Journal of Theoretical and Applied Information Technology
15th May 2025. Vol.103. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTING STUDENT ACADEMIC PERFORMANCE USING ENGAGEMENT FEATURES: A PROCESS MINING AND DEEP LEARNING APPROACH

DEENA M. HELAL[1], YEHIA M. HELMY[2], DOAA S. ELZANFALY[3]

[1]Department of Business Information Systems (BIS), Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt
[2]Department of Business Information Systems (BIS), Faculty of Commerce and Business Administration, Helwan University, Cairo Egypt.
[3]Department of Information Systems, Faculty of Computers & Artificial Intelligence, Helwan University, Cairo 11795, Egypt
Faculty of Informatics & Computer Science, British University in Egypt, Cairo Egypt.
Email :[1]dina.m.helal@commerce.helwan.edu.eg, [2]ymhelmy@commerce.helwan.edu.eg,[3]doaa.saad@fci.helwan.edu.eg

ABSTRACT

In the digital era, the increasing availability of data from online educational environments enables advanced analysis and prediction of student academic performance. As a key indicator of student progress and achievement, academic performance necessitates effective tools for analysis and intervention to enhance learning outcomes. This study integrates process mining, deep learning to predict academic performance with 99.86% accuracy for intermediate grades and 92.48% for final scores, using engagement features like mouse clicks and keyboard strokes from a widely recognized dataset spanning six sessions. Through novel feature extraction and various preprocessing techniques applied with process mining and deep learning approach , we identify that engagement behavior significantly correlate with academic success. The findings confirm the predictive strength of engagement features, providing actionable insights into student interactions and learning behaviors to inform targeted interventions.
Keywords • *Process Mining , Deep Learning , Machine Learning , Predicting Academic Performance.*

1.INTRODUCTION

The rapid expansion of digital education has generated vast datasets from online student activities, presenting new opportunities for analyzing and predicting academic performance [1].

Early identification of at-risk students is a critical challenge for academic institutions, as it supports timely interventions, enhances retention rates, and optimizes resource allocation. Traditional machine learning (ML) techniques have been widely used in educational research to predict student outcomes, yet they often fail to fully capture the complexity of engagement patterns and temporal dynamics in learning processes [2].

Process mining techniques, including discovery, conformance checking, and enhancement, provide valuable tools for analyzing event logs and understanding the processes underlying student engagement. These methods enable academic institutions to map student journeys, identify inefficiencies, and gain insights into engagement behaviors over time. As deep learning can capture complex patterns, while process mining provides a comprehensive approach to predicting academic performance using metrics such as participation rates, interaction data (e.g., keystrokes and mouse clicks), and session activities.". This integration can precisely foster personalized learning and supports early intervention strategies, addressing gaps in traditional machine learning approaches by enhancing the accuracy and reliability of performance predictions [2], [3].

Building on insights gained from process mining, machine learning —a branch of artificial intelligence— has become a widely used tool in educational research for analyzing student data and identifying learning patterns. Through various models, such as classification, regression, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ensemble methods, ML techniques can effectively predict student performance, track learning progress, and detect students at risk of underperforming [4].

While ML has provided significant advancements in the field, deep learning (DL) introduces a novel approach through its layered neural networks, which capture complex and intricate patterns within unstructured data allowing for a deeper understanding of student needs and behaviors [5].
This allows DL to capture complex student behaviors and needs, providing valuable insights for intervention. When combined with process mining, DL enables a more comprehensive analysis of engagement features, further supporting personalized intervention and enhancing academic outcomes [6], [7].

By leveraging process mining-derived trace fitness scores and temporal interpolation, our study fills a critical gap in the literature, providing a data-driven framework for understanding the behavioral and temporal characteristics that influence student outcomes. In this study, we propose a novel approach that integrates process mining with deep learning to accurately predict academic performance.
By utilizing trace fitness scores and time-based interpolations, our method effectively captures detailed engagement patterns. We evaluate multiple machine learning and deep learning models, achieving highest accuracy of 99.86%, demonstrating the strength of our approach. These findings provide educators with practical tools to identify at-risk students early and implement personalized learning strategies.
This paper organized as follows: section 2 reviews related studies, focusing on various approaches in process mining (PM), machine learning (ML), and deep learning (DL) that have been implemented in predicting student performance. Section 3 discusses the dataset used in this study and details the feature manipulation techniques applied. Section 4 details the study's findings and discusses their significance. Section 5 concludes with a summary of the results and suggestions for future research.

2.RELATED STUDIES

In today's fast-paced world, advanced algorithms are increasingly used to analyze student data, with the goal of predicting academic performance and providing timely support. the study of student performance prediction is grounded in theories related to temporal behavior modeling, self-regulated learning (SRL), and engagement metrics.

This section reviews existing literature across three pillars central to this study: process mining, machine learning, and deep learning. While these techniques have individually shown promise in educational data mining, limited work has explored their integration. This study addresses this gap by proposing a hybrid approach that combines process mining with advanced deep learning models.

In recent years, advancements in data analytics have led to a surge either in process mining with machine learning [7] or applying deep learning (DL) [9] to predict student performance. However, our literature review indicates that no studies have combined DL with process mining in this field.

Numerous studies have used machine learning and deep learning models to analyze student learning, behavior, and demographic characteristics. This review examines 9 studies for process mining 11 studies on student performance prediction, including 6 using deep learning models and 5 using machine learning techniques.

2.1 Process Mining

Process mining has emerged as a powerful tool for analyzing educational data, uncovering behavioral patterns, and predicting student performance. Techniques such as Fuzzy Miner, Heuristic Miner, and Inductive Miner are widely used in like [9], [10] [11], each offering unique advantages. Fuzzy Miner excels in managing ambiguous or unpredictable event logs, while Heuristic Miner is praised for its simplicity and effectiveness in recognizing frequent patterns [12] [13].Inductive Miner automates process discovery by examining event logs without relying on predefined rules, making it particularly suitable for complex datasets [14].
One of the core applications of process mining in education is to distinguish patterns in student behavior that correlate with academic success [15]. Fuzzy Miner has been used extensively for analyzing engagement behaviors, such as tutorial video viewing durations and frequencies, providing insights into differences between high and low-performing students. Studies using this technique have demonstrated its ability to simplify complex patterns and uncover active learning strategies, with accuracies exceeding 95% in modeling student behaviors [16].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Process mining also facilitates the discovery of learning pathways and self-regulated learning (SRL) processes [17].Techniques like Heuristic Miner and Inductive Miner have been employed to identify frequent activity sequences, enabling the analysis of how student behaviors align with instructional designs or self-regulation frameworks [18]. These approaches are particularly useful in blended and online learning environments, where understanding the reliance and sequence of activities can highlight factors contributing to academic performance [19].Beyond mining patterns, clustering and optimization techniques have expanded the scope of process mining in educational contexts. By applying clustering algorithms, researchers have identified key behavioral groups, correlating specific student actions with academic outcomes [20] , [21]. For instance, clustering has revealed strong associations between top-performing students and specific teaching resources [20], while Markov Chains and association rule mining have been used to model submission behaviors and predict academic success [21]. The integration of process mining with machine learning techniques, such as Support Vector Machines (SVM) and Naïve Bayes (NB), further enhances prediction accuracy, with studies reporting improvements over standalone ML methods [22].

Table 1:Overview of Process Mining techniques in predicting performance.

