TECHNOLOGY CONSULTING
IN THE GLOBAL COMMUNITY

Final Consulting Report

Peru SENATI

Kebing Bi

August 2024

Carnegie Mellon University

cmu.edu/tcingc

Table Of Content

1. Exclusive Summary

Overview of the Project

This project evaluates the performance of the uRetention system used by SENATI to predict student dropout risk. By analyzing key data points from student records and risk predictions over several semesters, this report aims to assess the effectiveness of uRetention in supporting SENATI's retention strategy and improving student outcomes.

Key Findings

- The dropout rate decreased significantly, from a pre-pandemic average of 6.5% to 3.6% in the 2024-2025 semester, reflecting a potential reduction of over 10% over the past three years.
- uRetention's overall accuracy is mediocre, particularly with low precision and sensitivity in predicting high-risk students, leading to a substantial number of false negatives.
- Inaccuracies in student records and the collection of data impacted the performance of the algorithm, with numerous cases of incorrect or missing student information affecting dropout prediction accuracy.
- Tutor involvement played a significant role in reducing dropout rates, with a substantial portion of high-risk students being retained due to active tutor intervention.

Recommendations Summary

- Refine the dropout prediction algorithm by incorporating more comprehensive data inputs, such as socioeconomic factors and academic performance.
- Improve the categorization of dropout reasons by adopting a more structured approach (MECE framework) to provide clearer insights.
- Collaborate with external experts and conduct a more thorough literature review to explore alternative algorithms better suited to SENATI's context.
- Focus on first-year students and early interventions to further decrease dropout rates.
- Investigate and address inaccuracies in data collection systems, including issues in the Blackboard and Banner platforms, to enhance uRetention's performance.

2. Introduction

Overview of SENATI

SENATI (Servicio Nacional de Adiestramiento en Trabajo Industrial) is a leading vocational training institution in Peru, dedicated to providing technical education in various industrial sectors. With campuses across the country, SENATI plays a critical role in equipping students with practical skills for the labor market. The institution faces challenges related to student retention,

which directly impacts its mission to train a skilled workforce for Peru's industrial sector. Addressing student dropout is essential for both the academic success of its students and the operational sustainability of SENATI.

Background and Motivation

In recent years, the dropout rate at SENATI has fluctuated, especially during the pandemic, which brought unique challenges to educational institutions worldwide. SENATI introduced the uRetention system as a predictive tool to identify students at risk of dropping out, allowing for timely intervention and support. However, concerns have arisen regarding the system's accuracy and its ability to effectively predict student behavior. Given that dropout rates can be influenced by a variety of personal, financial, and academic factors, it is essential to evaluate the effectiveness of the uRetention system and recommend improvements to optimize its performance. This evaluation aims to support SENATI's broader retention strategy and ultimately improve the student experience.

Objectives of the Project

The primary objective of this project is to assess the performance of the uRetention system in predicting student dropout risk at SENATI. Specific goals include:

- Evaluating the accuracy, precision, recall, and overall effectiveness of the uRetention algorithm.
- Identifying gaps in the system's current data collection and processing that might undermine its predictive power.
- Recommending improvements to both the algorithm and the data collection processes to enhance the system's reliability.
- Exploring the role of tutors in mitigating dropout rates and determining how they can be better integrated into the predictive system.

Overview of the Current System

The uRetention system is integrated with SENATI's Student Information System (SIS) and daily academic record platform (Blackboard). It uses a combination of student demographics, academic performance, and attendance data to predict dropout risk. uRetention classifies students into three risk categories: low, medium, and high. Tutors then review the predictions and meet with students classified as high-risk to provide additional support.

Despite being in use since 2019, uRetention has faced criticism due to concerns about its accuracy. Some tutors have reported that the system fails to correctly classify certain high-risk students, while others who were not predicted to be at risk ended up dropping out. These issues have prompted SENATI to reconsider its approach to retention and explore ways to improve the system's effectiveness.

Scope and Methodology

This report focuses on evaluating the performance of uRetention during the 2024-2025 academic year. The methodology includes:

- Analysis of historical dropout data and risk predictions from uRetention.
- Construction of a confusion matrix to assess the accuracy of the predictions, particularly focusing on false negatives (students who dropped out despite not being flagged as high risk).
- Surveys and interviews with tutors and students to gather qualitative insights into the system's performance and the role of human interventions.
- A literature review to benchmark SENATI's dropout prediction system against other similar systems used in educational institutions globally, identifying key factors that should be incorporated into the model.
- Recommendations for refining the system and enhancing the integration of tutor feedback and support mechanisms into the overall retention strategy.

3. Dropout Reasons and Factors

Overview

Background and Motivation

The current uRetention system's Dropout Reasons module lacks "Mutually Exclusive, Collectively Exhaustive (MECE)" categorization, which undermines its effectiveness in providing actionable insights. According to the report from last semester, the top three dropout reasons identified were financial issues, lack of awareness, and non-attendance despite having paid tuition. This lack of clarity in categorization means that SEANTI's reports cannot deliver meaningful insights or address the underlying issues effectively. To enhance the system's utility and accuracy, there is a pressing need to establish a new, well-structured system that provides comprehensive and distinct dropout reason categories. This will ensure that the insights derived are both actionable and relevant for improving retention strategies.

