A Digital Twin Framework For Analysing Students' Behaviours Using Educational Process Mining

Ambrose Azeta

Covenant University   https://orcid.org/0000-0001-7142-7405

Frank Agono

Covenant University

Adesola Falade (✉ adesola.falade@covenantuniversity.edu.ng)

National Open University Lagos

ea Azeta

Covenant University

Vivian Nwaocha

National Open University of Nigeria School of Science and Technology

---

Research

Abstract

Learning management systems (LMS) logs all actions taken on the system. These logs provide additional data about the activities and behaviour of users. Educational process mining techniques can use these data to unveil useful information to help instructors, educators and administrators accurately monitor, analyze and improve the online learning patterns of students. This research work presents a framework that uses process mining approach to analyse event log data generated within educational information systems, such as LMSs. In this framework, digital twin concept is employed to present a virtual representation of the students' activities on the LMS. This framework also used inductive and fuzzy miner algorithms to produce a process model which was represented using virtual model of student's learning patterns. This model was then evaluated for conformance with the activities observed in the log. The analysis conducted during this study showed the disparity between the behaviours of students that passed a particular course and students that failed the course. Findings also showed that the using the inductive and the fuzzy miner algorithms produced better fitness levels for the process model when compared with other previously used algorithms such as the heuristic miner and alpha miner algorithms. This paper concluded by recommending that the development of educational process mining specific tools can help domain experts better understand students' learning patterns.

I. Introduction

The use of electronic learning platforms in institutions of higher learning has increased rapidly in most recent years. This rapid adoption is sponsored by the flexibility of the system, allowing users, students, teachers and administrators to perform educational activities from any location. This form of electronic based education is independent of a specific hardware platform [1]. These systems are expensive to setup and maintain, thus it has become important for school administrators to measure its impacts by striving to understand students use and behaviour on these systems, this will help them determine the efficiency of these systems and determine if they are meeting their objectives or not [2]. Analyzing the event logs data generated and stored in the database by these systems can help to understand educational information system and ultimately understand the activities and behaviours of students using them. The results of such analysis can be used to optimize learning experiences for students and lecturers as well and also improve the quality and academic performance of students.

Most of the recent literatures in this study has used factors such as students' previous results, level of participation in forums, academic qualifications, race and even age to classify and predict students' retention rates, admission qualification, students' drop out risk and students' final semester grades. The need to improve the mode of course delivery on these educational information systems calls for a direct assessment of processes and activities carried out on these systems.

Process mining techniques brings a different form of analysis that's based on the actual behaviour exhibited by these students' as they learn and interact with the educational information systems [3]. As institutions face market and competitive pressures, they are forced to find new methods to reach

maximum efficiency while still increasing customer satisfaction. Many companies are exploring the idea of creating a "digital twin" to help them drive improvements [4]. Many organizations today have invested heavily on long-term interests in the predictions of their organization's development. In a fast-paced world as we have now, and to ensure recurring success, organizations must be able to adapt to changes as quick as possible [5]. Consider the possibility where the activities and processes in an organization is represented digitally, this would allow seamless monitoring and analysis of the factors affecting the performance of the organization. In digital twin, the researchers consider the concept of representing the entire organizational processes visually. Allowing that managers could try all possible scenarios and influences, and select the best combination of inputs that guarantees maximum expected success in the models.

Digital Twin was conceptualized in 2003 by Dr Michael Grieves, during his presentation on Product Lifestyle Management (PLM) [6]. The scope of digital twin at that time was largely limited owing to limited technological advancements. The academic and industrial domains have formulated different explanations for the digital twin phenomenon. DT is largely regarded as the simulation of systems and products which uses historical data and real time data to mirror the life and behaviour of a corresponding twin [7]. [8] defines DT as a model that easily combines the behaviours of an autonomous systems and their physical world environments. Digital twin can also be defined as mapping of all components in the product life cycle using physical data, virtual data and interaction data between them [9]. Digital twin framework helps us to build the virtual representation of the educational process engaged by students. This helps us scrutinize the model represented, allowing the detection of anomalies, bottlenecks and find out ways to improve the actual process.

This study aimed at providing a digital twin model for the analysing students electronic learning behaviors using educational process mining techniques in order to ultimately improve the experience of students using electronic learning systems. The aim will be achieved through the following objectives: (i) extraction of event logs from the Moodle LMS and pre-processing the logs acquired for anonymization, (ii) application of the Inductive and Fuzzy Miner algorithms on the ProM Framework to produce a process model from the recorded event logs, (iii) development of the digital twin model by performing comparison evaluation of the produced process model for deviations and conformity with observed activities in acquired event logs; and (iv) evaluation of the produced model and provision of process improvement recommendations. This paper structured as follows: section two contains literature review. Section three describes research methodology. System design, and data description/implementation are presented in section four and five respectively. Section six highlights results and discussion, while section seven concludes the paper.

