Annals of "Dunarea de Jos" University of Galati
Fascicle I. Economics and Applied Informatics
Years XXX – n°3/2024
ISSN-L 1584-0409 ISSN-Online 2344-441X
www.eia.feaa.ugal.ro

DOI https://doi.org/10.35219/eai15840409471

Data Analysis in Online Education: Tools and Techniques for Improving Academic Performance

Marian Stan*, Mihai Ciobotea, Mihaela Covrig***, Doina Liliana Badea******

ARTICLE INFO

Article history:
Accepted November 2024
Available online December 2024

JEL Classification
I21

Keywords:
data analysis; learning analytics; online education; academic performance; predictive models

ABSTRACT

The article explores the role of data analysis in improving student performance in online learning environments. With the expansion of distance learning during the COVID-19 pandemic, educational institutions faced growing challenges regarding increasing student engagement, predicting academic outcomes, and retaining students in school. These challenges remain relevant today. Data analysis techniques—from descriptive analyses to artificial intelligence-based predictive models—can help teachers and school administrations make informed decisions. Following a qualitative approach, through a literature review and the use of case studies, the research demonstrates how these tools can provide early intervention strategies for students at risk of falling behind and how they can create personalized learning experiences, analyzing key platforms such as Moodle and Canvas. The article also highlights the impact of third-party analytics integrations, which offer a more detailed understanding of student behavior and engagement. Although data analysis in online education has the potential to improve outcomes, there are challenges related to data interpretation, information overload, and institutional readiness to adopt these technologies. The study's conclusions indicate that, by using the right tools and techniques, data analysis can lead to more adaptive and personalized learning environments, ultimately contributing to better academic performance.

1. Introduction

Online learning experienced significant growth during the pandemic. This growth transformed the educational landscape, offering students access to an increasingly diverse range of online courses and educational resources. Online learning has generated a substantial volume of data from student interactions on learning platforms, tests, quizzes, clickstream data, as well as unstructured data such as texts, messages, feedback, and learning journals. As more educational institutions adopt online learning, new challenges have emerged regarding maintaining high levels of student engagement, student retention, early intervention, and more accurate measurement or prediction of academic performance. These challenges have driven the need for innovative solutions to improve learning experiences and to maximize or optimize academic outcomes.

One of the proposed solutions involves the use of educational data analysis, data analytics models, or learning analytics to identify and adapt data-driven decisions and to follow the best strategies for online education. Data analysis could help extract useful elements that can be used to improve academic performance. For example, data analysis can be used to predict dropout rates, provide personalized learning paths, and adjust learning resources.

The main objective of this study is to explore the role of data analysis in online education to improve student academic performance. By academic performance, we aim to reduce dropout rates, successfully complete online courses, provide a better understanding of the presented concepts, adequately assess student progress, and ensure early intervention when necessary. The current study starts with the concept of data analysis, then moves to data analytics as a broader field of research, and explores the concept of learning analytics (descriptive, diagnostic, predictive, and prescriptive analytics) and how these can be applied to improve online education.

Additionally, the presented case studies will explore the impact of educational data analysis on academic performance.

In this study, we used the definition of data analysis as defined by the Cambridge Dictionary, which refers to "*the process of examining information, especially using a computer, in order to find something out or to help with making decisions*" (Cambridge Dictionary, n.d.). Likewise, the concept of online learning, often used

*, **, ***, **** Bucharest University of Economic Studies, Bucharest, Romania. E-mail addresses: marian.stan.ro@gmail.com (M. Stan), mihai.ciobotea@gmail.com (M. Ciobotea), mihaelacovrig2018@gmail.com (M. Covrig), dlbadea@gmail.com (D. L. Badea).

interchangeably with e-learning, Internet-based learning, or web-based learning (Maddison, T., Doi, C., Lucky, S. and Kumaran, M., 2017) , was defined as "*any learning experience that involves interaction with or is mediated by the use of digital technologies*" (OECD, 2016).

This study aims to explore the intersection between online education and data analysis, investigating the role data analysis plays in ensuring the student success. The proposed research question is the following:

RQ: How can data analysis contribute to better student engagement in online learning environments?

The current paper is further structured in the following sections: Literature review, which presents an analysis of the main works addressing the topic under analysis; Methodology, which sets out the research methodology used; Case Studies, that includes examples of data analysis and data from Moodle and Canvas LMS; Discussions, where we performed an analysis of the current state of play; the Conclusions section and Study Limitations and Future Research Directions.

2. Literature review

Data analysis has become an essential component of any business. There is a growing interest from the academic community in this subject in multiple fields of research, such as medicine, computer science, and engineering. A query in the Web of Science (WoS) on the "data analysis" topic, with no other filters, has returned 276,513 results that included 215,979 articles (78%), 48,284 proceeding papers (17.4%), and 32,417 reviews (4.5%). The highest number of indexed studies (23,841) were published in 2022. Similarly, the same query in SCOPUS has returned 561,192 results that included 409,207 articles (72.9%), 84,395 proceeding papers (15%), and 12,609 reviews (5.8%). The highest number of indexed papers (39,658) were published in 2021.

The authors have reviewed the most cited papers from 2014 from both WoS and SCOPUS returned by our queries in these databases. After an initial review, the authors have removed a number of studies from the original list that were not entirely relevant to the topic of this research.

Most cited papers from the SCOPUS database

Integrating data analytics in online education has become increasingly vital as institutions aim to enhance academic performance through learning analytics.

Waheed et al. (Predicting academic performance of students from VLE big data using deep learning models, 2020) demonstrated how deep learning can be utilized to predict student performance based on data from virtual learning environments (VLEs). The model proposed by these authors has used clickstream data from VLEs and has achieved significant accuracy (84%-93%) in identifying at-risk students. Furthermore, the authors have shown that prediction models based on neural networks are superior comparing with the ones based on logistic regression and support vector machine (SVM). This supports early intervention and personalized support in educational settings.