Objective

To enhance the effectiveness of future reports for SENATI, it is crucial to improve the current system of dropout reason recorded by tutors. The objective is to refine the Dropout Reasons module to ensure that it categorizes data in a MECE manner, providing clear and actionable insights. This improvement will enable SENATI to receive more precise information, allowing for better understanding and addressing of dropout causes.

Methodology

The development of the new system involved consulting with key stakeholders and organizing the classification through interviews and discussions with relevant personnel. This process was influenced by feedback from the existing system and insights gathered from a tutor survey. The main contributors to this effort included Arnaldo Ramirez, the IT Coordinator; Neil Mesias, the Tutor Coordinator; Alicia Carmona, the Psychopedagogue; and Patricia Gongora, the Tutor Manager. Their collective input was instrumental in shaping a more effective classification system that addresses the shortcomings of the old system and aligns with the needs of SENATI.

Result

The system, newly designed and refined with tutor servey's feedback, is divided into voluntary and involuntary drop-outs. Voluntary drop-outs are divided into three broad categories: professional, academic, and personal. Involuntary dropout can be divided into force majeure and economic reasons.

Voluntary Withdrawal

Career Factor

Higher Education: Admission to another university
Work: Change of career, find work opportunity
Career Change: Dislike of chosen career, change of career
Dissatisfaction: Students are dissatisfied with the non-academic experience at SENATI, including career development, internship opportunities and other career-related factors

Academic Factor

Work-life balance: Balancing work, family, and study commitments can be overwhelming for some students, leading to burnout and dropout.
Confidence/ Failed in class: Failed in 1 or more courses
Lag in the curriculum: Having more than 5% unjustified absences in the semester or more than 20% justified
Mismatched Expectations: Students' expectations of a program or institution do not align with reality, leading to questions about their choice of institution or field of study, resulting in a decision to drop out.

Personal Factor

Social integration/ Adaptability: Feeling isolated or disconnected from the campus community can impact students' sense of belonging and their likelihood of persisting in their studies.
Family (R06): Parents' separation, Family member's death, change in living location
Relocation: Travel inside/outside of country, family relocation
Dissatisfaction: Students is dissatisfied with the non-academic experience at SENATI, including social life, food, support system and are dissatisfied with the non-academic experience at SENATI, including social life, food, support system, technology equipment etc

Enrolled but never attended class (R09): Once verified 5% absenteeism after the start of classes (enrolled until the first month of the semester)

Involuntary Withdrawal

Force Majeure Factor

Death (R03): Student's death
Victims(R13): Climatic disasters, Unexpected accidents, political issues, and so on
Mental Health: Mental health issues can impair students' ability to cope with academic demands and contribute to dropout
Physical Health(R02): Physical health issues can impair students' ability to cope with academic demands and contribute to dropouts, including pregnancy, car accidents, etc.
Military Service (R11): Compulsory Military Service

Financial Factors

Economic reasons (long existing): Financial constraints may force students to prioritize earning income over pursuing their education, leading to dropout. Include students who have full-time jobs, willing to continue but have time conflicts with the academic calendar
Financial constraints (short term): Unexpected expenses or lack of financial support may make it difficult for students to afford tuition or cover living expenses, forcing them to drop out. May include: robbery or loss of equipment/ laptop/ cellphone, Parents' loss job, family loss of main income stream, Unexpected loss due to natural disasters, etc

Discipline (R05)

Suspension(R12): Student is temporarily banned from attending classes or school activities
Expulsion: Student is permanently removed from their school due to breaking major rules or behaving in a way that is very disruptive or unsafe.

Recommendations & Action Points

Implement the Form

Tutor team should thoroughly review those reasons to refine it. and IT department could integrate it to the system for next semester to ensure comprehensive data collection.

Allow Multiple Selections

Given that dropout reasons often involve multiple factors, the system should be set up to allow for multiple selections in the dropout reasons section. This will provide a more accurate picture of the reasons behind student departures.

Record Unlisted Reasons

Make sure to document any reasons for dropout that are not included in the predefined options. This will help minimize unknown causes and provide SENATI with a complete understanding of the factors affecting student retention, ultimately enhancing their ability to support students more effectively.

4. Literature Review

Motivation and Objective

The literature review can support this project in several key ways:

1. **Identify Key Variables:** Determine the variables frequently used in dropout risk prediction to assess if uRetention is missing any critical inputs.
2. **Select Optimal Algorithms:** Identify the most effective algorithms for future development.
3. **Benchmark Evaluation Metrics:** Utilize evaluation metrics from existing literature as benchmarks to define the expected range of values for evaluating the effectiveness of algorithms.

Methodology

Research Rabbit and Google Scholar have been used in this literature review. Given that the majority of English-language results were focused on non-South American countries, and considering that different social and cultural contexts may impact dropout risk prediction, approximately 20 papers were selected in both English and Spanish.

