IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 006, LIR479
DOI: https://doi.org/10.52731/lir.v006.479

Design and Implementation of a Cyclic Dropout Prevention Model Using Institutional Research Data

Naruhiko Shiratori *

Abstract

This position paper proposes a cyclic model for dropout prevention in higher education, integrating data-driven prediction and student support practices. The model connects six key phases—data consolidation, time-series dropout risk prediction, student status monitoring, classification of student trajectories, targeted intervention, and evaluation of support effectiveness—into a continuous improvement cycle. Grounded in institutional research (IR), the model utilizes attendance records, academic performance, and pre-admission data to estimate dropout probabilities and classify students using clustering techniques such as X-means. Based on these classifications, tailored interventions including early alert systems and enhanced first-year education programs are implemented. The effectiveness of these interventions is evaluated through changes in attendance and academic outcomes, enabling feedback into the model for refinement. This framework aims to bridge the gap between predictive analytics and practical student support, offering a scalable and adaptable approach for universities seeking to reduce dropout rates and improve student success.

Keywords: Dropout Prediction, Institutional Research (IR), Student Support, Cyclic Intervention Model

1 Introduction

Student dropout in universities is a serious issue that not only results in the loss of individual learning opportunities but also leads to the waste of university management resources and broader societal assets. In Japan, in particular, the declining population of 18-year-olds has raised concerns about the assurance of educational quality and the effectiveness of student support systems. As a result, there is an increasing demand for the early identification of dropout risk and the provision of timely and appropriate interventions tailored to each situation.

According to a survey by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the number of students who dropped out from universities and junior colleges in Japan in the 2023 academic year was 56,710, accounting for 2.10% of all students [1]. Since the annual dropout rate remains around 2%, it means that approximately 8% of students leave school over a four-year period. While the overall dropout rate may appear low, this figure represents the average across all institutions. According to a survey by the Yomiuri Shimbun, some universities experience dropout rates of several tens of percent, imposing a significant burden on both the institutions and the students [2].

Against this backdrop, many universities in Japan have been introducing Institutional Re-

* Tokyo City University, Tokyo, Japan

search (IR) practices, attempting to utilize various types of student data for analysis and support. Among these efforts, research on dropout prediction using learning data such as attendance records and academic performance has advanced both domestically and internationally. By leveraging machine learning and statistical methods, many studies have reported attempts to estimate individual students' dropout risk with a certain degree of accuracy.

However, many previous studies on dropout prediction have not sufficiently addressed the mechanisms for organically connecting prediction results to practical student support measures. Additionally, the integration of dispersed data within universities and the construction of practical models that link student status monitoring and classification to support interventions remain significant challenges.

This study aims to address these issues by constructing a practical model that integrates dispersed data within universities, connects dropout prediction based on attendance and academic records to student status monitoring, implements concrete support measures, and evaluates the effectiveness of these interventions in a unified and cyclic manner. Specifically, this study classifies students based on risk estimation results from predictive models and implements support measures such as early alert systems and enhanced first-year education tailored to each group's characteristics in actual educational settings. Furthermore, data collected and analyzed for the evaluation of these interventions are fed back into the model, aiming for continuous improvement and advancement of student support through a cyclic approach.

2 Related Work

2.1 Theoretical Background of Dropout

University dropout is not only a personal issue for students but also a significant problem that leads to the loss of social resources and impacts the operations of educational institutions. According to Tinto's longitudinal model of student departure, students tend to drop out when they fail to achieve social and academic integration within the university environment [3]. Pre-admission academic ability, cultural capital, and experiences during enrollment interact to influence a student's decision to leave. Bean and Metzner's model takes a broader perspective, focusing on non-traditional students such as older students or those with significant work or family responsibilities, and points out that external environmental factors can also increase dropout risk [4]. Furthermore, Kerby's model, which utilizes Bourdieu's social theory, emphasizes that imbalances in cultural, social, and economic capital are key factors in determining academic achievement and educational exclusion [5].