Hernández-Lara et al. (2019) applied data mining and learning analytics techniques to explore interactions within business simulation games of 362 students, revealing how these interactions via online discussion forums correlate with learning outcomes. Their research offers deeper insights into student learning processes beyond traditional metrics like grades. The authors have shown that the learning results seem to be influenced by content details related to collaboratoin, communication, interaction, uncertainity and time. The authors have also highlighted the potential of data mining in personalizing educational content. Tao et al. (2022) have used a data set from AdelaideX platform and focused on how engagement and sentiment analysis can predict academic outcomes in large MOOC environments, using ensemble methods for analysis. The authors have stated that detecting stress levels among students from MOOC data turned out to be challenging.With techniques such as TensiStrength there may be additional possibilities to measure stress. The authors have found that stress and negative sentiment didn't impact much the students academic results. Furthermore, the study hasn't identified any signifficant increase in stress due to pandemic in 2020 data cmparing with 2021 data. The study emphasizes the importance of analyzing online engagement data to develop predictive models for student performance, providing insights that can aid in improving educational outcomes. Aljohani et al. (2019) have used an Open University Learning Analytics dataset and have used a long short-term memory (LSTM) model to analyze clickstream data to predict students' academic performance. With the new model, the authors have achieved a high precision (93.46%) and recall (75.79%). Analyzing the data on weekly segment can provide educators with insights regarding predicting at-risk students. This illustrates the effectiveness of machine learning models in classifying students at risk, thereby supporting data-driven decision-making. Ulfa and Fatawi (2020) investigated how various student activities, tracked through learning management systems (LMS), influence educational outcomes. The authors have used a linear regression method for data analysis. Their research confirmed that certain interactive tasks significantly enhance learning results, indicating that data mining techniques to track student behavior can guide the development of more effective instructional strategies. Rossi (2020) discussed the impact of self-regulatory processes and feedback on learning in online settings, highlighting the importance of these factors in student performance. The author has underscored the role of metacognitive strategies and feedback mechanisms in improving student outcomes in online learning environments. Zulfikri et al. (2021) have analyzed how the transition to online learning during the COVID-19 pandemic affected students' academic performance. The authors have used linear regression models in order to research this. Their research provide insights into the challenges and opportunities associated with online education during pandemic. Maraza-Quispe et al. (2022) implemented a predictive model for students at-risk using the KNIME platform. The authors have used 22 behavioral indicators and, based on Simple Regression Tree Learner training algorithm, they have forecasted students academic performance. This demonstrates how behavioral data from LMS interactions can enhance educational interventions and identify students at academic risk. Alsabhan (2023) developed a machine-learning model to detect cheating in online examinations, thus addressing a significant challenge in digital education. The proposed model has used the long short-term memory (LSTM) technique and it had an accuracy of 90%.

Most cited papers from the WoS database

Joksimovic et al. (2015) have shown the advantages and the potential of learning analytics and trace data in improving educational outcomes in online learning environments. They have used multilevel linear mixed modeling techniques and investigated the effect of the frequency and duration of various student interaction types on learning outcomes. The interaction types were defined as follows: student - student, student - instructor, student - system, and student – content. The authors have analyzed a large dataset that included trace data from six years and 29 courses and found that student-system interactions have positively influenced final course grades, while the frequency of student-content interactions was negatively corelated with performance. These research shows the importance of customized pedagogical strategies and suggest that institutions develop systematic approaches to foster effective online interactions. The study has shown the potential of learning analytics and trace data in improving educational outcomes in online learning environments.

Araka et al. (2020) conducted a systematic review to examine trends in tools for measuring and promoting self-regulated learning (SRL) in e-learning environments between 2008 and 2018. The review identified educational data mining (EDM) and learner analytics as emerging techniques for supporting SRL but noted that traditional classroom methods were still predominantly used in measuring SRL. The authors emphasized the need for further research into EDM to improve the accuracy of SRL interventions in online learning platforms.

Injadat et al. (2020) proposed an ensemble model selection approach for predicting student performance in e-learning settings. By utilizing various machine learning algorithms and feature selection techniques, the study demonstrated that ensemble models achieved high accuracy in predicting performance at different stages of course delivery. This research contributes to the development of robust predictive systems that aid educational decision-making.

Cerezo et al. (2017) focused on procrastination in computer-based learning environments (CBLEs) and its impact on academic performance. Using data mining techniques, the study found that procrastination was a key predictor of learning failure, especially in blended learning programs using platforms like Moodle. The authors suggested that interventions addressing time management and procrastination could improve learning outcomes.

Yousafzai et al. (2021) proposed the "Student-Performulator" a hybrid deep neural network model leveraging attention-based Bidirectional Long Short-Term Memory (BiLSTM) for student performance prediction. By utilizing historical academic data, their model achieved a prediction accuracy of 90.16%. The study highlights the importance of deep learning for automatic feature extraction, improving performance over conventional methods. The advanced sequence learning capabilities of the BiLSTM model, combined with an attention mechanism, significantly outperformed state-of-the-art models, offering valuable insights for early prediction of student success in e-learning platforms.

Akram et al. (2019) investigated academic procrastination in a blended learning course using data from homework submissions. Their model, called SAPE (Student Academic Performance Enhancement), applied k-means clustering and various classification methods to predict procrastination tendencies and academic performance. The study emphasized the impact of procrastination on student achievement, demonstrating that higher procrastination tendencies correlate with lower academic performance. The application of educational data mining (EDM) to analyze submission behavior provides a powerful tool for identifying students at risk of poor performance.