The final compilation includes an Excel spreadsheet with the following details for each paper:

- Name of the paper
- Year of publication
- Link to the paper
- Type of algorithm
- National context
- Sample size
- Key conclusions

Key Findings

Precision vs. Recall in Dropout Prediction

Two critical factors in predicting dropout rates are precision and recall. High dropout precision indicates that the model accurately predicts which students are likely to drop out, which is crucial for effectively guiding counseling resources. Conversely, high dropout recall is important because

it allows the university to identify all students at risk of dropping out. Issues arise when one metric is high but the other is not. For example, a model with 100% precision might only identify 20 students who are about to drop out, missing an additional 40 students, resulting in a 50% recall rate and missed counseling opportunities. Alternatively, a model with 100% recall might identify all students who will drop out but with low precision, wasting resources on students who might not actually drop out.

Algorithm Effectiveness and Accuracy

The scale of the student population affects the choice of the best algorithm. Preliminary analysis suggests that Decision Tree, Logistic Regression, and K-Nearest Neighbor are common effective algorithms, but they achieve only around 80% accuracy. Neural Networks can offer higher accuracy but lack the ability to provide specific reasons, making them less useful for providing targeted advice to tutors.

Impact of GPA and Socioeconomic Factors

In the first year, both overall GPA and socioeconomic factors significantly impact dropout rates. However, in the second and third years, only overall GPA remains a significant factor. This 0stay in the university system. Therefore, ongoing monitoring of academic performance and engaging students in academic support activities are crucial to preventing dropouts.

Recommendation

Collaborate with Experts

It is advisable to collaborate with Javier Fernando Vega García from Universidad Ricardo Palma, who is researching similar algorithms. Partnering with him could provide valuable firsthand insights and enhance the development of effective dropout prediction models.

Conduct a Comprehensive Literature Review

Perform a detailed literature review in Spanish to explore algorithm choices more thoroughly. Consider establishing collaborations with researchers who have published relevant articles and use their developed algorithms as prototypes to test their applicability within the SENATI context.

Focus on First-Year Students

To improve dropout prediction, collect demographic and socioeconomic data during the application process, ideally before students start their first year. Additionally, ensure that tutors pay special attention to first-year students, as early intervention can be crucial.

Evaluate Algorithms Using F1 Score

When assessing future algorithms, focus on balancing precision and recall by prioritizing the F1 score, which provides a measure of both metrics' balance and effectiveness.

Exercise Caution with Neural Networks

Be cautious when using Neural Networks due to their high computational requirements and lack of transparency in providing specific reasons for predictions. Consider whether the benefits outweigh the potential need for advanced hardware and longer processing times.

5. Algorithm Evaluation and Visualization

Background and Motivation

An evaluation of the uRetention system is crucial because the IT department has raised concerns about its accuracy, though no specific evidence has been provided to substantiate these doubts. SENATI has had an agreement for uRetention since 2019, but it was only piloted in the 202320 academic year and will be fully implemented in the 202410 semester. As the current agreement is set to expire at the end of the next semester, a decision must be made about whether to renew it. Therefore, conducting thorough research is essential to inform this business decision, ensuring that any continuation or adjustment to the system is based on solid evidence and aligns with SENATI's needs and expectations.

Objective of the Section

The objective of this evaluation is to assist SENATI in making an informed decision regarding the future use of the uRetention system. By quantifying uRetention’s performance and impact, the evaluation will provide critical insights into whether to continue using uRetention or to develop and implement a new algorithm tailored to SENATI's needs. This assessment aims to ensure that the choice made is based on a thorough analysis of the system’s effectiveness and its contribution to addressing dropout risks.

Impact of uRetention and Tutor Systems

A comparison of ten years of data, including pre-pandemic figures, reveals that the dropout rate has decreased from 6.5% to 3.6% over the past three years, representing an overall reduction of 11% in the dropout rate. This significant impact demonstrates the effectiveness of the uRetention and tutor systems in improving student retention. Detailed findings and further analysis of this impact can be found in the evaluation report.

Evaluation Metrics and Results

The evaluation utilized five key metrics—accuracy, precision, specificity, sensitivity, and F1 score—to assess the performance of the uRetention system. These metrics were examined across three different scenarios, and the findings concluded that the system falls short of effectiveness. For detailed calculations and methodologies of these key metrics, please refer to the appendix of the evaluation report.

Confusion Matrix

Table 1: Confusion Matrix Example

Actual \ Predicted	Predicted Positive (P)	Predicted Negative (N)
Actual Positive (P)	True Positive (TP)	False Negative (FN)
Actual Negative (N)	False Positive (FP)	True Negative (TN)

Evaluation Matrix

Table 2: Accuracy, Precision, Specificity, Sensitivity and F1 Score for Different Schemes

	H & L	HM & L	H & ML
True Positive	123	2299	123
False Negative	651	651	2827
False Positive	80411	1332787	80411
True Negative	272153	272153	1524529
Accuracy	77.06%	17.07%	94.82%
Precision	0.15%	0.17%	0.15%
Specificity	77.19%	16.96%	94.99%
Sensitivity	15.89%	77.93%	4.17%
F1 score	0.003025533	0.00343638	0.002946672

Key Findings

Improvement in Dropout Rates

In the 2024-2025 semester, the student dropout rate decreased to 3.6% (data up to June 3), showing a significant improvement from the pre-pandemic average of 6.5%. This reflects a 3% decrease and projects a potential 10% reduction over a three-year period.

