ISSN: 2473-4901
v10 n6193

Enabling Adaptive Learning through Data Analytics: A Literature Survey on Applying Machine Learning to Learning Management System Data

Cindy Zhiling Tu
cindytu@nwmissouri.edu

Gary Yu Zhao
zhao@nwmissouri.edu

Northwest Missouri State
Maryville MO

Omar El-Gayar
Omar.el.gayar@dsu.edu
Dakato State University
Madison SD

Abstract

This study presents a literature survey on the application of machine learning (ML) in learning management system (LMS) data analytics, aiming to provide insights into adaptive learning development and propose an agenda for future research. The literature survey is based on a proposed adaptive learning framework and critically analyzes the results within this context. The results reveal that machine learning methods can be used to evaluate the effectiveness of instructional interventions and combining online behaviors with textual data can improve the outcome of performance prediction. Key findings also highlight several open issues, including the use of small datasets and the need for comprehensive ML methods and algorithm development. Future research directions are suggested as improving the accuracy of student performance prediction, instructional intervention analysis and recommendation with ML methods, multimodal LMS data analytics, and big data ML approaches for learning behavior pattern detection.

Keywords: adaptive learning, machine learning, LMS, data analytics, literature survey.

1. INTRODUCTION

The digital transformation of education has brought significant advancements in how learning is delivered and managed. Adaptive learning has emerged as a promising technology and a new teaching paradigm in higher education (Xie et al., 2019). Adaptive learning refers to a pedagogical approach that uses technology to provide corresponding educational experiences to individual learners' needs (Li et al., 2021). Adaptive learning environment is personalized to meet the unique needs of individual learners by dynamically adjusting the instruction based on real-time data to optimize the learning process and make it more effective and efficient (Cavanagh et al., 2020). Adaptivity occurs in instructional activities such as the content, the assessment, and the instruction sequence (Castro, 2019) based on learner’s learning performance and learning characteristics. Higher education institutions need to use instructional contents and students’ learning data to conduct adaptive learning systems.

A Learning Management System (LMS) is a

 ISSN: 2473-4901
v10 n6193

software application for administering, documenting, tracking, reporting, and delivering educational courses, training programs, or learning and development programs (Elfeky & Elbyaly, 2021; Nizam Ismail et al., 2019). The data generated by an LMS includes learner-generated, teacher-generated, and system-generated data. LMS data contains a wealth of information about learning and teaching behavior and outcomes. As more LMS data becomes available, it is important to improve the capabilities for leveraging this data to gain insights into learning and teaching activities (Tenzin et al., 2020; Zhu et al., 2022). Accordingly, learning analytics using Machine learning (ML) techniques in analyzing LMS data has gained significant attention in the last few years. ML-based learning analytics can provide valuable insights and support for various learning theories and pedagogical interventions by analyzing data generated in educational contexts. Compared to traditional statistical analysis methods, ML methods can provide better accuracy and deal with complexity in data analytics, which provides powerful tools that can inform teaching practices and improve student learning experiences (Riestra-González et al., 2021; Villegas-Ch et al., 2020).

Research has been done on the use of ML in LMS data analytics for enhancing adaptive learning, including delivering learning content, adapting to the individual learner's needs, and providing recommendations for learning paths (Kabudi et al., 2021). In addition, previous studies on ML-based LMS data analytics focus on predicting student performance and analyzing student interactions with LMS platforms to attain perspectives into student discourse in online discussions, identifying at-risk students, and improving student engagement and teaching practices (Gasevic et al., 2014; Korkmaz & Correia, 2019; Tenzin et al., 2020). However, the evidence regarding the potential connection between challenges experienced by students and teachers and the effectiveness of ML-based learning analytics and interventions in resolving these issues, the grounding in relevant theories, the appropriateness of various techniques, and the suitability of the data remains unclear.

This literature survey aims to provide a comprehensive overview of the current state of research at the intersection of machine learning, LMS data, and adaptive learning. In that regard, this study addresses the following research questions: (1) Which ML methods and LMS data are used for various learning analytics objectives/outcomes in existing literature? (2) To what extent are ML-based LMS data analytics interventions grounded in adaptive learning? (3) What are the challenges and future research directions in leveraging ML in advanced learning analytics?

2. METHODOLOGY

Given the demonstrated potential by ML-based LMS learning analytics, we propose a literature survey framework adapted from Peng et al. (2019)'s personalized adaptive learning model. As shown in Figure 1, the adaptive learning route has three levels: "what to learn" - based on the learner's characteristics, "how to learn" - based on the learner's performance, and "how well learned" - based on the learner's personal development (Peng et al., 2019). In each level, three phases of data-driven pedagogical decisions based on ML-based learning analytics represent the ordinate. In the "what to learn" level, learning analytics focuses on learning content analysis and instructional design to tailor the learning resources that can match learners' characteristics. The ML-based analytics process must serve this objective, including LMS data collection, variable selection, ML model determination and training, model performance evaluation and optimization. Moreover, the content may undergo continuous refinement through multiple iterations and incremental adjustments to accommodate the variations and the evolving individual characteristics of learners. In "how to learn" adaptive learning level, the data-driven pedagogies focus on guiding learning activity based on learners' performance (Peng et al., 2019). At this level, LMS data, ML algorithms, and evaluation metrics are determined by learning performance prediction, risk-warning, and learning behavior detection. In "how well learned" level, the data-driven pedagogies focus on expanded learning tasks based on learners' learning progress and personal development (Peng et al., 2019). To achieve this goal, the ML analytics processes need to provide learning path analysis, course enrollment analysis, and dropout rate prediction.

Notably, the three levels of adaptive learning paths do not have the same weight and are not necessarily sequenced as in our model. Thus, our framework can be customized to fit different application contexts. Further, ML-based analytics are iterative processes, which means that based on the analytics outcomes, the LMS data, ML algorithms, and evaluation metrics need to be adjusted and refined multiple times to achieve better performance.

