
Sequential Pattern Mining for Predicting Student Dropout Risk From Learning Management System Interaction Logs

Karim Hassan El-Masry¹, Salma Ahmed Farouk²

¹Department of Computer Science, Faculty of Computers and Information Technology, Alexandria National University, Alexandria, Egypt

²Department of Information Systems, Faculty of Computing and Artificial Intelligence, Cairo International University, Cairo, Egypt
karimhassan2026e@gmail.com

ABSTRACT

The increasing availability of Learning Management System (LMS) interaction logs enables early identification of students at risk of dropout. However, conventional prediction approaches often do not consider the temporal relationships among sequential learning activities. In this study, we propose a sequential pattern mining approach for dropout risk prediction (SPM-DRP) to identify behavioral sequences and predict dropout risk from LMS interaction logs. The proposed method organizes student activities sequentially, uses the PrefixSpan algorithm to identify frequent and dropout-associated patterns, and finally transforms discriminative patterns into predictive features for dropout classification. The dataset includes 120,542 enrollment records and 8,157,277 learning-behavior records, with labels indicating dropout events and corresponding event data from seven LMS types. We measure performance using Lift, Pattern Growth Rate (PGR), PR-AUC, Brier Score (BS), and Matthews Correlation Coefficient (MCC), with the proposed SPM-DRP achieving values of 2.68, 33.2%, 0.914, 0.087, and 0.846, respectively. The findings illustrate the potential of the sequential behavioral patterns to capture dropout-related learning patterns and to assist in predicting dropout. The proposed SPM-DRP framework offers an actionable, interpretable method to identify deviations in student engagement and support timely academic intervention and retention strategies.

Keywords: Learning Management Systems, Student Dropout Prediction, Sequential Pattern Mining, PrefixSpan, Student Behavioral Analysis, Learning Analytics, Dropout Risk Prediction, Educational Data Mining, Temporal Activity Patterns, Student Retention.

1. Introduction

In AI-assisted education, association-based data mining, such as the Apriori algorithm, has been used to detect mental health risk patterns among students [1]. Some approaches based on explainable and privacy-preserving techniques have been proposed to explain and identify relevant attrition factors to predict dropouts transparently [2]. Machine learning analysis of LMS and Moodle activity data has been used to identify students at risk of dropping out [3]. Researchers have mined academic and behavioral patterns to identify those that correlate with dropouts and enable early interventions [4]. Studies on psychological and academic determinants of student persistence [5] have been conducted longitudinally. Research has also examined behavioral and engagement measures to identify patterns linked to student dropouts and outcomes [6].

Researchers have developed neural network models to predict student performance early and support intervention [7]. For learning analysis and recommendations, researchers have attempted to model interconnected behavioral patterns using graph-based recurrent representations [8]. Pattern-discovery techniques have identified learner behaviours associated with academic failure and used them to guide proactive educational intervention [9]. Data-driven classification methods have helped identify early attrition factors and inform student-retention strategies [10]. Predictive approaches have also been used to analyze students' related factors and detect dropout risk early, enabling timely intervention [11]. Sequential mining has been used to uncover patterns in sequences of behaviour and errors, which is helpful for learning analytics [12].

The main contributions of this work are:

- A sequential pattern mining-based dropout risk prediction (SPM-DRP) framework is proposed, which predicts dropout risk of a student based on their interaction logs from the LMS.
- The PrefixSpan algorithm is employed to find the sequential patterns in student learning activities that are present in a large number of students and are associated with dropouts.
- Discriminative patterns extracted are then converted to predictive features and used for capturing temporal changes in student engagement for the classification of dropouts.
- Evaluations of the proposed framework are based on the metrics of Pattern Strength (Lift), Pattern Growth Rate (PGR), PR-AUC, Brier Score (BS), and MCC, all of which assess pattern strength, predictive performance and classification reliability.

The proposed SPM-DRP approach predicts dropout risk from LMS interaction logs based on temporal order and learning behavior progression. The PrefixSpan algorithm discovers sequences associated with dropouts and transforms discriminative patterns into predictive features for classification. This sequential behavioral analysis provides interpretable indicators of changes in student engagement and helps identify students at risk of dropping out early.