Paper	Process Mining Task	Techniques used	Dataset size	Model Evaluation
[22]	Prediction of Student Academic Performance	Machine learning techniques such as SVM , NB, LR incorporated with process mining features	Coursera for Economics course of 167 students with demographic and interactive features	Naïve Bayesian achieved 89% accuracy after week 8 with process mining features integrated.
[23]	Assessing Student SRL skills	Inductive miner as Process mining algorithm	101 students in university in the North of Spain which after preprocessing led to 21,629 events	IM algorithm achieved 66% accuracy for all units and students.
[20]	Analyzing Student Behavior	Clustering and process Discovery algorithm	52 students in Indonesia that started with 24,589 events	Pearson correlation of 0.746 for Cluster 1; no evaluation metrics provided.
[24]	Analyzing Student behavior	Fuzzy Miner algorithm analyzes learning via course video length and duration	LMS Moodle of 247 students in Thonburi University with events produced to 40 learning behaviors	Dotted chart analysis highlighted group discrepancies: no numeric evaluation provided.
[21]	Analyzing SRL	Heuristic and inductive mining to discover the learning process of students'	LMS from 37 features with 4 features used in the process mining	Process models evaluated on fitness, precision, generalization, and simplicity for Alpha, Heuristic, ILP, Inductive Miner. Inductive Miner: fitness = 0.932, precision = 0.995. Heuristic Miner: fitness = 1, precision = 0.982.
[25]	Performance Prediction	Discovery methods and classification algorithms (SVM, decision trees, Naive	996 cases in 30,732 events collected from Learning System	Model developed using real-world student data to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

		Bayes) used to measure similarity between individual behaviors and shared processes.		validate and test its effectiveness.
[26]	Understanding Behavior	various clustering techniques such as K Means, DB scan, Gaussian Mixture Model and Trace Clustering Trace Clustering and Heuristic Miner.	Data from competition of 411 students with 110481 events	Silhouette score of 0.68 for a cluster size of 4 proved effective for clustering students by engagement behavior.
[27]	Resource Recommendation	Inductive Miner Heuristic Miner Alpha Miner	100students of 42,438 traces from Moodle LMS	Inductive Miner outperformed with fitness value (1) and generalization values of 0.6802 and 0.8191.
[28]	Verifying paths	Heuristic Miner algorithm	Data obtained from Programming course 24,334 events	No Evaluation was developed for the Heuristics Model

These findings highlight the strengths of process mining in revealing behavioral patterns, yet they fall short in predictive performance when used alone. This research leverages these strengths by integrating them with deep learning models to improve accuracy.

2.2 Machine Learning

There are several educational modes that serve different learning needs. The selection of an educational mode is often influenced by several variables, including personal preferences, educational objectives, availability of resources and technology, and particular criteria set out by the educational programs.

While machine learning models offer reliable classification, they often require extensive feature engineering. This study incorporates ML insights but enhances predictive power using DL techniques for automated feature extraction.

Zhang in [16] listed different modes and formats for education including Traditional Classroom-Based Education, Online Education, Blended or Hybrid Learning, Distance Learning, Self-Regulated Learning, Corporate Training, and Professional Development.

Classroom-based education is a widely used instructional approach in educational institutions, whereby students are required to be physically present in designated classrooms for their learning activities.

Online education encompasses a range of educational modalities, such as online courses, webinars, and virtual classrooms that use internet connectivity and digital technology. Blended learning is an instructional approach that integrates conventional in-person teaching methods with online components, enabling students to engage in face-to-face classes while utilizing digital resources at other timeframes.

Distance Learning refers to a kind of education when people are geographically separated from the educational institution. This mode of learning is made possible via the utilization of many technologies such as postal mail, electronic mail, video conferencing, and web-based platforms. Self-Regulated Learning enables people to autonomously choose their educational endeavors, supported by various resources such as printed books, online courses, and instructional aids. Corporate training and professional development programs are designed to meet the specific requirements of employees and professionals by providing focused training and skill upgrading that is customized to their respective positions or industries.

2.3 Deep Learning

In recent years, there has been a growing emphasis within the educational sector on using student records and data analytics to predict student performance and enhance their educational results. This led to the arise of a domain that examines the learners' data and their performance metrics called Learning Analytics (LA).

According to [17]Learning Analytics is defined as the collection and analysis of data pertaining to students' engagement with educational technology, coursework, and other educational resources, with the aim of acquiring valuable insights into their learning patterns and behaviors [18], [28]. According to [29]there are three fundamental components of LA:

a) Data: the entity upon which analytical insights are based.
b) Analysis: the logical deployment of methodologies and tools.
c) Action: the insights in which the use of data analytics techniques and technologies to education-related topics.

The use of learning analytics data facilitates decision-making based on empirical evidence, allowing educators to make well-informed decisions on the most efficacious teaching approaches. The use of learning analytics amplifies the capacities of educational process mining, therefore empowering educators, and institutions to make well-informed choices and enhance the entire educational experience for students. Many researchers have articulated early predictions as a fundamental task that can surely enhance students' performance by predicting either a category of learning outcome or by predicting a final score [25]. These predictions would be of more value if it could be done as early as possible.

The early prediction, as defined in [21], is the utilization of predictive techniques incorporated with performance metrics that accurately predict student outcome as early as possible. This early prediction often depends on the examination of diverse data sources, including student records, patterns of interaction, and past performance data and it is very critical as it affects the student's outcome related to their engagement in the course, engagement with the LMS.

Predicting the category of learning outcome is often a binary classification problem that distinguishes learners into "passed" and "failed" to estimate the likelihood of ensuring academic accomplishments in the future [16]. Since predictive learning is a significant advantage of machine learning technology, it is frequently used to train the learning performance prediction model using a straightforward way [8].

Although this form of predictor can produce accurate predictions, it has several drawbacks. Due to low generalizability and high computing complexity, a huge amount of e-learning behavior data of different dimensions is acquired and recorded throughout the e-learning process. Some of the predictors will evaluate the integrated influence of learning behavior data (i.e., perform feature fusion processing) on the same type of learning behaviors data and then utilize it for training.

Finally, crucial learning behavior markers are not standardized, and those discovered by various investigations vary. This field of study has been unable to find important behavioral variables that can accurately predict learning performance. The objective is to better fulfil students' requirements by providing personalized learning pathways, adaptive evaluations and suggestions, or adaptive and just-in-time feedback [24]. However, a deeper knowledge of how learning processes are connected to and may be captured by the data accessible in contemporary digital learning environments is necessary [25], [26].

Different papers have focused on the students' records using different data mining point of view disregarding the time as significant feature in their studies leaving very few studies that tackled student sequential behavior pattern from a process mining point of view which is the focus in our study [27], [30].

This study fills that gap by proposing an integrated approach by using various preprocessing techniques with process mining that addresses these limitations while maximizing predictive accuracy.

3.PROPOSED MODEL PM-DL PREDICT MODEL

This study employs a structured methodology to predict student academic performance by combining process mining and deep learning techniques. First, process mining was applied to create a Trace Fitness Score by comparing each student's engagement patterns to a normative model derived from top-performing students. This approach quantifies behavioral alignment, providing a unique feature indicative of high-performance traits.

Next, data preprocessing transformed raw logs into interpretable metrics such as total session time, idle periods, and categorized activities, enhancing data clarity. Advanced feature engineering techniques, including cumulative sums, quantile-based time points, and interpolations, were implemented to capture temporal engagement patterns across sessions. Finally, two deep learning models were developed to predict final and intermediate scores, with each model structure

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

designed to maximize accuracy through careful standardization and parameter optimization. This combined approach allows for a nuanced understanding of student behaviors and their impact on performance, making it ideal for predicting academic success.