Evaluation of uRetention System Performance

The evaluation of the uRetention Student Dropout Algorithm reveals mediocre accuracy, with low precision and low sensitivity highlighting areas for improvement. The presence of false negatives underscores the need for better identification of high-risk students and quantifying the contribution of tutor support.

Discrepancies in Student Records

The evaluation uncovered that 31,935 students have prediction records without corresponding enrollment records, suggesting possible financial losses for SENATI. Additionally, 8% of dropouts occurred before uRetention could generate data, indicating potential gaps in the system.

Issues with Post-Dropout Predictions

Even after students drop out, the remaining predictions by uRetention do not consistently classify them as high-risk. This suggests missing elements in the algorithm that need to be addressed to improve predictive accuracy and better identify at-risk students.

Recommendation

Financial Analysis

To better assess the financial benefits for the remaining 3% of students, I recommend conducting a detailed financial analysis. This analysis should focus on several key components: the revenue generated from additional tuition fees, the fixed costs associated with investments in the uRetention platform, and the variable costs, which include expenditures related to providing educational services such as tutor fees. This information is crucial for understanding uRetention's impact on SEANTI.

Refine Information Collection Systems

It is crucial to redesign the methods and channels for collecting information. During the evaluation process, numerous errors were identified, such as dropout records spanning five years within a single semester and future dates being inaccurately recorded. These issues may stem from a lack of oversight and review mechanisms or from the user-friendliness of the system.

6. Data Collection Methodology

Background and Motivation

The current platform, uRetention, relies on input data from the student information system (SIS) Banner and the daily academic record platform Blackboard. However, uRetention's claimed accuracy of 95% does not align with its actual performance, revealing a significant discrepancy. One possible reason for this shortfall is that the uRetention algorithm may overlook some critical

variables. Therefore, collecting additional information is essential for improving the accuracy of future algorithms. By gathering more comprehensive data, we can better identify input variables closely related to dropout risk, leading to more effective algorithm development.

Objective of the Survey

The survey aims to enhance the information collected from students at SENATI by refining the current information forms used during the admissions process. Through the improved form, we can gather more detailed and diverse information about students without altering SENATI's existing structure and operations. This will help achieve a more comprehensive understanding of students, which is crucial for developing better dropout prediction models and supporting student retention efforts.

Methodology

The newly designed pre-admission information collection questionnaire includes the following nine aspects

Basic Information:

- **Personal Details:** First name, last name, date of birth, gender, and residential status.

Contact Information:

- **Phone and Email:** Phone number and email address.
- **Addresses:** Permanent address, current/local address.
- **Relocation:** Whether you have relocated or plan to relocate to attend SENATI.
- **Living Arrangements:** Who you plan to live with during your time at SENATI (e.g., parents, relatives, friends, etc.).

Personal:

- **Marital Status:** Options include single, in a relationship, married, widowed, divorced, or separated.
- **Military Status:** Active duty, reserve, veteran, military spouse or dependent, planning to enlist, or no affiliation.
- **Education Status:** Whether you are a first-generation college student.
- **Career at SENATI:** Your chosen career path at SENATI.

Health:

- **Health Issues:** Chronic diseases, disabilities.
- **Emergency Contact:** Name, relationship, phone number, and email.
- **Medical Treatment:** Any treatments affecting class attendance.

SENATI Experience:

- **Reasons for Choosing SENATI:** Various motivations such as reputation, job placement, specialized programs, etc.
- **Academic Motivation:** Factors driving academic success.
- **Workshops and Clubs:** Interest in workshops and preferred club participation.
- **Concerns and Challenges:** Issues like academic workload, financial constraints, social integration.
- **Alternative Paths:** Consideration of other educational or career options.

- **Confidence Level:** How confident you are in managing academic workload.

Family Information:

- **Parental Details:** Names, phone numbers, job titles of parents.
- **Parental Education:** Highest degree obtained by parents.
- **Siblings:** Number of siblings and their educational statuses.

Economic Information:

- **Income and Employment:** Annual household income, employment status.
- **Financial Aid:** Receipt of financial aid or scholarships.
- **Financial Challenges:** Any financial issues affecting studies.

Family Responsibility:

- **Support Role:** Primary income source, financial responsibility for dependents.
- **Caregiving:** Any caregiving responsibilities affecting studies.
- **Family Factors:** Other family situations impacting college experience.

Technology:

- **Internet Access:** Frequency of high-speed internet access.
- **Device Access:** Reliability of access to a cell phone, computer, or laptop.

Digital Literacy:

- **Comfort with Technology:** Comfort level with computers, use of digital platforms, familiarity with learning management systems.
- **Digital Skills:** Confidence in using digital skills or software.
- **Additional Training:** Need for further digital literacy training.

Social Media:

- **Accounts:** Facebook, Instagram, Twitter, TikTok accounts.