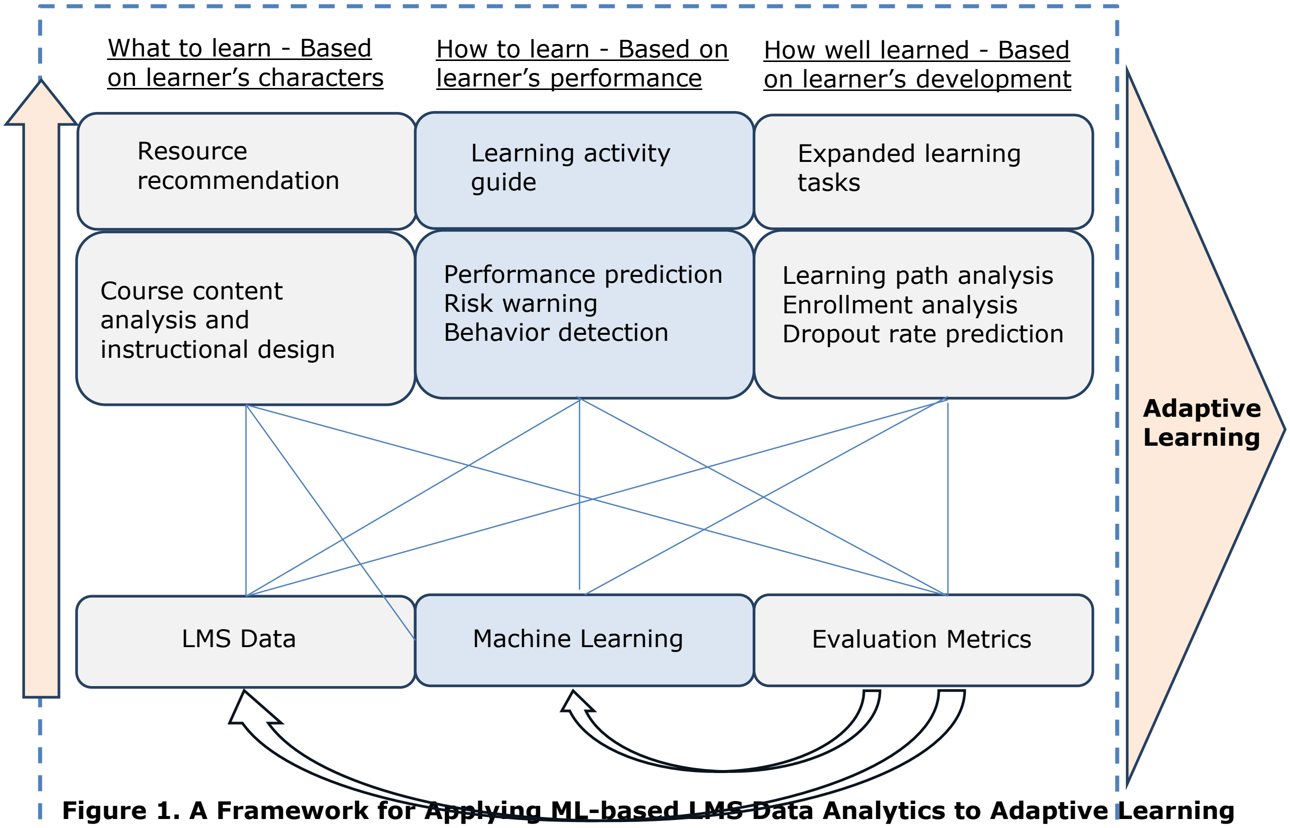


Figure 1. A Framework for Applying ML-based LMS Data Analytics to Adaptive Learning

We searched from five online databases: IEEE Xplore, ACM Digital Library, ProQuest Research Library, ABI/INFORM, and ScienceDirect (Elsevier) using two sets of keywords: ("machine learning" OR "ML" OR "analytics" OR "data analytics") and ("learning analytics" OR "learning management system" OR "LMS"). For each database, we combine these two keywords sets as the search string. Included studies must meet the following criteria: 1. Study of machine learning in LMS data analytics/learning analytics; 2. Full-text paper available; 3. Peer-reviewed paper; 4. Published between January 1, 2013 and January 31, 2023; 5. Written in English. Dissertations/theses, reviews, abstracts, books, book chapters, and reports are excluded for the purpose of this survey. Then, we manually scanned abstracts and filtered out irrelevant articles focusing on education curriculum, pedagogy, impacts, professional development, special external data sources, etc. In addition, we use the snowball technique to identify other relevant papers.

A total of 114 articles were extracted from all online databases. Two authors manually scanned titles and abstracts and filtered out irrelevant articles focusing on education curriculum, pedagogy, impacts, professional development, special external data sources articles based on the inclusion/exclusion criteria, etc. Then we conducted full-text screening. Two authors cross-checked those included articles. In addition, we used the snowball technique to identify other relevant papers in the full-text screening stage. Finally, 52 peer-reviewed academic articles are selected for analysis. These articles are numbered for analysis purpose (see Appendix A).

3. LITERATURE SURVEY RESULTS

Based on our survey framework, we identified the relevant information and extracted it from each paper. For synthesizing the extracted data, we divided the data form into (i) demographic and contextual attributes, (ii) adaptive learning analysis. The first data set was analyzed through statistical techniques and produced descriptive results. The second set of data items was analyzed with a thematic analysis method.

Demographic Distribution

Figure 2 shows the number of the selected papers published per year within the survey period. The number of published studies on the application of machine learning methods in LMS data analytics has been increasing since 2019 and reaches a peak in 2020. 38 papers out of 52 (73%) were published in the past three years signifying the increasing interest possibly due to the rise in online education since the pandemic and the increasing availability of learners' data.

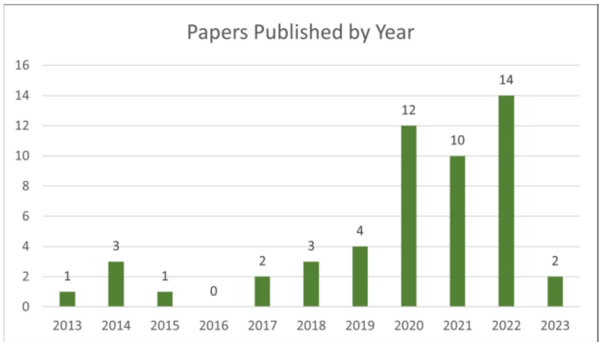


Figure 2: Papers Published by Year

As shown in Table 1, studies were reported from 27 countries. The United States accounted for most of the studies (17% or 9 studies), followed by the United Kingdom, Canada, India, Pakistan, and Greece (3 for each). Our findings are consistent with the prevalence of online education and technology in those countries.

Country	Paper Count	Country	Paper Count
United States	9	Bangladesh	1
Canada	3	Belgium	1
Greece	3	Ecuador	1
India	3	Hungary	1
Pakistan	3	Indonesia	1
United Kingdom	3	Kenya	1
China	2	Korea	1
Brazil	2	Malaysia	1
Croatia	2	Morocco	1
Japan	2	New Zealand	1
Spain	2	Philippines	1
Taiwan	2	Switzerland	1
Vietnam	2	Turkey	1
Australia	1		

Table 1: Distribution by Country

Analysis Based on Proposed Framework
Distribution of Papers by LMS Data Type. Table 2 shows the distribution of the 52 studies by LMS data type. Assessment data has been the most often utilized (27 papers or 52%). The following are learner's data (25 or 48%), user activity data (20 or 38%) and behavior log data (17 or 33%).