Table 2:Deep and Machine Learning techniques in predicting student performance.

paper	Aim of the study	Dataset size	ML /Deep Learning	Technique	Evaluation
[31]	Prediction next session difficulty	115 students with Six sessions with 13 features in every session	Machine Learning	artificial neural networks (ANNs), support vector machines (SVMs), logistic regression, Naïve bayes classifiers and decision trees	K-fold cross-validation, ROC curves, RMSE; ANN (75%) and SVM (65%) outperformed other models.
[38]	Prediction of student performance	Real World data in Indonesian university	Deep Learning	Long Short-Term Memory (LSTM) and Gate Recurrent Units (GRU) integrated and separate achieved good accuracy	LSTM and GRU achieved 90% accuracy for individual tasks.
[39]	Prediction of student performance	505 students over 13 weeks	Deep Learning	MultiSource fusion CNN and bidirectional LSTM with Multiple activity in addition to 4 Machine learning classifiers such as Logistic Regression , Naïve Bayesian , Random Forest and Decision trees	SPDN (73.51% accuracy), BLSTM_MA (70.30% accuracy).
[40]	Prediction of student performance	Open World's dataset with **32,593 students with** 22 courses OULAD dataset [44]	Deep Learning	CNN to quantify student engagement metrics and LSTM to predict sequence of activities based on students 's demographic and interaction data	SPEN showed an increased accuracy with 10% over all the compared model in the research
[7]	Prediction of student performance	The dataset comprises clickstream data collected from 66 students enrolled in a computer engineering course.	Machine Learning	KNN classifier and MLP (Multilayer perceptron)	In terms of accuracy K-NN was 60-67% while Multilayer Perceptron 57-67%)
[34]	Prediction of student performance	online and blended learning for 3 courses for 311,386 events	Machine learning Sequence mining	iBCM algorithm in addition to sequence classification algorithm	Achieved 90% accuracy for course-specific predictions.
[8]	Prediction of student performance	Open Worlds dataset with 1044 records that contain 33 attributes	Hybrid Deep Learning	Bidirectional Long Short-Term Memory integrated with the attention model.	Attention-based BiLSTM achieved 90.16% accuracy
[32]	Prediction of student performance	OULAD Open World's dataset with 32,593 students with 22 courses ' [44]	Machine Learning Technique	interpretable Hidden Markov Model	The developed model produced an accuracy ranged from 88% till 95%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[45]	Prediction of student performance	behavior records of 9207 students over 29 weeks that were categorized into 13 features	Machine Learning	Hybrid Recurrent Neural Network to encode their behaviours and and Support Vector machine as a data mining Classifier	The proposed method SPC was about 86.9% accuracy.
[42]	Prediction of student performance	Open World's dataset [44]	Deep Learning	Deep Learning model was developed and compared against machine learning techniques Artificial Neural Networks, Naive Bayes, and Support Vector Machines, to ensure it effectiveness	Deep learning model achieves a remarkable accuracy score of 98.94% and a specificity score of 93.10% in the first week
[46]	Prediction of student performance	ASSISTments: 1,011,079 gradings by 8,519 students	Deep Learning	Deep Biased Matrix Factorization,	the proposed approach achieves the smallest RMSE of 0.332 on the dataset this enhanced performance prediction

3.1 Dataset Overview

Several studies have focused on students' behavior discovery through process mining to detect students' behavioral patterns while interacting with LMS or online MOOC environment such as Coursera or Udemy, or to understand their SRL (Self-Regulated Learning) behavior while engaging in these online environments.
In our experiment, the main objective is to predict both student final academic performance and assignment-level outcomes. To achieve this, we utilize a set of interactive metrics alongside time-related features, providing a comprehensive view of student activities and engagement.
This is a quantitative study based on the prediction of student performance using advanced data mining and deep learning techniques. The data used for this study is derived from the DEEDS dataset, which includes detailed student activity logs. The study focuses on student performance data in an educational environment (e.g., students enrolled in Digital Electronics courses).

Table 3:Feature Description

Name of the feature	Description
Session	It shows the number of laboratory session from 1 to 6.
Student_Id	It shows the Id of student from 1 to 115.
Exercise	It shows the Id of the exercise the student is working on. Each session contains 4 to 6 exercises,
Activity	The activities are labeled based on the title of web pages that are on focus
start_time	It indicates the date and time when a specific activity begins, formatted as dd.mm.yyyy hh:mm:ss
end_time	Indicates the date and time when a specific activity begins, formatted as dd.mm.yyyy hh:mm:ss
idle_time	Indicates the date and time when a specific activity begins, formatted as dd.mm.yyyy hh:mm:ss.
mouse_wheel	Indicates the date and time when a specific activity begins, formatted as dd.mm.yyyy hh:mm:ss
mouse_wheel_click	Indicates the date and time when a specific activity begins, formatted as dd.mm.yyyy hh:mm:ss.
mouse_click_left	Indicates the date and time when a specific activity begins, formatted as dd.mm.yyyy hh:mm:ss

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Process Mining and Trace Fitness Score Calculation

The primary concept investigated in this study is student performance, which is operationalized using both final academic scores and interim performance scores. These scores are derived from students' grades in digital electronics courses, providing a direct measure of their academic success. The second core concept is student engagement, operationalized through the frequency, duration, and timing of specific interactions with the learning platform. These interactions include viewing instructional content, completing assignments, and participating in discussions. Engagement was measured using detailed event logs recorded in the DEEDS system.
The final concept involves process mining techniques. The Trace Fitness Score is an innovation introduced in this study, which measures how closely a student's engagement pattern aligns with that of high-performing peers. This score is calculated through the Inductive Miner algorithm, a process mining tool used to extract features from the engagement data. Additional engagement metrics include activity groups (grouping related activities) and time-related features (e.g., total time spent, cumulative session durations).

Process mining phase started with the application of the inductive miner method in ProM, as mentioned in [41], to construct a normative process model. This model was developed from activity logs of students who achieved grades above 90%, representing the ideal behavior of high-performing students, as inspired by [21]. To evaluate each student's adherence to this model, conformance checking was conducted, and the trace fitness score was calculated. This score quantifies how closely individual student activities align with the high standards set by their successful peers, offering a clear measure of their performance relative to the model.

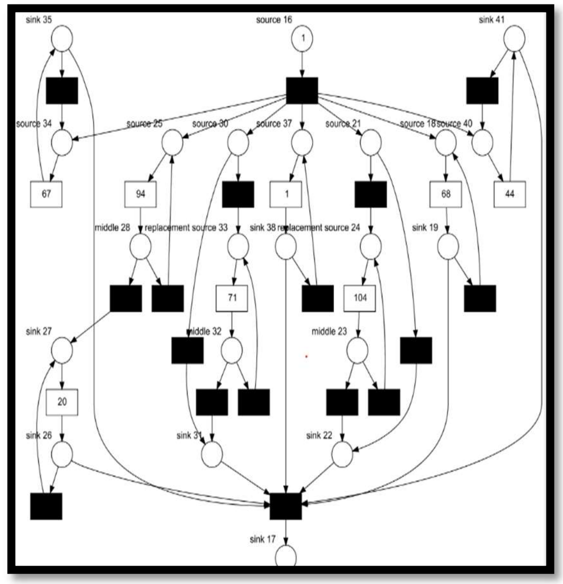


Figure 1 :Inductive Miner Petri Net Generation from XES Event Log Using Prom

The process mining phase included conformance testing, where each student's activity logs were replayed against the normative model using techniques described by *[3]*in ProM. This testing resulted in a log-model alignment, mapping each student's actions to the elements of the model, and highlighting deviations from expected behaviors. Following this, a Trace Fitness Score was calculated for each student based on these alignments, quantifying the conformity between each student's observed activity patterns and the normative model. These scores were then exported in CSV format and integrated as a key feature in the dataset, enhancing the analysis of student performance.

Session N	Student ID	Exercise	activity	start_time	end_time	idle_time	mouse_wh	Mouse_wh	mouse_cli	mouse_cli	mouse_m	key_stroke	Trace_Fitness
1	1	Es	Other	2.10.201	2.10.201	0	0	0	0	0	84	0	0.01692
1	1	Es	Aulaweb	2.10.201	2.10.201	218	0	0	4	0	397	0	0.01692
1	1	Es	Blank	2.10.201	2.10.201	0	0	0	0	0	59	0	0.01692
1	1	Es	Deeds	2.10.201	2.10.201	154117	6	0	8	0	1581	4	0.01692
1	1	Es	Other	2.10.201	2.10.201	0	0	0	2	0	103	0	0.01692
1	1	Es	Other	2.10.201	2.10.201	460	0	0	4	0	424	8	0.01692
1	1	Es	Blank	2.10.201	2.10.201	0	0	0	1	0	93	0	0.01692
1	1	Es	Deeds	2.10.201	2.10.201	0	0	0	1	0	75	0	0.01692
1	1	Es	Aulaweb	2.10.201	2.10.201	0	0	0	2	0	238	0	0.01692
1	1	Es	Deeds	2.10.201	2.10.201	4933	0	0	2	0	268	0	0.01692
1	1	Es	Other	2.10.201	2.10.201	3212	0	0	4	0	275	2	0.01692
1	1	Es	Aulaweb	2.10.201	2.10.201	1174	3	0	2	0	596	0	0.01692

Figure 2:Snippet Of Students In Session 1 With Trace Fitness Scores.