Ii. Literature Review

A Electronic Learning Processes

The explosion of the knowledge age has changed the context of what is learnt and how it is learnt – the concept of electronic learning processes is a manifestation of this knowledge revolution. Electronic educational processes refer to instruction in a learning environment where teachers and students are separated by time or space, or both, and the teacher provides course content through course management applications, multimedia resources, the Internet, video conferencing, etc [10,11]. Students receive the content and communicate with the teacher via the same technologies. Whether you have been looking to guide your own professional development or your organizations, you will undoubtedly have noticed the trend away from the traditional classroom learning experience. That tradition is being replaced by a wide range of technology-enabled learning modalities, from podcasts and webinars to formal, one-on-one coaching and a range of e-learning platforms. A key reason individuals and organizations are embracing technology-driven learning modalities today are the efforts of learning providers to ensure these newer formats deliver strong results.

Continuous learning and constant training have become vital for job holders and organizations alike to remain relevant and competitive in the global job market. 65% of higher education have been said to have identified electronic/online learning as a vital component of their long term plan for continued sustenance and relevance in providing quality education [12]. This wave of interest has led to massive introduction and extensive use of online learning management systems in various higher educational institutions. By the year 2012 the percentage of higher educational institutions in the US using some form of LMS had risen to a staggering 93% rising exponentially from the meagre 15% in 2000 [13]. A new and improved form of learning management systems is emerging in the market. These systems have superior functionalities over the previously used traditional LMS. These new LMS have embedded in them current forms of sharing information over social media, allowing more people to be reached and wider spread of information and course contents, a possibility which was limited in the old forms of LMS. These upgraded systems have extensive impact on traditional methods of running educational institutions and very much support current learning environments [14]. Non-traditional educational settings such as the military and corporate organizations has taken advantage of LMSs in order to train their members. Despite the different educational requirements spread across various industries and organizations, there seem to be a common ground regarding the expected outcomes from the use of learning management systems. LMSs have capabilities to integrate with other platforms or applications. These partnerships can be employed to meet certain business or organizational goals or objectives [15]. The topmost rated objectives for using LMS include content management, users assessment and finally, reporting [15].

The first iterations of learning management systems were the PLATO learning system which was developed at the University of Illinois during the 1960s, and the TICCIT System developed by the MITRE Corporation. As the earlier developed LMS moved from managing single lessons to a wider collection of lessons it became paramount to introduce a system to help properly manage these additional lessons. This need brought about the development of Integrated Learning Systems (ILS) and Course Management Systems (CMS). These new systems advertised new features such as reporting dashboards, pre-test and post-test techniques and students tracking as features that could improve learning experiences. Some

platforms had risen to market leader's status at the end of 2011. These platforms include Moodle, Pearson, Desire2Learn, Sakai and Blackboard [16].

B Moodle

Moodle, a very popular course management learning system helps educational institutions easily create, administer and manage electronic learning communities. Moodle is distributed free under open source licensing and has gathered countless university installations and use all around the world [17]. The Moodle application logs tracks of events and every activity performed on the system by students, instructors and all users of the system. Moodle has log viewing dashboard built into it allowing approved users view, manage and do other actions. The collected log data was also anonymized by removing all tables or identities that uniquely identifies users of the system. And then provide a random unique identifier for each row, which serves to represent the performers of the actions recorded on the event log.

C Educational Data Mining

One of the ways to enhance the quality of educational learning processes is the provision of useful knowledge previously hidden to the administrators of higher institutes of education. The use of interactive learning environments has resulted in the collection of huge volumes of data, as these environments log all user actions during the use of the system for learning [14]. This knowledge is useful to improve the quality of decision-making processes. This "knowledge" can be discovered from the enormous amount of data residing in the databases of these organizations and can be extracted for analysis use by applying data mining techniques.

Data Mining (DM) techniques have been widely used to find knowledge from huge datasets. Data mining is defined as a process that performs the extraction, classification and analysis of information from huge databases using statistical, mathematical and machine learning techniques [18]. DM helps to discover students' learning patterns [19]. Data mining is also known as a process of discovering or extracting useful knowledge from huge amounts of data stored in multiple data sources [20]. [21] defines DM as a process of discovering "tacit" knowledge and patterns within large amounts of data in order to make predictions of a possible future occurrence or outcome. The use of data mining methods in the educational domain has been termed Educational Data Mining (EDM).

EDM is a new body of research in the field of data mining and Knowledge Discovery in Databases (KDD) which is concerned with the mining relevant patterns and knowledge from educational information systems such as course management systems, registration and admission systems [22]. Educational data mining is defined as a growing body of study that is concerned with the development and execution of methods for exploring the unique data types that emerge from educational settings, and the use of this data to better understands student behaviours and the setting in which they learn the best [23,24,25]. EDM researchers focus on extracting valuable knowledge for use to help educational institutes manage students activities better, and to help improve the performance of its students [22]. The EDM process transforms raw data gotten from educational databases into useful information that could potentially

have huge impact on educational processes and practices. This process follows the same steps as the general Data Mining (DM) process; pre-processing and post processing. Traditional EDM techniques include classification, clustering, association-rule mining, sequential mining, text mining, etc. (Romero and Ventura, 2010). Classification is the most commonly used and also adjudged the most effective educational data mining technique; it is useful for the classification and prediction of values. Classification data mining technique was used in [22] for the analysis of students information collected through surveys, the goal was to provide classifications based on the collected data to predict students' performance in the upcoming semester.