LMS Data	Articles	#	%
Assessment data - question and grade related to assignment, test, quiz, exam, etc.	P1,P4,P6,P7,P9,P10,P12,P14,P18, P19,P21,P23,P24,P27,P30,P31,P32,P33,P36,P39,P40,P41,P43,P44, P45,P46,P50	27	52
Leaner Demographic data and socioeconomic data - age, gender, location, device, enrolment, income, etc.	P4,P6,P7,P9,P10, P11,P12,P13,P15,P19,P25,P27,P30,P31,P32,P33,P36,P39,P40,P41, P43,P45,P46,P48,P52	25	48
Activity data - submissions, comments, posts, etc.	P1,P3,P4,P6,P8,P10,P13,P16,P18, P22,P23,P24,P26,P30,P32,P42,P44,P46,P48,P51	20	38
User behaviour log - navigation, page views, time spent on the platform, etc.	P6,P11,P14,P15, P17,P22,P25,P26,P28,P29,P33,P34,P37,P38,P45,P49,P51	17	33
Course Information - content webpage, instructor, start/end date, number of students, etc.	P4,P15,P16,P18, P20,P21,P26,P28,P30,P32,P33,P35,P37,P38,P50,P51	16	31
Interaction data - discussions, forum posts, announcements, messages, etc.	P2,P3,P9,P10,P22,P24,P26,P30,P44,P47,P48,P49, P51	13	25
Multi-modal data - audio, video, presentation, sensor signal, body posture and hand gesture, etc.	P2,P5,P20,P29,P47,P48,P52	7	13
Learning progress data - time spent on each module, the percentage of course completion	P13,P33,P35,P38,P49	5	10

Table 2. Distribution by LMS Data Type

Distribution of Papers by Machine Learning Approach. Table 3 shows the distribution of the 52 studies by machine learning method. Overall, SVM and Random Forest are the most often used machine learning methods (24 papers or 46% for each), followed by Logistic Regression (17 or 33%), Decision Tree (16 or 31%), and MLP Neural Network (14 or 27%).

ISSN: 2473-4901
v10 n6193

Machine Learning	Articles	#	%
Support Vector Machine (SVM)	P3,P5,P6,P7,P9,P13,P15,P17,P22,P30,P31,P32,P34,P35,P36,P39,P40,P41,P43,P45,P46,P48,P52	24	46
Random Forest (RF)	P1,P6,P7,P9,P10,P11,P15,P17,P18,P20,P27,P29,P30,P32,P33,P35,P36,P38,P39,P40,P42,P48,P51,P52	24	46
Logistic Regression (LR)	P4,P6,P9,P12,P17,P22,P24,P25,P30,P33,P34,P40,P41,P42,P45,P46,P49	17	33
Decision Tree (DT)	P1,P3,P10,P11,P12,P15,P18,P26,P28,P34,P36,P39,P40,P46,P47,P51	16	31
MLP Neural Network (MLPNN)	P3,P9,P11,P13,P18,P19,P23,P28,P30,P34,P43,P44,P45,P46	14	27
KNN	P1,P7,P9,P11,P12,P22,P28,P30,P32,P33,P41,P46,P48	13	25
Naïve Bayes (NB)	P1,P6,P11,P18,P32,P33,P34,P40,P41,P42,P43,P49,P52	13	25
Clustering - K-means	P8,P16,P25,P27,P47,P49	6	12
Bayesian Network (BN)	P11,P21,P22,P28,P30	5	10
Gradient Boosting Machine (GBM)	P5,P11,P15,P40,P46	5	10
Linear Regression (LR)	P38,P39	2	4
BERT	P29,P50	2	4
Radial Basis Function Neural Network (RBFNN)	P3,P11	2	4
AdaBoost	P9	1	2
GPT3	P50	1	2
Long Short-term Memory (LSTM)	P46	1	2
Reinforcement Learning	P37	1	2
Bagging	P1	1	2
Convolutional Neural Network (CNN)	P15	1	2
RIPPER or JRIP	P18	1	2

Table 3. Distribution by ML Approach

Distribution of Papers by Analytics Outcome. Table 4 shows the distribution of selected papers by objectives/outcomes. Learning performance analysis/prediction is the most popular, being used in 27 studies (52%), followed by learning behavior/style detection and analysis, used in 17 studies (33%), learning path and recommendation (29% or 15 studies), then course delivery and instructional design (19% or 10 studies), and student enrollment, retention or dropout rate prediction, used in 13% of studies.

Analytics Outcome	Articles	#	%
Learning performance analysis/prediction	P1,P6,P7,P9,P10,P12,P13,P15,P17,P18,P20,P21,P22,P24,P25,P26,P27,P28,P30,P31,P32,P33,P34,P39,P43,P45,P46	27	52
Learning behavior/style detection and analysis	P2,P3,P8,P11,P14,P16,P23,P26,P29,P32,P43,P44,P47,P48,P49,P51,P52	17	33
Learning path and recommendation	P3,P9,P13,P14,P16,P21,P24,P29,P30,P35,p37,P38,P44,P48,P50	15	29
Course delivery and instructional design	P5,P14,P19,P20,P25,P35,P38,P42,P44,P50	10	19
Enrollment, retention, and dropout rate prediction/analysis	P4,P10,P17,P36,P40,P41,P46	7	13

Table 4. Distribution by Analytics Objective/Outcome

Distribution of Papers by Evaluation Method. As shown in Table 5, accuracy is the most often used evaluation metric in LMS data analytics with ML approaches (34 out of 52 or 65%). This is followed by F1-score, used in 27% (14 of 52), then AUC-ROC (21% or 11 papers), then Precision and Recall (19% or 10 papers), and RMSE/MSE/MAE metric (13%, 7 out of 52). Table 6 shows the use of analytics outcomes evaluation methods. 73% of studies do not clearly employ

 ISSN: 2473-4901
v10 n6193

the analytic outcomes evaluation method. Benchmarking is the most used (12%, 6 papers), followed by collecting learners' feedback (8%, 4 papers) and designated assessment (4%, 2 papers). Further, the statistical analysis and prescriptive analysis is used for analytic outcomes evaluation (one paper for each).

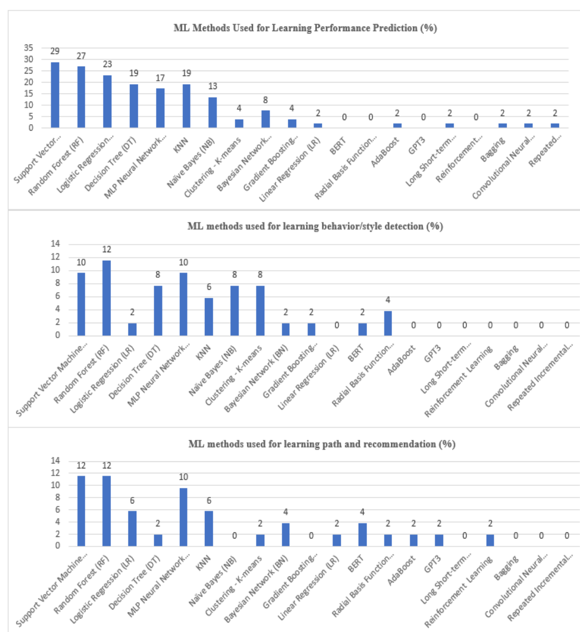
Evaluation Metric	Articles	#	%
Accuracy	P1,P3,P6,P7,P10,P11,P12,P13,P15,P18,P19,P20,P21,P23,P24,P26,P27,P28,P29,30,P32,P34,P35,P36,P38,P41,P43,P44,P45,P46,P47,P48,P51,P52	34	65
F1-score	P4,P7,P12,P17,P23,P29,P33,P36,P37,P40,P42,P48,P49,P51	14	27
AUC-ROC	P6,P7,P20,P23,p27,P30,P34,P41,P46,P48,P51	11	21
Precision and Recall	P7,P12,P17,P23,P27,P40,P46,P48,P49,P51	10	19
RMSE/MSE/MAE	P9,P22,P25,P31,P37,P39,P44	7	13

