

Learning Analytics of Timing-Based Enrollment Behaviors for Early Academic Attrition Prediction

Anwar Ali Yahya 

Department of Artificial Intelligence, Faculty of Engineering and Technology, Ar-Rasheed Smart University, Sana'a, Yemen

E-mail: anwar.althari@ar-rasheed.edu.ye

Received: 17 May 2026

Revised: 18 June 2026

Accepted: 21 June 2026

تحليلات التعلم لسلوكيات التسجيل القائمة على التوقيت للتنبؤ المبكر بالتسرب الأكاديمي

أنور علي يحيى 

قسم الذكاء الاصطناعي، كلية الهندسة والتكنولوجيا، جامعة الرشيد الذكية، صنعاء، اليمن

الإيميل: anwar.althari@ar-rasheed.edu.ye

تاريخ التقديم: 17 مايو 2026 تاريخ إرسال التعديلات: 18 يونيو 2026 تاريخ القبول: 21 يونيو 2026

Abstract

This paper investigates the predictive power of timing-based enrollment behaviors in response to a practical institutional need for an early warning system aimed at reducing academic attrition associated with late course drops. A learning analytics methodology has been adopted to examine three timing-based behavioral indicators: late enrollment, academic planning regularity, and enrollment procrastination, as early predictors of late course drop across three analytical levels, micro (course-level), meso (semester-level), and macro (program-level). More specifically, the three timing-based enrollment behaviors have been extracted from students' transactional enrollment logs of 13 semesters, and logistic regression analysis between each indicator and late course drop, at each level, has been conducted. The findings indicate that both late enrollment and enrollment procrastination are robust positive predictors of academic risk across all analytical scales ($p < 0.001$), with effect sizes becoming most pronounced at the programmatic level (Odds ratio=34.54 for late enrollment). In contrast, academic planning regularity emerged as a protective indicator (Odds ratio=0.404 at the meso level) associated with improved preparedness for semester workload demands. The findings support the transferability of timing-based behavioral indicators analytics while providing actionable intelligence for institutional early warning systems to enable earlier and more targeted student interventions before critical academic loss occurs.

Keywords: Learning analytics, transferability, academic attrition, late course drops, enrollment behaviors, transactional logs, early warning systems

المخلص

هدفت هذه الورقة إلى تقصي القدرة التنبؤية لسلوكيات التسجيل القائمة على التوقيت، والتي تأتي استجابةً لحاجة مؤسسية لعمل كنظام إنذار مبكر يهدف إلى الحد من التسرب الأكاديمي المرتبط بالانسحاب المتأخر من المقررات الدراسية. وقد تم اعتماد منهجية تحليل التعلم لدراسة ثلاثة مؤشرات سلوكية قائمة على التوقيت، وهي: التسجيل المتأخر، وانتظام التخطيط الأكاديمي، والمطالبة في التسجيل، باعتبارها مؤشرات تنبؤية مبكرة للانسحاب المتأخر من المقررات الدراسية عبر ثلاثة مستويات تحليلية: المستوى الجزئي (مستوى المقرر الدراسي)، والمستوى المتوسط (مستوى الفصل الدراسي)، والمستوى الكلي (مستوى البرنامج). وعلى وجه الدقة، تم استخلاص سلوكيات التسجيل الثلاثة المرتبطة بالتوقيت من سجلات الطلاب خلال 13 فصلاً دراسياً، ثم أُجري تحليل الانحدار اللوجستي بين كل مؤشر والانسحاب المتأخر من المقررات الدراسية على كل مستوى. تشير النتائج إلى أن كلا من التسجيل المتأخر والمطالبة في التسجيل مؤشران تنبؤيان إيجابيان قويان للمخاطر الأكاديمية عبر جميع المقاييس التحليلية ($p < 0.001$)، مع بروز تأثيرهما بشكل أوضح على مستوى البرنامج (نسبة الأرجحية = 34.54 للتسجيل المتأخر). في المقابل، برز انتظام التخطيط الأكاديمي كمؤشر وقائي (نسبة الأرجحية = 0.404 على المستوى المتوسط)، وارتبط بتحسين الاستعداد لمتطلبات عبء العمل الفصلي. تدعم هذه النتائج إمكانية تطبيق تحليلات المؤشرات السلوكية القائمة على التوقيت، مع توفير معلومات قابلة للتنفيذ لأنظمة الإنذار المبكر المؤسسية؛ مما يتيح تدخلات طلابية مبكرة وأكثر استهدافاً قبل حدوث خسائر أكاديمية جسيمة.

الكلمات المفتاحية: تحليلات التعلم، قابلية التعميم، التسرب الأكاديمي، الانسحاب المتأخر من المقررات، سلوكيات التسجيل، سجلات المعاملات، أنظمة الإنذار المبكر.

This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 3135-6842 | E-ISSN: 3135-6850 / © 2026 The Author. Published by Ar-Rasheed Smart University, Yemen.

Journal website: <https://journals.ar-rasheed.edu.ye/RUJBAS/en>

Ar-Rasheed University Journal of Basic and Applied Sciences 2026, 1(1), 24-48

<https://doi.org/10.67151/rujbas.v1i1.38>

1. Introduction

In response to the rapid growth of digital transformation in higher education, the two interdisciplinary fields of educational data mining (EDM) and learning analytics (LA) continue to develop computational methods for the effective analysis of large-scale educational datasets, which are being generated in increasing volumes and at an accelerating pace [1]. In achieving this common aim, researchers in both fields utilize various types of data on learners and their learning environments. A fundamental distinction among educational data is whether they reflect static or dynamic aspects of learners [2]. While static data provide a fixed historical profile, dynamic data, often termed 'digital trails' [3], capture the evolving nature of students' academic journeys as they interact with the learning environment.

Recent years have witnessed a growing shift in research focus toward the utilization of these dynamic behavioral data streams [4, 5], owing to their inherent ability to capture learners' behavioral characteristics and reveal their continuously evolving academic journeys—rather than fixed historical profiles [2]. This growth has been driven by the recognition that longitudinal behavioral data enable the construction of more accurate student models. In this respect, various types of dynamic educational data have been accumulated, including high-granularity interaction logs from learning management systems (LMSs) (e.g., login frequencies and resource access patterns), digital records of campus activities obtained through smart card check-ins (e.g., library entries and clinic visits), and Wi-Fi-based spatiotemporal trajectories that track students' movements across different study and recreational areas on campus.

A potent source of dynamic data capable of reflecting the behavioral traces of students' decision-making processes can be found in

transactional enrollment data. These data provide a process-oriented view of the meso-level enrollment trajectories that students navigate during the initial registration period before arriving at their finalized schedules [6]. In contrast to learning management system (LMS) activity logs, which primarily capture students' engagement and interactions after a course has commenced, transactional enrollment data reveal the sequence of enrollment decisions made before instruction begins. This level of granularity enables the identification of complex behaviors, such as course shopping and enrollment procrastination (EP), which remain largely invisible in other dynamic data streams [6]. Furthermore, as these records are extracted directly from administrative student information systems, they provide complete cohort coverage and are not subject to the self-reporting and non-response biases commonly associated with survey-based dynamic data [6].

A critical challenge facing higher education institutions, resulting in significant credit loss and financial penalties, is student attrition. In response to this problem, researchers have increasingly focused on developing early warning models capable of identifying at-risk students during the initial weeks of study [6, 7]. Fine-grained transactional enrollment data, in the form of timestamped records of course add and drop actions, have emerged as a valuable source of predictive indicators [6, 8]. The advantages of these data lie in their ability to provide timely and continuous predictors of student academic performance by revealing temporal decision-making processes weeks before official academic deadlines [8, 9]. This capability is of paramount importance for identifying and mitigating student attrition, particularly late course drops (LDs), which result in substantial yet often overlooked credit loss and financial penalties [6, 7, 10]. The identification of these digital trails during the enrollment phase enables institutions to detect at-risk students

early enough to implement pedagogical interventions and personalized support before academic loss occurs [5, 11, 12].

While existing literature has explored the impact of enrollment timing as an early predictor of student academic performance in diverse settings, a significant challenge in adopting reported findings persists due to the inherent and specific nature of LA as a heavily context-dependent field [1, 11]. More precisely, most research in LA originates from problems encountered at a specific institution and relies on data specific to that institution. Therefore, predictive models often lack external validity, generalizability, and practical utility when applied to diverse educational environments [11, 13]. For example, most research in enrollment-based analytics has been conducted at large US public institutions or online-only universities, which motivate the need for transferability research to investigate whether the established findings in these specific institutional contexts, remain valid and robust across different institutional settings [5, 4].

At Ar-Rasheed Smart University (RSU), to deal with the academic attrition problem, an early warning system (EWS) has been proposed, based on converting raw administrative logs into actionable intelligence required for effective decision-making [3]. This enables RSU to provide proactive support and intervention, such as targeted academic advising and workload awareness dashboards, before students reach a critical academic loss point [4, 12]. With this motivation, the present research seeks to examine the external validity of three timing-based student enrollment behaviors, specifically late enrollment (LE), academic planning regularity (APR), and EP, within the unique administrative context and student enrollment patterns of RSU [4, 12].

Despite the documented predictive power of enrollment timing in large US public universities, a significant research gap remains

regarding the transferability of these indicators to the unique administrative and cultural contexts of smart universities in the Middle East, such as RSU. While previous research often focuses on single-semester snapshots, this research establishes timing-based behaviors as objective behavioral proxies for unobserved psychometric traits like self-regulation and conscientiousness. Furthermore, the research introduces a novel multi-scale (micro–meso–macro) analytical framework that utilizes fine-grained transactional administrative logs to generate longitudinally updated behavioral profiles. This approach transforms routinely collected enrollment data into actionable intelligence for institutional EWS, enabling proactive pedagogical interventions before critical academic loss occurs.

2. Literature Review

Student academic performance has long been regarded as one of the key indicators of quality in higher education, around which institutional policies, teaching practices, and student support services are organized [14]. In this respect, academic performance reflects both the effectiveness of educational institutions in achieving their intended learning outcomes and the extent to which students successfully attain these outcomes [15]. Consequently, predicting future academic performance has become an important research objective, enabling institutions to move from reactive educational practices toward proactive intervention strategies. Such predictive capabilities allow timely implementation of academic advising, personalized feedback, and other support mechanisms that improve student success, enhance retention, and reduce academic failure [12]. Nevertheless, accurate prediction remains a challenging task because learning is inherently dynamic, cumulative, and influenced by a wide range of heterogeneous

factors that evolve throughout a student's academic journey [16].