2. Related Work

Leelaluk et al. [13] introduced an RNN-Attention model and knowledge distillation to predict student performance early on, but their main focus was performance rather than directly identifying dropouts. Temporal learner behavior was used in an explainable early-warning system for dropout prediction by Zhang and Duan [14], with the suggestion of wider generalisation over courses and learning environments. Okunola [15] assessed the performance of five machine learning algorithms (XGBoost, Random Forest, SVM, neural networks, and ensemble models) for predicting student outcomes; however, they tested them on institutional data that may not be generalizable.

Selvaraj et al. [16] presented a learning-based visual analytics approach to target detection for unmanned aerial vehicles (UAVs), which is not related to LMS-based student behaviour. Goh [17] examined LMS logs and found correlations between students' engagement persistence, consistency, and academic outcomes but did not predict sequential dropout. Shoukath [18] used temporal LMS interaction data for classroom co-regulation and orchestration but did not consider individual dropout risk. Ishaq et al. [19] analyzed LMS logs to understand student engagement but did not model dropout risk. Rahman and Karim [20]

created engagement profiles using the Moodle platform but did not integrate temporal patterns or dropout prediction.

Because they did not focus on longitudinal dropout-risk prediction, Ramli et al. [21] used Bayesian Networks to model associations among LMS engagement, resource usage, and academic performance. Angeioplastis et al. [22] applied BiLSTM temporal modeling of LMS behavioral features to early learning-outcome prediction, but did not directly classify dropout risk. Pethuraj et al. [23] presented an optimisation and neural learning approach for psychological assessment, which is not directly applicable to dropout prediction in LMS.

Existing LMS-based studies mainly use static or aggregated features for dropout prediction and engagement analysis, with limited consideration of the temporal order of student activities. Although learning analytics has explored sequential mining, its direct application to dropout-risk prediction remains limited. The proposed SPM-DRP framework addresses this gap by using PrefixSpan to extract dropout-associated sequences, converting them into predictive features, and evaluating Random Forest performance using Lift, PGR, PR-AUC, Brier Score, and MCC.

3. Methodology

The proposed methodology, called Sequential Pattern Mining-based Dropout Risk Prediction (SPM-DRP), aims to predict student dropout risk using chronological logs of students' interactions with the LMS. The framework is temporal; that is, student activities are organized along the time dimension. PrefixSpan extracts frequent behavioral and dropout-associated sequences from the activities, and the framework transforms discriminative patterns into predictive features. These features are then fed into a Random Forest (RF) model that estimates dropout probability and enables temporal learning behavior to be included in student retention analysis.

The proposed framework of SPM-DRP for prediction of student dropout risk from LMS interaction logs is shown in Fig. 1. LMS records are cleaned and chronologically sorted for enrollment IDs, event type and timestamps. Processed records are translated into activity sequences for each student, preserving the order of learning behavior orders. Then PrefixSpan extracts frequent and dropout-related sequential patterns. These discriminative patterns are converted to predictive features that represent students' engagement behavior. Finally, the Random Forest (RF) classifier predicts dropout risk based on the generated features. Lift, PGR, PR-AUC, Brier Score and MCC are used to evaluate the framework.

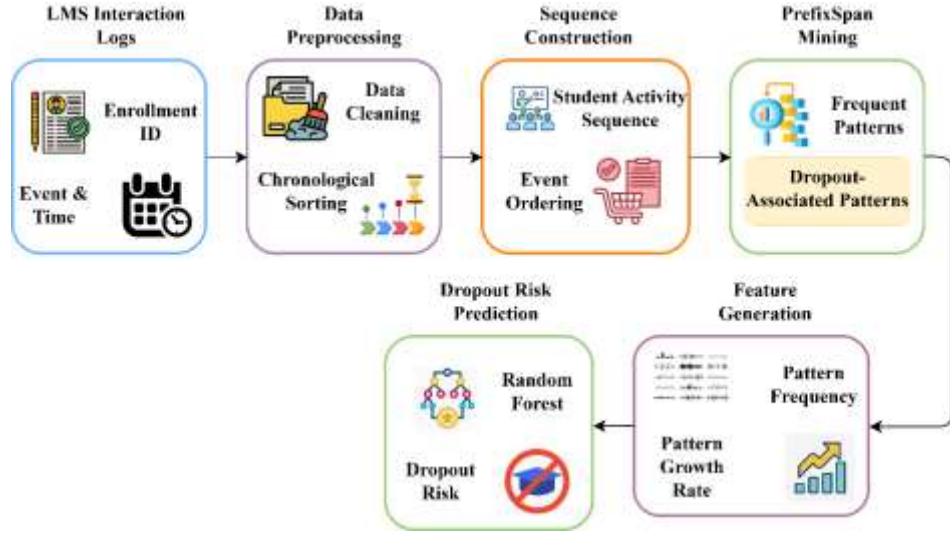


Figure 1: Architecture of the Proposed SPM-DRP Framework

The pre-processing step transforms raw LMS interaction records into a normalized temporal representation. The normalized interaction representation is formulated as expressed in Equation (1),