Table 5. Distribution by Evaluation Metric

Analytic Outcomes Evaluation Method	Articles	#	%
Benchmarking	P4,P14,P17,P21,P31,P37	6	12
Collecting learners' feedback	P2,P25,P44,P51	4	8
Specially designed assessment	P19,P48	2	4
Statistical analysis - ANOVA	P34	1	2
Prescriptive analysis	P32	1	2
No analytic outcomes evaluation methods clearly mentioned	P1,P3,P5,P6,P7,P8,P9,P10,P11,P12,P13,P15,P16,P18,P20,P22,P23,P24,P26,P27,P28,P29,P30,P33,P35,P36,P38,P39,P40,P41,P42,P43,P45,P46,P47,P49,P50,P52	38	73

Table 6. Distribution by Analytics Outcomes Evaluation Method

Distribution of ML Methods and LMS Data Used for Analytics Outcomes. As shown in Figure 2 and Figure 3, i) the most often used ML methods for learning performance prediction is SVM and Random Forest, and the most often used LMS data is learner demographic data and assessment data; ii) for learning behavior/style detection and analysis, Random Forest, MLP Neural Networks, and SVM are the most popular ML methods and the LMS data is learners' data and activity data; iii) for learning path and recommendation, top three most often used ML methods are SVM, Random Forest, and MLP Neural Network. The LMS data is assessment data, activity data, and course information; iv) for the outcome of course delivery/instructional design, Random Forest, MLPNN, LR, and SVM are commonly used with LMS assessment data, course information, and user behavior log data; v) for student retention/dropout rate prediction, SVM and LR are equally most popular used ML algorithms, and the LMS data is the combination of learner demographic data, assessment data, and activity data.



 ISSN: 2473-4901
v10 n6193

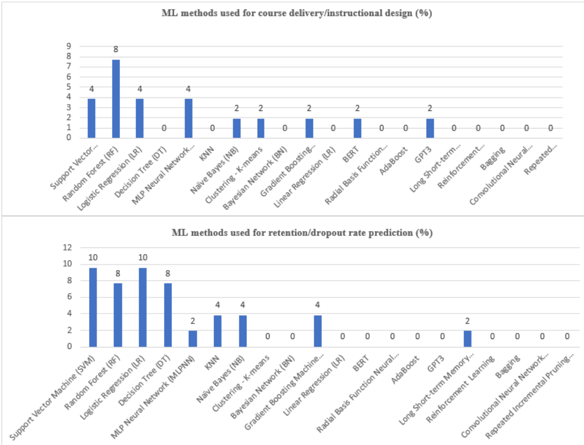


Figure 3. Distribution of ML Approaches Used for Analytics Outcomes

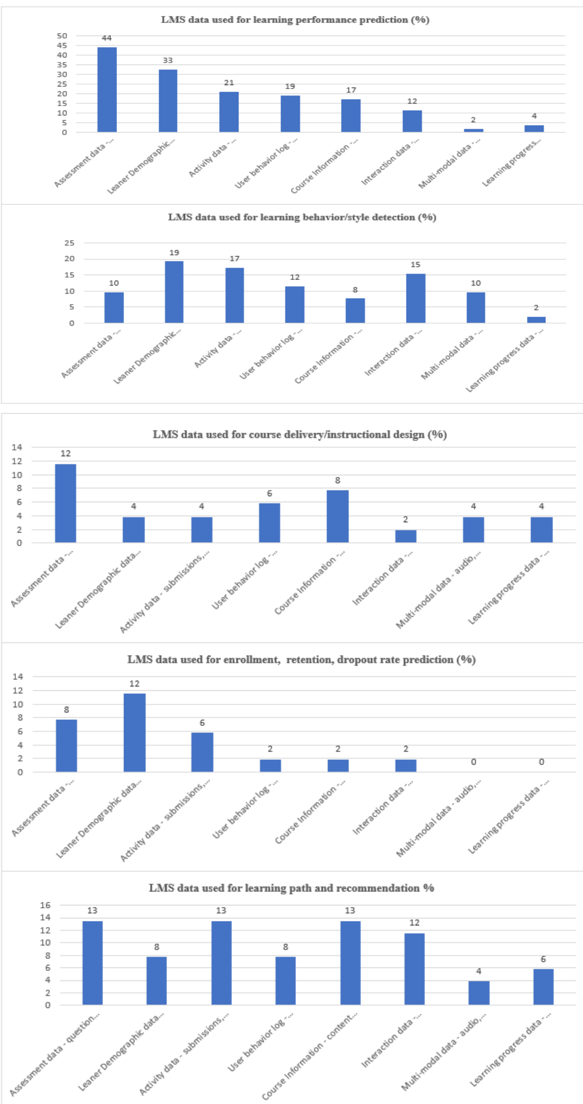


Figure 4. Distribution of LMS Data Used for Analytics Outcomes

Distribution of Evaluation Metrics by ML Methods. As shown in Table 7, accuracy is the most often used evaluation metric for all top six ML-algorithms.

ML Methods	Accuracy	Precision / Recall	F1- Score	AUC-ROC	RMSE / MSE	Cross-validatio
SVM	15	5	5	6	4	5
Random Forest (RF)	18	6	9	7	2	7
Logistic Regression (LR)	8	5	7	5	2	2
Decision Tree (DT)	14	4	4	3	0	2
MLP Neural Network (MLPNN)	13	2	1	4	2	2
KNN	12	4	4	5	2	4

Table 7. Distribution of Evaluation Metrics on Top Six ML Algorithms

In extant studies, researchers used multiple metrics instead of a single metric to assess the effectiveness and efficiency of utilized ML approaches. Notably, AUC-ROC is the second most popular utilized for SVM, followed by RMSE, Cross-validation, and F1-Score. for random forest, logistic regression, decision tree, neural networks, and KNN methods, F1-score, followed by Precision/Recall, AUC-ROC is the most popularly used evaluation metrics other than Accuracy.

4. DISCUSSION AND FINDINGS

Following the proposed review framework, we summarize primary challenges and issues in the current literature as shown in Figure 4. These findings cover the different levels of adaptive learning paths: LMS data (what to learn), machine learning methods (how to learn), analytics outcomes and evaluation (how well learned). Most extant studies use relatively small datasets to train ML models. Such datasets primarily focus on course-level cross-section numeric data (Du et al., 2020). However, the data generated from the LMS platform nowadays is large, multimodal longitudinal data. Machine learning in the context of big data presents unique challenges, and overcoming these obstacles requires approaches that differ from traditional learning methods. Scalable, multidomain, parallel, flexible, and intelligent

ISSN: 2473-4901
v10 n6193

learning methods are preferred in this context (Qiu et al., 2016).