To address this challenge, previous research has investigated numerous factors associated with student academic performance [4]. In general, these factors can be classified into four broad categories. The first category comprises demographic characteristics, including age, gender, ethnicity, and disability status. The second category includes prior academic records, such as high school grade point average (GPA), national examination scores, and university admission results. The third category encompasses psychometric characteristics, including self-efficacy, motivation, self-regulation, learning strategies, and personality traits such as the Big Five. Finally, socio-economic and environmental factors, including family income, parental education, and institutional support, constitute another important category of predictors [11, 17, 18]. Although these factors have demonstrated considerable predictive capability for traditional student populations, their effectiveness is often reduced when applied to non-traditional, mature, or socioeconomically diverse student groups, motivating researchers to investigate additional sources of predictive information [2]. In this respect, psychometric measures have received considerable attention because they capture non-cognitive dimensions of learning that are not reflected in conventional academic records. Previous studies have shown that variables such as conscientiousness, self-efficacy, and learning strategies explain a substantial proportion of the variation in academic performance. However, the dependence of these measures on self-reported questionnaires introduces several practical limitations, including response bias, limited scalability, and the difficulty of collecting such information continuously throughout students' academic progression [7].

More recently, research has increasingly shifted toward the utilization of dynamic

behavioral data, which consist of continuously evolving digital trails generated as students interact with institutional systems throughout their academic activities [3, 5]. Unlike static demographic or historical academic records, these data provide a real-time representation of students' behavioral patterns and learning activities. In this respect, various forms of dynamic behavioral data have been utilized, including high-granularity interaction logs collected from LMSs [4, 12, 19] and intelligent tutoring systems [20]. These systems capture detailed information regarding students' engagement, such as login frequency, resource access patterns, clickstream events, and system-specific activities, including saving, simulating, or modifying learning artifacts. Beyond instructional platforms, researchers have also utilized behavioral information derived from smart ID card transactions [21, 22] and Wi-Fi access logs [21] to characterize students' broader campus activities, including library utilization, consumption habits, and spatiotemporal movement across university facilities. Another important source of dynamic behavioral information is represented by transactional enrollment records, which provide fine-grained histories of students' administrative interactions during course registration [8, 9]. Unlike conventional learning activity logs, these records provide a process-oriented representation of enrollment behavior by capturing the sequence of course additions, drops, and waitlist actions that students perform before finalizing their semester schedules [6, 8]. Consequently, mining these transactional records enables the identification of complex enrollment behaviors, such as course shopping, through which students repeatedly modify their schedules in an attempt to optimize workload, course selection, and personal preferences [6, 10].

An interesting utilization of transactional logs is the detection of early behavioral signs, such as schedule mismanagement or

enrollment uncertainty, that remain entirely invisible in other forms of dynamic behavioral data [7-9]. A specific group derived from fine-grained transactional enrollment logs has emerged as a potent early signal of student attrition [6, 7]. One of the most extensively researched indicators is LE, typically defined as enrolling in a course on or after the first day of instruction [12, 23]. Extensive research across diverse institutional settings confirms a statistically significant negative correlation between registration timing and final academic marks, with LE students frequently earning as much as a full GPA point lower than their on-time peers [12, 23]. This delay is often interpreted as a behavioral proxy for student diligence and self-regulation; students who register late often begin the term already at a disadvantage, having missed foundational syllabus overviews and essential initial assignments [7, 12]. Closely linked to these temporal patterns is the construct of EP, which measures the relative time point at which a student initiates administrative actions in relation to institutional deadlines [6, 23]. Research indicates that high EP often triggers a vicious cycle of self-regulatory failure, where students inadvertently take on higher effective workloads and more difficult course sets that they are ultimately unprepared to manage [9, 24]. Conversely, APR, defined as the even spacing of enrollment actions throughout the registration window, serves as a critical protective indicator [6, 9]. Students who exhibit regular planning habits are statistically less likely to experience the schedule

mismanagement that leads to LDs and institutional credit loss [9, 10]. By mining these timing-based "digital trails" during the enrollment phase, institutions can develop proactive EWS that provide support at the point of registration rather than waiting for semester-end assessment failures [4, 12].

Despite the predictive power of enrollment behavioral indicators, a fundamental limitation often encountered in the existing models is their lack of external validity. More precisely, most existing LA models are heavily dependent on private, institution-specific datasets, which make them context-dependent models. Such models are trained on specific populations and frequently fail when applied to diverse educational environments [4]. Consequently, the existing LA models are restricted in their practical utility due to their inability to link their results to local educational policies and draw actionable insights required for practitioners, such as advisors and administrators, to make better decisions [4]. These inherent problems motivate transferability research that validates established enrollment behavioral indicators within specific institutional contexts to ensure that institutional EWS are both reliable and precisely tuned [6, 12].

To provide a clear overview of how the current study relates to and extends previous research in timing-based behavioral analytics, Table 1 summarizes the key dimensions of relevant prior studies in comparison to our multi-scale approach.

Table 1. Comparison of relevant timing-based predictive research

Ref.	Data source	Analytical level	Main outcome	Behavioral proxy
[6]	Transactional enrollment logs	Meso	Late course drops	Procrastination
[12]	Registration timestamps	Micro	Final course score	Registration timing
[7]	Transactional student records	Meso / Macro	Student attrition	Schedule changes
[25]	LMS activity logs	Micro	Course achievement	Learning persistence
This research	Transactional enrollment logs	Micro, Meso, Macro	Late course drops	Timing-based behaviors

3. Methodology

The primary objective of this research is to implement an LA framework using the transactional logs of RSU to investigate timing-based behavioral indicators as early predictors of student attrition, specifically LDs. Since an LA methodology has been adopted, the methodology of this research is viewed as a specialized adaptation of a knowledge discovery process that involves the following four steps [4, 6]: data acquisition; data processing and integration, feature engineering, correlation analysis, and actionable insights synthesis.

3.1. Data Acquisition

To achieve the objectives of this research, transactional enrollment records were extracted from the RSU Student Information System, which serves as the official repository of students' academic and administrative records. The study considers students enrolled in five undergraduate programs offered by the Faculty of Engineering and Technology (FET), namely Information Technology, Cybersecurity, Interior Design, Graphics and Multimedia, and Mechatronics. The dataset covers a period of 13 consecutive academic

semesters, providing a comprehensive longitudinal view of students' enrollment trajectories and enabling the analysis of behavioral patterns over an extended period.

The collected data comprise two primary categories. The first category consists of transactional course enrollment logs generated during the registration period of each academic semester. These records capture the complete sequence of enrollment activities through which students progressively construct their final semester schedules, thereby providing a process-oriented representation of enrollment behavior at the meso level. For each enrollment transaction, several fine-grained attributes are recorded. These include the transaction type documenting every course add and drop action initiated by the students; precise timestamps associated with each transaction, enabling the derivation of temporal behavioral indicators such as enrollment delay and procrastination; and administrative audit records, which preserve both the initial departmental decision and the final institutional approval, thereby distinguishing students' enrollment intentions from subsequent administrative adjustments. Figure 1 presents a sample of the raw transactional enrollment data.

Program Name	Student ID	Student Name	Course ID	Course Name	Entrance Semester	Semester No	Application Type	Department Approval	Application Status	Application Date	Final Approval date
تقنية المعلومات	202210101212		101244	تصميم وتطوير الويب	20221	20231	حذف	مرفوض	غير مسجل	14/07/2023	10/08/2023
الأن السياري	202210100106		101244	تصميم وتطوير الويب	20221	20231	حذف	معتد	غير مسجل	14/07/2023	
الأن السياري	202210100237		101244	تصميم وتطوير الويب	20221	20231	حذف	معتد	غير مسجل	14/07/2023	
تقنية المعلومات	202210101212		101247	أمن المعلومات	20221	20231	إضافة	معتد	مسجل	14/07/2023	08/08/2023
تقنية المعلومات	202210101193		101247	أمن المعلومات	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023
تقنية المعلومات	202210100385		101247	أمن المعلومات	20221	20231	حذف	معتد	غير مسجل	14/07/2023	
تقنية المعلومات	202210101212		101248	تحليل وتصميم النظم	20221	20231	إضافة	معتد	مسجل	14/07/2023	08/08/2023
تقنية المعلومات	202210101193		101248	تحليل وتصميم النظم	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023
تقنية المعلومات	202210101212		101253	إدارة الشبكات	20221	20231	إضافة	معتد	مسجل	14/07/2023	20/09/2023
تقنية المعلومات	202210101193		101253	إدارة الشبكات	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023
تقنية المعلومات	202210101212		101254	مشاريع تكنولوجيا المعلومات	20221	20231	إضافة	معتد	مسجل	14/07/2023	08/08/2023
تقنية المعلومات	202210100074		101254	مشاريع تكنولوجيا المعلومات	20221	20231	حذف	معتد	غير مسجل	14/07/2023	
هندسة ميكاترونكس	202210100056		101264	أساسيات البرمجة	20221	20231	إضافة	معتد	مسجل	14/07/2023	12/08/2023
هندسة ميكاترونكس	202210101262		101264	أساسيات البرمجة	20221	20231	إضافة	معتد	مسجل	14/07/2023	12/08/2023
جرافيكس وميديا	202210100819		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	12/08/2023
جرافيكس وميديا	202210100974		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	12/08/2023
الأن السياري	202210100897		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023
الأن السياري	202210100106		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023
جرافيكس وميديا	202210100387		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	12/08/2023
تقنية المعلومات	202210100385		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023
الأن السياري	202210100237		101266	الإحصاء والاحتمالات	20221	20231	إضافة	معتد	مسجل	14/07/2023	10/08/2023

20221 refers to the first semester of the academic year 2022.

Figure 1. Snapshot of RSU-FET transactional logs data.

The second form of raw data is a recording of student academic performance. This involves granular performance indicators

for every student-course pair: the course grade point, the raw mark standardized on a 4-scale point, and course status, which is the final

administrative status of the enrolled course, with categorical labels such as passed, fail, and late course drop (LD). The specific status "LD" (withdrawals after the official penalty-free deadline) is identified in literature as a key indicator of academic planning mismanagement. Figure 2 is a snapshot of the academic performance data.

As will be described in the subsequent subsections, the two forms of raw data are integrated into one record for each student-pair and serves as the foundation for the subsequent preprocessing and feature engineering stages (Figure 3). A summary of the dataset's scale is provided in Table 2.

Besides the forms of student data mentioned above, data on official study plans

for each of the five programs have been collected. These study plans serve as the normative baseline for optimal university curricula, defining the prescribed sequence of courses students are expected to pursue to achieve on-time graduation. For every academic program, the study plan dataset includes: curricular classification, detailed markers indicating whether a course is compulsory, university elective, or a program elective; academic prerequisites, the specific pre-requisite courses required for enrollment in a particular course; and course metadata including course hours, max score, and min score, which provide the necessary context for interpreting the Grade Point data.