Despite the many successes the use of classical EDM techniques have recorded in the educational domain, it however does not come without its shortcomings, a major one of such is that classical EDM techniques do not consider process flows and cannot be used for the discovery of workflows. There exist factors that limit the value of a data-driven approach if there is no theory that guides such process. Some authors, [26] and [27] regard recommendation as a major salient type of Process mining. Some authors however choose to see Recommendation as an extended for of enhancement.

Iii. Methodology

In order to achieve the first objective for this research work, the extraction of event logs from educational information systems was carried out on the Moodle platform. The administrator logs into the system, navigates to the administration tab, then course administration, click on reports and then logs. The administrator is provided with course title, participants' type, length of days, activity types, action types and event types options to choose from. After making preferred choices, the administrator proceeds to display logs by clicking on the "Get these logs" button. Once the logs have been displayed, the administrator selects his/her preferred download output, then clicks to download. The output can be formatted into different extensions (e.g. CSV, ODS, TXT). The collected log data was also anonymized by removing all tables or identities that uniquely identifies users of the system. And then provide a random unique identifier for each row, which serves to represent the performers of the actions recorded on the event log. The ProM framework cannot directly utilize the generated output from the Moodle system. The output from Moodle usually comes in CSV. Process mining software have preferred file formats and as such the CSV file must be into the MXML (Mining eXtensibleMarkup Language) or XES (eXtensible Event Stream).

For this research, our preferred format type is XES, because it is the most updated and comprehensible of the two available log types for process mining. This conversion is performed by first importing the log files into the ProM Framework tool. Clicking the import option available at the far-right side of the application interface opens a dialog that allows to select the intended log file from the computer storage. After a successful import, the ProM Framework then invokes the "Convert CSV to XES" plugin, which then converts the log file to XES, making it ready for further process mining analysis. Achieving the second objective involves the application of the Plugins for the two selected algorithms (inductive and fuzzy miner algorithms). This would serve as a form of comparison between the two selected algorithms and

the two widely used algorithms for process discovery, the heuristics miner algorithm and the alpha miner algorithm.

For the third objective, the conformance checking analysis of process mining will be performed on the process model gotten from achieving objective two to determine the conformity of the process model to the predefined curriculum and observed activities in the event logs. This action helps to detect bottlenecks and deviating instances in the use of the information system. The tool be the engaged is also the ProM Framework. Lastly, in achieving the final objective, the model obtained are exported, interpreted and used to proffer recommendations. The administrators use the acquired information to produce recommendations tailored towards improve the students' learning experience.

Iv. System Design

Figure 1 depicts the workflow of the overall system. It comprises of two (2) approaches, the existing approach, and the improved approach. The traditional approach involves collecting event log data from the Moodle LMS system. This activity is done by having a user with admin level access logging into the Moodle platform and navigating to extract the needed event logs captured by the system during interaction with users. The Alpha Miner, Heuristic Miner, Fuzzy Miner and Inductive Miner Algorithms on the ProM Framework were used to produce process models from the acquired event logs. Once this was achieved, the plugins for the conformance checking on the ProM Framework would be called to perform conformance analysis on the produced process model displaying the level of compliance of the produced process model with the extracted event logs.

The proposed framework in Figure 1 also begins with the extraction of event logs from the Moodle platform. The application of process discovery algorithms Fuzzy and Inductive miner algorithms will follow. Afterwards, the digital twin model would be developed by performing comparison evaluation of the produced process model for conformity and deviations with the expected activities. Once this is achieved, the results of the analysis would be used to generate recommendation based on the data processed. The following subsections further describes the methodology.

A Event Logs

This study used data from 115 undergraduate e-learning enthusiastic students, who completed an online course using the corporate LMS Moodle. The students participated in a trial course and test program with the aim of capturing the student's interactions with the system for analysis use. After data extraction from the Moodle LMS, and upon pre-processing which included identification of the students records to ensure anonymity, a total of 19,275 were gathered from the interactions of the 115 students with the system. The derived dataset was then used as input for the process discovery stages of process mining. The process models derived was then compared following the application of three process discovery algorithms to determine their fitness.

B Process Mining

Process mining allows educators, instructors and administrators to discover new forms of data analysis using educational datasets. Traditional methods of analysis (manual data analysis) provides limited insights, process mining can help trace the students' activities on the LMS, thus providing a more granular insight into the students learning behaviours.

C Process Discovery

This study compared the results of the four algorithms used for process discovery process mining analysis: these algorithms are the Alpha Miner, Heuristic Miner, Fuzzy Miner and the Inductive Miner algorithm.

- **Alpha Miner (AM)**: was the first recognized process discovery algorithm. Its deficiencies and shortcomings help informed the development of later and better performing algorithms. Its main recorded shortcoming was its inability to use frequencies, thus its unable to guarantee "sound" process models. Sound in this case means, is a property that a process model should to be able to reach the end state from its start state without bugs. The AM is also most suitable for noiseless event logs.
- **Heuristic Miner (HM)**: The HM which was introduced by (Bogarín *et al.*, 2014) recorded slightly better results than the Alpha Miner algorithm. The HM has three major advancements over the AM. It is able to filter out noisy and infrequent behaviors from event logs with its ability to use frequencies. It is also able to skip mis matching single activities and can detect short loops amongst event data. All of these makes the HM less sensitive to noise and incomplete logs. In spite of the HM's advancements over the AM, it also does not guarantee sound process models.
- **Fuzzy miner (FM):** this is one of the newer process discovery algorithms. It is the first algorithm to directly address the problems of large numbers of activities and highly unstructured behavior [25].
- **Inductive Miner (IM)**: The IM is clearly an improvement over the AM and HM algorithms, as it guarantees "sound" process models. It does this by first finding the most prominent split in an event log, then detects the relationship between the splits and then continues on both the split logs. It also has the ability to cope very well with large event logs and infrequent behaviour between them.