3.3 Feature Extraction

To further enrich the dataset, several new variables were derived, and data preprocessing techniques were applied:

• Creation of total_ms: This variable, representing the total time spent in each session, was calculated by taking the difference between end_time and start_time for each activity log entry.

• Creation of idle_ms: The idle_ms variable captures inactive periods within each session. It was derived by subtracting the total engagement time (from active interactions such as clicks or keystrokes) from total_ms, providing insight into periods of inactivity.

• Creation of actv_grp: To categorize students' activities, the actv_grp variable was derived by segmenting the actv_exer column, grouping activities into broader categories that represent different types of engagement. This feature enabled the aggregation of activities for more granular analysis.

3.4 Feature Engineering

During the feature extraction phase of the dataset preparation, several key steps were undertaken to enhance the predictive modeling capabilities:

1. Extraction of Original Values: Essential engagement variables such as mouse clicks, keystrokes, and trace fitness scores were extracted. These serve as direct indicators of student interaction and engagement.

2. Cumulative Sums: To capture the progression of student engagement over time, cumulative sums were calculated for each engagement variable. This helped to highlight patterns of activity and inactivity within the learning sessions.

3. Interpolation: Irregular time-series data from student interactions were standardized by interpolating missing values at consistent intervals, such as every 10% of the session time. This ensured that each dataset entry had uniform and complete data points, preserving the temporal dynamics of student behaviors.

4. Square Root Transformation: To normalize the distribution of features and minimize the impact of outliers, a square root transformation was applied to the interpolated data. This transformation helped in handling skewed data by compressing larger values more than smaller ones, making the dataset more suitable for the deep learning model.

Overall, these steps formed a comprehensive approach to preparing the dataset, making it robust for further analysis and modeling, capturing detailed engagement trends, and standardizing features for fair comparison across different students and sessions. After extracting features from your dataset, visualizing these features can provide valuable insights into their structure and relationships.

	sess	sid	start_dt	end_dt	exer	actv_grp	total_ms	idle_ms	mw	mwc	...	ks	TF	total_ms_cum	mw_cum	mwc_cum
0	1	1	2014-02-10 11:25:33	2014-02-10 11:25:34	Es	Other	1000.0	0.0	0	0	...	0	0.01692	1000.0	0	0
1	1	1	2014-02-10 11:25:35	2014-02-10 11:25:42	Es	Aulaweb	7000.0	218.0	0	0	...	0	0.01692	8000.0	0	0
2	1	1	2014-02-10 11:25:43	2014-02-10 11:25:43	Es	Blank	0.0	0.0	0	0	...	0	0.01692	8000.0	0	0
3	1	1	2014-02-10 11:25:44	2014-02-10 11:26:17	Es	Deeds	33000.0	154117.0	6	0	...	4	0.01692	41000.0	6	0
4	1	1	2014-02-10 11:26:18	2014-02-10 11:26:18	Es	Other	0.0	0.0	0	0	...	0	0.01692	41000.0	6	0
5	1	1	2014-02-10 11:26:19	2014-02-10 11:26:27	Es	Other	8000.0	460.0	0	0	...	8	0.01692	49000.0	6	0

Figure 3:Sample Data Preview: Processed Features And Cumulative Metrics

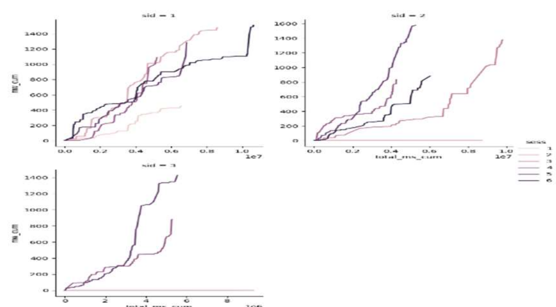


Figure 4:Cumulative Sum Of Mouse Movement Related Tototal Time Across Session For Five Students

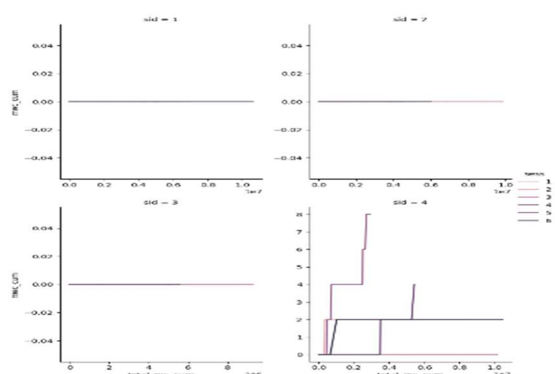


Figure 5:Cumalative Sum Of Mouse Wheel Click Related To Total Time Across Session For Five Students.

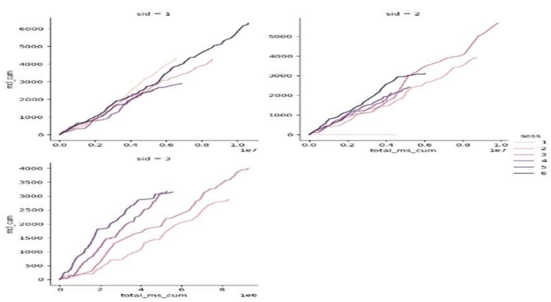


Figure 6: Cumulative Sum Of Mouse Click Left Related To Total Time Across Session For Five Students

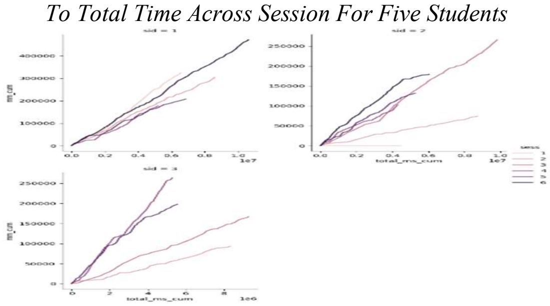


Figure 7:Cumalative Sum Of Mouse Click Left Related To Total Time Across Session For Five Students.

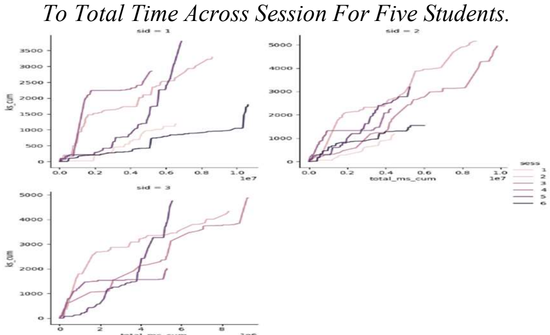


Figure 8:Cumalative Sum Of Mouse Movement Related To Total Time Across Session For Five Students.

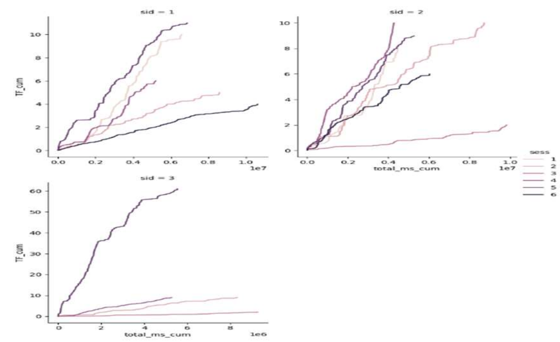


Figure 9:Cumalative Sum Of Keystrokes Related To Total Time Across Session For Five Students.