Next Step

Form a Cross-Departmental Team to Ensure Integration

Establish a cross-departmental team with representatives from each relevant department to ensure the effective implementation of the information collection plan. The designed questionnaire has been sent to the administrative manager of the academic department. However, to integrate the form into the system, involvement from additional departments is necessary:

- Academic Department: Needs to review and approve the questions. While currently supportive, final approval from the manager is required.
- Legal Department: Must ensure that the information collection complies with national regulations.
- Marketing Department: Should address potential queries from prospective students and parents, which may increase their workload.
- IT Department: Responsible for designing interfaces and integrating the final forms into the existing system.

Investigate Data Accuracy Issues

Many tutors believe that the accuracy of uRetention is compromised due to errors or inaccuracies in the academic data provided by Blackboard. Potential issues include students' difficulties with electronic devices, lack of personal devices, or mixed student accounts. Despite this, student interviews did not reveal similar problems, possibly due to the small sample size. Further investigation is needed to determine if these issues are prevalent and to address them effectively.

7. Tutor Survey Findings

Background and Motivation

After assessing the uRetention algorithm for the 2024-2025 semester, we found that its accuracy is notably low, reflecting poor performance overall. Despite this, the semester's dropout rate significantly decreased from a pre-pandemic average of 6.5% to 3.6%, suggesting a potential 10% reduction in the dropout rate over a three-year degree program. This seemingly paradoxical outcome is not yet fully understood. One plausible explanation could be the high-quality support provided by tutors, which may have helped retain students who otherwise might have dropped out. However, due to insufficient data on this aspect, we are conducting a survey to gather qualitative insights from tutors to better quantify their contributions.

Beyond examining this critical factor, the survey will also seek valuable firsthand feedback on the uRetention platform, understand the reasons behind student dropouts, and identify potential improvements.

Objective of the Survey

Quantify Tutor Contribution

The survey aims to measure the contribution of tutors and use this data to provide a more comprehensive evaluation of the uRetention dropout risk prediction system. By quantifying tutor input, we can better assess the effectiveness of the current system.

Gather Tutor Insights on uRetention

Collect insights and opinions from tutors regarding the uRetention platform. This information from direct users will help validate the accuracy of the algorithm and provide a clearer assessment of the uRetention dropout risk prediction system.

Identify Dropout Reasons Through Tutor Experience

Leverage tutors' experiences with potentially at-risk students to gather reasons why students might consider dropping out. This will help refine the information collected to improve the prediction of dropout risk.

Evaluate Acceptance of Potential System Changes

If a decision is made not to continue using the uRetention platform, future algorithms will need to be integrated into the SIS Banner system. It is essential to gauge tutors' acceptance of this change and gather suggestions for improvements to ensure a smooth transition and effective implementation.

Methodology

The survey was divided into four sections and comprised 22 questions. It was distributed via email to 98 tutors across 88 SENATI campuses, using the tutor manager as the point of contact. A total of 84 responses were received.

Questions in the Survey

Tutor Contribution Calculation

- What percentage of high-risk students do you talk to each week?
- What percentage of medium-risk students do you talk to each week?
- What percentage of the high-risk students you talk to are really likely to drop out?
- What percentage of the medium-risk students you talk to are really likely to drop out?
- What percentage of students who were initially about to drop out stayed after you talked to them?
- What percentage of students you talk to will remain?
- How many hours per week do you spend talking to students who might drop out?

uPlanner Algorithm Evaluation & Thought

- Do you think the uPlanner algorithm is accurate?
- Have you found any high-risk students who actually do not intend to drop out?
- Have you found any students predicted to be low-risk who ended up dropping out? If so, why do you think uPlanner failed to detect these students?

Potential Key Factors for Dropout

- What additional data do you think should be collected to better predict student dropout risk?
- During your conversations with students, have you discovered any dropout reasons that uPlanner does not cover?
- How can the dropout reason summary table be improved to make it more comprehensive and MECE?

Area of Improvement for Banner

- Some feedback suggests that using Banner is time-consuming. What features or improvements would you like to see added?
- Are there any issues with the Banner interface or navigation pages? How would you suggest they be changed, and are there any existing examples we should follow?
- What is your opinion on integrating the uPlanner dropout prediction algorithm into Banner?