All of the above actions were performed on the process mining tool ProM Framework, which is adjudged the most suitable for process mining projects in educational domains [26].

D Conformance Checking

The conformance checking activity describes the activity of checking if the behaviour in the produced process models matches with the behaviour observed in the event logs. Conformance analysis is useful to detect students' learning behaviour anomalies and possible bottlenecks encountered by students and other users during use of the learning management system. Results of the conformance analysis stage can help determine where exactly these bottlenecks are and their severity if any. After the process discovery stage, this project proceeded to first use the *Linear Temporal Logic Analysis* feature of the

ProM Framework. This activity is performed by calling the *LTL Checker plugin* that checks if the event log used satisfies a predefined set of properties expressed in LTL Logic. There exists a set of formula already predefined in the LTL Checker plugin. However, these formulas can be customized to fit user requirements as the need arises. A quick example of formulas in the LTL Logic. To perform a check to determine if students are required to perform the task "Quiz review" before the task "Quiz review summary", the formula will be defined as:

formulaquiz_review_is_a_prerequisite_of_quiz_review_summary
c1: ate.WorkflowModelElement,
c2: ate.WorkflowModelElement) :=
{<h2> Is the activity Quiz review a prerequisite for the activity Quiz Review Summary? </h2>}
(<>(activity==c2)/\(activity!=c2 _U
activity==c1));

Equation*1: LTL Logic for conformance checking*

The second activity in the conformance checking stage is the application of the *Conformance Checker* Plugin. This plugin allowed us analyze the produced model precision, fitness and structure via log replay (which is the replay of the produced model on the original log), state space analysis and structural analysis. To perform this action, the *Conformance Checker* Plugin on the produced model was applied from the inductive miner process discovery stage taking the event dataset as input. The results of this stage were two traces which represented conforming and non-conforming traces.

V. Data Description And Implementation

This study used data from 115 undergraduate e-learning enthusiastic students, who completed an online course using the corporate LMS Moodle. The students participated in a trial course and test program with the aim of capturing the student's interactions with the system for analysis use. After data extraction from the Moodle LMS, and upon preprocessing which included the identification of the students' records to ensure anonymity, a total of 19,275 were gathered from the interactions of the 115 students with the system. The event log file extracted from Moodle serves as the starting point into the analysis process in this study. It is vital to preprocess the log files by filtering it as a real event log that often contains noisy data [24].

The students' private data such as names and IDs were deleted and each student was assigned a random id number so as to ensure anonymity and data privacy. Some attributes from the log file were deleted leaving only four (Full name, Action, Information and Timestamp). Some students' actions gathered on the log were also not considered for this analysis. Only actions that weredirectly relevant to the learning process and academic performance. Next the Excel CSV log file was transformed using the "Convert CSV to XES" plugin on the ProM Framework into the XES (eXtensible Event Stream) format which is the required by the ProM Framework for process mining. Table 1 and 2 contains the attributes of a moodle event log and action attributes respectively.

Table 1: *Attributes of a Moodle event log*

Attributes of a Moodle event log file	
Attribute	**Description**
Course	The name of the course
IP Address	The IP of the device used to access Moodle
Time	The date Moodle was accessed
Full Name	The name of the student
Action	The action that student performed
Information	More information about the action

For thorough analysis of students' behaviour, this study used students' test grades to group these students into two categories. The test scores ranging from 0 to 100 was divided into two using the range, 0 to 49.9 classified as "Fail" and 50 to 100 classified as "Pass". Clustering by grades is useful for comparing the performance of different algorithms and for producing slimmer process models that's more likely to be accurate in replicating the observed events from the event logs [25]. At the end of this preprocessing step, the data was divided into three subsets; All (containing events for all the 115 students that took the Moodle course), Pass (containing events of 64 students considered to have passed the course) and then Fail (containing event actions of 51 students that failed the course).

Table 2: *Moodle Actions Attributes*

Moodle Attribute Actions	
Low level Moodle Action	**High Level Codification**
assign submit	EXECUTING
assign view	PLANNING
forum add discussion	FORUM PEER LEARNING
forum add post	FORUM PEER LEARNING
forum update post	FORUM PEER LEARNING
forum view discussion	FORUM PEER LEARNING
forum view forum	FORUM PEER LEARNING
page view	LEARNING
quiz attempt	EXECUTING
quiz close attempt	EXECUTING
quiz continue attempt	EXECUTING
quiz review	REVIEW
quiz view	PLANNING
quiz view summary	PLANNING
resource view	LEARNING
URL view	PLANNING

At the end of this process the study then goes ahead to perform the conformance checking analysis on the optimized process model and also evaluate the framework and the outcomes of the various algorithms used.