Program Name	Student ID	Student Name	Course ID	Course Name	Entrance Semester	Semester No	Score	Grade Point	Relative Score	Course Status	Final Approval	Credit Hours	Modification Date
جرافيكس وماتريديا	202210100841		101590	اللغة الإنجليزية (102)	20221	20223	0	0	0	راسب	موافقة	4	2023/07/20
جرافيكس وماتريديا	202210100849		101590	اللغة الإنجليزية (102)	20221	20223	0	0	0	راسب	موافقة	4	2023/07/20
تقنية المعلومات	202210101229		101005	الصراف العربي الإسرائيلي	20221	20223	92	3.7	92	ناجح	موافقة	2	2023/07/23
الامن السيبراني	202210100860		101589	اللغة الإنجليزية (101)	20221	20223	148	2.3	74	ناجح	موافقة	4	2023/07/25
تقنية المعلومات	202210100353		101263	رياضيات (2)	20211	20223	79	1	53	ناجح	موافقة	3	2023/07/26
تقنية المعلومات	202210100969		101589	اللغة الإنجليزية (101)	20221	20223	190	4	95	ناجح	موافقة	4	2023/07/27
التصميم الداخلي	202210101001		101589	اللغة الإنجليزية (101)	20221	20223	200	4	100	ناجح	موافقة	4	2023/07/27
التصميم الداخلي	202210100422		101590	اللغة الإنجليزية (102)	20221	20223	0	0	0	راسب	موافقة	4	2023/07/27
الامن السيبراني	202210100897		101590	اللغة الإنجليزية (102)	20221	20223	0	0	0	راسب	موافقة	4	2023/07/27
الامن السيبراني	202210100858		101262	رياضيات (1)	20221	20223	81	1	54	ناجح	موافقة	3	2023/07/27
التصميم الداخلي	202210100422		101002	لغة عربية (1)	20221	20223	0	0	0	طالب عن النهائي	موافقة	2	2024/01/22
التصميم الداخلي	202210100422		101262	رياضيات (1)	20221	20223	20	0	13	راسب	موافقة	3	2024/01/22
التصميم الداخلي	202210100999		101263	رياضيات (2)	20221	20223	84	1.3	56	ناجح	موافقة	3	2024/01/22
تقنية المعلومات	202210100152		101262	رياضيات (1)	20221	20223	0	0	0	طالب عن النهائي	موافقة	3	2024/01/22
جرافيكس وماتريديا	202210100796		101589	اللغة الإنجليزية (101)	20221	20223	0	0	0	طالب عن النهائي	موافقة	4	2024/01/24
جرافيكس وماتريديا	202210100805		101589	اللغة الإنجليزية (101)	20221	20223	158	2.7	80	ناجح	موافقة	4	2024/01/27
جرافيكس وماتريديا	202210101178		101589	اللغة الإنجليزية (101)	20221	20223	162	3	81	ناجح	موافقة	4	2024/01/27
هندسة ميكاترونكس	202210100193		101264	أساسيات البرمجة	20211	20223	90	1.7	60	ناجح	موافقة	3	2024/01/27
تقنية المعلومات	202210100311		101019	لغة الإنجليزية للتخصص 3	20221	20223	0	0	0	طالب عن النهائي	موافقة	2	2024/01/27
تقنية المعلومات	202210100311		101243	نظم قواعد البيانات	20221	20223	0	0	0	طالب عن النهائي	موافقة	3	2024/01/27
تقنية المعلومات	202210100762		101243	نظم قواعد البيانات	20221	20223	0	0	0	طالب عن النهائي	موافقة	3	2024/01/27

Figure 2. Snapshot of the RSU-FET student academic performance.

Table 2. Quantitative summary of the collected RSU-FET data

Data aspect	Magnitude / Details
Institutional context	FET, RSU
Observation period	13 Semesters
Academic programs	Information Technology, Cybersecurity, Interior Design, Graphics and Multimedia, Mechatronics)
Primary data source	RSU student information system
Transaction types	Add, Drop, Department decision, Final approval
Performance identifiers	Grade point (GP), Course status (Pass/Fail/LD)
Data format	Student-Period (one row per student-course-semester)
Size of enrollment data	52890 records
Size of academic data	51771 records

Program Name	Study Plan Semester	Course Semester	Course ID	Course Code	Course Name - Arabic	Course Name - English	Course Hours	Pre-requisite IDs	Max Score	Min Score	Course Type-I	Course Type-II	# Required Electives	# Electives	Compulsory	University Elective	Program Elective
تقنية المعلومات	20241	3	101266	FET201	الإحصاء والاحتمالات	Probability & Statistics	3	[]	150	75	3	1	3	9	TRUE	FALSE	FALSE
التصميم الداخلي	20231	3	101570	INTD234	التصميم الداخلي بواسطة الكمبيوتر	Computer Aided Interior Design	3	[]	150	75	1	1	1	1	TRUE	FALSE	FALSE
التصميم الداخلي	20241	3	101570	INTD234	التصميم الداخلي بواسطة الكمبيوتر	Computer Aided Interior Design	3	[]	150	75	1	1	1	1	TRUE	FALSE	FALSE
جرافيكس وملتيميديا	20211	3	101274	BMG261	التصوير الفوتوغرافي	Digital Photography	3	['101270']	150	75	1	1	3	8	TRUE	FALSE	FALSE
جرافيكس وملتيميديا	20231	3	101274	BMG261	التصوير الفوتوغرافي	Digital Photography	3	['101270']	150	75	1	1	3	8	TRUE	FALSE	FALSE
جرافيكس وملتيميديا	20241	3	101274	BMG261	التصوير الفوتوغرافي	Digital Photography	3	['101637']	150	75	1	1	3	8	TRUE	FALSE	FALSE
التصميم الداخلي	20222	3	101274	INTD415	التصوير الفوتوغرافي	Digital Photography	3	[]	150	75	1	1	1	1	TRUE	FALSE	FALSE
جرافيكس وملتيميديا	20241	3	101469	BMG249	التفكير التصميمي والابتكار	Design Thinking & Innovation	2	['101637']	100	50	1	1	3	8	TRUE	FALSE	FALSE
جرافيكس وملتيميديا	20211	3	101001	UNR101	الثقافة الإسلامية	Islamic Culture	2	[]	100	50	2	1	3	8	TRUE	FALSE	FALSE
جرافيكس وملتيميديا	20231	3	101001	UNR101	الثقافة الإسلامية	Islamic Culture	2	[]	100	50	2	1	3	8	TRUE	FALSE	FALSE
التصميم الداخلي	20231	3	101001	UNR101	الثقافة الإسلامية	Islamic Culture	2	[]	100	50	2	1	1	1	TRUE	FALSE	FALSE
التصميم الداخلي	20241	3	101001	UNR101	الثقافة الإسلامية	Islamic Culture	2	[]	100	50	2	1	1	1	TRUE	FALSE	FALSE
التصميم الداخلي	20231	3	101008	UNR108	التقانة الوطنية	National Education	2	[]	100	50	2	1	1	1	TRUE	FALSE	FALSE
التصميم الداخلي	20241	3	101008	UNR108	التقانة الوطنية	National Education	2	[]	100	50	2	1	1	1	TRUE	FALSE	FALSE
تقنية المعلومات	20231	3	101266	FET201	الإحصاء والاحتمالات	Probability & Statistics	3	[]	150	75	3	1	3	9	TRUE	FALSE	FALSE
تقنية المعلومات	20241	3	101266	FET201	الإحصاء والاحتمالات	Probability & Statistics	3	[]	150	75	3	1	3	9	TRUE	FALSE	FALSE
التصميم الداخلي	20231	3	101570	INTD234	التصميم الداخلي بواسطة الكمبيوتر	Computer Aided Interior Design	3	[]	150	75	1	1	1	1	TRUE	FALSE	FALSE
التصميم الداخلي	20241	3	101570	INTD234	التصميم الداخلي بواسطة الكمبيوتر	Computer Aided Interior Design	3	[]	150	75	1	1	1	1	TRUE	FALSE	FALSE
الآمن الصناعي	20231	3	101019	FET202	اللغة الإنجليزية للتخصص	English for Specific Purposes	2	[]	100	50	3	1	3	12	TRUE	FALSE	FALSE
الآمن الصناعي	20241	3	101019	FET202	اللغة الإنجليزية للتخصص	English for Specific Purposes	2	[]	100	50	3	1	3	12	TRUE	FALSE	FALSE

Figure 3. Snapshot of the RSU-FET programs study plans.

3.2. Data Preprocessing and Integration

The raw data extracted from the RSU student information system constitute a large repository of administrative digital trails that require extensive preprocessing before they can be utilized for predictive modeling. This stage is essential for ensuring that the information collected from the three primary data sources—transactional enrollment records, academic performance records, and study plans—is complete, consistent, and suitable for subsequent analysis. In this respect, preprocessing establishes a reliable representation of students' academic behavior by resolving inconsistencies across the different data sources while preserving the integrity of the original administrative records. Consequently, the resulting dataset provides an objective basis for behavioral analysis without the response and recall biases commonly associated with self-reported data.

Each data source was subjected to a series of preprocessing operations to improve data quality prior to integration. Student privacy was preserved through the anonymization of all personally identifiable information, including the removal of student names, thereby ensuring compliance with institutional requirements for research ethics. Temporal attributes, including application dates, modification dates, and final approval dates, were subsequently standardized into a unified date format to ensure consistency across all transactional records. Missing values were

handled according to the availability of historical information; records with sufficient academic history were completed where possible, whereas observations with insufficient information were excluded to preserve the reliability of the subsequent analysis. In addition, abnormal temporal values that fell outside the official registration period were identified and corrected according to institutional registration rules. Categorical variables, including program name, entrance semester, and semester number, were encoded into numerical representations to facilitate computational processing, while duplicate records were identified and removed to ensure that repeated administrative transactions did not distort the derived behavioral indicators.

Following the preprocessing stage, the three cleaned datasets were systematically integrated into a single longitudinal dataset representing each student's academic trajectory throughout the study period. This integration process established a unified analytical repository that combines enrollment behavior, academic performance, and curriculum information, thereby providing the foundation for subsequent feature engineering and predictive modeling. This phase began by unifying column nomenclature across the sources, such as reconciling 'Course Hours' from the program study plans with 'Course Credit Hours' from the academic performance records. The enrollment data was then linked with the academic performance data using unique combinations of 'Program ID', 'Student

ID', 'Entrance Semester', 'Semester No', and 'Course ID'. This combined data source was further integrated with the program study plan data using the 'Program ID', 'Study Plan Semester', and 'Course ID' columns as relational keys. To enhance the granularity of the longitudinal analysis, a new feature, 'Student Sem No', was derived by calculating

the student's relative semester standing based on their entry point ('Entrance Semester') and the current 'Semester No'. In summary, the preprocessing stage transformed raw, fragmented administrative records into a unified longitudinal student-period format as illustrated in Figure 4.

