE data is merged with Student Info data for a comprehensive analysis.

Prediction efficiency and the handling capacity of large datasets are the main advantages of RFEDA work. Single Algorithm focus causes sometimes causes misleading predictions – which is identified as the limitation of the RFEDA work.

2.9. Building Learning Analysis System with GQM Methodology and ELK Stack (GQMELK)

Nien-Lin Hsueh et.al., proposed GQMELK work [16] in 2023 for the purpose of addressing the challenge of extracting valuable information from the vast and complex log data generated by online learning platforms, this research proposes a structured data analysis process. The methodology is guided by the GQM (Goal Question Metric) method, which is combined with the Banerjee analysis model to create a series of questions and metrics that evaluate students' learning behavior. The ELK Stack (Elasticsearch, Logstash, Kibana) is used as the analysis environment to efficiently process the data. This approach is applied in a case study of a programming course on the OpenEdu e-learning platform. The analysis helps educators transform log data into actionable insights, enabling them to understand students' learning behaviors better. This ultimately supports the development of effective strategies for improving learning outcomes.

The GQMELK architecture consists of four main components: the OpenEdu MOOC system, the ELK Stack system, external learning systems, and external analysis modules. The data flow between these modules is indicated by arrows. The OpenEdu platform generates extensive log data reflecting student learning behaviors, which is collected by Filebeat within the ELK Stack system. The data is then transferred to Logstash for parsing and filtering before being stored in the Elasticsearch repository. Logstash also extracts diverse data from external systems like Online Judge and Star Trek. Elasticsearch integrates, counts, and processes this data, displaying the analysis results visually on the Kibana dashboard. For more complex computations and specialized analyses that exceed Elasticsearch's capabilities, an advanced analysis module is used in an external environment.

To facilitate deployment across various machines, the ELK Stack system is built using Docker Compose. Initially, the Linux host's “vm.max_map_count” parameter must be set to 262,144 to ensure sufficient virtual memory for the system. The setup includes four systems: Filebeat, Logstash, Elasticsearch, and Kibana. The Elasticsearch cluster is composed of one primary node and two data nodes, with the JVM heap sizes configured at 1G for the primary node and 2G for the data nodes, which can be adjusted based on the machine's memory capacity. It is important to balance the JVM heap allocation to avoid prolonged garbage collection times. Filebeat's output and Logstash's batch transfer settings are fine-

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

tuned to ensure smooth data transfer. Filebeat collects log files and sends them to Logstash for parsing, where YAML files dictate the parsing rules, extracting relevant data from the disorganized logs. This parsed data is then stored in the Elasticsearch repository for further analysis. The extracted event logs, detailed in Table 1, include information on videos, quizzes, and forums, providing insights into student activities within the online course.

Elasticsearch is structured as a cluster with nodes that play different roles, including essential master and data nodes. The master node manages cluster metadata, while data nodes handle storage and data-related operations like searching and aggregating. To ensure high availability, each node carries both its master and replica slices, allowing the system to maintain functionality even if a node fails. Data in Elasticsearch is stored in JSON format as indices, which can be monitored, managed, and organized using Kibana. The system uses a three-layered data warehouse architecture as Operational Data Store (ODS), Data Warehouse Detail (DWD), and Application Data Store (ADS) to handle varying levels of data, enabling advanced analysis and visualization. Kibana uses a simple query language (KQL) to search data stored in Elasticsearch. It allows users to find specific information by searching keywords in fields and then visualize the results using various charts like pie charts, histograms, and line charts. Kibana has an easy-to-use interface, which makes it simpler to work with Elasticsearch data without needing to write complex commands. This tool helps users quickly see important data on a dashboard and identify any issues. The visual templates provided make understanding and analyzing data much faster and easier which are utilized in GQMELK work.

User-oriented strategy, Data collection efficiency, and integration flexibility are the observed advantages of GQMELK work, whereas, ELK stack dependency restricts the application area which is noted as the limitation.