$$x_{i,j} = \left[\text{onehot}, (e_{i,j}) \frac{t_{i,j} - t_i^{\min}}{t_i^{\max} - t_i^{\min} + \epsilon} \frac{t_{i,j} - t_{i,j-1}}{t_i^{\max} - t_i^{\min} + \epsilon} \right] \quad (2)$$

where $x_{i,j}$ denotes the normalized representation of the j -th interaction in enrollment i , and $e_{i,j}$ is the corresponding LMS event. The timestamp $t_{i,j}$ records when the event occurred, while $t_i^{\min}$ and $t_i^{\max}$ represent the earliest and latest timestamps in enrollment i , respectively. The function $\text{onehot}[\cdot](\cdot)$ converts the categorical event into a binary vector, and ϵ is a small positive constant that prevents division by zero. The second captures the normalized time gap from the preceding interaction. For the first interaction, $t_{i,j-1}$ should be assigned a predefined value, such as $t_{i,1}$, so that the initial time-gap feature is zero. Each enrollment's ordered LMS activity sequence is represented by equation (2),

$$S_i = \langle (e_{i,\pi_i(1)}, t_{i,\pi_i(1)}), (e_{i,\pi_i(2)}, t_{i,\pi_i(2)}), \dots, (e_{i,\pi_i(n_i)}, t_{i,\pi_i(n_i)}) \rangle, t_{i,\pi_i(1)} \leq \dots \leq t_{i,\pi_i(n_i)} \quad (2)$$

where S_i denotes the chronological LMS activity sequence for enrollment i , and n_i is the total number of interactions in that enrollment. The term $e_{i,\pi_i(j)}$ represents the event at chronological position j , while $t_{i,\pi_i(j)}$ denotes its corresponding timestamp. The permutation π_i specifies the ordering that rearranges the original interactions chronologically. PrefixSpan is used to analyze constructed LMS sequences to find recurrent subsequences of behavioral patterns. To determine the minimum-support condition for retaining frequent behavioral patterns, the support of a sequential pattern is defined by Equation (3),

$$\text{Sup}(P) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}[P \leq S_i], P \in \mathcal{F} \Leftrightarrow \text{Sup}(P) \geq \sigma_{\min} \quad (3)$$

where P denotes a candidate sequential pattern, N is the total number of enrollment sequences, and $\mathbb{I}[\cdot]$ is an indicator function. The notation $P \leq S_i$ indicates that pattern P occurs as a subsequence of an enrollment sequence S_i . The set $\mathcal{F}$ contains the frequent patterns, and $\sigma_{\min}$ is the minimum-support threshold. The support value measures the proportion of enrollment sequences containing P , with $0 \leq \text{Sup}(P) \leq 1$. A pattern is retained in $\mathcal{F}$ if its support

is at least $\sigma_{\min}$. The association can be quantified via a smoothed log-lift measure oriented towards dropouts as expressed in equation (4),

$$\text{DLift}(P) = \log \left(\frac{(n_{P,1} + \alpha)/(N_1 + \alpha + \beta)}{(n_{P,0} + \alpha)/(N_0 + \alpha + \beta)} \right) \quad (4)$$

where $\text{DLift}_{f_0}(P)$ denotes the dropout-associated log-lift of pattern P . The quantities $n_{(P,1)}$ and $n_{(P,0)}$ represent the numbers of dropout and non-dropout enrollments containing P , respectively. Similarly, N_1 and N_0 denote the total numbers of dropout and non-dropout enrollments. The parameters $\alpha, \beta > 0$ provide smoothing to avoid zero-frequency estimates.