Secondly, existing literature lacks a comprehensive machine learning (ML) method or a combination of methods designed to achieve specific analytics outcomes (Islam & Mahmud, 2020). The focus has primarily been on using existing ML methods and comparing the performance, with a lack of new algorithm development. For example, over 40% of reviewed studies use SVM but rarely combine it with other ML methods or optimize it to achieve better analysis accuracy. Notably, neural networks and some emerging ML methods such as LSTM, BERT, and GPT are gaining more and more attention and need deeper investigation in the domain (Pan et al., 2020; Yang, 2021). Furthermore, the development of robust methods for feature selection and assessing their effectiveness is a promising direction in the domain (Coussement et al., 2020; Soleimani & Lee, 2021).

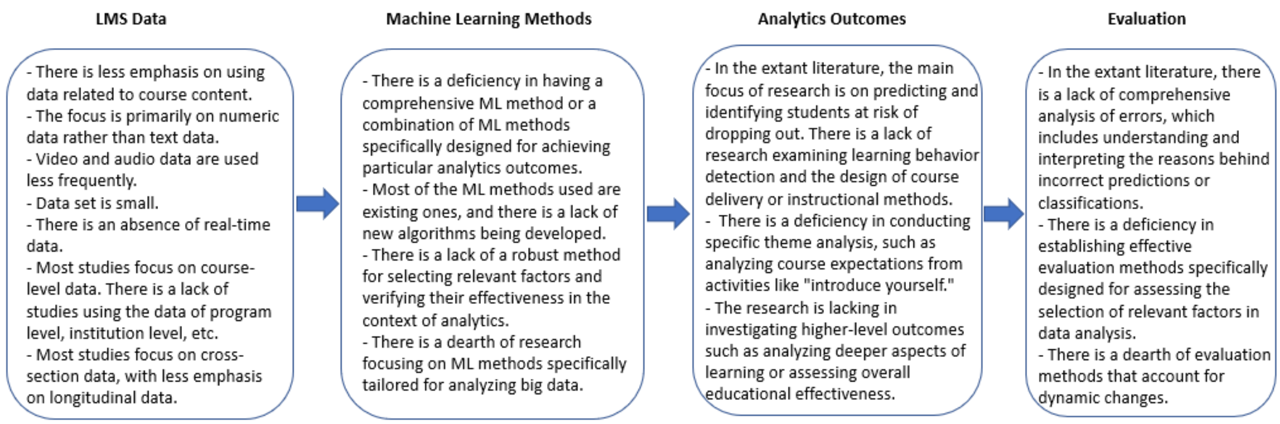


Figure 4. Challenges and Issues in Current Literature

Regarding analytics objective/outcome, the primary emphasis is on forecasting learners' performance and detecting learners who are likely to discontinue their studies (Villegas-Ch et al., 2020). Less attention has been given to investigating the identification of learning behaviors and the development of instructional techniques for course delivery. There is a dearth of research regarding the text analysis of specific themes, such as evaluating course expectations through activities like "introduce yourself" (Tran et al., 2022). Additionally, the current body of research is limited regarding the exploration of various outcomes, including the examination of more profound facets of learning and the assessment of overall educational effectiveness (Yang, 2021).

With respect to evaluation metrics, extant literature lacks a comprehensive examination of errors, including a thorough understanding and interpretation of the underlying causes for inaccurate predictions or classifications. Moreover, there is a gap in defining evaluation methods that effectively assess the selection of pertinent features in data analysis (Lan et al., 2014). Further, there is a scarcity of evaluation methods that adequately consider dynamic changes (Yang, 2021).

Based on the abovementioned challenges and issues, there are several important directions for future research in this domain. First, student performance prediction remains a viable research topic in the domain (Jiao et al., 2022; Riestra-González et al., 2021). Even though many studies have been done on the utilization of various ML methods in student performance prediction, there remains a need for robust predictive models with good generalizability for different courses, programs, and institutions.

A general predictive model can be developed using existing machine learning algorithms, or combining multiple algorithms, or newly created algorithms. For example, recent advancements in text mining with ML methods have prompted researchers to utilize social media to predict learning performance (Shahbazi & Byun, 2020). However, in the field of LMS data analytics, there is a scarcity of studies that have effectively integrated online behaviors with textual data to enhance prediction accuracy. Therefore, it is crucial to combine online behaviors with textual data to improve the outcome of performance prediction.

 ISSN: 2473-4901
v10 n6193

Second, machine learning methods can be used to evaluate the effectiveness of instructional interventions including course content delivery and instructional design. Due to the lack of an educational framework, there is no consistent results that can be extracted from the studies of instructional content and design (Lee, 2021; Tran et al., 2022). Well-designed instructional content has a significant impact on enhancing learning effectiveness. Consequently, it is anticipated that more researchers will focus on identifying content design patterns in the future. However, current studies have not emphasized the automated support for teachers and learners to improve their teaching and learning experiences, such as offering automatic suggestions for instructional design or adjustments to learning strategies (Du et al., 2020).

Third, newer LMS platforms include many user interaction features, such as discussion boards, announcement portal, conversation tools, collaboration tools, etc., aimed at improving student engagement and teaching performance. These advanced tools generate large volumes of text, video, and audio data. Extant literature is limited in regard to the use of ML methods with text or other types of data (Shahbazi & Byun, 2020). One of the important directions for future studies is to investigate the utilization of appropriate ML methods for LMS multimodal data analytics. For example, one study investigated students' learning motivation and predicted their learning performance using video-viewing data in a flipped statistic course (Liao & Wu, 2023).

Fourth, the application of big data ML approaches in the detection of learning style and behavior. LMS platforms have been used for more than ten years. An individual institution possesses a large volume of longitudinal learners' activity and log data. Also, this big data can include various types and formats. How to gain valuable insights into learners' behavior patterns by using ML methods in this big data analytics context is an attractive direction for future research. For example, how many learning patterns are needed to train a classifier depends on achieving a balance between cost and accuracy when dealing with overfitting issues (Bird et al., 2022).

5. CONCLUSIONS

This study identified, organized and discussed challenges and issues related to the application of ML in analyzing data from LMS into four perspectives according to the proposed literature analysis framework. Findings indicate that extant research often uses small datasets and focused on numeric data, while LMS platforms generate large multimodal longitudinal data. Future research directions include student performance prediction, instructional intervention analysis, multimodal data analytics, and big data ML approaches for learning style and behavior detection.

Overall, the application of ML in LMS data analytics has significant potential to improve teaching and learning outcomes. Institutions can implement adaptive learning platforms that adjust the delivery of content based on student data collected from the LMS. ML models can dynamically suggest content adjustments (e.g., additional resources for struggling students or advanced materials for high-performing students). The proposed research agenda focuses on a range of research questions and machine learning methods that can be used to advance the field. By addressing these questions, researchers can develop a deeper understanding of the utilization of machine learning approaches in learning analytics as well as the development of advanced solutions. While there is a natural inclination to rely on existing methods in the field, it is essential to pursue a parallel line of research that develops new methods and systems.