2.10. Analysis of student's academic performance based on their time spent on extra-curricular activities using machine learning techniques (ASAP-MLT)

In 2022, Neeta Sharma et.al., debuted ASAP-MLT [17] work for analyzing students' academic performance in relation to time spent on extracurricular activities, this research employs Decision Tree, Random Forest, and KNN machine learning algorithms. For understanding the shortcomings of each technique, comparisons of prediction outcomes are conducted. For the dataset used, the Decision Tree algorithm achieved the highest performance, with an F1 score of 84 and an accuracy of 85%. For future research, this study serves as an introductory exploration. For more complex and specialized analyses, this research aims to provide a foundation in predicting academic performance.

For ASAP-MLT methodology, the authors gather and integrate student data, convert and normalize it in a .csv file, extract patterns using classifier methods, and compare results. For achieving the goals, the work forecasts academic success and selects the best algorithm for predicting student performance. For processing, the team converts the dataset to numerical values, scales features, and completes normalization. For model assessment, they use a 70:30 ratio between training and test datasets and evaluate models using the Precision matrix, RoC curve, and Confusion matrix. After thorough investigation, it was found that few universities in India track students' extracurricular activities, limiting the availability of large datasets. To gather information, questionnaires were used, with surveys administered physically to 500 students. Of the 415 responses received, 395 were valid and utilized for further research.

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

The dataset undergoes several key processing steps such as, cleaning removes instances and irrelevant features not pertinent to the study. Next, text values are converted to numeric or binary formats for model training. Features are then scaled to improve classification outcomes, especially for large datasets. Finally, the data is normalized, ensuring it is cleaned, consistent, and ready for use by the training model. Random Forest, KNN, and Decision Tree methods are used as the core components in ASAP-MLT work.

The combination of Decision Tree with Gini Index produces accurate classifications for specific metrics – which is the advantage of ASAP-MLT method. However, Limited model scope, and Mono Evaluation metric focus affects the prediction specificity – which is identified as the limitation of ASAP-MLT work.

2.11. Design of student success prediction application in online learning using Fuzzy-KNN (FKNN)

In 2023, Kharis, S et.al., submitted FKNN [18] work to effectively evaluate student performance, a range of techniques, including statistics, physical examinations, and data mining, are utilized. Data mining, known as Educational Data Mining (EDM), helps in uncovering hidden patterns within student datasets by collecting, processing, and reporting data. EDM employs machine learning to extract valuable insights from various data sources, such as intelligent computer tutors, online classes, academic data, and assessments. The adoption of open and distance learning (ODL) by universities has led to the accumulation of extensive student and learning data in Learning Management Systems (LMS). Students' interactions with LMS, including material access and forum participation, provide valuable data for EDM. This study focuses on designing applications that use machine learning techniques and EDM to predict student performance in online learning environments.

The FKNN predictive model is developed using Fuzzy-KNN. This classifier assigns a membership value to each unlabeled signature to determine how closely the observations in the test data align with a particular classification. The model's accuracy is evaluated using test data, and if it fails to deliver optimal classification results for student graduation, adjustments are made by reconfiguring the training data composition. This process of refinement continues until the model achieves the best possible results. The model is simulated using Python, and a dashboard interface named LeADS (Learning Analytics Dashboard System) is created to predict student graduation rates for courses.

Predictive accuracy, and Comprehensive dashboard are the advantages of FKNN work, whereas, Data dependency and increased computational complexity are observed as the limitations.

2.12. Bayesian Model for Academic Performance Prediction in Learning Analytics (BMAP)

In 2024, Siti Salwa Salleh et.al., proposed BMAP [19] work to train the Bayesian prediction model, the collected data was used, resulting in an accuracy rate of 81%. The F1 scores ranged from 0.74 to 0.98, reflecting moderate to excellent performance. To identify risks of misclassification accurately, a sensitivity score of 0.8 was achieved. This performance level is deemed satisfactory within the study's scope and compares favorably to manual calculations. n the BMAP work, probabilities represent all unknowns about the input and output parameters. The model's posterior predictive distribution simulates outcome values and represents the distribution of future, unseen data based on observed data. This distribution serves as a highly accurate predictor for forecasting future trends. The approach allows for the seamless integration of prior knowledge and data into the prediction framework. Additionally, the

model provides precise inferences based on the available information.