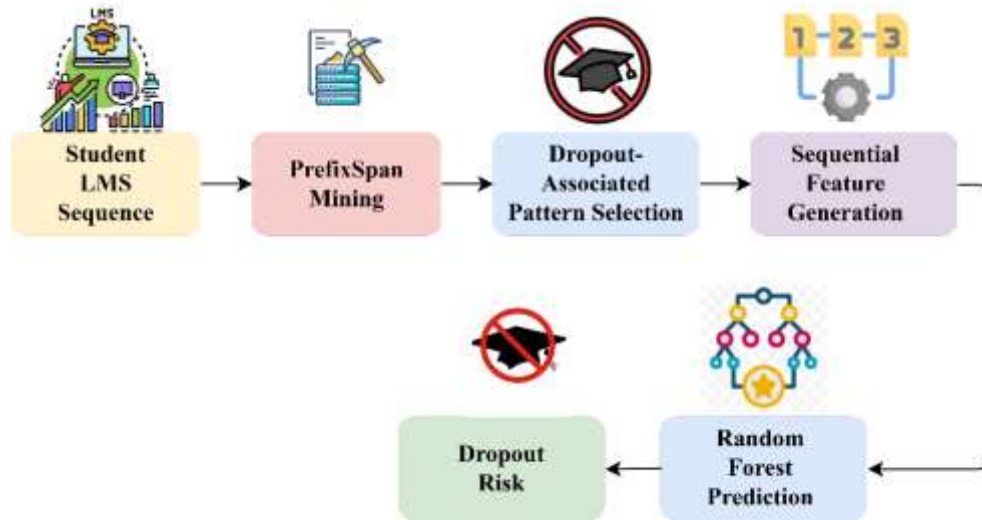


Figure 2: Core Processing Workflow of the Proposed SPM-DRP Method

The fig. 2 shows the main processing procedure of the proposed SPM-DRP method. First, PrefixSpan is used to process the student LMS activity sequence and identify frequent sequential patterns. The method then identifies dropout patterns and converts them into sequential behavioral features, which are fed to a Random Forest classifier to predict dropout risk. This represents the extent and evolving levels of student engagement. This creates a feature vector which is represented as in Equation (5),

$$Z_i = \left[\underbrace{\frac{\sum_{j=1}^{n_i} \mathbb{I}(P_k \leq S_i^{(j)})}{\max(1, n_i - |P_k| + 1)}}_{\text{Pattern occurrence}}, \underbrace{\frac{f_{i,k}^{(r+1)} - f_{i,k}^{(r)}}{f_{i,k}^{(r)} + \epsilon}}_{\text{Temporal growth}} \right]_{\substack{k=1, \dots, K \\ r=1, \dots, R-1}} \quad (5)$$

where Z_i denotes the sequential feature vector for enrollment i . The pattern P_k is the k -th selected dropout-associated pattern, and $S_i^{(j)}$ represents the sequence suffix beginning at position j . The variable n_i denotes the number of interactions in enrollment i , while $f_{(i,k)}^{(r)}$ is the frequency of pattern P_k in temporal window r . The quantities K and R represent the number of selected patterns and temporal windows, respectively, and ϵ is a small positive constant used to avoid division by zero. The resulting sequential feature vector is used as input to the Random Forest classifier to predict the probability of student dropout. The average of the ensemble dropouts is calculated as in Equation (6),

$$\hat{p}_i = \frac{1}{T} \sum_{t=1}^T \frac{\sum_{m \in B_{t, \ell_i}} y_m}{|B_{t, \ell_i}|} \quad (6)$$

where $\hat{p}_i$ denotes the estimated dropout probability for enrollment i , and T represents the total number of decision trees in the ensemble. The variable $y_m \in \{0, 1\}$ is the observed dropout label of training enrollment m . The set $B_{(t, l_i)}$ contains the training samples that reach the terminal leaf l_i of tree t , while $|B_{(t, l_i)}|$ denotes the number of samples in that leaf. For each tree, the inner fraction calculates the proportion of dropout cases among the samples in the corresponding leaf. The final estimate $\hat{p}_i$ is obtained by averaging these leaf-level dropout proportions across all T trees.

Algorithm 1. Proposed SPM-DRP

Input: LMS interaction logs D , dropout labels Y , minimum support $\sigma_{\min}$, temporal windows R , number of trees T , and decision threshold θ .

Output: Dropout probability $\hat{p}_i$ and risk prediction $\hat{y}_i$.