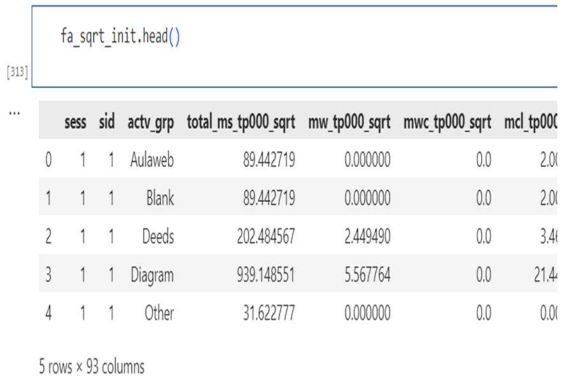
The presented plots in Figures[6 -10] provide an insightful visualization of cumulative user interactions across sessions, focusing on features such as mouse movement (ms_cum), clicks, keystrokes, and trace fitness scores, all plotted against the cumulative elapsed time (total_ms_cum). These plots allow for a temporal comparison of student engagement across sessions and between individual users.

Sharp inclines in cumulative interaction curves typically reflect bursts of high engagement, possibly during task-intensive periods or moments requiring increased cognitive effort. For example, in Figure 6, the steep slopes in ms_cum across multiple sessions for sid 1 and sid 3 indicate a concentrated period of activity. In contrast, plateaus or horizontal segments suggest periods of disengagement, inactivity, or passive learning, possibly when students are reading or watching a video without interacting with the interface.

Furthermore, the comparison of sessions (indicated by color) reveals how different instructional designs or temporal progression (e.g., early vs. late sessions) may affect student behavior. Notably, sessions 5 and 6 often show increased cumulative interaction in reading events (rd_cum, Figure 11), which may correspond to increased course difficulty or proximity to assessments.

Using cumulative sum plots helps visualize the intensity and distribution of actions over time and enables us to compare not only between users but also across sessions. This visualization technique effectively highlights behavioral patterns that could inform predictive modeling and instructional design.

While visual insights are powerful, combining them with engagement or performance labels would allow for more direct interpretation of how these patterns relate to *learning outcomes.*



```
[313] fa_sqrt_init.head()
```

	sess	sid	actv_grp	total_ms_tp000_sqrt	mw_tp000_sqrt	mwc_tp000_sqrt	mcl_tp00(
0	1	1	Aulaweb	89.442719	0.000000	0.0	2.0(
1	1	1	Blank	89.442719	0.000000	0.0	2.0(
2	1	1	Deeds	202.484567	2.449490	0.0	3.4(
3	1	1	Diagram	939.148551	5.567764	0.0	21.4
4	1	1	Other	31.622777	0.000000	0.0	0.0(

5 rows × 93 columns

Figure 9:Input Features For Data Learning Dataset After Interpolation And Square Root.

3.5 PM-DL Predict Model (Process Mining and Deep Learning)

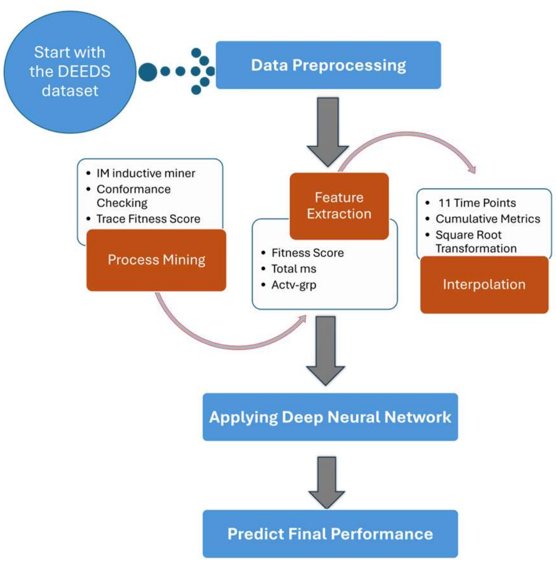


Figure 10:Proposed Model For Predicting Final Performance Using Process Mining (PM-DL Predict Model)

3.6 Predicting Final Scores (Multiclass Classification)

In this experiment, a deep learning in the PM-Score model was developed and evaluated for a multiclass classification task. The dataset was first divided into training and testing subsets, with 80% of the data allocated for training and 20% reserved for testing. This split provided a robust test set to evaluate the model's generalization on unseen data.

To ensure consistent feature scaling, the data was standardized so that each feature had a mean of 0 and a standard deviation of 1, which facilitated efficient model convergence. Following this, the standardized data was converted into PyTorch tensors, with target labels specifically converted to the long integer type to ensure compatibility with the CrossEntropyLoss function.

The model architecture consisted of a single linear layer, where the input size was aligned with the number of features and the output size was set to 3, corresponding to the three classes in the classification task. A CrossEntropyLoss function inherently applied a SoftMax function. The model's parameters were optimized using the Adam optimizer with a learning rate of 0.01, chosen for its ability to adapt the learning rate during training. The training was conducted over 100 epochs, during which the model was trained on small batches of data, allowing the loss function to be minimized iteratively. Finally, the model was evaluated on the test set, where the test loss was found to be 0.2519 and the test accuracy was 0.9248, indicating a high level of effectiveness in predicting the final academic performance of students based on their session activities as captured in the EPM dataset.

3.7 Predicting Intermediate Scores (Binary Classification)

To prepare the dataset for predicting intermediate grades using the deep learning model, several preprocessing steps were conducted to ensure the inclusion of only relevant and consistent data. Data from sessions 4, 5, and 6 were excluded from the analysis due to inconsistencies. Specifically, sessions 4 and 5 were excluded because no valid entries for the actv_grp variable, which is considered the categorical variable were present, resulting in a lack of usable data points. Similarly, session 6 was excluded because the data for actv_grp contained anomalies that differed from those in other sessions, rendering it incompatible with the model.

For sessions 2 and 3, relevant features were extracted based on specific variables, all of which had been square root-transformed (indicated by _sqrt suffix). These variables included total_ms, mw, mwc, mcl, mcr, mm, ks, and TF at various time points, leading to the selection of 88 features for the model. The integrity of the data was further ensured by verifying the unique categories within the actv_grp variable for both sessions. It was confirmed that each session contained 10 consistent activity groups, making them suitable for comparison and subsequent model training.

By integrating process mining, specifically the Trace Fitness Score and newly derived variables such as (total_ms) and (actv-grp), this study was able to capture comprehensive engagement behaviors, resulting in robust predictive models for student performance.

•Additionally, the sid and sess variables, initially stored as integers, were converted to object data types to preserve their categorical nature during the modeling process. This conversion was necessary to ensure that session identifiers were treated appropriately as categorical variables rather than numerical values.

•The resulting datasets from sessions 2 and 3 were then used as input for the deep learning model, ensuring that the data utilized for prediction was both consistent and relevant to the analysis.

Following the preparation of the intermediate grades data, a deep learning model was developed and implemented to perform logistic regression, aiming to predict the binary outcome of student performance. The features (all columns except the target) were assigned to the variable X, while the target variable, interim pass, was assigned to y. Label encoding was applied to the categorical variable actv_grp to convert it into a numeric format suitable for the model.

•The dataset was then split into training and testing sets, with 80% of the data allocated for training and 20% for testing, ensuring a robust evaluation of the model's performance. To standardize the data, a Standard Scaler was applied, transforming the features so that each had a mean of 0 and a standard deviation of 1, which is essential for optimizing the learning process of the model. Both the training and testing data were subsequently converted into PyTorch tensors, enabling their use in the deep learning framework.

4. RESULTS AND DISCUSSION

1-This study introduces a novel approach to student performance prediction by integrating process mining techniques with deep learning architectures, offering a behavioral and temporal lens that extends beyond traditional academic predictors. While prior studies have either focused on behavioral pattern extraction using process mining or employed deep learning for performance prediction, none have effectively combined the two into a unified predictive framework — a gap this research addresses.