The first phase involves knowledge acquisition, including problem formulation, LAD requirements, and construct development. The second phase focuses on data collection, which includes instrument development with constructs derived from Pintrich et al. (1993), expert validation, and a pilot test. Independent attributes, such as sociodemographic, motivation, cognitive strategy use, and personality data, were gathered through questionnaires. Academic results, attendance, and continuous assessments like quizzes, assignments, tests, and project progress were represented as synthetic data. In Phase 3, data treatment, cleansing, and exploratory data analysis (EDA) were performed to identify significant patterns, relationships, and anomalies, with outliers replaced by approximate values. Phase 4 involved designing and developing the prediction model and dashboard, where the model was iteratively trained for improved precision with weighted elements added. In Phase 5, the prediction model was evaluated using a confusion table, followed by simulations of the entire process.

Based on interviews with three academics in Phase 1, a set of requirements for the Learning Analytics Dashboard (LAD) was established, consisting of five basic features and access for two user types. The features include linking, updating, and viewing socio-demographic, profile, behavior, cognitive skills, and personality data. It also allows updating ongoing assessment data, such as quizzes, assignments, projects, and tests, and viewing grades, interventions, and various analytical outcomes. The prediction outcomes lead to recommendations for interventions, monitoring, disciplinary actions, or escalation to student affairs. These recommendations are based on a formative procedure calculated by target class prediction and iteratively tested score indicators.

Microsoft Power BI was used to create the BMAP dashboard, while Microsoft Excel and Data Analysis Expressions (DAX) were utilized to calculate the Bayesian prediction probability values. The Learning Analytics Dashboard consists of three main pages: (a) an individual academics page displaying scales of behavior, cognitive ability, and personality traits, along with suggestions for intervention if needed; (b) another individual academics

page showing grade predictions, average scales for behavior, cognitive ability, and personality traits, and an average score in a Bubble chart. Additionally, the page includes student feedback on their weekly performance.

Inherited Bayesian model strengths, and Interactive visuals are the major advantages of BMAP work. Dependency over predictive indicators affects the overall precision scores - which is identified as the limitation of BMAP work

Key points about the discussed methods are listed in the table – given below

S. No	Title	Year	Author	Finding	Limitations
1	Examining university teachers' self-regulation in using a learning analytics dashboard for online collaboration	2023	Huang, L., Zheng, J., Lajoie, S.P. et al.	Self-Regulated Learning (SRL) improves Learning Analytics	Higher time consumption of SRL

				Dashboards (LAD)	
2	EX-LAD: an Explainable Learning Analytics Dashboard in Higher Education	2024	Tesim Khelifi, Nourhène Ben Raba, Bénédicte Le Grand, Ibtissem Daoudi	Improvements in performance, engagement, and perseverance using EX-LAD	Complexity in building interpretations of LAD
3	Developing a multimodal classroom engagement analysis dashboard for higher-education	2023	Alpay Sabuncuoglu and T. Metin Sezgin	Multi-model Class room Engagement improves Student centric performance architecture baseline	Sensitivity towards individual performance refracts global performance
4	Learning analytics and the Universal design for learning (UDL): A clustering approach	2023	Marvin Roski, Ratan Sebastian, Ralph Ewerth, Anett Hoppe, Andreas Nehring	Exploratory Factor Analysis based K-Means Clustering enhances LAD	Four factor suggestion making system diminishes LAD accuracy
5	Cluster-based performance of student dropout prediction as a solution for large scale models in a moodle lms	2023	Louis-Vincent Poellhuber, Bruno Poellhuber et.al.	Supports Learning Management System Big data operability	Framework requires threshold fine tuning
6	Content-Focused Formative Feedback Combining Achievement, Qualitative and Learning Analytics Data	2023	Martinez C, Serra R, Sundaramoorthy P et.al.	Implementation of Quantitative achievement indicators improves LAD accuracy	Dependency on large LMS dataset
7	Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT	2023	Susnjak, T.	Transparent predictive analysis is performed using Explainable AI	Inferior impact over high performance students