Hooda et al. (2022) examined the use of artificial intelligence (AI) in assessment and feedback to enhance student success in higher education. The study compared several AI techniques, with the I-FCN (Improved Fully Connected Network) outperforming other machine learning models like ANN, SVM, and Random Forest. The research underscores the potential of AI-driven assessment strategies to provide immediate, data-driven feedback, improving learning outcomes. This shift, driven by the rise of online education during the COVID-19 pandemic, highlights the growing role of AI in education, enabling more reliable and constructive assessment processes.

Aljohani et al. (2019) have used an Open University Learning Analytics dataset and have used a long short-term memory (LSTM) model to analyze clickstream data to predict students' academic performance. With the new model, the authors have achieved a high precision (93.46%) and recall (75.79%). Analyzing the data on weekly segment can provide educators with insights regarding predicting at-risk students. This illustrates the effectiveness of machine learning models in classifying students at risk, thereby supporting data-driven decision-making.

He et al. (2020) developed a novel joint neural network using Recurrent Neural Networks (RNN) and Gated Recurrent Units (GRU) to predict student performance in virtual learning environments. By combining static and sequential data, the model achieved over 80% prediction accuracy by the end of the semester. The study highlighted that simpler models like GRU and RNN outperformed the more complex LSTM in some cases, suggesting the importance of model selection based on the nature of the data and task.

Poudyal et al. (2022) proposed a hybrid 2D Convolutional Neural Network (CNN) model for predicting student academic performance. By transforming 1D educational data into 2D image data, the model outperformed traditional machine learning models like k-nearest neighbor, decision trees, and logistic regression in terms of accuracy. This innovative application of CNNs in educational data mining suggests that deep learning techniques can be adapted to the pedagogical domain, providing new avenues for improving student performance prediction.

Based on the papers retrieved from WoS and SCOPUS, we have used VOSviewer, version 1.6.19, Leiden University, The Netherlands (Van Eck and Waltman, 2010), to perform co-word analysis in order to identify notable key terms associated with our research topic.

In Figures 1 and 2 below, we have used VOSViewer to display the co-occurrence keywords analysis. The illustration of co-occurrence keywords has been done separately for WoS and SCOPUS in order to be able to illustrate the differences among these two major databases.

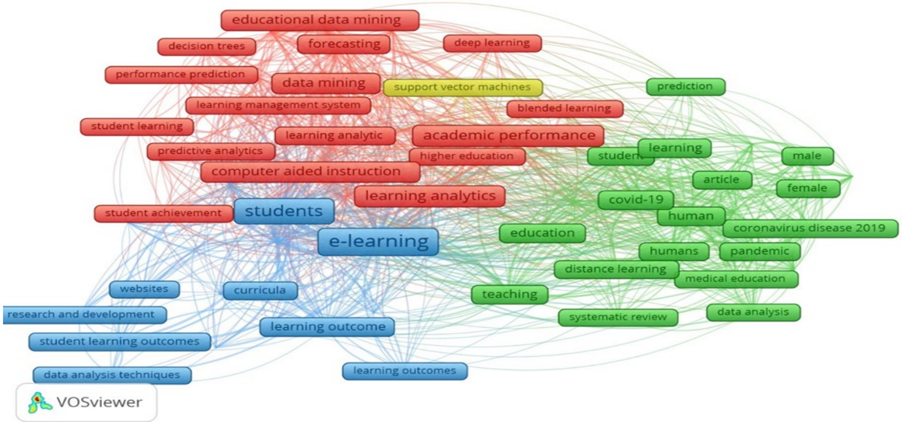

Figure 1. The keyword co-occurrence map for the SCOPUS papers

Source: generated by authors with VosViewer

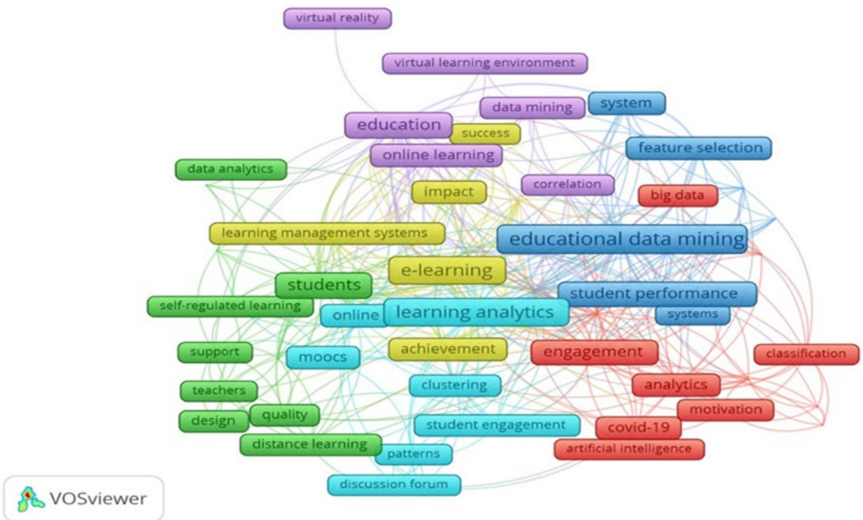

Figure 2. The keyword co-occurrence map generated for the WoS papers

Source: generated by authors with VosViewer

The literature review offers an overview of recent progresses in applying data analytics, machine learning, and educational data mining to enhance online learning and predict student performance. The studies use various methodologies and focus areas however, that seem to suggest the important role of data analysis and analytics in improving educational outcomes in virtual learning environments.

Integrating data analysis, analytics and machine learning techniques with online learning can have a positive impact on enhancing academic performance and student engagement.