This approach adds a behavioral dimension to the predictive model, offering insights that may help educators more effectively identify and support students based on their engagement patterns.

2-In our study, we employed advanced preprocessing techniques to enhance the predictive accuracy of student performance models. By utilizing process mining features extracted through the Inductive Miner algorithm, we developed trace fitness scores that quantify the alignment between observed student activities and predefined process models. This approach allowed us to create new variables, such as activity groups and total milliseconds spent, providing a detailed classification of engagement features. Additionally, data interpolation offered a closer examination of engagement behaviors over time. Integrating these enriched features into a deep learning framework led to a prediction accuracy of 92.48% for final academic performance, surpassing traditional machine learning methods that achieved lower accuracies.

3- One of the key contributions of this work is the introduction of the Trace Fitness Score — a unique metric derived from process mining that quantifies how closely a student's engagement behavior aligns with high-performing peers. This score adds a qualitative behavioral dimension to the prediction model, enabling a more nuanced understanding of how specific engagement sequences relate to academic success. Unlike conventional features, the Trace Fitness Score captures structural and temporal coherence in student activity, offering deeper insight into the nature of learning engagement.

4-Advanced preprocessing and feature engineering were employed to support this integration. Using the Inductive Miner algorithm, features such as activity groups, cumulative time metrics, and interpolated engagement behaviors were extracted and analyzed. These enriched features allowed the deep learning model to capture both the intensity and progression of engagement across sessions, offering standardized comparisons despite varying log lengths. This focus on temporal engagement dynamics, such as cumulative sums and quantile-based time segmentation, highlighted not only how much students engage, but how consistently and effectively they do so over time.

5-The deep learning model, trained with these features, significantly outperformed traditional machine learning models. It achieved 92.48% accuracy in predicting final academic performance and 99.86% accuracy for interim assessments, surpassing ensemble methods and standalone classifiers previously tested on the DEEDS dataset [47]. This underscores the model's capacity to capture complex behavioral trends and validate the utility of process mining-derived features in neural network architectures.

From a practical standpoint, the study offers institutions a data-driven method to monitor and predict student performance early and accurately. The behavioral insights generated — particularly through the Trace Fitness Score and time-based activity grouping — can inform early warning systems, guide personalized learning interventions, and enhance educational decision-making. Educators can leverage these findings to encourage sustained, effective engagement, which the model identifies as a key indicator of success.

In summary, the novelty of this study lies in its interdisciplinary fusion of process mining and deep learning, the development of new behavioral engagement metrics, and the demonstration of their effectiveness through significantly improved predictive performance. These contributions offer a valuable framework for future educational data mining research and practical implementation in academic support systems.

5.CONCLUSION

This study highlights the efficacy of integrating deep learning with process mining-derived features, such as trace fitness scores, to predict student performance in higher education. by achieving 99.86% accuracy for interim scores and 92.48% for final scores, our approach demonstrates the advantages of advanced feature engineering and temporal interpolation techniques in capturing nuanced engagement patterns.

These findings address the gap in existing literature, which often overlooks the role of behavioral and temporal dynamics in performance prediction, by offering a comprehensive, data-driven framework. practically, the study provides a reliable model for early identification of at-risk students, enabling academic institutions to implement timely interventions. the use of tailored neural networks to uncover nonlinear relationships in engagement data suggests practical applications for personalized learning and enhanced student support systems. future research could extend this approach by incorporating additional process mining features, testing across diverse academic settings, and exploring other engagement metrics to further improve adaptability and predictive power.

Table 4:Comparison OF Different Models Accuracies On DEEDS dataset with our proposed model in the field of prediction

Paper	*ML/DL*	*Model*	*Accuracy*
[31]	ML	ANN Artificial Neural Network	75%
		LR Logistic Regression	73%
		NBC Naïve Bayes Classifier	75%
		SVM Support Vector Machine	75%
[42]	ML	MLP Multilayer Perceptron	95.7%
		RF Random Forest	97.4%
		SVM Support Vector Machine	94%
		LR Logistic Regression	92.1%
		NB Naïve Bayesian	0.826%
[43]	ML	Ensemble	
		RF	96.49%
		Ensemble	
		FDT Fuzzy Decision Tree	98.25%
PM-DL predict - our Proposed model	Process Mining-DL	Interim score	92.48%
		(Process Mining + Logistic Regression Deep Learning	99.86 %

Journal of Theoretical and Applied Information Technology
15th May 2025. Vol.103. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

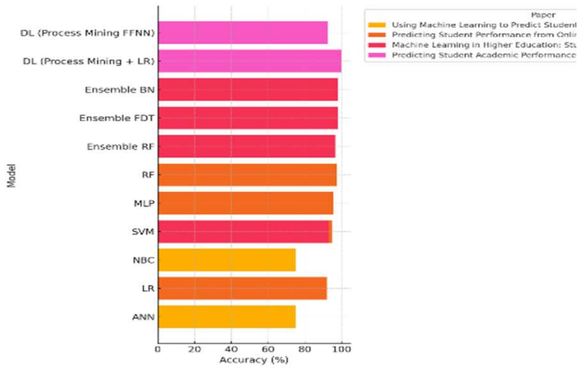


Figure 11:Comparsion of accuracy on ML/DL prediction models on DEEDS dataset.

Bar chart comparing the accuracy of different machine learning and deep learning models applied to the EPM DEEDS educational process mining Digital Electronics Education and Design Suite dataset [44].

6.LIMITATIONS AND FUTURE WORK

While this study provides valuable insights into the integration of process mining and deep learning techniques for predicting student performance, there are several limitations that should be considered. Addressing these limitations in future research can improve the generalizability and effectiveness of predictive models in diverse educational contexts.

4.1 Blended Data Integration

One limitation of the current study is its reliance solely on online engagement data from the DEEDS platform. This dataset, while rich in information, does not capture the full spectrum of student engagement, particularly in offline learning environments (e.g., face-to-face interactions, group activities, and independent study). As such, future research should focus on integrating both online and offline engagement data to create a more comprehensive understanding of student behavior. The inclusion of offline data could help account for a broader range of student activities, which may further enhance the generalizability and accuracy of predictive models across various educational settings. By incorporating offline interactions, such as classroom participation or extracurricular engagement, future models could capture more holistic patterns of student engagement.

4.2 Generalizability Across Courses and Educational Contexts

This study focuses on predicting student performance within a single course—digital electronics—which may limit the generalizability of the findings to other academic subjects or educational contexts. The engagement patterns observed in this course may not necessarily apply to students in other fields, where the nature of content delivery, student interactions, and engagement behaviors could vary. To enhance the applicability of the predictive model, future studies should incorporate data from multiple courses across different domains (e.g., humanities, social sciences, or engineering) and educational settings (e.g., high school, undergraduate, or online learning). By expanding the scope of the study, researchers can determine if the identified engagement patterns hold across different subject areas and student populations, improving the robustness of predictive models in diverse educational contexts.

4.3 Real-Time Monitoring and Intervention

While the study demonstrated the effectiveness of process mining and deep learning in predicting student performance, it does not address the potential for real-time monitoring and intervention. One promising avenue for future research is the development of real-time systems capable of continuously analyzing student engagement during their learning activities. These systems could provide immediate feedback to students and educators, enabling timely interventions that support students who may be at risk of underperforming. Real-time analysis could help educators identify patterns of disengagement or struggling students as they occur, allowing for proactive measures such as personalized tutoring, additional resources, or behavioral nudges. Future research could explore the implementation of such systems, testing their effectiveness in real-time prediction and intervention.