8	Utilizing random forest algorithm for early detection of academic underperformance in open learning environments	2023	Balabied SAA, Eid HF.	Achievement of 90% accuracy in identifying risky students	Random Forest method sometimes produce misleading predictions
9	Building Learning Analysis System with GQM Methodology and ELK Stack	2023	Nien-Lin Hsueh, Jun-Jie Wang, Daramsene Bilegjargal	Goal Question Metric based LAD approach improves the reliability	ELK framework dependability decreases the application area
10	Analysis of student's academic performance based on their time spent on extra-curricular activities using machine learning techniques	2023	Sharma, N., Appukutti, S., Garg, U., Mukherjee, J., & Mishra, S.	Evaluation based on Multiple Machine Learning Models produces compelling predictions	Lacking of predictions based on popular ML models such as AVM and ANN
11	Design of student success prediction application in online learning using Fuzzy-KNN	2023	Kharis, S., Hertono, G., Wahyuningrum et.al.	Improved LAD accuracy of 92.5% by applying Fuzzy-KNN method	Increased computational complexity of Educational Data Mining Process (EDM)
12	Bayesian Model for Academic Performance Prediction in Learning Analytics	2024	Salleh, S. S., & Yassin, Y	Implementation of Bayesian Prediction model in LAD secured around 80%	Dependency over social context learning impacts the convergence

CONCLUSIONS

While significant advancements have been made in learning analytics and educational data mining, several critical gaps remain that need addressing. There is a pressing need for methodologies that enhance both the adaptability and transparency of predictive models, as well as those that effectively handle large and diverse datasets. Improving the accuracy and precision of predictions, along with better managing data privacy concerns, is essential for more reliable and actionable insights. Future research should focus on developing models that integrate seamlessly with various educational contexts and provide comprehensive, user-friendly dashboards for educators. Addressing these challenges will enable more effective

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

support for student success and more precise educational interventions. As the field progresses, continued innovation and refinement are crucial to fully realize the potential of learning analytics in enhancing educational outcomes.

REFERENCES

1 Ramaswami, G., Susnjak, T., Mathrani, A. et al. Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies. Tech Know Learn 28, 959–980 (2023). https://doi.org/10.1007/s10758-022-09613-x

2 Huang, L., Zheng, J., Lajoie, S.P. et al. Examining university teachers' self-regulation in using a learning analytics dashboard for online collaboration. Educ Inf Technol 29, 8523–8547 (2024). https://doi.org/10.1007/s10639-023-12131-7

3 Damien S. Fleur, Wouter van den Bos, Bert Bredeweg, Social comparison in learning analytics dashboard supporting motivation and academic achievement, Computers and Education Open, Volume 4, 2023, 100130, ISSN 2666-5573, https://doi.org/10.1016/j.caeo.2023.100130

4 Liu, Y., Huang, L. & Doleck, T. How teachers' self-regulation, emotions, perceptions, and experiences predict their capacities for learning analytics dashboard: A Bayesian approach. Educ Inf Technol 29, 10437–10472 (2024). https://doi.org/10.1007/s10639-023-12163-z

5 Riordan Alfredo, Vanessa Echeverria, Yueqiao Jin, Lixiang Yan, Zachari Swiecki, Dragan Gašević, Roberto Martinez-Maldonado, Human-centred learning analytics and AI in education: A systematic literature review, Computers and Education: Artificial Intelligence, Volume 6, 2024, 100215, ISSN 2666-920X, https://doi.org/10.1016/j.caeai.2024.100215