Program ID	Student ID	Entrance Semester	Semester No	Student Sem No	Study Plan Semester	Course ID	Application Type	Department Approval	Application Status	Application Date	Approval Date	Score	Grade Point	Relative Score	Course Status	Final Approval	Course Semester #	Credit Hours	Modification Date	Course Code	Pre-requisite IDs	Max Score	Min Score	Compulsory	University Elective	Program Elective
2	202210100892	4	5	2	4	101006	إدارة	مهند	مستج	1/9/2023	1/11/2023	80	3	80	ناجح	موافقة	2	2	3/28/2023	UNR105	["101002"]	100	50	TRUE	FALSE	FALSE
4	202210100898	4	5	2	4	101006	إدارة	مهند	مستج	1/9/2023	1/25/2023	84	3	84	ناجح	موافقة	2	2	1/25/2023	UNR105	["101002"]	100	50	TRUE	FALSE	FALSE
4	202210100905	4	5	2	4	101006	إدارة	مهند	مستج	1/9/2023	1/16/2023	97	4	97	ناجح	موافقة	2	2	1/16/2023	UNR105	["101002"]	100	50	TRUE	FALSE	FALSE
5	202110100200	1	5	5	1	101010	إدارة	مهند	مستج	1/9/2023	1/15/2023	61	1.7	61	ناجح	موافقة	4	2	4/5/2023	UNR202	[""]	100	50	TRUE	FALSE	FALSE
5	202110100201	1	5	5	1	101010	إدارة	مهند	مستج	1/9/2023	1/11/2023	100	4	100	ناجح	موافقة	4	2	4/5/2023	UNR202	[""]	100	50	TRUE	FALSE	FALSE
5	202110100215	1	5	5	1	101010	إدارة	مهند	مستج	1/9/2023	1/11/2023	94	3.7	94	ناجح	موافقة	4	2	4/5/2023	UNR202	[""]	100	50	TRUE	FALSE	FALSE
5	202110100218	1	5	5	1	101010	إدارة	مهند	مستج	1/9/2023	1/11/2023	69	2	69	ناجح	موافقة	4	2	4/5/2023	UNR202	[""]	100	50	TRUE	FALSE	FALSE
1	202110100327	1	5	5	1	101010	حرف	مهند	مستج	1/9/2023	2/8/2023						4			UNR202	["101003"]	100	50	TRUE	FALSE	FALSE
1	202110100346	1	5	5	1	101010	حرف	مهند	مستج	1/9/2023							4			UNR202	["101003"]	100	50	TRUE	FALSE	FALSE
1	202110100392	1	5	5	1	101010	إدارة	مهند	مستج	1/9/2023	1/11/2023	82	3	82	ناجح	موافقة	4	2	4/4/2023	UNR202	["101003"]	100	50	TRUE	FALSE	FALSE
5	202110100480	1	5	5	1	101010	إدارة	مهند	مستج	1/9/2023	1/11/2023	95	4	95	ناجح	موافقة	4	2	4/5/2023	UNR202	[""]	100	50	TRUE	FALSE	FALSE

Figure 4. Snapshot of the integrated RSU-FET dataset.

3.3. Feature Engineering

In this phase, the three timing-based enrollment behavioral indicators, as manifestations of underlying student traits of conscientiousness, decisional procrastination, and academic self-efficacy are quantified to investigate their prevalence within RSU-FET dataset, across three hierarchical levels: micro (course), meso (semester), and macro (program) levels.

3.3.1. Late Enrollment

LE is formally defined as the act of enrolling in an academic course on or after the first day of the instructional term. However, for the RSU context, LE is defined as the act of enrolling in an academic course after the deadline of the allowed enrollment period, which span two weeks for course add and four weeks for course drop. Within LAs literature, LE indicator is widely utilized as a behavioral proxy for a student's academic diligence and momentum, with longitudinal research consistently linking registration delay to a decline of approximately one full GPA point in performance. In this research, the quantification of LE is extended across three hierarchical levels—micro (course), meso (semester), and macro (program)—to

differentiate between situational course-level delays and a longitudinal trait of chronic administrative procrastination. At the micro level, the indicator is defined relative to the specific enrollment window of an individual course. This metric identifies students at risk of "Syllabus Shock," as they may miss critical introductory material, orientation sessions, or early assignment deadlines for a specific subject. To ensure comparability across courses with varying enrollment periods, the behavior is quantified using the following formula:

$$micro_LE_c = \frac{T_{add} - T_{enroll_start}}{N_{enroll_days}}$$

Where T_{add} is the date the student applied to enroll in a specific course C , T_{enroll_start} is the opening date of the enrollment period, and N_{enroll_days} represents the total number of days in the enrollment window.

At meso level, the metric evaluates a student's overall diligence in navigating the institutional registration system for an entire semester. This captures the student's ability to secure a preferred schedule before sections become impacted or closed. The indicator for a student at semester S is quantified by averaging $micro_LE$ across all courses in

which the student enrolled during that semester:

$$meso_LE_S = \frac{1}{C} \sum_{i=1}^C micro_LE_i$$

Finally, at macro level, the indicator represents the longitudinal behavioral consistency of the student across their entire academic study at RSU. This level is designed to isolate LE students for whom administrative delay is a stable behavioral trait rather than a one-time incident. The macro indicator for a particular student is computed as the average of all semester-level meso-indicators:

$$macro_LE_x = \frac{1}{CS} \sum_{s=1}^{CS} meso_LE_{x,s}$$

3.3.2. Academic Planning Regularity

APR is defined as the degree to which a student exhibits a consistent and even distribution of registration activities throughout the enrollment window, rather than frantic or sporadic bursts of action. First introduced by Borchers [9], this index serves as a behavioral proxy for a student's conscientiousness and organizational self-regulation. Students with high regularity scores demonstrate a deliberate and sustained engagement with their academic trajectory, which the literature links to higher academic performance and a lower incidence of late-term course withdrawals. Conversely, irregular or episodic planning behavior often signals a lack of preparedness or decisional anxiety, which can lead to suboptimal course selection.

As established in the hierarchical framework of this report, APR is conceptually and mathematically restricted to the meso and macro levels, because it requires a longitudinal distribution of multiple timestamped actions to establish a tendency or spread of activity, making it impossible to derive from a single course interaction.

At the meso level, the regularity index captures how a student manages their registration tasks within a single academic term's enrollment window. The metric is calculated as the standard deviation (SD) of the difference in days (Δ_i) between consecutive enrollment actions (Adds and Drops), normalized by the total duration of the enrollment period (N_{enroll_days}):

$$meso_APR_S = \frac{SD(\Delta_2, \Delta_3, \dots, \Delta_n)}{N_{enroll_days}}$$

Where $\Delta_i = T_i - T_{i-1}$ denotes the time gap in days between the i -th and $(i-1)$ -th enrollment action, and n represents the total number of actions. In RSU-FET dataset, it is derived by isolating all application dates for a specific student within a semester. The day-gaps between these sequential dates are calculated, and the standard deviation formula is applied to produce a semester-based regularity score.

At the macro level, the indicator evaluates the longitudinal stability of a student's planning behavior over their entire semesters at RSU. This level distinguishes between students who may have had one frantic semester due to external life stressors and those for whom irregular planning is a persistent behavioral trait. The macro regularity indicator of a student x is quantified by calculating the longitudinal mean and variance of the student's semester-based APR scores:

$$macro_APR_x = \frac{1}{CS} \sum_{s=1}^{CS} meso_APR_{x,s}$$

3.3.3. Enrollment Procrastination

Procrastination in the context of course enrollment is defined as a student's tendency to wait until the final stages of the registration window to complete course enrollment. While procrastination has traditionally been measured using subjective, self-reported survey scales,

recent advancements by Borchers [6] pioneered the use of fine-grained transaction logs to operationalize this behavior objectively. Within the LA literature, EP serves as a behavioral proxy for decisional procrastination and task avoidance, traits frequently linked to lower academic self-efficacy and a higher probability of experiencing adverse outcomes, such as LDs. Research indicates that students with high procrastination indices are more likely to mismanage their course selection, often enrolling in higher overall workloads—particularly more difficult courses—than their earlier-registering peers.

It is essential to distinguish the EP from the closely related indicator of LE. While both metrics describe the timing of student actions, they capture different dimensions of the registration process. LE focuses on the delay from the opening of the enrollment window to a student's first successful enrollment action, serving as a measure of academic diligence and initial momentum. In contrast, EP focuses on the proximity of a student's actions (including adds, drops, and swaps) to the final administrative deadline. Under this framework, a student might enroll on time (at the start of their window) but still be identified as a procrastinator if they wait until the final hours of the add-drop period to finalize their schedule or make critical changes.

At micro level, the EP index quantifies the relative timing of a student's action for enrolling a single course. The micro EP for a course C , $micro_EP_C$ is normalized to a scale of 0 to 1 using the following formula:

$$micro_EP_C = \frac{T_{action} - T_{enroll_start}}{T_{enroll_end} - T_{enroll_start}}$$

Where T_{action} is the timestamp of course add or drop action, T_{enroll_start} is the opening date of the registration phase, and T_{enroll_end} is the institutional add-drop deadline. The final semester-based index is derived by computing

the median of these relative values for all actions initiated by the student. In the RSU-FET dataset, this is derived by comparing the application date of every transaction against the start and end dates of enrollment window.

At meso level, the metric for a student at semester S is quantified by averaging $micro_EP$ across all courses the student enrolled in during that semester:

$$meso_EP_S = \frac{1}{C} \sum_{i=1}^C micro_EP_i$$

At macro level, the indicator evaluates longitudinal behavioral consistency across the student's entire semesters or degree path to identify chronic procrastination. This distal view distinguishes between students who may procrastinate in a single semester due to external life stressors and those for whom task avoidance is a stable, persistent trait. The macro procrastination indicator for student x is computed as the mean of their semester-based procrastination scores across all semesters observation window:

$$macro_EP_x = \frac{1}{CS} \sum_{s=1}^{CS} meso_EP_{x,s}$$

3.4. Statistical Modeling

In this phase of the methodology, a statistical modeling is performed to examine the correlation between each one of the three timing-based enrollment behavior indicators, LE, APR, and EP, and student attrition, represented by the LD [6, 12]. In this statistical model, a behavior indicator is the independent variable, which is a continuous variable and the LD is the dependent variable, which is a binary variable indicating whether a student withdrew from a course after the permitted withdrawal deadline [6]. Therefore, logistic regression (LR) is employed for statistical modeling [18, 19]. LR is particularly appropriate for this research because the

dependent variable, the LD, is binary in nature [4, 18]. This approach enables quantitative enrollment timing indicators to be transformed into actionable predictive insights by estimating the likelihood of attrition before critical academic withdrawal points are reached [4, 19]. Moreover, unlike advanced machine learning approaches such as Random Forest, XGBoost, or Support Vector Machines, functioning as 'black boxes' that lack the theoretical transparency required for pedagogical application, LR belongs to the category of 'white-box' models, providing interpretable coefficients and odds ratios (OR) that clarify both the direction and magnitude of the relationship between behavioral predictors and academic risk.