Initialize

$$S_i \leftarrow \emptyset, \mathcal{F} \leftarrow \emptyset, \mathcal{P} \leftarrow \emptyset, Z \leftarrow \emptyset.$$

for each enrollment $i = 1, \dots, N$ **do**

Construct the chronological LMS sequence:

$$S_i = \langle (e_{i,1}, t_{i,1}), (e_{i,2}, t_{i,2}), \dots, (e_{i,n_i}, t_{i,n_i}) \rangle, t_{i,1} \leq t_{i,2} \leq \dots \leq t_{i,n_i}.$$

end for

Apply PrefixSpan to identify frequent sequential patterns:

$$\mathcal{F} = \{P_k \mid \text{Sup}(P_k) \geq \sigma_{\min}\},$$

where

$$\text{Sup}(P_k) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}[P_k \leq S_i].$$

for each $P_k \in \mathcal{F}$ **do**

Calculate dropout association:

$$L_k = \log \left(\frac{(n_{k,1} + \alpha)/(N_1 + \alpha + \beta)}{(n_{k,0} + \alpha)/(N_0 + \alpha + \beta)} \right).$$

if $L_k > 0$ **then**

$$\mathcal{P} \leftarrow \mathcal{P} \cup \{P_k\}.$$

end for

for each enrollment $i = 1, \dots, N$ **do**

Generate the sequential feature vector:

$$Z_i = \left[\frac{\sum_{j=1}^{n_i} \mathbb{I}(P_k \leq S_i^{(j)})}{\max(1, n_i - |P_k| + 1)}, \frac{f_{i,k}^{(r+1)} - f_{i,k}^{(r)}}{f_{i,k}^{(r)} + \epsilon} \right]_{\substack{k=1, \dots, K \\ r=1, \dots, R-1}}.$$

end for

Train the Random Forest using Z and Y , with T decision trees.

for each enrollment $i = 1, \dots, N$ **do**

Calculate the probability from tree t :

$$p_i^{(t)} = \frac{\sum_{m \in B_{t, \ell_i}} y_m}{|B_{t, \ell_i}|}.$$

Aggregate the tree predictions:

$$\hat{p}_i = \frac{1}{T} \sum_{t=1}^T p_i^{(t)}.$$

if

$$\hat{p}_i \geq \theta$$

then

$$\hat{y}_i \leftarrow 1;$$

else

end for
Return

$\hat{y}_i \leftarrow 0.$

Algorithm 1 presents the proposed SPM-DRP approach, which mines LMS activity sequences chronologically with PrefixSpan, extracts dropout-related patterns, creates sequential behavioral features, and then estimates dropout probability and classifies dropout risk with Random Forest.

4. Results and Discussion

a. Dataset Description

The MOOC Dropout Prediction dataset [24], derived from KDD Cup 2015, contains 120,542 enrollment records and 8,157,277 learning-behavior records. Each enrollment includes anonymous enrollment, user, and course IDs, while activity logs contain timestamps and seven LMS interaction types: problem, video, access, wiki, discussion, navigation, and page_close. The dataset also provides dropout labels through dropout_prob, where 1 indicates dropout and 0 indicates non-dropout. The chronological interaction logs support sequential pattern mining and dropout-risk prediction based on evolving student engagement behavior.

b. Experimental Setup

Table 1: *Experimental Setup and Hyperparameter Configuration.*

Parameter	Setting
Dataset	MOOC Dropout Prediction
Training / Testing	80% / 20%
Sequence Mining	PrefixSpan
Minimum Support	0.01
Minimum Confidence	0.60
Maximum Sequence Length	10
Classifier	Random Forest
Number of Trees	200
Maximum Depth	15
Cross-Validation	5-fold
Random Seed	42
Metrics	Lift, PGR, PR-AUC, Brier Score, MCC
Software	Python 3.11, Scikit-learn 1.5
Hardware	Intel Core i7, 32 GB RAM, NVIDIA RTX 3060 12 GB

Table 1 presents the experimental setup and hyperparameter configuration, splitting the dataset into 80% training and 20% testing sets and sorting the LMS events chronologically with the enrollments. Frequent sequential patterns are mined using PrefixSpan and fed into a Random Forest model to predict dropout risk. Various measures are used to evaluate the proposed approach, including Lift, PGR, PR-AUC, Brier Score and MCC.

c. Pattern Strength and Growth Analysis — Lift and PGR

The evaluated methods are assessed using the extracted sequential patterns, and comparisons are based on two evaluation parameters: Lift (L) and Pattern Growth Rate (PGR). Higher values indicate a stronger association with dropping out and more patterns related to dropping out. The proposed SPM-DRP is likely to show more prominent sequential pattern characteristics because PrefixSpan explicitly captures the sequential nature of LMS activities.