Key Findings

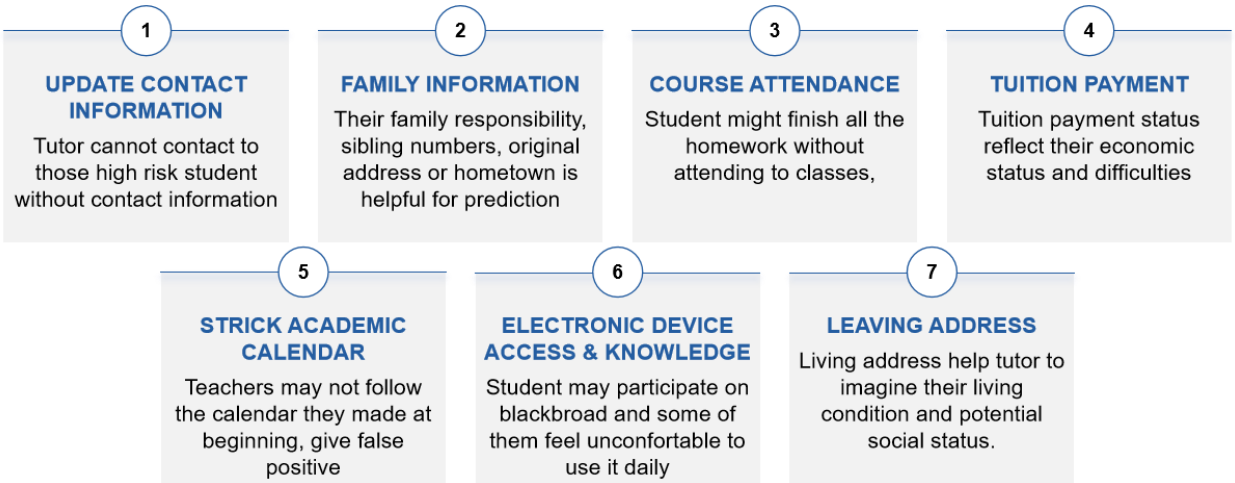
Tutor Engagement and Effectiveness

According to the survey data, most tutors engage with 20-30% or 80-90% of high-risk students, but only 10-20% of these interactions are considered true positives. For medium-risk students, tutors primarily talk to 20-30%, with less than 10% of these students actually at risk of dropping out from SENATI. Tutors believe they successfully prevent 20-30% of true positive students from dropping out. Overall, 20,000 to 26,000 students were retained who would otherwise have dropped out, thus preventing the dropout rate from rising above 20%.

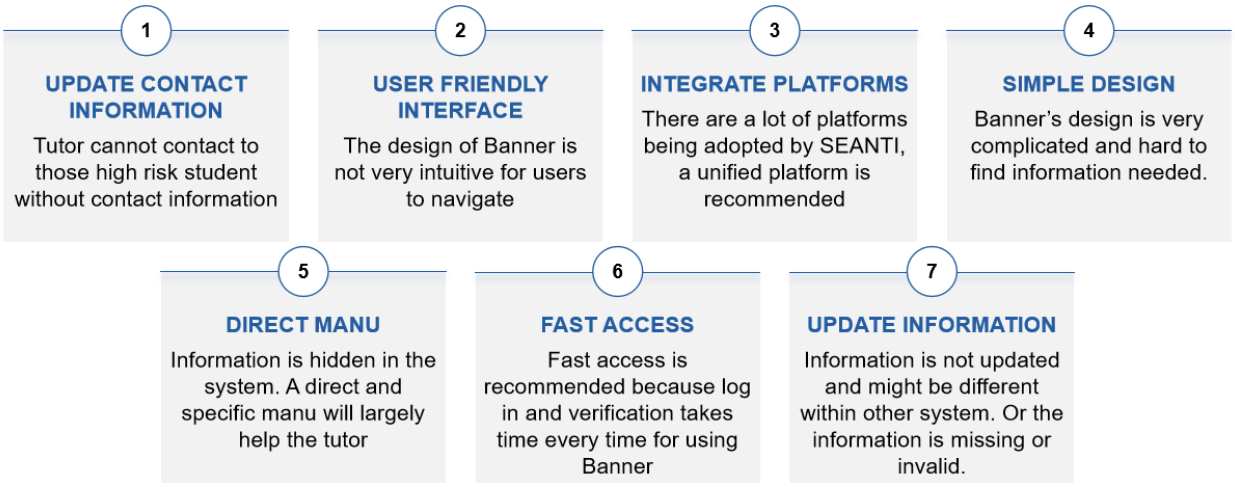
Evaluation of uPlanner Algorithm

Regarding the uPlanner algorithm, 55% of tutors feel it is not very accurate, while only 4% consider it accurate. Additionally, 84% of tutors have observed false negatives, with 10% indicating that these occur very frequently.

Top 7 Factors Contributing to Dropout



Top 7 Banner Improvement suggestion



Operational Issues Affect Algorithm Performance

Feedback from tutors reveals that several operational issues, unrelated to the algorithm's inherent effectiveness, are impacting its accuracy. For instance, inaccuracies in Blackboard data, instructors not following the academic calendar, outdated contact information preventing tutors from reaching at-risk students, and the algorithm's failure to account for students who were previously marked as resolved after discussions with tutors all contribute to reduced accuracy. Addressing these operational issues is as crucial as refining the algorithm itself to ensure comprehensive and effective performance.

8. Student Interviews

Background and Motivation

To better understand student perspectives on the dropout issue and SENATI's system logistics in general, student interviews have been conducted, organized, and supported by the Tutor Coordinator Neil Mesias (lima campus), and IT Coordinator Arnaldo Ramirez (other compasses).

Objective of the Interview

Objectives of the interviews are to understand tutor-student relationships, SENATI's support system, orientation setup, dropout-preventing methodology, technology accessibility, and general suggestions.

Methodology

There were a total of six questions, and responses were received from six students at the Lima campus and five students from other campuses, demographic information including names, contact email, major, and number of semesters has been collected..