The major themes identified were: Predictive Modeling for Student Performance; Importance of Student Interactions; Psychological Factors Affecting Learning; Stress and Sentiment Analysis; Technological Advancements in Education and Impact of External Factors like pandemic. The use of co-word analysis with VOSviewer further identifies key terms and trends associated with the research topic, illustrating the evolving landscape of online education and data analytics.

3. Methodology

The objective of this research is to explore the data analysis in online learning environments with an aim to ensure student success. The authors adopted a qualitative research approach, utilizing a literature review and case study analysis. The goal is to better understand the effectiveness of various data analytics tools and their impact when paired with relevant data analysis techniques.

The qualitative approach enabled an in-depth exploration of existing research and case studies focusing on the intersection of data analytics and education. A literature review was conducted to identify key academic sources on data analysis in the context of online education, complemented by case study analysis of real-world LMS applications.

Literature review: The literature review involved querying the academic databases Scopus and Web of Science (WoS) to identify peer-reviewed open-access journal articles, conference papers, and review articles discussing data analysis in online education, published in English. The focus was on literature from 2014 onwards, capturing the latest developments in data analysis for online learning.

The search in WoS database has returned 70 documents, using the following query:
"data analysis" OR "data analytics" OR "learning analytics" OR "educational data mining" (Topic) AND "online education" OR "e-learning" OR "distance learning" OR "virtual learning" OR "web-based learning" (Topic) AND "tools" OR "software" OR "platforms" OR "methods" OR "techniques" (Topic) AND "academic performance" OR "student achievement" OR "learning outcomes" (Topic) AND Article or Proceeding Paper or Review Article AND 2014-2024 AND Open Access AND English.

Similarly, the SCOPUS database has returned 163 results, using the following query:
(TITLE-ABS-KEY("data analysis" OR "data analytics" OR "learning analytics" OR "educational data mining") AND "online education" OR "e-learning" OR "distance learning" OR "virtual learning" OR "web-based learning") AND "tools" OR "software" OR "platforms" OR "methods" OR "techniques" AND "academic performance" OR "student achievement" OR "learning outcomes") AND PUBYEAR > 2014 AND PUBYEAR < 2024 AND (DOCTYPE "cp" OR DOCTYPE "ar" OR DOCTYPE "re") AND LANGUAGE "English" AND OA "all".

The top 10 most cited documents were analyzed, and irrelevant ones were excluded. Previous studies have used citation counts as a metric for assessing research impact, but the validity of this metric is still debated (Benda & Engels, 2011; Kapoor & Jain, 2024; Coryn, 2007; Aksnes, Langfeldt, & Wouters, 2019).

Case Studies: Case studies were selected based on their relevance to data analysis techniques and their demonstrated impact on academic performance. Each case was analyzed to extract insights into how these tools and techniques were applied, and what outcomes were achieved.

The design framework of this study is outlined in Table 1 below.

Table 1. The design framework

Data Type(s)	Unit(s)	Variables	Longitudinal study (yes / no)	Themes
Qualitative: Literature review based on the query in the Web of Science (WoS) and the SCOPUS databases	Research papers from WoS and SCOPUS databases: Articles, Proceeding Papers, Review Articles	Main findings and themes in data analysis for online learning; Major tools and techniques related data analysis.	No	Data analysis role; Data analytics frameworks; Learning analytics applications; Prediction models
Qualitative: case studies of two LMS providers	Learning Management Systems (LMS) case studies (Canvas, Moodle)	Tools and software used; Methods of using analytics in the educational process	No	Real-world application of educational data mining
Qualitative: content analysis of LMS providers websites	Websites of major LMS providers (e.g., Instructure - Canvas, Moodle)	Major features and offerings of LMS related to data analytics, educational data mining, and learning analytics	No	Data-driven insights into LMS functionalities; Impact on academic performance and student outcomes

Source: Elaborated by authors

Explanation:

Data Type(s): Refers to the nature of the information collected and analyzed in the study. This could include qualitative data derived from a literature review, case studies, or content analysis of online sources. In this study, the data types are primarily qualitative, focusing on literature analysis and case studies of Learning Management Systems (LMS) like Canvas LMS and Moodle.

Unit(s): Represents the sources or subjects from which data is gathered. For example, in the context of this study, the units could be research papers (articles, proceeding papers, review articles) found in databases like Web of Science (WoS) and SCOPUS, LMS platforms like Canvas and Moodle for case studies, or LMS provider websites for content analysis.

Variables: Variables are the key factors or metrics that are analyzed in the study. For this research, variables include findings related to data analysis in online learning, tools and techniques used in data analytics, and their impact on student performance. These could be specific methods of data analysis, the types of platforms used, or measures of academic performance such as learning outcomes or student achievements.

Longitudinal Study: This column indicates whether the study tracks changes over time. As clarified in the explanation, this study is not longitudinal—it does not examine trends or developments over a period of time. Instead, it provides a snapshot based on the current research and content available on e-learning platforms. The focus is on the present tools, techniques, and outcomes rather than how they evolve over time.

Themes: Themes are the overarching topics or categories of analysis that emerge from the research. In this study, the themes relate to how data analytics is applied in online learning, the effectiveness of different data analysis techniques in predicting academic performance, the role of learning analytics frameworks, and the use of educational data mining in real-world LMS environments. Other possible themes might include the comparison of LMS platforms, AI literacy outcomes, and student success metrics.

4. Case Studies

Moodle Analytics

This case study explores how Moodle utilizes data analysis, data analytics, and learning analytics to improve academic performance. Moodle is one of the most customizable and trusted e-learning solutions in the world. Additionally, Moodle is an open-source LMS. By using Moodle, high-quality online educational experiences can be provided (Moodle, 2023a). The company Moodle was founded in 1999 by Martin Dougiamas. This solution was created to offer educators a way to deliver quality courses via the internet, overcoming the limitations of a "brick-and-mortar" classroom (Moodle, 2023b). The software solution has been widely adopted globally and currently serves over 393 million users, with 2.3 billion course enrollments across more than 46 million courses in 42 languages, hosted on 167,000 Moodle sites. Additionally, Moodle offers a range of built-in reports primarily based on descriptive log data—they tell participants what happened, but not necessarily why (Moodle Docs, 2023a).