6 Hasnine MN, Nguyen HT, Tran TTT, Bui HTT, Akçapınar G, Ueda H. A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learners' Affective States. Sensors. 2023; 23(9):4243. https://doi.org/10.3390/s23094243

7 Cobos R. Self-Regulated Learning and Active Feedback of MOOC Learners Supported by the Intervention Strategy of a Learning Analytics System. Electronics. 2023; 12(15):3368. https://doi.org/10.3390/electronics12153368

8 Huang, L., Zheng, J., Lajoie, S.P. et al. Examining university teachers' self-regulation in using a learning analytics dashboard for online collaboration. Educ Inf Technol (2023). https://doi.org/10.1007/s10639-023-12131-7

9 Khelifi, T., Rabah, N. B., Le Grand, B., & Daoudi, I. (2024). EX-LAD: explainable learning analytics dashboard in higher education. In Proceedings of 36th International Conference on Computer Applications in Industry and Engineering (pp. 38-51).

10 Alpay Sabuncuoglu and T. Metin Sezgin. 2023. Developing a Multimodal Classroom Engagement Analysis Dashboard for Higher-Education. Proc. ACM Hum.-Comput. Interact. 7, EICS, Article 188 (June 2023), 23 pages. https://doi.org/10.1145/3593240

11 Marvin Roski, Ratan Sebastian, Ralph Ewerth, Anett Hoppe, Andreas Nehring, Learning analytics and the Universal Design for Learning (UDL): A clustering approach, Computers & Education, Volume 214, 2024, 105028, ISSN 0360-1315

ISSN: 2581 - 9380
TOJSR. 2025, 4 (1)

https://doi.org/10.1016/j.compedu.2024.105028

12 Louis-Vincent Poellhuber, Bruno Poellhuber, Michel Desmarais, Christian Leger, Normand Roy, and Mathieu Manh-Chien Vu. 2023. Cluster-Based Performance of Student Dropout Prediction as a Solution for Large Scale Models in a Moodle LMS. In LAK23: 13th International Learning Analytics and Knowledge Conference (LAK2023). Association for Computing Machinery, New York, NY, USA, 592–598. https://doi.org/10.1145/3576050.3576146

13 Martinez C, Serra R, Sundaramoorthy P, Booij T, Vertegaal C, Bounik Z, van Hastenberg K, Bentum M. Content-Focused Formative Feedback Combining Achievement, Qualitative and Learning Analytics Data. Education Sciences. 2023; 13(10):1014. https://doi.org/10.3390/educsci13101014

14 Susnjak, T. Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT. Int J Artif Intell Educ (2023). https://doi.org/10.1007/s40593-023-00336-3

15 Balabied SAA, Eid HF. 2023. Utilizing random forest algorithm for early detection of academic underperformance in open learning environments. PeerJ Computer Science 9:e1708 https://doi.org/10.7717/peerj-cs.1708

16 Nien-Lin Hsueh, Jun-Jie Wang, Daramsene Bilegjargal, "Building Learning Analysis System with GQM Methodology and ELK Stack," Journal of Internet Technology, vol. 24, no. 2 , pp. 379-387, Mar. 2023.

17 Sharma, N., Appukutti, S., Garg, U., Mukherjee, J., & Mishra, S. (2023). Analysis of student's academic performance based on their time spent on extra-curricular activities using machine learning techniques. International Journal of Modern Education and Computer Science, 15(1), 46-57.

18 Kharis, S., Hertono, G., Wahyuningrum, E., Yumiati, Y., Irawan, S., Danial, T., & Saputra, D. (2023). DESIGN OF STUDENT SUCCESS PREDICTION APPLICATION IN ONLINE LEARNING USING FUZZY-KNN. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 17(2), 0969-0978. https://doi.org/10.30598/barekengvol17iss2pp0969-0978

19 Salleh, S. S., & Yassin, Y. Bayesian Model for Academic Performance Prediction in Learning Analytics.