Here are 6 questions included:

1. How would you describe your relationship with your tutors? Do you feel comfortable seeking help from them? Do you proactively approach your tutors for assistance? Do you find their support helpful for your studies and daily life?
2. If you encounter academic, psychological, or personal difficulties, do you know which support systems at school you should approach for help?
3. How do you feel about your participation in and opinions regarding the school's freshman orientation activities? Do you have any suggestions on how these activities could be improved? (A student interviewed earlier mentioned not receiving email notifications to attend orientation and felt that SENATI should provide better orientation to help students adjust to the environment.)
4. What are your thoughts on the algorithm used to predict student dropout rates? Do you think the school should rely on algorithms to predict high-risk students and attempt intervention, or should struggling students actively seek help? (Currently, the school has a system called uRetention to predict student dropout risks. Tutors review this weekly and meet with high-risk students to offer possible assistance to prevent dropout.)
5. In school, do you often need to use electronic devices? If so, what problems do you encounter when using these devices? How do you think the school could better assist students?
6. What are your overall thoughts and suggestions on how the school can improve to reduce student dropout rates?

Key Findings

Relationship with Tutors

Most students describe their relationship with tutors as good or excellent. They feel comfortable seeking help and find the tutors' support useful for their studies and daily life.
Some students indicated that they rely more on peers for help before approaching tutors, and a few expressed that not all tutors are equally helpful.

Comfort in Seeking Help

A majority of students know where to seek help for academic and personal issues. However, there are a few students who are unaware of the available support systems.

Orientation Activities

The responses indicate mixed feelings about orientation activities. While some students find them helpful, others feel that more personalized or tailored activities could improve the experience. Suggestions include better communication about these activities and offering more interactive and engaging sessions.

Opinion on Predictive Algorithms

There is a varied awareness of the algorithm used to predict student dropouts. Some students support its use, believing it is crucial in a technological era, while others are unaware of it or express concern about relying solely on technology. A common suggestion is to combine algorithmic predictions with personal interactions to better identify and support at-risk students.

Use of Electronic Devices

Most students use electronic devices frequently, mainly for educational purposes. However, they face challenges such as poor maintenance of school-provided devices, lack of access to reliable internet, and sometimes outdated software.
Recommendations include improving Wi-Fi access, ensuring devices are well-maintained, and providing workshops on the effective use of technology.

Suggestions to Reduce Dropout Rates

Students suggest offering more financial support, improving communication and support services, fostering parental involvement, and promoting mental health education.
Other ideas include implementing mentorship programs, offering more career-related support, and increasing the availability of practical training opportunities

Recommendations

Orientation Design

Tailor orientation activities to specific student groups, and include more interactive sessions that help build community and engagement.

Technology Support

Invest in better maintenance of electronic devices and improve campus Wi-Fi to support students' academic needs. Provide training sessions on using technology effectively.
Lending electronic devices is an option that SENATI could consider to better support that students from low-income families.

Tutor Management

At some SENATI campuses, a single tutor is responsible for several thousand students, making it difficult to meet the needs of each student effectively. When students request assistance and do not receive a timely response, their trust in the support system can be significantly undermined. To address these challenges, further research is necessary to explore potential solutions. Possible

approaches include hiring additional tutors to improve the student-to-tutor ratio, or redistributing some of the tutoring responsibilities to professors. These measures could help alleviate the workload and enhance the support provided to students, ultimately strengthening the overall effectiveness of the support system.

Appendices

Detailed Tables and Charts

	Algorithm Prediction				Record (Actual situation)				
Week	Total Count	Predicted Low-risk Count	Predicted Med-risk Count	Predicted High-risk Count	Total Dropout Count	Dropout Low-risk Count	Dropout Med-risk Count	Dropout High-risk Count	Dropout No Data Count
2.19	90210	22013	68197	0	254	57	147	0	49
2.26	90904	23720	67143	41	339	76	211	0	52
3.5	95934	26123	67021	2790	283	66	184	7	26
3.11	101863	29141	69356	3366	418	126	258	16	18
3.18	104310	19870	80853	3587	275	52	214	3	6
3.25	105802	17872	83638	4292	310	44	244	11	11
4.2	107152	16356	85221	5575	199	31	146	15	7
4.8	109273	19718	83343	6212	342	72	206	19	45
4.16	112518	19286	87260	5972	183	30	126	14	13
4.23	113152	14722	91995	6435	140	25	104	5	6
4.29	113818	15021	92054	6743	165	21	127	7	10
5.6	114587	14507	92118	7962	125	23	92	4	6
5.13	115905	13730	93086	9089	96	18	64	11	3
5.2	116202	10675	96264	9263	49	5	35	9	0
5.27	116260	10050	97003	9207	25	5	18	2	0

	H & L	HM & L	H & ML
TP	123	2299	123
FN	651	651	2827
FP	80411	1332787	80411
TN	272153	272153	1524529
Accuracy	77.06%	17.07%	94.82%
Precision	0.15%	0.17%	0.15%
Specificity	77.19%	16.96%	94.99%
Sensitivity	15.89%	77.93%	4.17%
F1 score	0.003025533	0.00343638	0.002946672