The different interfaces engaged in the analysis process have been displayed as follows: Figure 2 shows the event log dataset used for this analysis in CSV format. Figure 3 is the process of converting CSV file to the expected file format XES for PRoM.

Vi. Results And Discussion

Figure 4 and 5 shows the behavior of both the students that passed and failed in attempt to watch course video and submit course quiz before deadline respectively. Figure 5 shows the students that passed the course follow a sequential order when submitting responses to their quizzes every week. However, there are few exceptions with quiz 1 and 2, quiz 6 and the final quiz. Future information from the trail revealed that some students jumped to quiz 6 right after taking the first quiz. This helped to explain why they failed the course as it shows that have spent wasted doing tests, they shouldn't have done just yet while failing to do the ones they should be doing. Our system also reveals the behavior of the clusters of students while watching a course video and submitting the corresponding quiz attached to the course within the specified deadlines for that week.

A. Process Analysis Evaluation

Its worthy to note that although Fuzzy miner produces very good and understandable process models, it does not provide enough information useful for evaluation and as such was not considered for evaluation. Table 3 shows the evaluation table.

Table 3: Evaluation table

Algorithm	Cluster	Trace Fitness
Alpha Miner	All	0.789
Heuristic Miner	All	0.567
Inductive Miner	All	0.867
Alpha Miner	Pass	0.651
Heuristic Miner	Pass	0.599
Inductive Miner	Pass	0.935
Alpha Miner	Fail	0.774
Heuristic Miner	Fail	0.632
Inductive Miner	Fail	0.927

This analysis was conducted using the Dotted Chart Analysis Technique in the ProM Framework. The x axis of the chart indicates the timestamp of the event (the period an event was executed) while the x axis

indicates the performers of the event (in this case, the students). The blue dots signify the first time a lecture video was watched, while the white dot signifies the time a quiz was submitted on first attempt. Events reoccurrence such as students watching a video again and students submitting quizzes in subsequent attempts is indicated using Yellow and Red respectively. The black line across the face of the chart indicates the specified soft deadline, while the white line indicates the hard deadline which is the final day of submission for quizzes after the initial period of grace after the blackline.

The study outcome shows that the cluster of students that passed the course tends to watch course videos first and then attempt the corresponding quiz before the hard deadline elapses. The study also found that after the deadline, some students in this cluster go ahead to watch those videos again for better understanding of what the video teaches. On the other hand, students of the fail cluster failed to keep to the set deadline. The dotted chart analysis also shows that their behavior could largely be unpredictable as some even watch videos after the related quiz has closed. Finally, the analysis further shows the behavior of students while submitting quizzes amongst the cluster that passed the course, this analysis reveals a fine pattern of students who strive to submit quizzes before the hard deadline. While noticing the trend from the analysis of students that failed, the course shows that a lot of these students fail to meet the soft deadline and a few of them still struggle to meet the hard deadline.

Vii. Conclusion

This research is focused on providing higher educational institutions with the ability to see through its data opportunities to effectively improve its educational processes, thus impacting on the quality of students it turns out as graduates. Expected at the end is an optimized process model useful for recommendation and improving the student, instructor interaction with educational information systems. The results in this current research are useful for studies in Open and Distance Learning [29, 30]. As part of recommendation and future works, it is essential to enhance the diagnosis of students' behavior through the development of more specific EPM tools in order help domain experts (i.e., educational specialists and researchers) properly analyze educational processes.

Abbreviations

LMS: Learning Management System

ODL: Open and Distant Learning

EPM: Educational Process Mining

DM: Data Mining

CSV: Comma Separated Values

Declarations

Availability of data and material

Not applicable.

Competing interests

The authors (Ambrose Azeta1, Frank Agono2, Falade Adesola3, A. E. Azeta4, Vivian O. Nwaocha5) declare that they have no competing interests.

Funding

Author's contributions

In this study, Azeta Ambrose was the main contributor in all phases from start to finish except for data collection. Frank Agono and Falade Adesola were assisting Azeta Ambrose during all phases of this study. Moreover, A. E. Azeta and Vivian O. Nwaocha had a vital contribution to data collection and drafting of the article. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Author information

Affiliations: Covenant University, Ota, Ogun State, Nigeria

Azeta Ambrose, Frank Agono, Falade Adesola, A. E. Azeta & Vivian O. Nwaocha

Corresponding author

Correspondence to Falade Adesola

References

[1] Romero C, Pedro G, Amelia Zafra, Jose Raul Romero, and Sebastian Ventura. (2013) 'Web usage mining for predicting final marks of students that use Moodle courses', *Computer Applications in Engineering Education*, 21(1), pp. 135–146. doi: 10.1002/cae.20456.

[2] Lorenzen S. and Alstrup, S. (2018) 'Tracking Behavioral Patterns among Students in an Online Educational System', *Proceedings of the 11th International Conference on Educational Data Mining*, pp. 2–7.

[3] Mukala P, Buijs J. Aalst W. (2015) 'Exploring Students ' Learning Behaviour in MOOCs using Process Mining Techniques'.

[4] Mellon, M. (2018) 'How Process Mining Enables the Digital Twin of an About the author'.