2.2 Dropout Prediction Using Pre-Admission Data

Early warning models that use data available at the time of university admission have attracted attention as important methods for identifying high-risk students before the start of the academic term. Carballo-Mendívil et al. constructed an XGBoost model using data from approximately 40,000 new students at a public university in Mexico, predicting dropout risk based on academic ability, family background, and socioeconomic indicators at the time of admission, achieving a sensitivity of 88% [6]. Nagy and Molontay also demonstrated that early risk identification is possible using CatBoost models with pre-entry data [7].

However, models that rely solely on pre-admission information have limitations in accuracy. In response, Shiratori et al. constructed a hierarchical Bayesian model that combines pre-entry variables such as the number of absences in high school and type of high school with post-entry variables such as GPA and credits earned after university admission, calculating dropout probabilities for each semester [8]. This study revealed that pre-entry factors, such as coming from a correspondence high school and having a high number of absences, significantly increase dropout probability. Furthermore, by examining the relationship between semester-by-semester academic performance and pre-entry factors using a hierarchical Bayesian model, the study showed that more precise risk assessment is possible.

2.3 Dropout Prediction Using In-Semester Data

As the semester progresses, a variety of behavioral and academic data accumulate, and integrating these data appropriately can dramatically improve prediction accuracy. Goren et al. collected attendance rates, LMS logs, and academic performance at fixed points such as the 4th and 8th weeks and before and after final exams, and used XGBoost and neural networks for dropout prediction. They found that while the AUC was 0.62–0.73 in the early semester, it improved to over 0.90 by the end of the semester [9]. Notably, using a "studentship" indicator that divides LMS activity into cognitive and social elements improved the performance of the XGBoost model even in the early semester.

Recent research by Shiratori et al. proposed a method for identifying high-risk semesters by calculating dropout probabilities using logistic regression on students' academic data for each semester. Analysis of 173 actual dropouts revealed that, on average, students were already in a high-risk state about 2.97 semesters (approximately 18 months) before dropping out. The study also showed that dropout risk tends to be particularly high in the spring and fall semesters of the first year, highlighting the importance of early intervention [10]. Furthermore, Shiratori et al. proposed a clustering method using time-series dropout probability data. After calculating dropout probabilities for each semester using logistic regression, they used the X-means method to classify student trajectories into five patterns, clarifying how dropout students transition until graduation. This enables quantitative identification of appropriate intervention points and contributes to the effective design of dropout prevention measures.

Shiratori et al. also conducted research classifying students into several learning behavior patterns by predicting weekly GPA in the first spring semester using a random forest model and performing cluster analysis with the X-means method [11]. This study showed that the learning state in the first spring semester is extremely important for graduation within the standard period, and that understanding state transitions within the semester enables the identification of optimal intervention points.

In online learning environments, Early Warning Systems (EWS) that utilize learning logs from LMS have developed as effective dropout prevention measures. Bañeres et al. analyzed daily learning behavior data, focusing on the number of days assignments were not submitted, and constructed a system for timely intervention for high-risk learners at an online university. This significantly reduced dropout rates and promoted active engagement with assignments [12]. The research results of Shiratori et al. also suggest that integrating LMS activity data and semester-by-semester academic and external relationship data can be applied to dropout prevention measures at Japanese universities.

2.4 Integration with Student Support

The process leading to dropout is closely related not only to academic ability and attendance but also to psychological fatigue and socioeconomic factors. A survey study in Portugal analyzed variables in five domains, including family educational background, economic status, and learning fatigue, and found that "academic exhaustion" is a major factor in dropout intention [13]. This suggests that comprehensive interventions, including psychological and economic support, are necessary for future dropout prevention measures.

To put early prediction models into practice, a support system for students identified as at risk is essential. In a randomized field experiment by Plak et al., risk prediction by machine learning models was incorporated into student counseling, and tutors provided guidance based on risk information. However, simply notifying students of their risk did not lead to significant improvements in dropout rates or academic performance [14]. This indicates that, in addition to early warnings, comprehensive interventions such as specific feedback, skill development support, and psychological and economic assistance are indispensable.