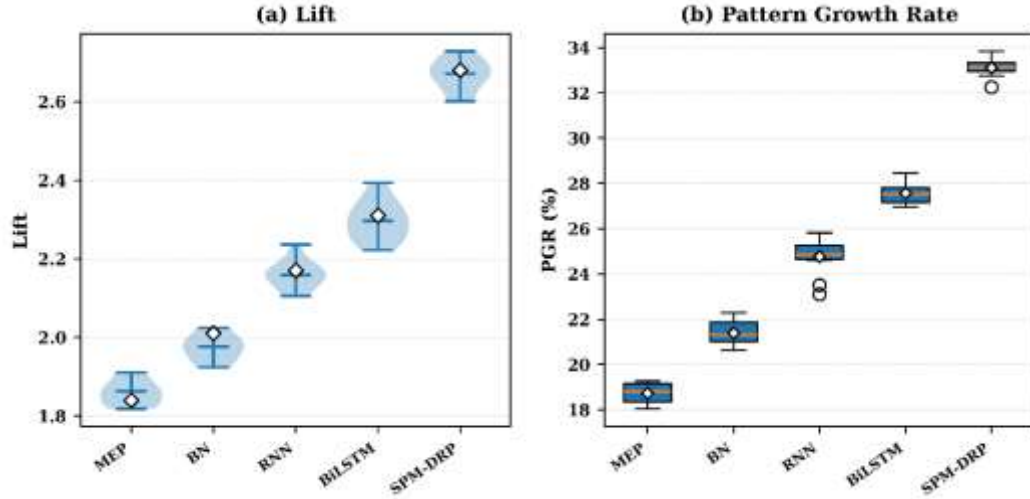


Figure 3: Distribution of Lift and Pattern Growth Rate Across the Evaluated Methods

The distribution of Lift values is presented by a violin plot in Fig. 3(a), and the distribution of Pattern Growth Rate (PGR) is presented by a box plot in Fig. 3(b) between MEP, BN, RNN, BiLSTM and the proposed SPM-DRP. The plots show the differences and stability of pattern-discovery performance across runs.

d. Predictive Performance Analysis — PR-AUC and Brier Score

PR-AUC and Brier Score assess predictive performance. Precision-Recall AUC measures precision-recall performance in the presence of dropouts, while Brier Score measures the accuracy of predicted probabilities. Higher PR-AUC and lower Brier Score indicate better dropout-risk prediction.

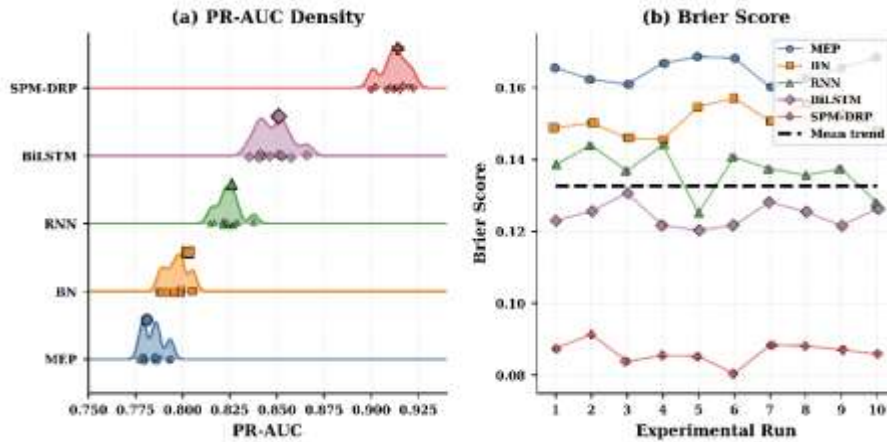


Figure 4: Distribution and Run-Wise Variation of PR-AUC and Brier Score Across the Evaluated Methods