4.4 Advancements in Preprocessing Techniques

Another area for future research is the refinement and expansion of the preprocessing techniques used in educational data mining. This study highlights the potential of process mining methods to extract valuable features from student engagement data. However, additional advanced preprocessing strategies, such as incorporating multi-source data (e.g., combining online, offline, and student demographic data), could significantly enhance model accuracy and generalizability. Future studies should focus on developing and testing preprocessing techniques that can handle diverse and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

complex datasets, making the predictive models more adaptable to various educational environments. These enhanced preprocessing methods could lead to improved accuracy in predicting student outcomes across different contexts, such as varying course formats, institutions, and cultural settings.

REFERENCES

[1] E. a. M. S. Osmanbegovic, "Data mining approach for predicting student performance," Economic Review: Journal of Economics and Business, vol. 10, no. 1, pp. 3-12., 2012.

[2] Peach,R. L., Yaliraki, S. N., Lefevre, D., & Barahona, M, "Data-driven unsupervised clustering of online learner behaviour," Science of Learning, vol. 4(1), p. 14, (2019).

[3] W. A. A. D. M. A. K. A. A. F. B. T. B. T. .. & W. M. Van Der Aalst, "Process mining manifesto. In Business Process Management," International Workshops, Clermont-Ferrand, France, vol. Springer Berlin Heidelberg., no. BPM 2011, pp. 169-194, August 29, 2011.

[4] Chen, W., Shen, Z., Pan, Y., Tan, K., & Wang, C., "Applying Machine Learning Algorithm to Optimize Personalized Education Recommendation System," Journal of Theory and Practice of Engineering Science, vol. 4, no (01), pp. 101-108, 2024.

[5] Y. a. L. Z. Chen, "A comparative study on student performance prediction using machine learning," Education and Information Technologies, vol. 9, no. 12039-12057, p. 28, 2023.

[6] El Aouifi, Houssam, Mohamed El Hajji, Youssef Es-Saady, and Hassan Douzi, "Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining," Education and Information Technologies, vol. 5, p. 26, 5799-5814.

[7] Yousafzai, Bashir Khan, Sher Afzal Khan, Taj Rahman, Inayat Khan, Inam Ullah, Ateeq Ur Rehman, Mohammed Baz, Habib Hamam, and Omar Cheikhrouhou, "Student-performulator: student academic performance using hybrid deep neural network," Sustainability 13, vol. no 17, no. 9775, 2021.

[8] A. P. P. a. W. P. Nammakhunt, "Creating and Collecting e-Learning Event Logs to Analyze Learning Behavior of Students through Process Mining," International Journal of Information and Education Technology, Vols. 13, no. 2, pp. 211-222, 2023.

[9] L. R. A. a. R. F. Valensia, "Implementation of Process Mining to Discover Student Learning Patterns using Fuzzy Miner Algorithm (Case Study: Learning Management System (LMS) Telkom University)," in 3rd International Conference on Electronics Representation and Algorithm (ICERA) pp.126-131, 2021.

[10] C. W. N. P. P. P. A. S. I. a. W. P. Theptudborvomnun, "Analysis of Learners Participative Behavior from Active Learning Management by Process Mining Technique," in International Conference on ICT and Knowledge Engineering (ICT&KE), pp. 1-4. IEEE, 2020.

[11] R. F. S. a. K. R. S. Sarno, "Anomaly detection in business processes using process mining and fuzzy association rule learning.," Journal of Big Data , vol. 7, no. no 1, p. 5., 2020.

[12] Chanifah, Sabila, Rachmadita Andreswari, and Rokhman Fauzi, "Analysis of student learning pattern in learning management system (LMS) using heuristic mining a process mining approach," in 3rd International Conference on Electronics Representation and Algorithm (ICERA), pp. 121-125. IEEE, 2021., 2021.

[13] Cerezo, Rebeca, Alejandro Bogarín, María Esteban, and Cristóbal Romero., "Process mining for self-regulated learning assessment in e-learning," Journal of Computing in Higher Education , Vols. 32,, no. no. 1 , pp. 74-88. (2020).

[14] Bogarín, Alejandro, Rebeca Cerezo, and Cristóbal Romero, "A survey on educational process mining." Wiley Interdisciplinary Reviews:," Data Mining and Knowledge Discovery , p. e1230., 2018.

[15] Bakar, M. H. B. A., Shahrinaz Ismail, and Siti Haryani Shaikh Ali, "A process mining approach to understand self regulated-learning in moodle environment," International Journal of advanced trends in computer science and engineering, Vols. 8, no. 1.3, pp. 195-200, 2019.

[16] R. A. B. M. E. a. C. R. Cerezo, "Process mining for self-regulated learning assessment in e-learning," Journal of Computing in Higher Education , vol. 32, no. 1, pp. 74-88, 2020.

[17] Xu, Wei, Ye-Feng Lou, Hang Chen, and Zhi-Yi Shen., "Exploring the interaction of cognition and emotion in small group collaborative discourse by Heuristic Mining Algorithm

(HMA) and Inductive Miner Algorithm (IMA),' Education and Information Technologies , vol 28, no. 10, pp. 13153-13178., 2023.

[18] Moemi, Thusoyaone Joseph, and Bukohwc Michael Esiefarienrhe. "Analysis of students behaviour in Java programming class in blendec learning environment using Process Mining techniques., ""Analysis of students' behaviou in Java programming class in blended learning environment using Process Mining techniques,' in In Proceedings of 52nd Annual Conference of the Southern African Computer Lecturers Association (SACLA 2023), (SACLA 2023) 2023.

[19] Cenka, Baginda Anggun Nan, Harry B Santoso, and Kasiyah Junu, "Analysing studen behaviour in a learning management systen using a process mining approach.," Knowledge Management & E-Learning , vol. no. 1, pp. 62-80., 2022.

[20] Thiyagarajan, Gomathi, and S. Prasanna "Process Mining-Based Behavioral Modeling o Learners in Self-paced Learning Environment,' in In International Conference on Signal & Data Processing, pp. 121-132. : Springer Nature Singapore, ., Singapore, 2022.

[21] R. T. S. A. M. a. S. S. Umer, "On predicting academic performance with process mining ir learning analytics," Journal of Research ir Innovative Teaching & Learning, Vols. 10, no 2, pp. 160-176, 2017.

[22] Cerezo, Rebeca, Alejandro Bogarín, Maria Esteban, and Cristóbal Romero., "Process mining for self regulated learning assessment,' Journal of Computing in Higher Education, vol (2020), no. no. 1, pp. 74-88, 2020.

[23] Nammakhunt, Anake, Parham Porouhan, anc Wichian Premchaiswadi., "Creating anc Collecting e-Learning Event Logs to Analyze Learning Behavior of Students through Process Mining.," International Journal of Informatior and Education Technology, Vols. 13, no. 2, pp 211-222, 2023.

[24] Bao, Yunxia, Faming Lu, Yanxiao Wang Qingtian Zeng, and Cong Liu, "Studen performance prediction based on behavio process similarity," Chinese Journal o Electronics (, Vols. 29, no. 6, pp. 1110-1118. 2020).

[25] Thiyagarajan, Gomathi, S. Prasanna, and V Vijayalakshmi, "Process Mining basec Student's Behavioural modeling for Online Assessment," in In 2022 3rd Internationa Conference on Communication, Computing anc Industry 4.0 (C2I4), pp. 1-6. IEEE, 2022.

[26] Hachicha, Wiem, Leila Ghorbel, Ronar Champagnat, Corinne Amel Zayani, and Ikram Amous, "Using process mining for learning resource recommendation: A Moodle case study.," Procedia Computer Science, vol. 192 pp. 853-862., 2021.

[27] Real, Eduardo Machado, Edson Pinheiro Pimentel, Lucas Vieira de Oliveira, Juliana Cristina Braga, and Itana Stiubiener. "Educational process mining for verifying student learning paths in an introductory programming course," in In 2020 IEEE Frontiers in Education Conference (FIE), pp. 1-9. IEEE, 2020.

[28] A. D. G. a. P. K. Gupta, "Mining sequentia learning trajectories with hidden markov models for early prediction of at-risk students ir e-learning environments.," IEEE Transactions on Learning Technologies, vol. 15, no. no.6, pp 783-797., 2022.