Key features include: 1) Grades and Gradebook: Educators have the ability to monitor students' progress and performance on assignments and tests using the Gradebook. It provides direct access to educational data analysis regarding student progress (Moodle Docs, 2023b). 2) Tracking Progress: Moodle offers badges as gamification elements to track and motivate student progress. 3) Course Reports: By offering built-in reports such as activity reports, participation reports, and general course logs, these help educators analyze which educational resources are most accessed and which activities generate the most interaction (Moodle Docs, 2023a).

Example of application: An educator notices through the Gradebook that several students are underperforming in quizzes. By analyzing activity completion reports, the teacher identifies that these students have not engaged with key learning materials. The educator intervenes by providing additional resources and support, leading to improved student performance.

Data Analytics at Moodle

Using the Gradebook, a teacher observes that a certain number of students are underperforming in the online course tests. Through this solution, the teacher can analyze the activity completion reports. Following this analysis, the teacher identifies that these students have not interacted with the essential learning materials provided on the online learning platform. As an intervention, the teacher could offer these students additional educational resources and support. These interventions can lead to improved student performance.

Example of application: A university integrates the Moodle solution with a third-party analytics tool to analyze the level of student engagement in online courses. By examining patterns of access to educational resources and participation in online discussion forums, the university identifies topics that require additional instructional support. Additionally, the university or educators can adjust the curriculum accordingly.

Learning Analytics in Moodle: Learning analytics uses data analysis specifically to improve online learning processes and outcomes. Moodle's Learning Analytics API offers machine-learning algorithms. When implemented or used correctly, these algorithms aim to predict and enhance student performance.

The key features of this implementation may include: 1) Predictive Models: Models like "Students at risk of dropping out" analyze historical and current data to predict student outcomes. These models can be customized to match the educational priorities of the institution. 2) Insights and Actions: Educators receive proactive notifications (insights) about students who may be at risk, along with suggested actions such as sending personalized messages or offering additional resources. 3) Customization: Institutions can create custom analytics models using Moodle's open system, allowing for tailored predictive learning analytics (Moodle Docs, 2023c).

Case Scenario

Context: An online educational institution aims to reduce student dropout rates and enhance academic performance across its programs.

Implementation

1. Enabling Learning Analytics

The institution activates Moodle's Learning Analytics feature, enabling predictive models like "Students at risk of dropping out" (Moodle Docs, 2023c).

2. Configuring Models

Models are configured to match the institution's educational priorities, focusing on key indicators of student engagement and performance.

3. Training the Models

Machine-learning models are trained using historical site data, including activity logs, course completions, and assessment results (Moodle Docs, 2023e). The machine-learning models require sufficient historical data to generate accurate predictions.

4. Generating Predictions

As courses progress, the models generate insights by predicting which students are at risk, allowing timely interventions.

Outcomes

- ✧ Early Intervention Educators receive insights about at-risk students and use Moodle's messaging system to offer support, guidance, and additional resources (Moodle Docs, 2023d).
- ✧ Monitoring and Adjustment: The institution monitors model performance and adjusts configurations to improve accuracy and effectiveness (Moodle Docs, 2023c).
- ✧ Positive Impact: The proactive approach leads to a reduction in dropout rates and improved academic performance, demonstrating the effectiveness of learning analytics.

Impact

- ✧ Enhanced Student Support: Students benefit from personalized interventions, fostering a supportive learning environment.
- ✧ Data-Driven Decision Making: Analytics data informs strategic decisions on curriculum design, resource allocation, and pedagogical strategies.
- ✧ Continuous Improvement: The institution refines its learning analytics models, contributing to ongoing enhancements in educational delivery.

Conclusion: Moodle's integration of data analysis, data analytics, and learning analytics exemplifies how online education platforms can leverage data to improve academic performance. By providing tools for basic data inspection, supporting advanced analytics through plugins and integrations, and offering robust learning analytics capabilities with predictive modeling, Moodle enables educators to make informed, data-driven decisions. This comprehensive approach aligns with Moodle's mission to provide accessible, quality education worldwide.

Canvas LMS

Instructure, the company behind Canvas LMS, has been at the forefront of this transformation since its inception in 2008. Founded by two enterprising graduate students, Instructure has grown into a leading education technology company, supporting over 30 million educators and learners across over 6,000 organizations worldwide (Instructure, 2024a). Canvas LMS, launched in 2011, disrupted the traditional LMS landscape by offering a cloud-native, open, and extendable platform designed to meet the diverse needs of K-12 and Higher Education institutions. Instructure's commitment to continuous innovation is evident through its strategic acquisitions and the expansion of its product suite, which now includes robust assessments, actionable analytics, and dynamic course content solutions (Instructure, 2024a). The company has over 7000 global customers (Instructure, 2024b). Besides Canvas LMS, the company offers the following solutions:

Offering	Description
Canvas LMS	The core platform providing flexible tools for creating, managing, and delivering content for all forms of instruction (online, blended etc).
Canvas Studio	An online platform for video materials that is integrated with Canvas LMS, enabling video content management and interactive courses.