3 Proposal of a Cyclic Dropout Prevention Model

In this study, we propose a "Cyclic Dropout Prevention Model" that aims to enhance dropout prevention and student support in universities by connecting a series of processes—data integration → dropout prediction (risk estimation) → student status monitoring → student status classification → implementation of support measures → evaluation of effectiveness → reintegration of data—in a continuous cycle.

Figure 1 shows an overview of the cyclic dropout prevention model. This model centrally manages and integrates diverse data dispersed within the university, predicts dropout risk based on the integrated information, and visualizes and classifies student status. Based on the classification results, support measures are implemented for individuals or groups, and the effectiveness of these interventions is evaluated and reintegrated into the system, enabling continuous improvement and optimization.

(1) Data Integration

To efficiently and effectively promote dropout prevention measures, it is necessary to centrally collect student data related to dropout, which is dispersed across various departments within the university. Such data includes, as the target variable, information on whether a student has dropped out or not, and as explanatory variables, basic student information, number of absences and grade point average in high school, entrance examination data, placement test results, attendance records, academic performance, and various survey results.

(2) Dropout Risk Prediction

Using the integrated data, a model is constructed to predict the dropout risk for each student. Prediction can be performed either as a time series or as a one-time risk assessment. Referring to previous studies, risk estimation is conducted by combining machine learning algorithms and statistical methods.

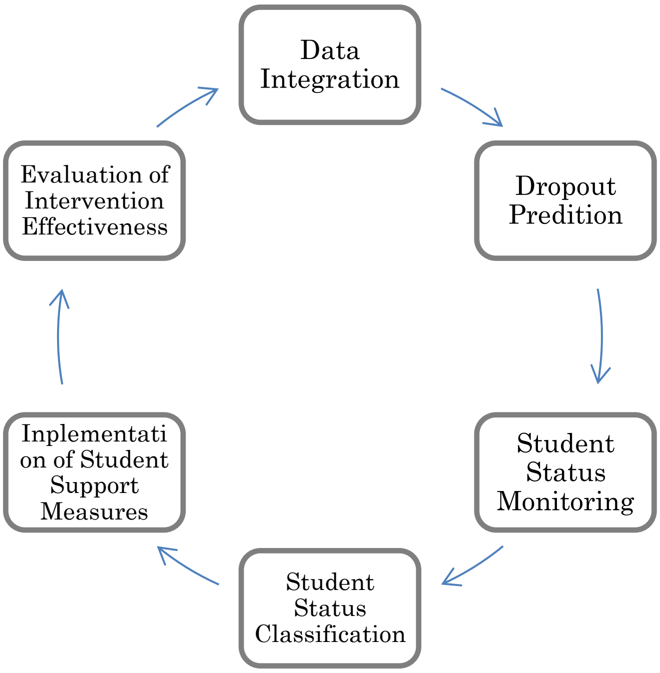

Figure 1: Cyclic Dropout Prevention Model

(3) Student Status Monitoring

Based on the predicted dropout risk, the status of each student is monitored. By analyzing trends such as increasing or decreasing risk and the timing of status changes, the timing and targets for support interventions can be clarified.

(4) Student Status Classification

Based on the results of status monitoring, students are classified and grouped according to their status patterns. Using clustering techniques such as X-means, the following patterns can be identified:

- A pattern in which risk remains persistently high, leading to dropout
- A pattern in which risk temporarily increases but subsequently decreases, resulting in graduation
- A pattern in which risk remains consistently low, resulting in graduation

This classification can be applied not only to students who drop out but also to those who graduate, allowing for the understanding of learning trajectories from admission to graduation as groups.

(5) Implementation of Student Support Measures

Based on the classification results, support measures are implemented for individuals or groups. Representative measures include:

- Early Alert System: Notifying high-risk students of their risk status and connecting them to individual counseling or follow-up support. The design can utilize institutional research (IR) data.
- Enhancement of First-Year Education: As support for adapting to the new environment of university, seminars and workshops are conducted to foster learning habits and a sense of belonging.

These measures are designed and implemented based on an understanding of the size and composition of high-risk groups, enabling effective support.