[5] Kathuria, S. (2018) 'TECHNICAL UNIVERSITY OF MUNICH TUM Data Innovation Lab A Company ' s Digital Twin'.

[6] Grieves, M. (2014) 'Digital Twin : Manufacturing Excellence through Virtual Factory Replication A Whitepaper', *Digital Twin White Paper*.

[7] Glaessgen, E. and Stargel, D. (2012) 'The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles', pp. 1–14. doi: 10.2514/6.2012-1818.

[8] Dazhong Wu, Lane J, David W, Dirk S. (2015) 'Towards a Cloud-Based Design and Manufacturing Paradigm: Looking Backward, Looking Forward', (May 2014), pp. 315–328. doi: 10.1115/detc2012-70780.

[9] Fei Tao, Fangyuan Sui, Ang Liu, Qinglin Qi, Meng Zhang, Boyang Song, ZirongGuo, Stephen C.-Y. Lu & A. Y. C. Nee. (2018) 'Digital twin-driven product design framework', *International Journal of Production Research*. Taylor & Francis, (February), pp. 1–19. doi: 10.1080/00207543.2018.1443229.

[10] Cole, J. D. R. and Foster, H. (2008) *Using Moodle, 2nd Edition*, *Journal of Chemical Information and Modeling*. doi: 10.1017/CBO9781107415324.004.

[11] Rice, W. (2006) *Moodle E-Learning Course Development*, *Communications of the ACM*.

[12] Elaine A. (2011) 'Going the Distance: Online Education in the United States', *Modern Healthcare*, 41(40), pp. 1–44. doi: 10.1049/el.2012.2018.

[13] Green, K. (2013) 'Campus Computing 2013', p. 35. Available at: http://www.campuscomputing.net.

[14] Watson, W. R. and Watson, S. L. (2007) 'An Argument for Clarity: What are Learning Management Systems, What are They Not, and What Should They Become?', *TechTrends: Linking Research and Practice to Improve Learning*, 51(2), pp. 28–34.

[15] Ellis R. K. (2009) 'Learning Management Systems Field Guide to Learning Management Systems', *ASTD learning circuits*, pp. 1–7. doi: 10.1097/AIA.0b013e3181e5c1d5.

[16] Hill, P. (2014) 'State of the Higher Education LMS Market: A Graphical View (2014 edition)', *e-Literate*, p. 2017. Available at: http://mfeldstein.com/state-of-the-higher-education-lms-market-a-graphical-view/.

[17] Manolis C, Stefanos G, Manolis M, Cleo S and Anastasios T. (2014) 'Improving Quality of Educational Processes Providing New Knowledge Using Data Mining Techniques', *Procedia - Social and Behavioral Sciences*. Elsevier B.V., 147, pp. 390–397. doi: 10.1016/j.sbspro.2014.07.117.

[18] Batware, J. B. (2007) 'Real-time data mining for e-learning'.

[19] Baker, R. S. J. (2008) 'Data Mining for Education'.

[20] Job, M. A. (2018) 'Data Mining Techniques Applying on Educational Dataset to Evaluate Learner Performance Using Cluster Analysis', *European Journal of Engineering Research and Science*, 3(11), pp. 25–31. doi: 10.24018/ejers.2018.3.11.966.

[21] Luan, J. (2002) 'Data Mining and Its Applications in Higher Education', *New Directions for Institutional Research*, 2002(113), pp. 17–36. doi: 10.1002/ir.35.

[22] Abu, A. (2016) 'Educational Data Mining & Students' Performance Prediction', *International Journal of Advanced Computer Science and Applications*, 7(5), pp. 212–220. doi: 10.14569/IJACSA.2016.070531.

[23] Ihantola, P. (2015) 'Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies', *Proceedings of the 2015 ITiCSE on Working Group Reports*., 16, pp. 41–63. doi: 10.1145/2858796.2858798.

[24] Romero, C. and Ventura, S. (2010) 'Educational data mining: A review of the state of the art', *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 40(6), pp. 601–618. doi: 10.1109/TSMCC.2010.2053532.

[25] Bogarín, A., Cerezo, R. and Romero, C. (2018b) 'Discovering learning processes using inductive miner: A case study with learning management systems (LMSs)', *Psicothema*, 30(3), pp. 322–329. doi: 10.7334/psicothema2018.116.

[26] Cairns, A. H. (2014a) 'Towards Custom-Designed Professional Training Contents and Curriculums through Educational Process Mining', *The Fourth International Conference on Advances in Information Mining and Management Towards (IMMM)*, (july), pp. 53–58. doi: 10.13140/2.1.3060.5760.

[27] Khodabandelou, G. (2013) 'Process mining versus intention mining', *Lecture Notes in Business Information Processing*, 147 LNBIP, pp. 466–480. doi: 10.1007/978-3-642-38484-4_33.

[28] Aguirre, S. and Parra, C. (2011) 'Combination of process mining and simulation techniques for process redesign: a methodology approach', *Industrial Engineering Department. Pontificia ...*, pp. 24–43.