Offering	Description
Canvas Catalog	A course catalog and registration solution that helps organizations to create branded marketplaces for course offerings delivered via internet.
Canvas Network	A platform for hosting and delivering Massive Open Online Courses (MOOCs) to a larger population of students beyond the institution's students.
Canvas Credentials	A digital badging solution providing learners with verified, portable credentials and stackable skill pathways.
Canvas Student Pathways	A solution for guiding students via personalized learning paths leading to digital credentials, certifications and employment readiness.
Mastery Assessment	A suite of assessment tools including Mastery Connect for assessment management, and content solutions for classroom instruction and regulatory exams.
Impact	Software solutions to promote the usage of new technology tools and measure their impact on the engagement levels and learning outcomes.
Elevate Data Sync	A data synchronization tool ensuring seamless data, grade, and roster integration between educational products and students information systems.
Elevate K-12 Analytics	A data analytics software that combines multiple data sources to provide visual insights for program and student performance improvement.
Elevate Data Quality	A solution ensuring district data is accurate, complete, and up-to-date by quickly identifying and resolving data issues.
Elevate Standards Alignment	Tools for content providers to align educational content with standards, improving content discoverability and market reach.
LearnPlatform	Tools that enable educational institutions to assess, manage, and measure the effectiveness of online learning applications.

Source: Instructure Holdings, Inc. 10-K Report, 2024

Canvas' platform is designed to support lifelong learning, offering scalable and adaptable solutions for institutions from K-12 to Higher Education, with a particular focus on continuing education and online learning initiatives. The company's SaaS model allows customers to rapidly deploy the system, avoid costly IT infrastructure, and enjoy regular updates with minimal disruptions. Their platform's reliability and scalability have positioned Canvas as a leading LMS in the U.S. and globally (Instructure Holdings, Inc. 10-K Report, 2024).

Analytics capabilities of Canvas LMS: Analytics assess various aspects of a course and analyze student performance by taking a comprehensive, three-pronged approach to generate meaningful insights for Canvas users.

Justification: This component focuses on evaluating system reports to understand how the platform is being utilized, providing insights into the system's overall effectiveness.

Intervention: Analytics aims to identify at-risk students early and provide data-driven strategies to address their needs and improve their chances of success.

Learning: This aspect examines learning outcomes, evaluates the effectiveness of teaching methods, and monitors the time distribution between students who achieve competence and those who may be falling behind, ensuring a more balanced and effective learning environment.

Instructors can access course analytics to track and analyze the activities of students, observers, and course designers within the course. These analytics complement course statistics, providing a comprehensive view of engagement and performance.

Course analytics in Canvas has been illustrated in Figure 3 below.

Figure 3. Course analytics in Canvas

Source: Canvas Community, 2024

Course analytics provide valuable insights that help instructors enhance the learning experience. By using analytics, instructors can predict how students will respond to various course activities, identify those who are at risk and may need additional support, and assess the effectiveness of their teaching strategies in fostering student learning. Additionally, course analytics offer a quick overview of student progress and achievements, enabling instructors to make informed decisions and personalize their approach to better meet their students' needs.

5. Discussions

Analyzing data from online learning environments, mainly through learning management systems (LMS) such as Moodle and Canvas, reveals several critical insights into how data analytics can improve educational performance. As the adoption of e-learning has accelerated, particularly during the COVID-19 pandemic, educational institutions rely on data to understand and improve student engagement, performance, and retention. The purpose of the above case studies was to indicate the potential of data analysis, data analytics, and learning analytics to improve students' performance. Although the conclusions cannot be extended to all online learning platforms, the case studies have their merits in highlighting the role of data analysis in supporting more informed decisions. Identifying at-risk students early is one of the most important benefits of data analysis in online learning environments. Predictive models in Moodle and Canvas, such as "Students at risk of dropping out," offer educators insights that allow for early interventions. These interventions are crucial for improving student retention rates, as they offer targeted support to those struggling. Early detection and intervention are especially effective in addressing the challenges posed by distance learning, where students often face more isolation than traditional in-person education. Another critical point of discussion is the enhanced personalization of the learning experience. Data analytics enables educators to customize and customize their teaching strategies subject to students performance and engagement levels. Tools like Moodle's competencies, course completion, and badges help track student progress and motivate learners by recognizing their achievements.

Additionally, Canvas LMS offers real-time analytics that allows instructors to monitor engagement, predict responses to course activities, and adapt their teaching methods accordingly. This personalized approach not only improves student motivation but also improves academic results. Integrating third-party analytics tools within these platforms, such as plugins in Moodle or Canvas's Elevate Analytics solutions, expands the capabilities and the features of basic data analysis. These integrations offer a more comprehensive view of student behavior, engagement patterns, and performance. For example, institutions using third-party analytics tools in Moodle can track heatmaps of student clicks, time spent on resources, and engagement across different activities. This allows for a deeper understanding of which areas of the curriculum need improvement or additional support. Moreover, data analytics promotes data-driven decision-making at both the classroom and institutional levels. Educators and administrators can use insights from data analytics to make informed decisions regarding curriculum design, resource allocation, and instructional methods. For example, Canvas LMS's "Justification" feature evaluates system reports to provide insights into the effectiveness of the platform's usage. These insights inform decisions on optimizing the use of educational technologies and improving student outcomes. However, it is important to raise awareness of the challenges associated with implementing data analytics in online learning. While these platforms provide robust tools for data analysis, educators often face barriers such as data interpretation and actionable insights. Many institutions may need more expertise or training to effectively interpret complex data sets and convert them into meaningful interventions. Additionally, although learning analytics offers substantial potential, there remains the risk of data overload, where the large volume of collected data becomes overwhelming and challenging, making it difficult to prioritize interventions.

Finally, even if the tools available in online learning environments offer valuable insights and even nice visualizations, it is ultimately the task of the educator or decision-maker to find the best way to implement the decision to achieve the desired outcome. Even the most advanced technique, the most accurate model, and the most scalable solution may not substitute the ability of the educator to implement the system recommendations according to educational practices.