Time Period	Semester	Total Student	Dropout Student	Percentage
Before Pandemic	201410	56462	4393	7.78%
	201420	59525	3651	6.13%
	201510	61932	4295	6.94%
	201520	67278	4406	6.55%
	201610	71093	5170	7.27%
	201620	71610	4556	6.36%
	201710	75619	5392	7.13%
	201720	75260	4413	5.86%
	201810	81273	5722	7.04%
	201820	75376	4284	5.68%
	201910	84262	5553	6.59%
	201920	79176	4030	5.09%
After Pandemic	202010	87337	44977	51.50%
	202020	55126	7711	13.99%
	202110	82015	11147	13.59%
	202120	81827	8165	9.98%
	202210	101236	9970	9.85%
	202220	91486	6269	6.85%
	202310	115451	8410	7.28%
uRetention -Trial	202320	107185	5450	5.08%
uRetention - Full	202410	116363	4224	3.63%

Code snippets

```
import pandas as pd

# read file
df                =                pd.read_csv(r'C:/Users/KBi/OneDrive                -
senati.edu.pe/Desktop/Prediccion_uRetention_Senati_202410.csv')

# split and save
chunk_size = 1000000  # column number
for i in range(0, len(df), chunk_size):
    df_chunk = df.iloc[i:i + chunk_size]
   df_chunk.to_csv(f'C:/Users/KBi/OneDrive - senati.edu.pe/Desktop/{i//chunk_size
+ 1}.csv', index=False)
print("done")
```

Literature Review

Literature Review

	Main Categories	Sub Categories	Detail	Potential solution
Voluntary Withdrawal	Career Factor	Higher Education	Admission to another university	Long-term plan: Upgrade SENATI to university
		Work	Change of career,find work opportunity	
		Career Change	Dislike of chosen career, change of career	Change study program/major
		Dissatisfication	Students is dissatisfied with the non-academic experience at SENATI, including career development, internship opportunity and other career related factors	Change study program/major
	Academic Factor	Work-life balance	Balancing work, family, and study commitments can be overwhelming for some students, leading to burnout and dropout.	Support system from tutor, professor, peers, department
		Confidence/ Failed in class	Failed: In 1 or more courses	Retake courses, make up opportunity by summer courses/online courses
		Lag in curriculum	Having more than 5% unjustified absences in the semester or more than 20% justified	Tutoring, TA, Office hour, additioanl resource support
		Mismatched Expectations	Students' expectations of a program or institution do not align with reality, led to question their choice of institution or field of study, resulting in a decision to drop out.	
	Personal Factor	Social integration/ Adaptability	Feeling isolated or disconnected from the campus community can impact students' sense of belonging and their likelihood of persisting in their studies.	Community build within department, club, campus
		Family (R06)	Parents' separation, Family member's death, change in living location	Temporary suspension of school membership
		Relocation	Travel inside/outside of country, family relocation	
		Dissatisfication	Students is dissatisfied with the non-academic experience at SENATI, including social life, food, support system and are dissatisfied with the non-academic experience at SENATI, including social life, food, support system, technology equipment etc	
		Enrolled but never attended class (R09)	Once verified 5% absenteeism after the start of classes (enrolled until the first month of the semester)	
Involuntary Withdrawal	Force Majeure Factor	Death (R03)	Student's death	N/A
		Victims(R13)	Climatic disasters, health, political, among	N/A
		Mental Health	Mental health issues can impair students' ability to cope with academic demands and contribute to dropout	
		Physical Health(R02)	Physical health issues can impair students' ability to cope with academic demands and contribute to dropout, including pregnancy, car accidents, etc.	Provid concealing service
		Military Service (R11)	Compulsory Millitary Service	N/A
	Financial Factors	Economic reasons (long existing)	Financial constraints may force students to prioritize earning income over pursuing their education, leading to dropout.Include student who have full-time job, willing to continue but have time conflicts of academic calendar	on-campus part-time opportunities, scholarships, fellowship, flexible class schedule (night/weekend class)
		Financial constraints (short term)	Unexpected expenses or lack of financial support may make it difficult for students to afford tuition or cover living expenses, forcing them to drop out. May include: robbery or loss of equipment/ laptop/ cellphone, Parents loss job, family loss main income stream,Unexpected loss due to natural disasters, etc	Student Loan Assistance
	Discipline (R05):	Suspension(R12)		N/A
		Expulsion		N/A
Other	Uncovered Factor	Reason Unlisted		
	No Contact Info	No reason Record		

Accuracy, Precision, Specificity, Sensitivity, F1-Score

Here is the definition, and calculation methodology of each one of them:

- Accuracy: The ratio of correctly predicted observations to the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision: It measures the proportion of true positive predictions (correctly predicted dropouts) among all positive predictions (all predicted dropouts).

$$precision = \frac{TP}{TP + FP}$$

- Specificity: The ability of the model to correctly identify negative cases, measured by TN / (TN + FP). Specificity measures the ratio of actual negative cases that are correctly identified.

$$Specifity = \frac{TN}{TN + FP}$$

- Sensitivity/ Recall: The ability of the model to correctly identify positive cases. It can be calculated as TP / (TP + FN). Sensitivity measures the ratio of actual positive cases that are correctly identified.

$$Sensitivity = \frac{TP}{TP + FN}$$

- F1 score: It is the harmonic mean of precision and recall. It provides a balance between precision and recall.

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall}$$