It is worth mentioning the statistical modeling is applied at the micro, meso, and macro levels for LE and EP, whereas it is applied at meso and macro levels only for APR. As mentioned above, the computation of the APR requires a sequence of enrollment-related actions over time [6]. At micro level, the analysis focuses on individual student-course interactions to assess whether a specific enrollment behavior of a course can serve as immediate indicators of risk associated with that course. At meso level, student behavior is aggregated at the student-semester level, to capture broader enrollment patterns and identify behaviors across an academic semester. At macro level, the analysis extends longitudinally across multiple academic semesters and student cohorts.

3.5. Actionable Insights Synthesis

The findings derived from the statistical modeling are interpreted in light of the RSU contextual specifications, allowing for the synthesis of actionable insights that can be directly integrated into institutional EWS to enhance student success and retention [4, 26]. In this manner, the point of intervention can be

shifted to the very beginning of the semester, in many forms such as proactive academic advising focused on workload management [4, 12]. Additionally, the actionable insights can be integrated for the development of course recommender systems or workload awareness dashboards to assist students in selecting course sets that align with their realistic credit-hour capacity [9, 27].

4. Results and Discussion

In this section, the results obtained from the application of LA approach to the enrollment logs of RSU are presented and the drawn findings are discussed.

RSU-FET Dataset Statistics

A quantitative overview of the population in the RSU-dataset is illustrated in Figure 5(a), where 2,065 unique students are distributed across the five programs in the FET. The largest share of students is enrolled in Information Technology, representing 28.6% (591 students) of the total population, followed by Cybersecurity with 23.1% (478 students) and Graphics and Multimedia with 23.0% (475 students). Mechatronics accounts for 16.0% (330 students), while Interior Design represents the smallest proportion at 9.2% (191 students).

The temporal distribution of unique students across the 12 academic semesters, Figure 5(b) further reveals noticeable fluctuations in student volume between university semesters, which can be attributed to the university's structure of the academic year that consists of two regular semesters and an optional summer semester. Moreover, according to the institutional admission and enrollment system, new student intake occurs primarily during the first regular semester of each academic year. That is why the highest numbers of unique students are observed during these semesters, whereas substantially lower counts appear during the summer semesters.

In addition, Figure 5(c) shows the distribution of unique students by student semester number (i.e., progression through the degree pathway), which demonstrates a clear concentration of students in the early stages of study, particularly in Student Semester No. 1

across all programs. The number of unique students then progressively declines in the higher student semesters, indicating a gradual reduction in cohort size as students advance toward the later stages of their programs.

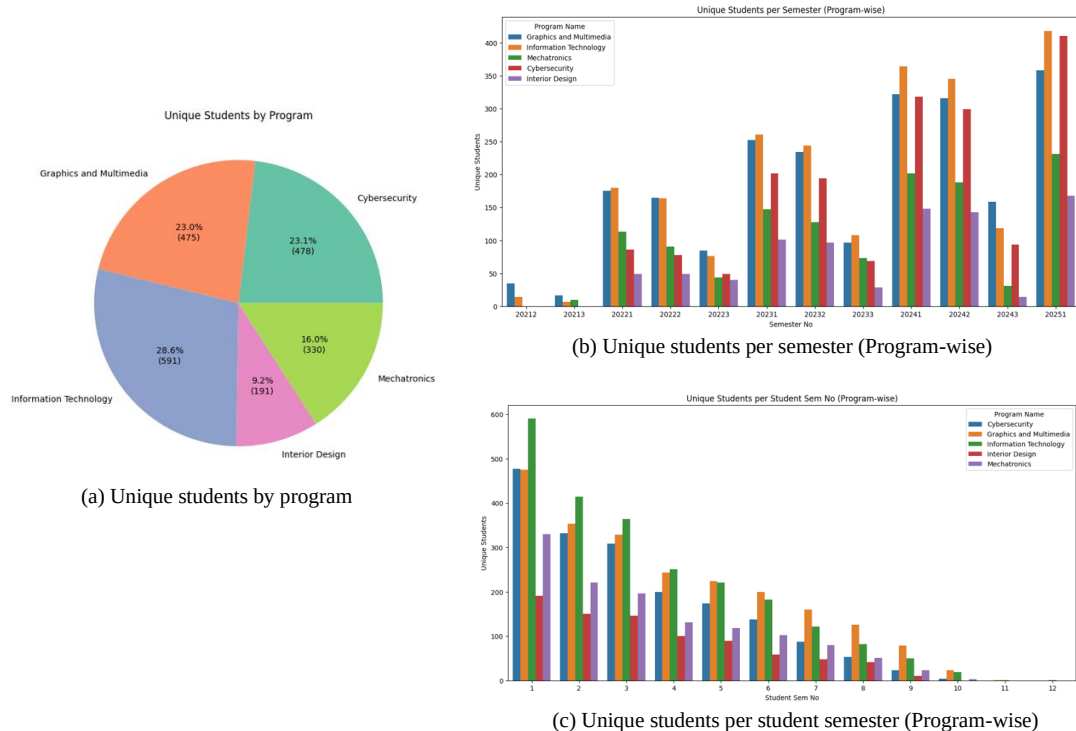


Figure 5. Statistics of RSU-FET dataset.

Similarly, Figure 6(a) presents the distribution of LD cases across the five academic programs. Graphics and Multimedia records the highest proportion of LDs, accounting for 32.3% (367 cases), followed by Information Technology with 28.5% (324 cases). Mechatronics contributes 17.8% (202 cases), whereas Cybersecurity and Interior Design account for 11.4% (130 cases) and 9.9% (114 cases), respectively. Further insight is obtained by examining the temporal distribution of LD cases across university semesters and student semesters, as illustrated in Figures 6(b) and 6(c). Across university semesters, the frequency of LDs exhibits noticeable fluctuations that closely follow the

university's academic calendar. Higher numbers of LDs are generally observed during the first regular semester of each academic year, coinciding with periods of maximum student enrollment following the admission of new cohorts, while substantially fewer cases occur during the summer semester. This pattern indicates that institutional enrollment cycles have a direct influence on the overall incidence of late course withdrawals. From the perspective of student progression, Figure 6(c) shows that LDs are concentrated predominantly during the early stages of study, particularly within Student Semester Nos. 1–4, after which their frequency gradually decreases as they progress through their academic

programs. This trend suggests that students in the initial stages of their university education are more likely to encounter challenges associated with academic adjustment, course selection, workload management, and familiarity with program requirements, increasing their likelihood of withdrawing from courses after the official drop deadline. Although the overall distribution of LD cases broadly follows the distribution of student enrollment across the five programs, several program-specific differences are evident. In

particular, Graphics and Multimedia exhibits a disproportionately higher proportion of LD cases relative to its student population, whereas Cybersecurity demonstrates a comparatively lower proportion. These observations suggest that factors specific to individual academic programs, including curriculum structure, course difficulty, workload intensity, and students' adaptation to program requirements, may contribute to differences in late course withdrawal behavior.

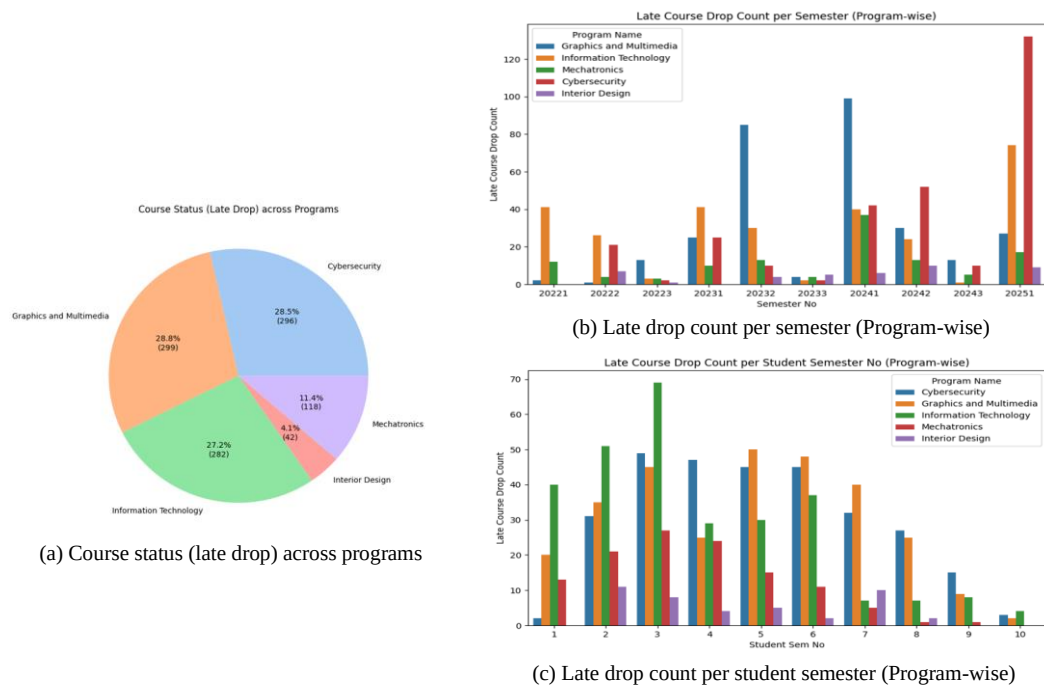


Figure 6. Statistics of RSU-FET dataset.

4.1. Results of Statistical Modeling

This section reports the findings of LR applications to test the correlation between each timing-based enrollment behavior and the LD at micro, meso, and macro levels.

4.1.1. Late Enrollment

LR is used for modeling the association between the LE behavioral indicator and the LD at micro, meso and macro levels. The

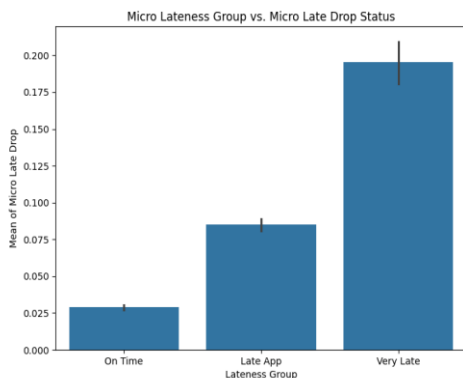
corresponding regression estimates across all levels are summarized in Table 3.