6. Conclusions

This study highlighted the interesting potential of data analytics in online learning. Data analysis may show its value in analyzing student engagement levels, academic performance, and making data-driven decisions. Through the utilization of various data analysis techniques, such as predictive modeling, machine learning, and educational data mining, it becomes more and more apparent the value that these tools can bring, in enhancing both learning experiences and academic performance.

From the case studies of Moodle and Canvas we have presented earlier, one of the key insights is the ability of these software solutions to integrate data analytics to predict and support student engagement levels and academic performance. Tools such as Moodle's Learning Analytics API and Canvas's comprehensive analytics features may help the professors to early identify students at risk of dropout from the course cohort. Another utilization is identifying the students with a poor academic performance. These makes possible a timely intervention. This represents a proactive approach that may enhance student retention rates and academic outcomes, addressing one of the primary challenges of online learning environments, which is student isolation and disengagements.

Another point to consider is that the integration of third-party analytics solutions further expands the capabilities of LMS platforms by providing detailed insights regarding student behavior, such as engagement patterns and resource utilization. These insights allow the professors to adjust instructional strategies, tailor their interventions, and optimize the course design for better educational experience.

It is important to mention that effective utilizations of these data analytics tools do not come without any challenges. As this study has shown, while LMS offer powerful data collection, analysis and visualization capabilities, the interpretation of this data and its concrete applications in educational strategies require a well-developed skillset. Educators and institutions must invest in continuous training and support to leverage the full potential of data analytics. Without this expertise, there is a risk of information overload, where vast amounts of data may hinder decision-making rather than enhance it.

Last but not least, data analysis can offer significant opportunities for creating more adaptive, personalized, and effective online learning environments. With the help of predictive models, machine-learning, and learning analytics, educational institutions can improve both student engagement and academic performance. Moving forward, a more strategic and informed use of data analytics tools, combined with the continuous professional development of educators, will become critical in maximizing the benefits of these technologies for online learning environments.

7. Study Limitations and Future Research Directions

This paper contributes to the research community by helping academics, researchers, decision-makers, and practitioners in enhancing their understanding on the possibilities offered by data analysis techniques to generate insights from the online learning environments. There are a number of inherent limitations in this study.

(1) A major limitation of this study is the absence of primary data collection. Due to time constraints and limited access to real-time educational data, this research relied solely on secondary data from academic literature and case studies. While this approach provides valuable insights, future research could benefit from the inclusion of primary data collection to validate and expand upon the findings presented here.

(2) The utilization of data from the WoS and SCOPUS databases may lead to the omission of pertinent publications that are not cataloged in it.

(3) Limiting the study to English-language open access articles may result in disregarding significant contributions that are not open access or are made in other languages.

(4) The ever-evolving nature of the research could cause a slight modification in the list of chosen papers to be included in the literature review process and new papers to be included in the topmost cited papers.

There is a need for future exploration in the field of data analysis and online learning, focusing on machine learning, deep learning, and advanced statistical models. This requires interdisciplinary studies integrating artificial intelligence, generative models, and advanced statistical modelling.

References

1. *Akram, A., Fu, C., Li, Y., Javed, M.Y., Lin, R., Jiang, Y. and Tang, Y. (2019). Predicting Students' Academic Procrastination in Blended Learning Course Using Homework Submission Data. IEEE Access, 7, 102487-102498. doi:10.1109/ACCESS.2019.2930867.*
2. *Aksnes, D., Langfeldt, L. and Wouters, P. (2019). Citations, citation indicators, and research quality: an overview of basic concepts and theories. SAGE Open, 9(1). doi:10.1177/2158244019829575.*
3. *Aljohani, N. R., Hassan, S. U., Ahmed, A. N. & Hardman, J. (2019). A deep learning-based approach for predicting academic performance using online learning systems. IEEE Access, 7, 106557-106570. doi:10.1109/ACCESS.2019.2924009.*
4. *Aljohani, N.R., Fayoumi, A. and Hassan, S.U. (2019). Predicting At-Risk Students Using Clickstream Data in the Virtual Learning Environment. Sustainability, 11(24), 7238. doi:10.3390/su11247238.*
5. *Alsabhan, A. (2023). Machine learning-based detection of cheating in online assessments: Enhancing academic integrity in virtual classrooms. Computers & Security, 115, 102544. doi:10.1016/j.cose.2022.102544.*
6. *Araka, E., Maina, E., Gitonga, R. and Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments-systematic review (2008-2018). Research and Practice in Technology Enhanced Learning, 15(1), p.6. doi:10.1186/s41039-020-00129-5.*
7. *Benda, G.G. and Engels, T.C.E. (2011). The predictive validity of peer review: a selective review of the judgmental forecasting qualities of peers, and implications for innovation in science. International Journal of Forecasting, 27(1), 166-182. doi:10.1016/j.ijforecast.2010.03.003.*
8. *Cambridge Dictionary. (n.d.). Retrieved September 20, 2024, from https://dictionary.cambridge.org/dictionary/english/data-analysis*
9. *Canvas Community. (2024). What Are Analytics in Canvas? Retrieved September 20, 2024, from https://community.canvaslms.com/t5/Canvas-Basics-Guide/What-are-Analytics/ta-p/88*
10. *Cerezo, R., Esteban, M., Sánchez-Santillán, M. and Núñez, J.C. (2017). Procrastinating behavior in computer-based learning environments (CBLEs). Frontiers in Psychology, 8, 1403. doi:10.3389/fpsyg.2017.01403.*
11. *Coryn, C.L.S. (2007). The use and abuse of citations as indicators of research quality. Journal of Multidisciplinary Evaluation, 3(4). doi:10.56645/jmde.v3i4.82*
12. *He, Y., Chen, R., Li, X., Hao, C., Liu, S., Zhang, G. and Jiang, B. (2020). Online At-Risk Student Identification using RNN-GRU Joint Neural Networks. Information, 11(10), 474. doi:10.3390/info11100474.*