6. REFERENCES

Ahmed, M. R., Tahid, S. T. I., Mitu, N. A., Kundu, P., & Yeasmin, S. (2020, 7/2020). A Comprehensive Analysis on Undergraduate Student Academic Performance using Feature Selection Techniques on Classification Algorithms. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT),

Anderson, K. (2022, November 7, 2022). Real-time Feedback for Developing Conversation Literacy.ICMI '22

Babić, I. Đ. (2017). Machine learning methods in predicting the student academic motivation. Croatian Operational Research Review, 8(2), 443-461. https://doi.org/10.17535/crorr.2017.0028

Bird, K. A., Castleman, B. L., Song, Y., & Yu, R. (2022). Is big data better? LMS data and predictive analytic performance in postsecondary education. EdWorking Paper. https://doi.org/10.26300/8XYS-YM74

Bognár, L., & Fauszt, T. (2022). Factors and conditions that affect the goodness of machine learning models for predicting the success of learning. Computers and

 ISSN: 2473-4901
v10 n6193

Education: Artificial Intelligence, 3, 100100. https://doi.org/10.1016/j.caeai.2022.100100

Castro, R. (2019). Blended learning in higher education: Trends and capabilities. Education and Information Technologies, 24(4), 2523-2546. https://doi.org/10.1007/s10639-019-09886-3

Cavanagh, T., Chen, B., Lahcen, R. A. M., & Paradiso, J. R. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. International review of research in open and distributed learning, 21(1), 173-197. https://doi.org/10.19173/irrodl.v21i1.4557

Chen, L., Leong, C. W., Feng, G., & Lee, C. M. (2014, November 12, 2014). Using Multimodal Cues to Analyze MLA'14 Oral Presentation Quality Corpus: Presentation Delivery and Slides Quality.MLA '14

Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. Decision Support Systems, 135, 113325. https://doi.org/10.1016/j.dss.2020.113325

Deeva, G., De Smedt, J., Saint-Pierre, C., Weber, R., & De Weerdt, J. (2022). Predicting student performance using sequence classification with time-based windows. Expert Systems with Applications, 209, 118182. https://doi.org/10.1016/j.eswa.2022.118182

Dervenis, C., Kyriatzis, V., Stoufis, S., & Fitsilis, P. (2023, January 30, 2023). Predicting Students' Performance Using Machine Learning Algorithms.ICACS '22

Du, X., Yang, J., Jui-Long, H., & Shelton, B. (2020). Educational data mining: A systematic review of research and emerging trends. Information Discovery and Delivery, 48(4), 225–236. https://doi.org/10.1108/IDD-09-2019-0070

Elfeky, A. I. M., & Elbyaly, M. Y. H. (2021). The use of data analytics technique in learning management system to develop fashion design skills and technology acceptance. Interactive Learning Environments, 31(6), 3810–3827. https://doi.org/10.1080/10494820.2021.1943688

Elliott, R., & Luo, X. (2022, 2022-10-8). Learning Management System Analytics to Examine the Behavior of Students in High Enrollment STEM Courses During the Transition to Online Instruction. 2022 IEEE Frontiers in Education Conference (FIE),

Gasevic, D., Rose, C., Siemens, G., Wolff, A., & Zdrahal, Z. (2014). Learning analytics and machine learning. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, 287–288. https://doi.org/10.1145/2567574.256763

Gkontzis, A. F., Kotsiantis, S., Tsoni, R., & Verykios, V. S. (2018, November 29, 2018). An effective LA approach to predict student achievement.PCI '18

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. Computers & Education, 131, 22-32. https://doi.org/10.1016/j.compedu.2018.12.006

Imhof, C., Comsa, I.-S., Hlosta, M., Parsaeifard, B., Moser, I., & Bergamin, P. (2022). Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models. IEEE Transactions on Learning Technologies, 1-15. https://doi.org/10.1109/TLT.2022.3221495

Islam, S., & Mahmud, H. (2020, June 18, 2020). Integration of Learning Analytics into Learner Management System using Machine Learning.ICMET '20

Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. Artificial Intelligence Review, 55(8), 6321-6344. https://doi.org/10.1007/s10462-022-10155-y

Joseph, B., & Abraham, S. (2022, 2022-2-12). Analyzing the Cognitive Process Dimension and Rate of Learning to Identify the Slow Learners in e-Learning. 2022 International Conference on Innovative Trends in Information Technology (ICITIIT),

Jui-Long, H., Rice, K., Kepka, J., & Yang, J. (2020). Improving predictive power through deep learning analysis of K-12 online student behaviors and discussion board content. *Information Discovery and Delivery*, *48*(4), 199-212. https://doi.org/10.1108/IDD-02-2020-0019

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017. https://doi.org/10.1016/j.caeai.2021.100017

Kokoç, M., Akçapınar, G., & Hasnine, M. N. (2021). Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics. *Journal of Educational Technology & Society*, *24*(1).

Kondo, N., Okubo, M., & Hatanaka, T. (2017, 7/2017). Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data. 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI),

Korkmaz, C., & Correia, A.-P. (2019). A review of research on machine learning in educational technology. *Educational Media International*, *56*(3), 250–267. https://doi.org/10.1080/09523987.2019.1669875

Kumar, M., Sharma, C., Sharma, S., Nidhi, N., & Islam, N. (2022, 2022-3-23). Analysis of Feature Selection and Data Mining Techniques to Predict Student Academic Performance. 2022 International Conference on Decision Aid Sciences and Applications (DASA),

Lagman, A. C., Alcober, G. M. I., Fernando, M. C. G., Goh, M. L. I., Lalata, J.-a. P., Ortega, J. H. J. C., . . . Claour, J. P. (2020, 2020-06-14). Integration of Neural Network Algorithm in Adaptive Learning Management System. ICRSA 2020: 2020 3rd International Conference on Robot Systems and Applications,

Lamb, R., Neumann, K., & Linder, K. A. (2022). Real-time prediction of science student learning outcomes using machine learning classification of hemodynamics during virtual reality and online learning sessions. *Computers and Education: Artificial Intelligence*, *3*, 100078. https://doi.org/10.1016/j.caeai.2022.100078

Lan, A. S., Waters, A. E., Studer, C., & Baraniuk, R. G. (2014). Sparse factor analysis for learning and content analytics. *The Journal of Machine Learning Research*, *15*(1), 1959–2008.