[29] C. K. Ayo, Jonathan A. Odukoya, Ambrose Azeta (2014), "A Review of Open and Distance Education and Human Development in Nigeria". International Journal of Emerging Technologies in Learning. Volume 9 Issue 6, 2014. pp. 63-67. Index in Scopus. ISSN: 1863-0383

[30] Azeta, A. A., Inam I. A. and Daramola O. (2018).'Voice-based e-Examination Framework For Visually Impaired Students in Open and Distance Learning'. *Turkish Online Journal of Distance Education-TODJE*. 19(2), pp. 34-46. **Index in Scopus.**
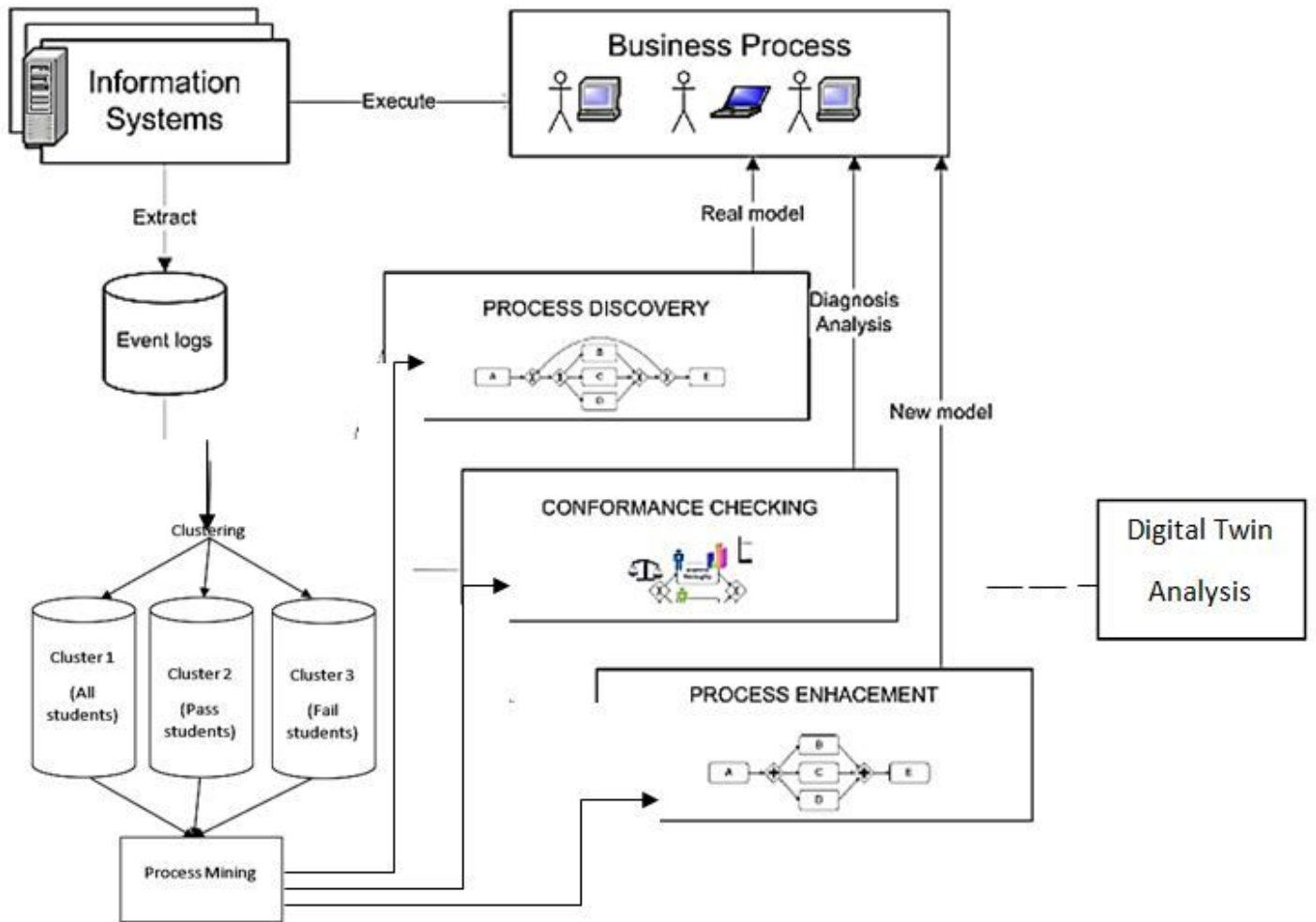
Figures

Figure 1

Overview of the proposed framework Adapted from [28]