13. *Hernández-Lara, A. B.-L.-L. (2019). Applying learning analytics to students' interaction in business simulation games: The usefulness of learning analytics to know what students really learn. Computers in Human Behavior, 92, 464-479. doi:10.1016/j.chb.2018.03.001*
14. *Hooda, M., Rana, C., Dahiya, O., Rizwan, A. and Hossain, M.S. (2022). Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education. Mathematical Problems in Engineering, 5215722. doi:10.1155/2022/5215722.*
15. *Injadat, M., Moubayed, A., Nassif, A.B. and Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. Knowledge-Based Systems, 200, 105992. doi:10.1016/j.knosys.2020.105992.*
16. *Instructure. (2024a.). Canvas LMS Overview. Retrieved September 20, 2024, from https://www.instructure.com/canvas*
17. *Instructure. (2024b). Our Story. Retrieved September 20, 2024, from https://www.instructure.com/about/our-story*
18. *Instructure Holdings, Inc. (2024). 0-K Report 2024. Retrieved September 20, 2024, from https://s28.q4cdn.com/381552498/files/doc_financials/2023/ar/INSTRUCTURE-HOLDINGS-INC-_10K_2024_V1.pdf*
19. *Joksimovic, S., Gasevic, D., Loughin, T.M., Kovanovic, V. and Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. Computers & Education, 87, 204-217. doi:10.1016/j.compedu.2015.07.002.*
20. *Kapoor, D. and Jain, A. (2024). Sustainable tourism and its future research directions: a bibliometric analysis of twenty-five years of research. Tourism Review, 79(3), 541-567. doi:10.1108/TR-11-2022-0540.*
21. *Maddison, T., Doi, C., Lucky, S. and Kumaran, M. (2017). Chapter 2 - Literature Review of Online Learning in Academic Libraries. In Maddison, T. and Kumaran, M. (eds.) Distributed Learning (pp. 13-46). Chandos Publishing. doi:https://doi.org/10.1016/B978-0-08-100598-9.00002-7*
22. *Maraza-Quispe, W., Lira, L. & Choquehuanca, J. A. (2022). Predictive models for academic performance using KNIME analytics platform: A case study of online learning behavior. Journal of Educational Technology Development and Exchange, 15(3), 90-107. doi:10.18785/jetde.1503.07.*
23. *Moodle. (2023a). Moodle LMS Overview. Retrieved September 20, 2024, from https://moodle.com/lms/*
24. *Moodle. (2023b). History of Moodle. Retrieved September 20, 2024, from https://moodle.com/about/history/*
25. *Moodle Docs. (2023a). Course Reports. Retrieved September 20, 2024, from https://docs.moodle.org/311/en/Course_reports*
26. *Moodle Docs. (2023b). Gradebook. Retrieved September 20, 2024, from https://docs.moodle.org/311/en/Gradebook*
27. *Moodle Docs. (2023c). Learning Analytics. Retrieved September 20, 2024, from https://docs.moodle.org/311/en/Learning_analytics*
28. *Moodle Docs. (2023d). Insights and Actions for At-Risk Students. Retrieved September 20, 2024, from https://docs.moodle.org/311/en/Using_insights*
29. *Moodle Docs. (2023e). Machine Learning Models in Moodle. Retrieved September 20, 2024, from https://docs.moodle.org/311/en/Machine_learning_models*
30. *Moodle Plugins Directory. (2023). Third-Party Plugins. Retrieved September 20, 2024, from https://moodle.org/plugins/*
31. *OECD. (2016). Innovating Education and Educating for Innovation: The Power of Digital Technologies and Skills. (O. Publishing, Ed.) Retrieved September 20, 2024, from https://www.oecd.org/education/ceri/GEIS2016-Background-document.pdf*
32. *Poudyal, S., Mohammadi-Aragh, M.J. and Ball, J.E. (2022). Prediction of Student Academic Performance Using a Hybrid 2D CNN Model. Electronics, 11(7), 1005. doi:10.3390/electronics11071005.*
33. *Rossi, P. G. (2020). The role of feedback and metacognition in enhancing academic performance in online learning environments. Journal of Learning Analytics, 7(7), 45-62. doi:10.18608/jla.2020.7351.*
34. *Tao, Y., Zhang, H. & Liang, T. (2022). Engagement and sentiment analysis in large MOOC environments: A predictive model of academic performance. Journal of Educational Computing Research, 60(4), 1234-1254. doi:10.1177/07356331211015343.*
35. *Ulfa, S. & Fatawi, M. H. (2020). Understanding student activity and interaction through learning analytics: A case study in a learning management system (LMS). Educational Technology Research and Development, 68(2), 145-164. doi:10.1007/s11423-019-09670-w.*
36. *Van Eck, N.J. and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538. doi:10.1007/s11192-009-0146-3.*
37. *Waheed, H. H. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189. doi:10.1016/j.chb.2019.106189*
38. *Yousafzai, B.K., Khan, S.A., Rahman, T., Khan, I., Ullah, I., Rehman, A.U., Baz, M., Hamam, H. and Cheikhrouhou, O. (2021). Student-Performulator: Student Academic Performance Using Hybrid Deep Neural Network. Sustainability, 13(17), 9775. doi:0.3390/su13179775*
39. *Zulfikri, M. A., Nordin, M. S. & Ariffin, A. (2021). Impact of COVID-19 on students' academic performance: An analysis using learning analytics and regression models. International Journal of Advanced Computer Science and Applications, 12(5), 244-255. doi:10.14569/IJACSA.2021.0120532.*