(6) Evaluation of Intervention Effectiveness

After implementing interventions, data such as attendance, academic performance, and changes in risk trends are collected to evaluate effectiveness. For example, by analyzing the impact of attendance in first-year seminars on absenteeism in other classes, the ripple effects of interventions can be assessed.

The above processes (1) to (6) form a cyclic support model, where data flows from prediction, monitoring, classification, intervention, and evaluation, and the evaluation data is reintegrated at the end. This model integrates previously separate research on student data, high-risk student prediction, status monitoring, pattern classification, intervention, and evaluation, resulting in an effective and efficient support model that contributes to substantial dropout prevention.

4 Design of Intervention Phase

In the cyclic dropout prevention model, the intervention phase is a crucial stage where support is provided to students identified as high-risk based on the classification of their status. This chapter describes the methods for classifying student status and targeting, the design and implementation of specific support measures, and the evaluation of intervention effectiveness.

4.1 Classification and Targeting of Student Status

Based on the estimated dropout risk, the patterns of risk transitions for each student are analyzed and classified. Clustering techniques such as X-means are used to extract the following patterns:

- Persistently High-Risk Pattern: The risk remains above a certain threshold for an extended period, and the student is highly likely to drop out.
- Temporarily Elevated Pattern: The risk temporarily increases but subsequently decreases, leading to graduation.
- Consistently Low-Risk Pattern: The risk remains low and stable, and the student graduates.

This classification can be applied not only to students who drop out but also to those who graduate, allowing for the understanding of learning trajectories from admission to graduation as groups. The classification results are utilized for selecting intervention targets and designing support content.

4.2 Design and Implementation of Support Measures

Based on the classification results, support measures are implemented for individuals or groups. The main interventions include:

- Early Alert System:

  A mechanism that notifies high-risk students or faculty members of elevated dropout risk. After notification, individual counseling or follow-up support is provided. Many universities are developing environments where students can check their own learning status and risk trends, and the design of these systems leverages institutional research (IR) data.

- Enhancement of First-Year Education:

  Systematic design of first-year education as support for adapting to the new environment of university life. Based on theories such as Tinto's, seminars and workshops are conducted to foster learning habits and a sense of belonging. These are designed as interventions for high-risk groups as a whole, allowing for effective implementation based on an understanding of the group's size and composition.

Support measures are designed from both individual (faculty) and organizational perspectives, and are classified according to the target (individual students or the entire student body). For example, individual support includes counseling and academic advising, while organizational support includes the institutionalization of alert systems and first-year education programs.状態

4.3 Evaluation and Feedback of Intervention Effectiveness

After implementing interventions, data such as attendance, academic performance, and changes in risk trends are collected to evaluate effectiveness. Although evaluation indicators differ by university, comparing results before and after interventions allows for the assessment of changes in risk trends and academic outcomes among targeted students.

For example, by analyzing how attendance in first-year seminar classes affects absenteeism in other courses, the ripple effects of interventions can be evaluated. Such evaluation results are reflected in the design of interventions for subsequent years, contributing to the improvement of the cyclic model.

5 Future Perspectives and Challenges

In this study, we proposed a cyclic practical model that structurally connects the processes of "data integration → dropout prediction → status monitoring and classification → support interventions → evaluation of effectiveness → reutilization of data" in order to unify and continuously improve dropout prevention and student support in universities. Specifically, we established a foundation for integrating and visualizing data dispersed within the university, constructed a time-series dropout risk estimation model using attendance and academic performance data, and designed a framework that cyclically connects student status monitoring and classification based on risk patterns with concrete support interventions and their evaluation.

Through this approach, we demonstrated that it is possible to move beyond one-off risk estimation and support implementation, and to build a continuous improvement cycle based on data, thereby advancing dropout prevention and student support.

Future challenges include the following points:

- Full-scale Operation and Effectiveness Verification of the Model:

  It is necessary to continuously operate the proposed model in actual educational settings and quantitatively and qualitatively verify the outcomes of the interventions.