[29] A. Peña-Ayala, "Learning Analytics fundaments, applications, and trends.," A view of the current state of the art to enhance e-learning (2017).

[30] Arpasat, Poohridate, Nucharee Premchaiswadi Parham Porouhan, and Wichian Premchaiswad "Applying process mining to analyze the behavior of learners in online courses.,' International Journal of Information anc Education Technology 11, no. 10 (2021): 436-443..

[31] Hussain, Mushtaq, Wenhao Zhu, Wu Zhang Syed Muhammad Raza Abidi, and Sadaqat Ali "Using machine learning to predict studen difficulties from learning session data,' Artificial Intelligence Review, vol. 52, pp. 381-407., 2019.

[32] Kurniawati, Gisela, and Nur Ulfa Maulidevi "Multivariate sequential modelling for studen performance and graduation prediction," in In 2022 9th international conference or information technology, computer, anc electrical engineering (ICITACEE) pp. 293-298, 2022.

[33] X. X. Z. X. Z. Y. J. a. X. T. Li, "Studen academic performance prediction using deep multi-source behavior sequential network," ir In Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD , , May 11–14, 2020, Proceedings, Part I 24, pp

567-579. Springer International Publishing, Singapore, 2020.

[34] Song, Xiangyu, Jianxin Li, Shijie Sun, Hui Yin Phillip Dawson, and Robin Ram Mohan Doss. "SEPN: a sequential engagement basec academic performance prediction model," IEEE Intelligent Systems, vol. 36, no. no. 1, pp. 46-53, 2020.

[35] Kuzilek, Jakub, Martin Hlosta, and Zdenek Zdrahal, "Open university learning analytics dataset," in Scientific data 4, no. 1 (2017): 1-8.

[36] Deeva, Galina, Johannes De Smedt, Cecilia Saint-Pierre, Richard Weber, and Jochen De Weerdt, "Predicting student performance using sequence classification with time-basec windows," Expert Systems with Applications 209, p. 118182, 2022.

[37] A. D. G. a. P. K. Gupta, "Mining sequentia learning trajectories with hidden markov models for early prediction of at-risk students in e-learning environments," in IEEE Transactions on Learning Technologies 15, no. 6 (2022)783-797., 2022.

[38] X. X. Y. L. G. F. L. a. L. X. Wang, "Studen performance prediction with short-tern sequential campus behaviors," Information 11 no. 4, p. 201, 2020.

[39] Almahdi, Abdulla A., and Baraa T. Sharef "Deep learning based an optimized predictive academic performance approach," in In 2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD), pp. 1-6. IEEE, 2023 .

[40] Huynh-Ly, Thanh-Nhan, Huy-Thap Le, and Nguyen Thai-Ngh, "Deep Biased Matrix Factorization for Student Performance Prediction," EAI Endorsed Transactions on Context-aware Systems and Applications, p. 9 2023.

[41] Wen, Lijie, Jianmin Wang, Wil MP van der Aalst, Biqing Huang, and Jiaguang Sun, "A novel approach for process mining based on event types.," Journal of Intelligent Information Systems 32 (2009): 163-190..

[42] Brahim, Ghassen Ben, "Predicting studen performance from online engagement activities using novel statistical features," Arabian Journal for Science and Engineering, Vols. 47 no. 8, pp. 10225-10243., 2022.

[43] Latif, Ghazanfar, Sherif E. Abdelhamid, Khaled S. Fawagreh, Ghassen Ben Brahim, and Runna Alghazo, "machine learning in highe education: students' performance assessmen considering online activity logs," in IEEE access 11 (2023): 69586-69600..

[44] Donzellini,G, "Digital Electronics Deeds https://www.digitalelectronicsdeeds.com/deeds .html#," 2018.

[46] Lntayoad, W., Kamyod, C., &. Temdee and P. "Process mining application for discovering student learning paths.," In 2018 Internationa Conference on Digital Arts, Media and Technology (ICDAMT) IEEE, pp. pp. 220-224 Febraury 2018.

[47] S. M. N. S. a. V. J. Prabhu, "Enhancing Citizen Satisfaction Using Citizen-Facing Process Mining," Transforming Public Services—Combining Data and Algorithms to Fulfi Citizen's Expectations, vol. Cham: Springer Nature Switzerland, pp. pp 29-86, 2024.

[48] Bernacki, Matthew L., Linyu Yu, Shelbi L Kuhlmann, Robert D. Plumley, Jeffrey A Greene, Rebekah F. Duke, Rebekah Freed Christina Hollander-Blackmon, and Kelly A Hogan, "Using multimodal learning analytics to validate digital traces of self-regulated learning in a laboratory study and predict performance in undergraduate courses," Journal of Educationa Psychology, 2024.

[49] Patil, D., N. L. Rane, P. Desai, and J. Rane "Machine learning and deep learning: Methods techniques, applications, challenges, and future research opportunities," Trustworthy Artificia Intelligence in Industry and Society, pp. 28-81 2024.

[50] S. a. S. R. P. Nayani, "Combination of Deep Learning Models for Student's Performance Prediction with a Development of Entropy Weighted Rough Set Feature Mining," Cybernetics and Systems, pp. 1-43, 2023.

[51] Kurniawati, Gisela, and Nur Ulfa Maulidevi "Multivariate sequential modelling for studen performance and graduation prediction," in In 2022 9th international conference on information technology, computer, and electrical engineering (ICITACEE), pp. 293-298. IEEE, 2022, 2022.

[52] X. X. Y. L. G. F. L. a. L. X. Wang, "Studen performance prediction with short-tern sequential campus behaviors," Information 11 vol. 11, no. no 4 , p. 201, 2020.

[53] Arpasat, Poohridate, Nucharee Premchaiswadi Parham Porouhan, and Wichian Premchaiswadi, "Applying process mining to analyze the behavior of learners in online

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

courses," International Journal of Informatior and Education Technology , vol. no. 10, pp 436-443, 2021.

[54] Wen, Lijie, Jianmin Wang, Wil MP van de Aalst, Biqing Huang, and Jiaguang Sun, "A novel approach for process mining based or event types," Journal of Intelligent Informatior Systems 32, pp. 163-190, 2009.

[55] G. S. E. A. K. S. F. G. B. B. a. R. A. Latif "Machine learning in higher education students' performance assessment considering online activity logs.," IEEE access 11, pp 69586-69600, 2023.

[56] Alshamaila, Yazn, Hamad Alsawalqah, Ibrahin Aljarah, Maria Habib, Hossam Faris, Mohammac Alshraideh, and Bilal Abu Salih., "Alshamaila, Yazn Hamad Alsawalqah, Ibrahim Aljarah, Maria Habib Hossam Faris, Mohammad Alshraideh, and Bilal Abu Salih.," Multimedia Tools and Applications, Vols 46369-46396, pp. 83, no. 15, 2024.

[57] Lai, Xiaochen, Sheng Zhao, Zheng Zhang, anc Xiaodi Pan, "Research on Student Performance Prediction Based on Clustered Graph Neura Networks," in In 2024 4th International Conference on Machine Learning and Intelligent System Engineering (MLISE) pp. 192-195. IEEE, 2024 2024.

[58] Yin, Mengjiao, Hengshan Cao, Zuhong Yu, anc Xianyu Pan. "Manual label and machine learning ir clustering and predicting student performance, "A practice based on web-interactive teaching systems," in International Journal of Web-Based Learning anc Teaching Technologies (IJWLTT) 19, no. 1 (2024) 1-33., Yin, Mengjiao, Hengshan Cao, Zuhong Yu and Xianyu Pan. "Manual label and machine learning in clustering and predicting student performance.

[59] Pelima, Lidya R., Yuda Sukmana, and Yusep Rosmansyah., "Predicting university studen graduation using academic performance and machine learning," in a systematic literature review." IEEE Access (2024)., 2024.