	activity_name	course_id	module	idnumber	added	score	visible	visibleold
1	activity_name	course_id	module	idnumber	added	score	visible	visibleold
2	assign submit	10464	9	user186977097674129409	1372670270	94.5	1	1
3	resource view	10464	19	user4786049973974204417	1378189004	44	1	1
4	forum view forum	10464	7	user7133053338412646401	1372678410	85	1	1
5	assign submit	10464	9	user3543713697010548737	1372677682	30	1	1
6	quiz view	10464	16	user1626801571545219073	1372678281	38.5	1	1
7	quiz view	10464	16	user4549114572627247105	1372679905	82	1	1
8	forum view forum	10464	7	user6741987709889282049	1372678410	52.5	1	1
9	forum add discussion	10464	3	user1752183332046635009	1372677766	8.5	1	1
10	forum add discussion	10464	3	user8345756041840951297	1372678056	18.5	1	1
11	forum update post	10464	5	user2125677738260955137	1372678792	59	1	1
12	quiz review	10464	15	user3180865810348376065	1377792996	60	1	1
13	quiz review	10464	15	user3098829117496229889	1435737732	40.5	1	1
14	quiz review	10464	15	user1639510899465125889	1377683443	90	1	1
15	quiz close attempt	10464	12	user7739766835326746625	1449566049	64	1	1
16	forum add discussion	10464	3	user8627213778552881153	1372744095	67.5	1	1
17	quiz attempt	10464	10	user2420451385418973185	1377242031	67	1	1
18	quiz close attempt	10464	12	user2896319635668336641	1449570716	97	1	1
19	assign submit	10464	9	user2216540094985142273	1377095848	62	1	1
20	quiz view	10464	16	user8371538219417731073	1372679905	50	1	1
21	quiz continue attempt	10464	13	user2986645744250781697	1372680029	58.5	1	1
22	forum view discussion	10464	6	user1395770998610984961	1377243045	97.5	1	1

mdl_course_modules ⊕

Figure 2

Event log used for analysis

Figure 3

Process of converting CSV Log file to XES format

Figure 4

Behaviour of pass students in attempt to watch course video and submit course quiz before deadline
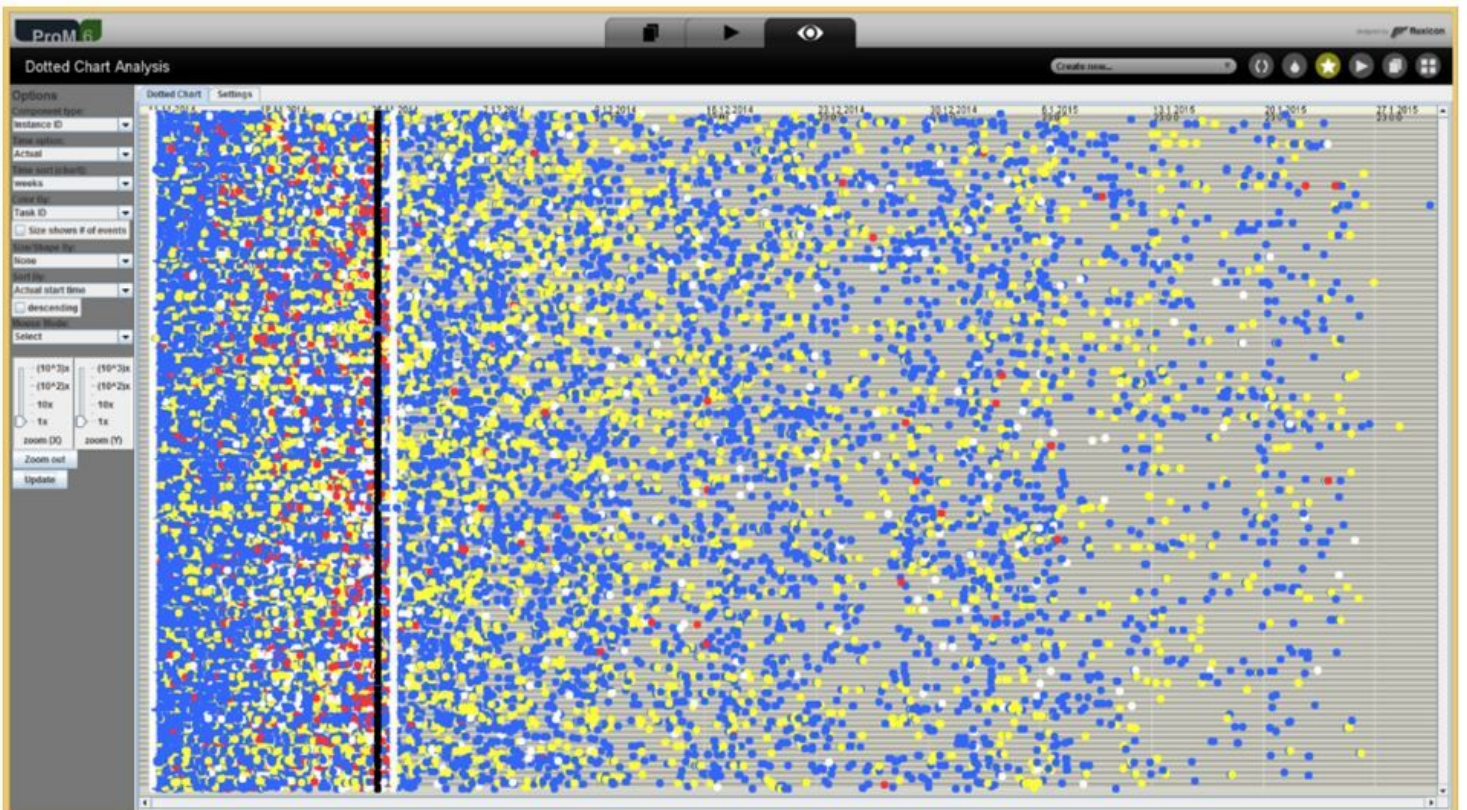
Figure 5

Behaviour of fail students in attempt to watch course video and submit course quiz before deadline