- Accumulation of Multifaceted Data:

  By collecting and analyzing multifaceted data, including the implementation status of support interventions, student responses, and behavioral changes, the accuracy and effectiveness of the model can be further enhanced.

- Integration of Psychological and Social Factors:

  Dropout is influenced not only by academic performance and attendance but also by non-cognitive factors such as motivation, sense of isolation, and family environment. It is essential to establish methodologies for incorporating such information into the model.

- Strengthening Collaboration among Faculty and Staff:

  It is important to develop a system in which IR departments, academic affairs, and student support offices collaborate and share and utilize the model.

For future development, we aim to further improve the model based on operational results and to verify its applicability to other universities, with the goal of establishing a more practical and versatile student support model.

Acknowledgement

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research C 23K02668.

References

[1] Ministry of Education, Culture, Sports, Science and Technology. [Survey on the status of student enrollment (drop-outs and leaves of absence) (as of the end of the 2022 academic year)] “Gakusei no shugakujokyo (Chutai, Kyugaku) tou ni kansuru Chosa no Kekka ni Tsuite (in Japanese)”. 2023.

[2] Yomiuri Shimbun Kyoiku Network, [University Competencies 2019] ”Daigaku no Jitsuryoku 2019 (in Japanese)”,Chuo Koron Shinsho, 2018.

[3] V. Tinto, “Dropout from higher education: A theoretical synthesis of recent research,” Review of Educational Research, vol. 45, no. 1, pp. 89–125, 1975.

[4] J. P. Bean and B. S. Metzner, “A conceptual model of nontraditional undergraduate student attrition,” Review of Educational Research, vol. 55, no. 4, pp. 485–540, 1985.

[5] E. Kerby, “Toward a new theoretical model of student persistence in higher education,” Review of Higher Education, vol. 40, no. 3, pp. 327–356, 2017.

[6] B. Carballo-Mendívil, A. Arellano-González, N. J. Ríos-Vázquez, and M. del P. Lizardi-Duarte, “Predicting student dropout from day one: XGBoost-based early warning system using pre-enrollment data,” Applied Sciences, vol. 15, no. 16, p. 9202, 2025.

[7] M. Nagy and R. Molontay, “Interpretable dropout prediction: Towards XAI-based personalized intervention,” International Journal of Artificial Intelligence in Education, vol. 34, pp. 274–300, 2024.

[8] N. Shiratori, S. Tairi, T. Oishi, M. Mori, and M. Murota, “Analysis of the influence of learning state before university admission to college dropout using hierarchical Bayesian model,” presented at NPSE Int. Conf. on Science Education, 2019.

[9] O. Goren, L. Cohen, and A. Rubinstein, “Early prediction of student dropout in higher education using machine learning models,” in Proc. 17th Int. Conf. Educational Data Mining (EDM), 2024, pp. 349–361.

[10] N. Shiratori et al., “Identifying the high-risk duration to the semester to drop out of college using dropout probability,” IIAI Letters on Institutional Research, vol. 3, 2024.

[11] N. Shiratori, S. Tajiri, T. Oishi, M. Mori, and M. Murota, “Making dropout patterns using transition of dropout probability,” Japan Journal of Educational Technology, vol. 44, no. 1, pp. 11–22, 2020.

[12] D. Bañeres, M. E. Rodríguez-González, A.-E. Guerrero-Roldán, and P. Cortadas, “An early warning system to identify and intervene online dropout learners,” International Journal of Educational Technology in Higher Education, vol. 20, no. 3, 2023.

[13] B. Gonzalez, T. P. Mendes, R. Pinto, S. V. Correia, S. Albuquerque, and P. Paulino, “Predictors of higher education dropout intention in the post-pandemic era: The mediating role of academic exhaustion,” PLOS One, vol. 20, no. 7, p. e0327643, 2025.

[14] S. Plak, I. Cornelisz, M. Meeter, and C. van Klaveren, “Early warning systems for more effective student counselling in higher education: Evidence from a Dutch field experiment,” Higher Education Quarterly, vol. 76, pp. 131–152, 2022.