Le, M.-D., Nguyen, H.-H., Nguyen, D.-L., & Nguyen, V. A. (2020, July 13, 2020). How to Forecast the Students' Learning Outcomes Based on Factors of Interactive Activities in a Blended Learning Course.*ICFET '20*

Lee, A. V. Y. (2021). Determining quality and distribution of ideas in online classroom talk using learning analytics and machine learning. *Journal of Educational Technology & Society*, *24*(1), 236-249. https://www.proquest.com/pqrl/docview/2515025065/abstract/D46CC4C0E1E44CCDPQ/24

Li, F., He, Y., & Xue, Q. (2021). Progress, challenges and countermeasures of adaptive learning. *Educational Technology & Society*, *24*(3), 238-255. https://www.jstor.org/stable/27032868

Liao, C.-H., & Wu, J.-Y. (2023). Learning analytics on video-viewing engagement in a flipped statistics course: Relating external video-viewing patterns to internal motivational dynamics and performance. *Computers & Education*, *197*, 104754. https://doi.org/10.1016/j.compedu.2023.104754

Lwande, C., Link to external site, t. l. w. o. i. a. n. w., Oboko, R., & Lawrence, M. (2021). Learner behavior prediction in a learning management system. *Education and Information Technologies*, *26*(3), 2743-2766. https://doi.org/10.1007/s10639-020-10370-6

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588-599. https://doi.org/10.1016/j.compedu.2009.09.008

Nguyen, V. A., Nguyen, Q. B., & Nguyen, V. T. (2018, August 13, 2018). A Model to Forecast Learning Outcomes for Students in Blended Learning Courses Based On Learning Analytics.*ICSET 2018*

Nizam Ismail, S., Hamid, S., & Chiroma, H. (2019). The utilization of learning analytics to develop student engagement model in learning management system. *Journal of Physics: Conference Series*, *1339*(1), 012096. https://doi.org/10.1088/1742-6596/1339/1/012096

Omiros, I., Link to external site, t. l. w. o. i. a. n. w., Savvas, I. K., Panos, F., & Gerogiannis, V. C. (2021). A two-phase machine learning

ISSN: 2473-4901
v10 n6193

approach for predicting student outcomes. *Education and Information Technologies*, *26*(1), 69-88. https://doi.org/10.1007/s10639-020-10260-x

Oreski, D., & Hajdin, G. (2019, 5/2019). A Comparative Study of Machine Learning Approaches on Learning Management System Data. 2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO),

Pan, Z., Li, C., & Liu, M. (2020, 2020-08-12). Learning Analytics Dashboard for Problem-based Learning. L@S '20: Seventh (2020) ACM Conference on Learning @ Scale,

Peng, H., Ma, S., & Spector, J. M. (2019). Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment. *Smart Learning Environments*, *6*, 9 (2019). https://doi.org/10.1186/s40561-019-0089-y

Pérez Sánchez, C. J., Calle-Alonso, F., & Vega-Rodríguez, M. A. (2022). Learning analytics to predict students’ performance: A case study of a neurodidactics-based collaborative learning platform. *Education and Information Technologies*, *27*(9), 12913-12938. https://doi.org/10.1007/s10639-022-11128-y

Pimentel, J. S., Ospina, R., Link to external site, t. l. w. o. i. a. n. w., Anderson, A., & Link to external site, t. l. w. o. i. a. n. w. (2021). Learning Time Acceleration in Support Vector Regression: A Case Study in Educational Data Mining. *Stats*, *4*(3), 682. https://doi.org/10.3390/stats4030041

Purwoningsih, T., Santoso, H. B., & Hasibuan, Z. A. (2020, 2020-11-3). Data Analytics of Students' Profiles and Activities in a Full Online Learning Context. 2020 Fifth International Conference on Informatics and Computing (ICIC),

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, *2016*(7), 67. https://doi.org/10.1186/s13634-016-0355-x

Ramaswami, G. S., Susnjak, T., Mathrani, A., & Umer, R. (2020, 2020-12-16). Predicting Students Final Academic Performance using Feature Selection Approaches. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE),

Riestra-González, M., Paule-Ruíz, M. D. P., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, *163*, 104108. https://doi.org/10.1016/j.compedu.2020.104108

Segarra-Faggioni, V., & Ratte, S. (2021, March 6, 2021). Computer-based Classification of Student's Report.*ICETC '20*

Sghir, N., Adadi, A., El Mouden, Z. A., & Lahmer, M. (2022, 2022-3-3). Using Learning Analytics to Improve Students' Enrollments in Higher Education. 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET),

Shahbazi, Z., & Byun, Y. C. (2020). Toward Social Media Content Recommendation Integrated with Data Science and Machine Learning Approach for E-Learners. *Symmetry*, *12*(11), 1798. https://doi.org/10.3390/sym12111798

Soleimani, F., & Lee, J. (2021, June 8, 2021). Comparative Analysis of the Feature Extraction Approaches for Predicting Learners Progress in Online Courses: MicroMasters Credential versus Traditional MOOCs.*L@S '21*

Suleiman, R., & Anane, R. (2022, 2022-5-4). Institutional Data Analysis and Machine Learning Prediction of Student Performance. 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD),

Tamada, M. M., Giusti, R., & De Magalhaes Netto, J. F. (2021, 2021-10-13). Predicting Student Performance Based on Logs in Moodle LMS. 2021 IEEE Frontiers in Education Conference (FIE),

Tenzin, D., Lemay, D. J., Basnet, R. B., & Bazelais, P. (2020). Predictive analytics in education: a comparison of deep learning frameworks. *Education and Information Technologies*, *25*(3), 1951-1963. https://doi.org/10.1007/s10639-019-10068-4

Tran, T. P., Jan, T., & Kew, S. N. (2022). Learning Analytics for Improved Course Delivery: Applications and Techniques. *Proceedings of the 6th International Conference on Digital*

ISSN: 2473-4901
v10 n6193

Technology in Education, 100–106. https://doi.org/10.1145/3568739.3568758

Umer, R., Mathrani, A., Susnjak, T., & Lim, S. (2019, March 30, 2019). Mining Activity Log Data to Predict Student's Outcome in a Course.*ICBDE '19*

Veluri, R. K., Patra, I., Naved, M., Prasad, V. V., Arcinas, M. M., Beram, S. M., & Raghuvanshi, A. (2022). Learning analytics using deep learning techniques for efficiently managing educational institutes. *Materials Today: Proceedings*, *51*, 2317-2320. https://doi.org/10.1016/j.matpr.2021.11.416

Villegas-Ch, W., Román-Cañizares, M., & Palacios-Pacheco, X. (2020). Improvement of an Online Education Model with the Integration of Machine Learning and Data Analysis in an LMS. *Applied Sciences*, *10*(15), 5371. https://doi.org/10.3390/app10155371

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, *104*, 106189. https://doi.org/10.1016/j.chb.2019.106189

Waheed, H., Hassan, S.-U., Nawaz, R., Aljohani, N. R., Chen, G., & Gasevic, D. (2023). Early prediction of learners at risk in self-paced education: A neural network approach. *Expert Systems with Applications*, *213*, 118868. https://doi.org/10.1016/j.eswa.2022.118868

Worsley, M. (2018, March 7, 2018). (Dis)engagement matters: identifying efficacious learning practices with multimodal learning analytics.*LAK '18*

Wu, J.-Y. (2021). Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement. *Computers & Education*, *163*, 104066. https://doi.org/10.1016/j.compedu.2020.104066

Xie, H., Chu, H.-C., Hwang, G.-J., & Wang, C.-C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Xing, W., & Goggins, S. (2015, March 16, 2015). Learning analytics in outer space: a Hidden Naïve Bayes model for automatic student off-task behavior detection.*LAK '15*

Yang, S. J. H. (2021). Precision Education - A New Challenge for AI in Education. *Journal of Educational Technology & Society*, *24*(1).