At the micro level, Relative Lateness is significantly associated with LD outcomes ($\beta = 1.4174$, $p < 0.001$, $OR = 4.13$, 95% CI [1.330, 1.505], Pseudo $R^2 = 0.04315$). As illustrated in Figure 7(a), course-level patterns show a clear monotonic increase in late-drop frequency across enrollment categories, ranging from On-Time (0.03) to

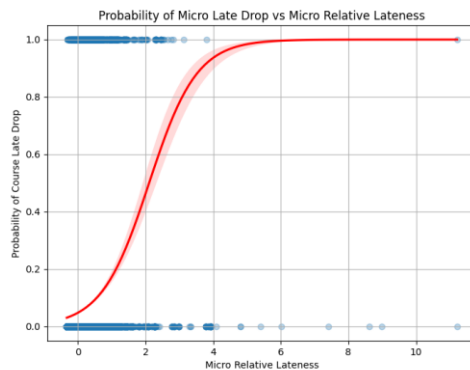
Late (0.085) and Very late (0.20). The LR curve in Figure 7(b) further confirms a nonlinear relationship, where the probability of a LD increases sharply with rising Relative Lateness, approaching near certainty at extreme values.

Table 3. LR parameters for LE and LD associations

Analysis level	No. of Obs. (N)	Coefficient (β)	Std. error	Z-score	p-value	Odds ratio ($\text{Exp}(\beta)$)	95% CI	Pseudo R ²
micro LE	56,102	1.4174	0.045	31.76 < 0.001		4.13 [1.330, 1.505]		0.04315
meso LE	7,799	1.6128	0.119	13.61 < 0.001		5.02 [1.380, 1.845]		0.02647
macro LE	2,065	3.542	0.362	9.79 < 0.001		34.54 [2.833, 4.251]		0.07482



(a) Micro late course drop frequency across enrollment categories

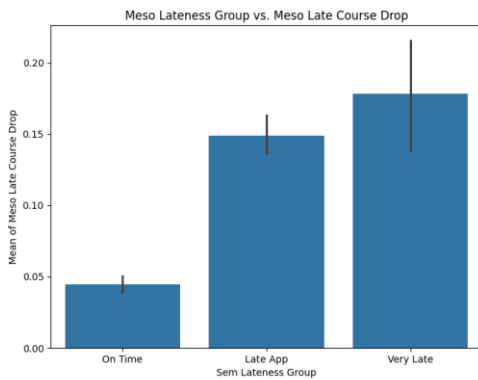


(b) Micro LE logistic regression curve

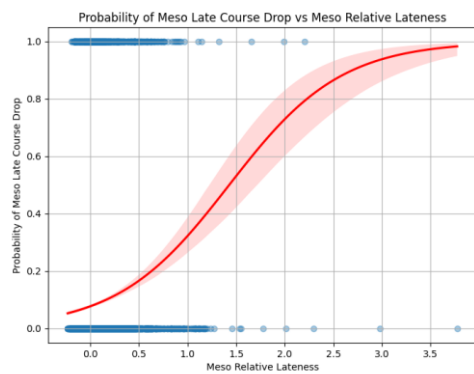
Figure 7. Results of micro LE analysis.

As shown in Figure 8, longitudinal semester trends indicate that increases in mean lateness are frequently accompanied by corresponding increases in late-drop incidence, with pronounced spikes observed in specific terms (e.g., semester 20221). Figure 8(a) shows that students classified as “Very late” at the semester level exhibit a mean late-drop probability of 0.29, compared to 0.11 for “On-Time” students. The corresponding LR curve Figure 8(b) confirms a consistent upward shift in late-drop probability as semester-level lateness increases, reinforcing the predictive value of aggregated enrollment behavior.

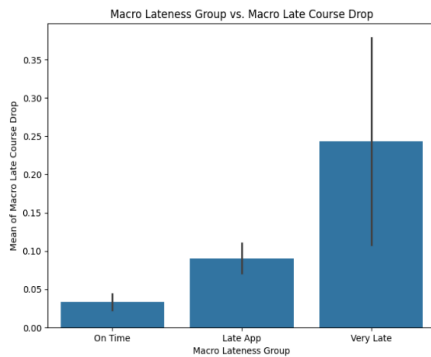
At macro level, LE demonstrates a strong and highly significant association with LD outcomes ($\beta = 3.542$, $p < 0.001$, OR = 34.54, 95% CI [2.833, 4.251], Pseudo R² = 0.07482). Figure 9 (a) further highlights this stratification, where programs characterized by chronic “Very late” registration exhibit an average late-drop probability of 0.38, compared to 0.06 for “On-Time” groups. The program-level LR curve in Figure 9(b) confirms a strong nonlinear increase in institutional risk as aggregate lateness increases.



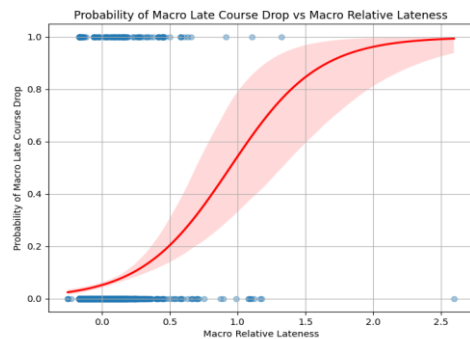
(a) Meso late-drop frequency across enrollment categories



(b) Meso LE logistic regression curve

Figure 8. Results of meso LE analysis.

(a) Macro late-drop frequency across enrollment categories



(b) Macro LE logistic regression curve

Figure 9. Results of macro LE analysis.

Across the three analytical levels, LE demonstrates a statistically significant association with LD outcomes ($p < 0.001$). As shown in Table 3, the magnitude of the relationship increases progressively from the micro level ($OR = 4.13$) to the meso level ($OR = 5.02$), reaching its highest value at the macro level ($OR = 34.54$). This trend suggests that the influence of enrollment timing becomes more pronounced as students' behavioral patterns accumulate over successive semesters. Although the pseudo R^2 values are relatively modest, they are comparable to those reported in behavioral prediction studies within

Educational Data Mining, with the highest explanatory power observed at the macro level (Pseudo $R^2 = 0.0748$). Collectively, these findings indicate that enrollment timing represents an informative behavioral indicator for predicting academic attrition and highlight its potential value as a component of institutional Early Warning Systems.

The findings obtained from the RSU FET dataset are consistent with previous studies that identify enrollment timing as an important behavioral predictor of academic performance and student attrition. In particular, the

progressive increase in LD frequency from the "On-Time" to the "Very late" enrollment categories is consistent with the findings reported by Bohler [12], who observed that students registering after the commencement of classes were more likely to experience academic difficulties and lower course performance. Similar observations have been reported in [6, 9], where fine-grained transactional enrollment records were used to demonstrate that delayed registration is positively associated with late course withdrawals, possibly reflecting ineffective workload planning and greater uncertainty during course selection. Furthermore, the substantially larger effect observed at the macro level ($\beta = 3.542$, $OR = 34.54$) supports the view that enrollment timing may serve as a

behavioral proxy for latent characteristics, including diligence, motivation, and self-regulation, as suggested by Lenchner [7] and Siefken [23]. Taken together, these findings provide additional evidence that behavioral digital trails derived from transactional enrollment data can contribute to the early identification of students at risk of academic attrition, thereby enabling institutions to implement timely academic support before significant academic loss occurs [4, 6, 12].

4.1.2. Academic Planning Regularity

LR is used for modeling the association between APR behavioral indicator and the LD at meso and macro levels. The corresponding regression estimates across all levels are summarized in Table 4.

Table 4. LR parameters for APR and LD associations

Analysis level	No. of Obs. (N)	Coefficient (β)	Std. error	Z-score	p-value	Odds ratio ($Exp(\beta)$)	95% CI	Pseudo R ²
Meso APR	6,391	-0.9071	0.093	-9.772	< 0.001	0.404	[-1.089, -0.725]	0.0056
Macro APR	1,991	-1.6947	0.227	-7.474	< 0.001	0.184	[-2.139, -1.250]	0.0890

At the meso level, APR demonstrates a statistically significant negative association with LD ($\beta = -0.9071$, $SE = 0.093$, $z = -9.772$, $p < 0.001$, $OR = 0.404$, 95% CI [-1.089, -0.725]). The LR results indicate that students exhibiting more regular and evenly spaced enrollment actions have lower odds of being classified into LD (Figure 10). This finding suggests that consistent academic planning and systematic enrollment behavior may function as a protective behavioral indicator against subsequent late course withdrawal and associated credit loss.

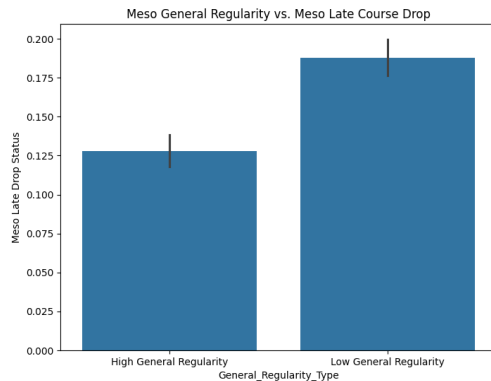
At the macro level, the negative association between APR and LD remains statistically significant and is stronger in

magnitude ($\beta = -1.6947$, $SE = 0.227$, $z = -7.474$, $p < 0.001$, $OR = 0.184$, 95% CI [-2.139, -1.250]). The corresponding ordinal regression results indicate that students with higher aggregate planning regularity have lower odds of experiencing LD status.

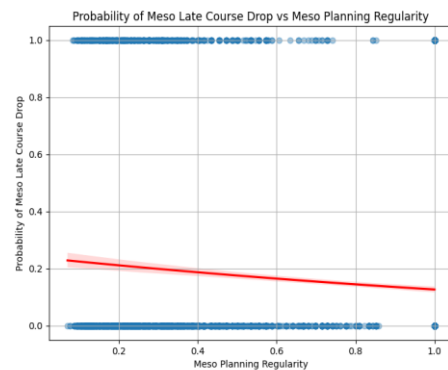
Across both analytical levels, APR demonstrates a statistically significant negative association with LD behavior. As shown in Table 4, higher APR values—representing greater regularity in the temporal spacing of enrollment actions—are associated with lower odds of occupying a more LD status category at both the semester (meso) and program/student longitudinal (macro) levels. The stronger coefficient at the macro level

suggests that accumulated planning regularity may provide a more pronounced representation

of students' longer-term enrollment behavior than semester-specific regularity.



(a) Meso APR frequency across enrollment categories



(b) Meso APR logistic regression curve

Figure 10. Results of meso APR analysis.