Fig. 4(a) shows the densities of PR-AUC values from multiple runs of MEP, BN, RNN, BiLSTM, and the proposed SPM-DRP. The run-wise variation of the Brier Score is shown in Fig. 4(b) in which multiple Brier Score lines are shown with observations scattered around with jitter. The results indicate that the predictive performance is consistent across the experimental runs.

e. Overall Performance Comparison

Table 2: Performance Comparison Across Five Metrics

Method	Lift ↑	PGR (%) ↑	PR-AUC ↑	BS ↓	MCC ↑
MEP	1.84	18.6	0.781	0.164	0.684
BN	2.01	21.4	0.803	0.151	0.712
RNN	2.17	24.8	0.826	0.137	0.741
BiLSTM	2.31	27.5	0.851	0.124	0.773
SPM-DRP	2.68	33.2	0.914	0.087	0.846

Table 2 compares the evaluated methods across five metrics. Temporal models generally provide stronger predictive performance than engagement profiling and probabilistic modeling. SPM-DRP further incorporates explicit sequential pattern mining to convert dropout-associated activity sequences into predictive features, improving Lift, PGR, PR-AUC, Brier Score, and MCC.

Overall, the results indicate that SPM-DRP outperforms across the analysed metrics. The proposed framework captures dropout-associated learning patterns by integrating sequential pattern mining and temporal behavioral features, leading to better predictive performance than the baseline methods.

5. Conclusion

In this paper, we propose a new interpretable sequential pattern mining-based Dropout Risk Prediction (DRP) framework based on student interaction logs from the Learning Management System (LMS). The framework organizes activities chronologically, extracts sequential patterns associated with dropouts using PrefixSpan, and converts discriminative patterns into predictive features, capturing temporal changes in student engagement. The proposed approach achieved a Lift of 2.68, Pattern Growth Rate (PGR) of 33.2%, PR-AUC of 0.914, Brier Score of 0.087, and MCC of 0.846. The findings suggest that sequential patterns of behavior can support early identification of students at risk of dropping out and inform classroom interventions. However, the study's results may be limited in generalizability because the data came from a single LMS and the dropout labels were predefined, which may not generalize across institutions and learning environments. We plan to assess SPM-DRP on several educational datasets, study real-time sequence mining, add more contextual and demographic features and investigate adaptive deep sequential models for more timely and personalized prediction of the dropout risk.

Each manuscript must include a 250-450-word conclusion that encapsulates the major outcomes of the work, underscores their significance, acknowledges limitations, discusses relevance and applications, and offers recommendations.