Zhao, F., Liu, G.-Z., Zhou, J., & Yin, C. (2023). A Learning Analytics Framework Based on Human-Centered Artificial Intelligence for Identifying the Optimal Learning Strategy to Intervene in Learning Behavior. *Journal of Educational Technology & Society*, *26*(1).

Zhou, J., Hang, K., Oviatt, S., Yu, K., & Chen, F. (2014, November 12, 2014). Combining empirical and machine learning techniques to predict math expertise using pen signal features.*MLA '14*

Zhu, M., Sari, A. R., & Lee, M. M. (2022). Trends and issues in MOOC learning analytics empirical research: A systematic literature review (2011–2021). *Education and Information Technologies*, *27*(7), 10135-10160. https://doi.org/10.1007/s10639-022-11031-6

ISSN: 2473-4901
v10 n6193

APPENDIX A
Articles Coded for Literature Survey

Citation	Paper Title	#
(Ahmed et al., 2020)	A Comprehensive Analysis on Undergraduate Student Academic Performance using Feature Selection Techniques on Classification Algorithms	P1
(Anderson, 2022)	Real-time Feedback for Developing Conversation Literacy	P2
(Babić, 2017)	Machine learning methods in predicting student academic motivation	P3
(Bognár & Fauszt, 2022)	Factors and conditions that affect the goodness of machine learning models for predicting the success of learning	P4
(Chen et al., 2014)	Using Multimodal Cues to Analyze MLA'14 Oral Presentation Quality Corpus: Presentation Delivery and Slides Quality	P5
(Deeva et al., 2022)	Predicting student performance using sequence classification with time-based windows	P6
(Dervenis et al., 2023)	Predicting Students' Performance Using Machine Learning Algorithms	P7
(Elliott & Luo, 2022)	Learning Management System Analytics to Examine the Behavior of Students in High Enrollment STEM Courses During the Transition to Online Instruction	P8
(Gkontzis et al., 2018)	An effective LA approach to predict student achievement	P9
(Gray & Perkins, 2019)	Utilizing early engagement and machine learning to predict student outcomes	P10
(Imhof et al., 2022)	Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models	P11
(Islam & Mahmud, 2020)	Integration of Learning Analytics into Learner Management System using Machine Learning	P12
(Jiao et al., 2022)	Artificial intelligence-enabled prediction model of student academic performance in online engineering education	P13
(Joseph & Abraham, 2022)	Analyzing the Cognitive Process Dimension and Rate of Learning to Identify the Slow Learners in e-Learning	P14
(Jui-Long et al., 2020)	Improving predictive power through deep learning analysis of K-12 online student behaviors and discussion board content	P15
(Kokoç et al., 2021)	Unfolding Students’ Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics	P16
(Kondo et al., 2017)	Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data	P17
(Kumar et al., 2022)	Analysis of Feature Selection and Data Mining Techniques to Predict Student Academic Performance	P18
(Lagman et al., 2020)	Integration of Neural Network Algorithm in Adaptive Learning Management System	P19
(Lamb et al., 2022)	Real-time prediction of science student learning outcomes using machine learning classification of hemodynamics during virtual reality and online learning sessions	P20
(Lan et al., 2014)	Sparse factor analysis for learning and content analytics	P21
(Le et al., 2020)	How to Forecast the Students' Learning Outcomes Based on Factors of Interactive Activities in a Blended Learning Course	P22
(Lwande et al., 2021)	Learner behavior prediction in a learning management system	P23
(Macfadyen & Dawson, 2010)	Mining LMS data to develop an “early warning system” for educators: A proof of concept	P24
(Nguyen et al., 2018)	A Model to Forecast Learning Outcomes for Students in Blended Learning Courses Based On Learning Analytics	P25

 ISSN: 2473-4901
v10 n6193

(Nizam Ismail et al., 2019)	The utilization of learning analytics to develop student engagement model in learning management system	P26
(Omiros et al., 2021)	A two-phase machine learning approach for predicting student outcomes	P27
(Oreski & Hajdin, 2019)	A Comparative Study of Machine Learning Approaches on Learning Management System Data	P28
(Pan et al., 2020)	Learning Analytics Dashboard for Problem-based Learning	P29
(Pérez Sánchez et al., 2022)	Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform	P30
(Pimentel et al., 2021)	Learning Time Acceleration in Support Vector Regression: A Case Study in Educational Data Mining	P31
(Purwoningsih et al., 2020)	Data Analytics of Students' Profiles and Activities in a Full Online Learning Context	P32
(Ramaswami et al., 2020)	Predicting Students Final Academic Performance using Feature Selection Approaches	P33
(Riestra-González et al., 2021)	Massive LMS log data analysis for the early prediction of course-agnostic student performance	P34
(Segarra-Faggioni & Ratte, 2021)	Computer-based Classification of Student's Report	P35
(Sghir et al., 2022)	Using Learning Analytics to Improve Students' Enrollments in Higher Education	P36
(Shahbazi & Byun, 2020)	Toward Social Media Content Recommendation Integrated with Data Science and Machine Learning Approach for E-Learners	P37
(Soleimani & Lee, 2021)	Comparative Analysis of the Feature Extraction Approaches for Predicting Learners Progress in Online Courses: MicroMasters Credential versus Traditional MOOCs	P38
(Suleiman & Anane, 2022)	Institutional Data Analysis and Machine Learning Prediction of Student Performance	P39
(Tamada et al., 2021)	Predicting Student Performance Based on Logs in Moodle LMS	P40
(Tenzin et al., 2020)	Predictive analytics in education: a comparison of deep learning frameworks	P41
(Umer et al., 2019)	Mining Activity Log Data to Predict Student's Outcome in a Course	P42
(Veluri et al., 2022)	Learning analytics using deep learning techniques for efficiently managing educational institutes	P43
(Villegas-Ch et al., 2020)	Improvement of an Online Education Model with the Integration of Machine Learning and Data Analysis in an LMS	P44
(Waheed et al., 2020)	Predicting academic performance of students from VLE big data using deep learning models	P45
(Waheed et al., 2023)	Early prediction of learners at risk in self-paced education: A neural network approach	P46
(Worsley, 2018)	(Dis)engagement matters: identifying efficacious learning practices with multimodal learning analytics	P47
(Wu, 2021)	Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement	P48
(Xing & Goggins, 2015)	Learning analytics in outer space: a Hidden Naïve Bayes model for automatic student off-task behavior detection	P49
(Yang, 2021)	Precision Education - A New Challenge for AI in Education	P50
(Zhao et al., 2023)	A Learning Analytics Framework Based on Human-Centered Artificial Intelligence for Identifying the Optimal Learning Strategy to Intervene in Learning Behavior	P51
(Zhou et al., 2014)	Combining empirical and machine learning techniques to predict math expertise using pen signal features	P52