The APR findings are consistent with previous research in enrollment-based LA, which identifies planning regularity as a protective behavioral indicator associated with reduced academic risk [6]. The negative relationship observed at both analytical levels suggests that students who maintain consistent enrollment planning behaviors may be less likely to experience the schedule instability and repeated enrollment adjustments that can ultimately contribute to late course withdrawals and credit loss [6]. This observation is also consistent with previous studies that regard timing-based digital traces as behavioral proxies for latent characteristics such as conscientiousness, self-regulation, and effective planning skills, which have been associated with student persistence and academic success [4, 6, 23]. Furthermore, the presence of a significant negative association at both the meso and macro levels suggests that longitudinal enrollment behavior provides meaningful information about students' academic risk beyond individual semester-level behavior. These findings support the use of longitudinal administrative enrollment records as a potential source of predictive information for the early identification of

students who may be at risk of late course withdrawal and subsequent academic attrition [4, 7].

4.1.3. Enrollment Procrastination

As for EP, LR modeling its association with the LD at micro, meso and macro levels and the corresponding regression estimates across all levels are summarized in Table 5.

At micro level, EP is a significant predictor of LD outcomes ($\beta = 0.6863$, $p < 0.001$, OR = 1.986, 95% CI [0.650, 0.723], Pseudo $R^2 = 0.0602$). As shown in Figure 11, disciplinary variation is evident during semester 20241, where technical programs such as Mechatronics and Graphics and Multimedia exhibit both higher mean procrastination and higher mean late-drop frequencies across several courses (e.g., Course IDs 101019 and 101600). The LR curve in Figure 12 demonstrates that as the procrastination index increases, the probability of a LD rises sharply toward 1.0. These findings suggest that delays in finalizing course enrollment actions function as an immediate behavioral indicator of course-level academic risk.

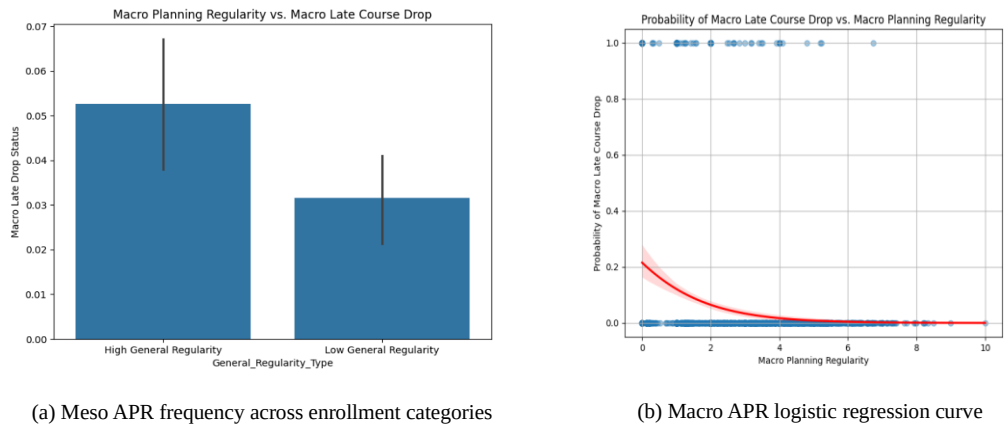


Figure 11. Results of macro APR analysis.

Table 5. LR parameters for modeling EP and LD associations

Analysis Level	No. of Obs. (N)	Coefficient (β)	Std. Error	Z-score	p-value	Odds Ratio ($Exp(\beta)$)	95% CI	Pseudo R ²
Micro (Course)	56,104	0.6863	0.019	37.06	< 0.001	1.986	[0.650, 0.723]	0.0602
Meso (Semester)	7,799	0.6216	0.089	6.95	< 0.001	1.862	[0.446, 0.797]	0.0267
Macro (Program)	2,065	0.8549	0.190	4.49	< 0.001	2.351	[0.482, 1.228]	0.0337

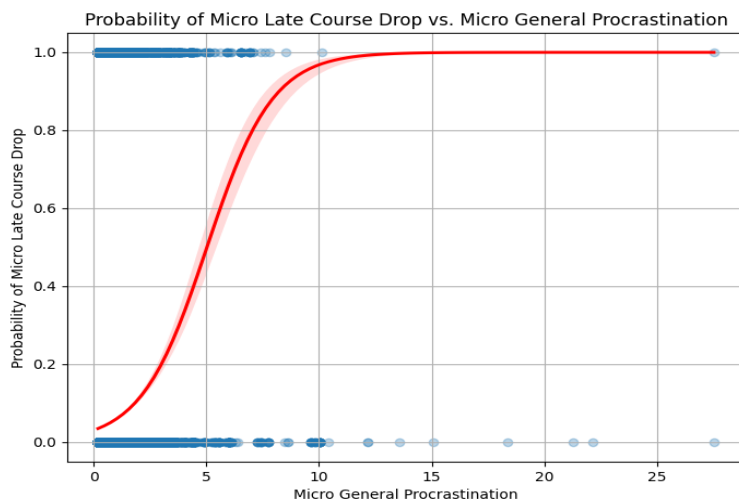
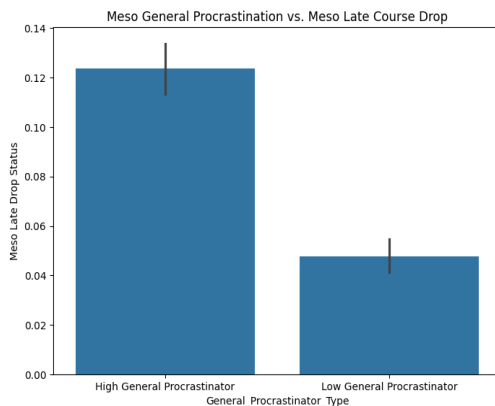


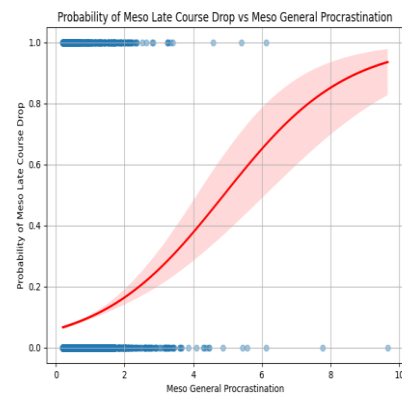
Figure 12. Logistic regression of micro EP.

At the meso level, meso EP remains a statistically significant predictor of LD behavior ($\beta = 0.6216$, $p < 0.001$, $OR = 1.862$, 95% confidence interval (CI) [0.446, 0.797], Pseudo $R^2 = 0.0267$). Figure 13 identifies notable longitudinal fluctuations in procrastination behavior across semesters, including a substantial institutional spike during semester 20221, particularly within Information Technology and Mechatronics programs. As illustrated in Figure 13, students

classified as “high general procrastinators” exhibit a late-drop probability of approximately 0.029, compared to 0.009 among “low general procrastinators.” The logistic probability curve in Figure 13 further confirms that increasing aggregate procrastination across a student’s semester schedule is associated with a higher probability of experiencing at least one LD during the term.



(a) Meso EP frequency across enrollment categories



(b) Meso EP logistic regression curve

Figure 13. Results of meso EP analysis.

At macro level, EP demonstrates a strong and statistically significant association with LD outcomes ($\beta = 0.8549$, $p < 0.001$, $OR = 2.351$, 95% CI [0.482, 1.228], Pseudo $R^2 = 0.0337$). Figure 14(a) shows that student groups characterized by persistently high procrastination maintain consistently higher probabilities of late-drop outcomes compared to lower procrastination groups. The LR curve in Figure 14(b) confirms a strong positive relationship between aggregate procrastination behavior and institutional-level attrition risk.

Across the three analytical levels, EP demonstrates a statistically significant positive association with LD outcomes ($p < 0.001$). As summarized in Table 5, the observed effect is consistent across the micro ($OR = 1.986$), meso ($OR = 1.862$), and macro ($OR = 2.351$)

levels, indicating that delayed enrollment decision-making is associated with an increased likelihood of late course withdrawal at both the individual and institutional levels. Although the pseudo R^2 values are relatively modest, they are comparable to those reported in previous behavioral prediction studies within Educational Data Mining. Among the three analytical levels, the micro-level model exhibits the highest explanatory power (Pseudo $R^2 = 0.0602$), suggesting that procrastination behaviors observed during individual course enrollment provide stronger predictive information than aggregated behavioral patterns. Overall, these findings indicate that EP represents a meaningful behavioral indicator that can contribute to the early identification of students who may be at risk of late course withdrawal.

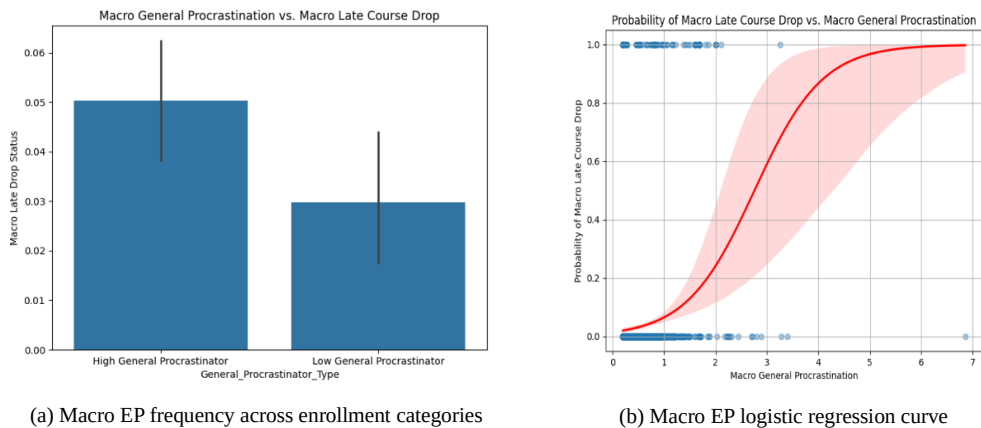


Figure 14. Results of macro EP analysis.

The findings obtained from the RSU FET dataset are consistent with previous studies that identify EP as an important behavioral predictor of academic attrition. In particular, the positive association observed across all analytical levels agrees with the findings reported by Borchers [6], who demonstrated, using fine-grained transactional enrollment records, that students who postpone course add and drop decisions are more likely to encounter schedule instability and course selection difficulties that subsequently lead to late course withdrawals. Previous research has also suggested that repeated delays in enrollment decision-making may contribute to a cycle of ineffective academic planning and reduced self-efficacy, thereby increasing the likelihood of unfavorable academic outcomes [24]. Furthermore, the consistency of the EP indicator across the three analytical levels supports the interpretation of timing-based digital trails as behavioral proxies for latent characteristics, including diligence and self-regulation, which have previously been associated with academic achievement and successful course completion by Siefken [23] and You [25]. Collectively, these findings provide further evidence that transactional enrollment behavior contains valuable

predictive information that can support institutional Early Warning Systems by enabling earlier identification of students who may require academic intervention.