REFERENCES

- [1] Y. Fu, F. Ren, and J. Lin, "Apriori algorithm based prediction of students' mental health risks in the context of artificial intelligence," *Frontiers in Public Health*, vol. 13, Art. no. 1533934, 2025.
- [2] H. Liu, M. Mao, X. Li, and J. Gao, "Model interpretability on private-safe oriented student dropout prediction," *PLOS ONE*, vol. 20, no. 3, Art. no. e0317726, 2025.
- [3] M. Rebelo Marcolino, T. Reis Porto, T. Thompsen Primo, R. Targino, V. Ramos, E. Marques Queiroga, et al., "Student dropout prediction through machine learning optimization: Insights from Moodle log data," *Scientific Reports*, vol. 15, no. 1, Art. no. 9840, 2025.
- [4] N. Roslan, J. M. Jamil, I. N. M. Shaharane, and S. J. S. Alawi, "Prediction of student dropout in Malaysian's private higher education institute using data mining application," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 45, no. 2, pp. 168–176, 2025.
- [5] W. T. Trusty, B. E. Scofield, A. E. Christensen, T. D. White, Y. E. Murphy, R. A. Janis, et al., "Psychological symptoms and academic dropout in higher education: A six-year cohort study," *Journal of College Student Mental Health*, vol. 39, no. 4, pp. 776–796, 2025.
- [6] A. Bettahi, F. Z. Belouadha, and H. Harroud, "A modular and explainable machine learning pipeline for student dropout prediction in higher education," *Algorithms*, vol. 18, no. 10, Art. no. 662, 2025.
- [7] N. U. R. Junejo, M. W. Nawaz, Q. Huang, X. Dong, C. Wang, and G. Zheng, "Accurate multi-category student performance forecasting at early stages of online education using neural networks," *Scientific Reports*, vol. 15, no. 1, Art. no. 16251, 2025.
- [8] A. S. D. Alluhaidan, P. Krishnan, K. Jain, P. Prabu, S. Baskar, and M. S. Pethuraj, "GraphNetNLP-enhanced graph recurrent mapping framework for social computing recommendation," *IEEE Internet of Things Journal*, vol. 12, no. 18, pp. 37695–37704, 2025.
- [9] F. Santiago-Muñoz, M. Larrañaga, J. A. Elorriaga, and A. Arruarte, "Identification of patterns to prevent the failure of engineering students," *International Journal of Artificial Intelligence in Education*, vol. 35, no. 5, pp. 3356–3405, 2025.
- [10] A. Bakariwie, D. Asamoah, and A. B. Duwiejuah, "Prevention of student attrition: A data-backed approach to school counselling using Delphi technique and multiple classification algorithms," *Discover Education*, vol. 4, no. 1, Art. no. 259, 2025.
- [11] Ö. Korkmaz and M. N. Aydın, "Detection of early school drop out in vocational and technical high schools in Turkey," *SAGE Open*, vol. 15, no. 3, Art. no. 21582440251370443, 2025.
- [12] F. V. Roque, L. C. Junior, C. Cechinel, M. Z. Marcon, A. Kuhnen, R. Munoz, and M. Grellert, "Learning analytics for virtual industrial labs: Performance segmentation and error pattern discovery via sequential mining," *IEEE Access*, vol. 13, pp. 194401–194420, 2025.
- [13] S. Leelaluk, C. Tang, V. Švábenský, and A. Shimada, "Knowledge distillation in RNN-attention models for early prediction of student performance," in *Proc. 40th ACM/SIGAPP Symp. Appl. Comput.*, Mar. 2025, pp. 64–73.
- [14] L. Zhang and S. Duan, "Artificial intelligence-driven early warning for online learner dropout: Temporal explainability and theory-grounded intervention," *Scientific Reports*, 2026.
- [15] A. Okunola, "Leveraging student outcome predictive analytics to mitigate institutional drop-out and performance risks in higher education curriculum projects," *The Science Post*, vol. 2, no. 3, 2026.
- [16] R. Selvaraj, V. M. Kuthadi, A. Duraisamy, B. Selvaraj, and M. S. Pethuraj, "Learning optimizer-based visual analytics method to detect targets in autonomous unmanned aerial vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 16, no. 6, pp. 72–85, 2024.
- [17] T. T. Goh, "Learning management system log analytics: The role of persistence and consistency of engagement behaviour on academic success," *Journal of Computers in Education*, vol. 13, no. 1, pp. 283–306, 2026.
- [18] T. K. Shoukath, "Modelling classroom co-regulation from LMS interaction logs for real-time classroom orchestration," *Journal of Intelligent Decision Making and Information Science*, vol. 3, no. 4s, pp. 116–149, 2026.
- [19] I. Ishaq, S. Rufaida, and S. Saleng, "Analyzing student engagement through LMS log data in an applied statistics course," *Journal of Educational Sciences*, vol. 10, no. 3, pp. 117–128, 2026.

- [20] A. Rahman and A. Karim, "Learning analytics-based student engagement profiling using Moodle activity logs across multiple courses," *Journal of Nexural Intelligence*, vol. 1, no. 1, pp. 1–9, 2026.
- [21] N. Ramli, M. T. Ismail, and H. Z. Naji, "Analysis of academic performance using learning management system (LMS) data and Bayesian network," *Journal of Science and Mathematics Letters*, vol. 14, no. 2, pp. 192–202, 2026.
- [22] A. Angeioplastis, M. Konstantakis, J. Aliprantis, K. Ordoumpozanis, P. Vourla, D. Varsamis, et al., "LEAP: A deep learning pipeline for early prediction of student learning outcomes from LMS logs," in *Proc. 2026 Int. Conf. Advances in Artificial Intelligence and Machine Learning (AAIML)*, Mar. 2026, pp. 945–953.
- [23] M. S. Pethuraj, M. A. Burhanuddin, and N. R. Dzakiyullah, "MV-DUO: Multi-variate discrete unified optimization for psychological vital assessments," *Neural Computing and Applications*, vol. 36, no. 31, pp. 19777–19793, 2024.
- [24] Kaggle, "MOOC dropout prediction," Kaggle. [Online]. Available: <https://www.kaggle.com/competitions/mooc-dropout-prediction-17/data>. Accessed: Sep. 28, 2026.