4.2. Synthesis for Early Warning Application

The significant relationships identified across the micro, meso, and macro analytical levels demonstrate the potential of administrative transactional enrollment records as a valuable source of information for RSU EWS. By utilizing timing-based behavioral indicators, particularly LE, institutions can identify students who exhibit delayed enrollment behaviors during the initial weeks of the semester and implement timely monitoring strategies. Such a process-oriented approach enables academic advisors to detect early patterns of enrollment behavior and provide appropriate interventions, including academic advising, workload planning support, and personalized reminders, before students are more likely to experience late course withdrawal.

The practical contribution of this research lies in transforming routinely collected administrative records into actionable information that can assist universities in

addressing student attrition, credit loss, and the associated financial consequences [4, 12]. Evaluating these timing-based behavioral indicators within the RSU context also addresses the context-dependent nature of LA by examining whether predictive methodologies developed in other educational settings remain applicable under the university's administrative policies and student enrollment practices [11, 13]. Furthermore, as these behavioral indicators are generated during the registration period, they enable institutions to identify students who may require additional academic support before conventional performance indicators, such as assessment results or course grades, become available [12, 23].

Integrating these behavioral indicators into an automated Early Warning System provides a practical framework for supporting evidence-based academic decision-making [6, 9]. Early identification of students who exhibit high-risk enrollment behaviors enables universities to implement timely interventions that may reduce the likelihood of late course withdrawal while improving student persistence and academic success. Consequently, the findings of this study provide empirical support for incorporating transactional enrollment behavior into institutional Early Warning Systems as an additional source of predictive information available at the beginning of the academic semester [6, 26].

While this study provides robust evidence for the predictive power of timing-based enrollment behaviors, several limitations must be acknowledged. First, the data utilized in this research are derived from a single institution (Ar-Rasheed Smart University), which may limit the direct applicability of the results to institutions with significantly different administrative structures or cultural contexts. Second, the analysis focused exclusively on five undergraduate programs within the FET. Consequently, the behavioral signatures

identified here may reflect the unique workload demands of engineering, computing, and applied design disciplines, and might not be fully generalizable to the humanities or pure social sciences. Third, this study considered a limited set of behavioral predictors (LE, APR, and EP). Future research should incorporate additional variables, such as student interactions within LMS or socio-economic indicators, to provide a more holistic view of student risk. Fourth, at the micro and meso levels, the models treat repeated observations from the same student as independent behavior, which does not account for the individual-level variance typically captured by mixed-effects modeling. Finally, the current model does not account for differences in course difficulty or curriculum structure across programs. Future iterations of this Early Warning System would benefit from weighting behaviors against program-specific academic rigors to further enhance classification accuracy.

5. Conclusion

This study employed a LA approach to evaluate the effectiveness of timing-based enrollment behaviors as early predictors of academic attrition, with particular emphasis on LDs at RSU. Using fine-grained transactional enrollment records extracted from the university's administrative systems, three behavioral indicators - LE, APR, and EP - were examined across the micro, meso, and macro analytical levels. The results demonstrate that both LE and EP are positively associated with LD behavior, whereas APR exhibits a significant negative association, indicating that students with more consistent enrollment planning practices are less likely to withdraw from courses after the official drop period. The findings also suggest that timing-based enrollment behaviors capture aspects of students' self-regulatory practices that are not directly observable through conventional academic records. In particular, the stronger

predictive relationships observed at the macro level indicate that long-term enrollment behavior provides a more stable representation of students' academic planning than semester-specific enrollment activities. These results further demonstrate the value of longitudinal transactional enrollment records as a complementary source of behavioral information for identifying students who may be at risk of academic attrition. From a practical perspective, this research demonstrates that routinely collected administrative enrollment records can support the development of institutional Early Warning Systems by enabling the identification of at-risk students during the registration period, before traditional academic performance indicators become available. Such an approach allows universities to implement timely academic support and advising interventions at the beginning of the semester, thereby contributing to improved student retention and greater effectiveness in using institutional support resources.

Acknowledgment

The authors gratefully acknowledge Ar-Rasheed Smart University for providing the data used in this study.

Data Availability

The data are available from the corresponding author upon request.

Funding

This research received no external funding.

AI-tools Declaration

AI-assisted tools were used to improve the grammar, spelling, clarity, and overall readability of this manuscript.

References

- [1] Papamitsiou Z, Economides AA. Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Educational Technology & Society*. 2014;17(4):49-64.
- [2] Gray G, McGuinness C, Owende P, Carthy A. A review of psychometric data analysis and applications in modelling of academic achievement in tertiary education. *Journal of Learning Analytics*. 2014;1(1):75-106.
- [3] Albreiki B, Zaki N, Alashwal H. A systematic literature review of student' performance prediction using machine learning techniques. *Education Sciences*. 2021;11(9):552.
<https://doi.org/10.3390/educsci11090552>
- [4] Alyahyan E, Düşteğör D. Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*. 2020;17(1):3.
<https://doi.org/10.1186/s41239-020-0177-7>
- [5] López-Zambrano J, Torralbo JAL, Romero C. Early prediction of student learning performance through data mining: a systematic review. *Psicothema*. 2021;33(3):456-65.
<https://doi.org/10.7334/psicothema2021.62>
- [6] Borchers C, Xu Y, Pardos ZA. Are you an early dropper or late shopper? Mining enrollment transaction data to study procrastination in higher education. In: *Proceedings of the 17th International Conference on Educational Data Mining (EDM 2024)*; 2024 Jul 14-17; Atlanta, GA. 2024. p. 426-33.
<https://doi.org/10.35542/osf.io/zud7q>
- [7] Lenchner E. Mining transactional student-level data to predict community college student outcomes [PhD Thesis]. New York, NY: Columbia University; 2017.

- [8] Xu Y, Pardos ZA. Mining detailed course transaction records for semantic information. In: Feng M, Käser T, Talukdar P, editors. Proceedings of the 16th International Conference on Educational Data Mining (EDM 2023); 2023 Jul; Bengaluru, India. International Educational Data Mining Society; 2023. p. 388-395. <https://doi.org/10.5281/zenodo.8115794>
- [9] Borchers C, Xu Y, Pardos ZA. Workload overload? late enrollment leads to course dropout. Journal of Educational Data Mining. 2025;17(1):126-156. <https://doi.org/10.5281/zenodo.14907388>
- [10] Hagedorn LS, Maxwell WE, Cypers S, Moon HS, Lester J. Course shopping in urban community colleges: an analysis of student drop and add activities. The Journal of Higher Education. 2007;78(4):464-85. <https://doi.org/10.1080/00221546.2007.11772324>
- [11] Hellas A, Ihtantola P, Petersen A, Ajanovski VV, Gutica M, Hynninen T, et al. Predicting academic performance: a systematic literature review. In: Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2018 Companion); 2018 Jul 2-4; Larnaca, Cyprus. New York, NY: ACM; 2018. p. 175–199. <https://doi.org/10.1145/3293881.3295783>
- [12] Bohler J, Sherman S, Larson B, Mills H. Using registration timing as an early indicator of at-risk students in an online stem course. In: Proceedings of the Informing Science and Information Technology Education Conference (InSITE 2023); 2023 Jul 5-6. Santa Rosa (CA): Informing Science Institute; 2023. Article 12. <https://doi.org/10.28945/5151>
- [13] Namoun A, Alshantiti A. Predicting student performance using data mining and learning analytics techniques: a systematic literature review. Applied Sciences. 2021;11(1):237. <https://doi.org/10.3390/app11010237>
- [14] Admass WS. Review on predicting student academic performance using data mining classification algorithm. Journal of Computer Engineering & Information Technology. 2021;10(11).
- [15] Summers MD. Enrollment and registration behaviors as predictors of academic outcomes for full-time students in a rural community college [PhD Thesis]. Urbana, IL: University of Illinois at Urbana-Champaign; 2000.
- [16] Deeva G, Snoeck M, De Weerd J. Discovering the impact of students' modeling behavior on their final performance. In: Buchmann R, Karagiannis D, Kirikova M, editors. The Practice of Enterprise Modeling. PoEM 2018. Lecture Notes in Business Information Processing. Cham: Springer; 2018. p. 335-50. https://doi.org/10.1007/978-3-030-02302-7_21
- [17] Shahiri AM, Husain W, Rashid NA. A review on predicting student's performance using data mining techniques. Procedia Computer Science. 2015;72:414–22. <https://doi.org/10.1016/j.procs.2015.12.157>
- [18] Ameen AO, Alarape MA, Adewole KS. Students' academic performance and dropout predictions: a review. Malaysian Journal of Computing. 2019;4(2):278–303. <https://doi.org/10.24191/mjoc.v4i2.6701>

- [19] Macfadyen LP, Dawson S. Mining LMS data to develop an “early warning system” for educators: a proof of concept. *Computers & Education*. 2010;54(2):588–99.
<https://doi.org/10.1016/j.compedu.2009.09.008>
- [20] Baker RSJD, Yacef K. The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining*. 2009;1(1):3–17.
<https://doi.org/10.5281/zenodo.3554657>
- [21] Wang X, Yu X, Guo L, Liu F, Xu L. Student performance prediction with short-term sequential campus behaviors. *Information*. 2020;11(4):201.
<https://doi.org/10.3390/info11040201>
- [22] Sun H, He S, Esho O. Students performance prediction by mining behavioural pattern. *Journal of Physics: Conference Series*. 2021;1881:042051.
<https://doi.org/10.1088/1742-6596/1881/4/042051>
- [23] Siefken J. Registration delay and student performance. *Journal of the Scholarship of Teaching and Learning*. 2017;17(2):45–52.
<https://doi.org/10.14434/josotl.v17i2.20754>
- [24] Wäschle K, Allgaier A, Lachner A, Fink S, Nückles M. Procrastination and self-efficacy: tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction*. 2014;29:103–14.
<https://doi.org/10.1016/j.learninstruc.2013.09.005>
- [25] You JW. Examining the effect of academic procrastination on achievement using LMS data in e-learning. *Educational Technology & Society*. 2015;18(3):64-74.
- [26] Cui Y, Chen F, Shiri A, Fan Y. Predictive analytic models of student success in higher education: a review of methodology. *Information and Learning Sciences*. 2019;120(3-4):208–27.
<https://doi.org/10.1108/ILS-10-2018-0104>
- [27] Tan FZ, Chan WH. Interpreting student performance through predictive learning analytics. *International Journal of Innovative Computing*. 2024;14(2):1–13.
<https://doi.org/10.11113/ijic.v14n2.434>