PREDICTIVE Modeling of Student Retention Across Different Student Cohorts Using Machine LEARNING

Edmund V. Enopia[1], Ma. Nanette S. Casquejo, PhD[2]

[1]Faculty, Pilar College of Zamboanga City, Inc.
edmundenopia@gmail.com

[2]University of the Immaculate Conception, Davao City
mncasquejo@uic.edu.ph

ABSTRACT

This study aimed to predict student retention using machine learning, particularly logistic regression, integrated into the Guidance Office Records Management System (GORMS) of the participating catholic institution in Zamboanga City. It focused on system development, predictive performance, technology acceptability evaluation, and integration planning. A quantitative research design with a hybrid model as a system development approach was utilized. Student records were used to develop the logistic regression model, while system users, including guidance personnel, administrators, and IT experts, evaluated the system's technology acceptance using an adopted instrument. Data were analyzed using mean and standard deviation. The application was fully developed in PHP and Python. Major modules were treated as prototypes, including student profiling, student retention prediction, the institutional integration survey, and the exit interview. User-specific access rights were implemented to ensure data privacy. Logistic regression showed satisfactory predictive performance, as indicated by accuracy, recall, precision, and F1-measure. A train-test split and k-fold cross-validation were used to evaluate the model's performance. Furthermore, the technology acceptance evaluation revealed very high levels of acceptance across perceived usefulness, perceived ease of use, and intention to use, suggesting a strong likelihood of adoption. An integration plan was proposed, with the institution serving as the pilot school and future implementation across RVM schools in Southern Mindanao. The institution and other RVM schools will be able to use the results to implement early interventions to improve student retention.

KEYWORDS: *Education, predictive model, logistic regression, machine learning, Philippines.*

INTRODUCTION

Student retention is a critical challenge for academic institutions, serving as a primary indicator of instructional quality and the effectiveness of student support services (Armstead, 2023). High graduation rates are essential for attracting prospective students, as they signify robust academic success and institutional stability (Armstead, 2023). However, many learners, particularly freshmen, encounter unexpected financial burdens—including tuition, books, and transportation—that frequently lead to discontinuance (Watermark Insights, 2023). Students from low-income families are especially vulnerable, often forced to transfer to state-funded institutions or drop out entirely due to these economic constraints (Gabana & Madrigal, 2021).

Globally, the scale of this issue is significant and persistent. In Malaysia, approximately 30% of freshmen leave before their second year due to financial concerns (Shafiq et al., 2022). Similarly, in Ecuador, institutions continue to struggle with retention due to a reliance on manual data acquisition and limited analytical tools, which hinders the strategic identification of at-risk students (Villegas-Ch et al., 2023; Morris, 2016). In the United States, millions of undergraduates with federal student loan debt have dropped out of school due to academic failure, family problems, and emotional strain (Barshay & Barshay, 2021; ASO Staff writers, 2023).

In the Philippines, the shift to modular learning during the COVID-19 pandemic exacerbated these challenges, particularly in remote areas, where limited technology and connectivity led to high dropout rates (Bustillo & Aguilos, 2022). Studies in Cagayan and at De La Salle Araneta University indicate that students often leave due to feelings of academic alienation, financial constraints, or uncertainty regarding their career paths (Bravo, 2023; Dalangin, 2018). Locally, a Catholic institution in Zamboanga City has observed concerning enrollment declines during transitions from Grade 10 to Senior High School and from Senior High School to college. Initial interviews suggest these departures are driven by university preferences and localized financial pressures, underscoring an urgent need for a proactive early-identification system to replace reactive manual record-keeping.

While predictive modeling is established in global literature, there is a notable gap in localized Philippine research that applies data analytics at the institutional level across multiple student cohorts (Biescas et al., 2022; Solo, 2021). Existing local studies are often limited in scope, focusing only on specific degree programs or examining employee rather than student retention (Biescas et al., 2022; Solo, 2021). This study addresses this gap by applying machine learning to a broader, context-specific population—encompassing Grade 10, Grade 12, and early college years—to provide a comprehensive institutional framework for intervention.

This research is theoretically anchored in Tinto’s Student Integration Model and Astin’s Student Involvement Theory, which emphasize that social and academic integration are vital for persistence (Tinto, 1993; Astin, 1984). Furthermore, it utilizes the Theory of Reasoned Action (TRA) and the Technology Acceptance Model (TAM) to evaluate how users perceive and adopt the proposed system (Davis, 1989; Fishbein & Ajzen, 1975). The study aligns with United Nations Sustainable Development Goals (SDG #4 for quality education, SDG #8 for productive employment, and SDG #9 for institutional innovation)

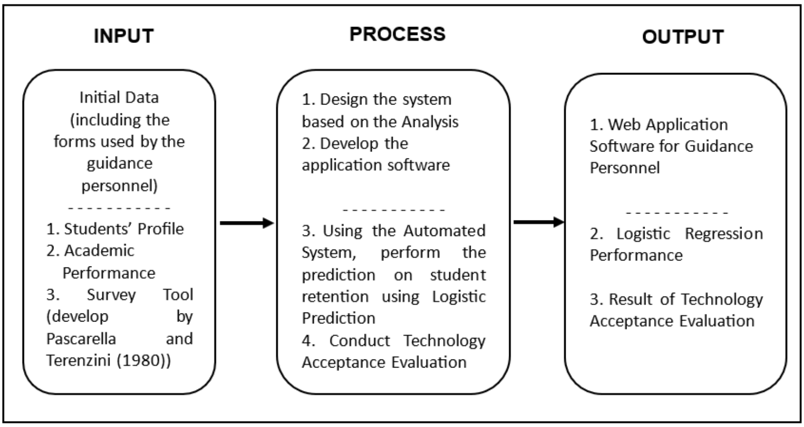
The conceptual framework of this study is anchored on the Input-Process-Output (IPO) model, which systematically guides the development of the Guidance Office Records Management System (GORMS) and the evaluation of its predictive capabilities and user acceptance. According to Figure 1, the "Input" phase involves identifying and collecting manual guidance forms, student records, and institutional requirements that define the system's parameters. These inputs transition into the "Process" stage, characterized by a hybrid development approach that includes incremental prototyping, logistic regression to predict retention, and stakeholder receptivity assessment using the Technology Acceptance Model. The final "Output" phase results in a fully functional, web-based GORMS, alongside validated predictive performance metrics and a strategic integration plan for institutional implementation.

Statement of the Problem

1. How should the application software be designed and developed to effectively collect and analyze relevant data for predicting student retention across diverse student cohorts?
2. How does logistic regression perform in predicting student retention rates among first and second-year college, Senior High School, and Grade 10 students based on the performance metrics: Accuracy, Precision and Recall, and F1 Scores?

3. What is the level of technology acceptance of the developed application software based on the Technology Acceptance Model (TAM) in terms of Perceived usefulness; Perceived ease of use and 3.3 Intention to Use;
4. What is the integration plan for implementing the developed application software for predicting student retention?

Figure 1
Conceptual Framework for GORMS Development and Evaluation

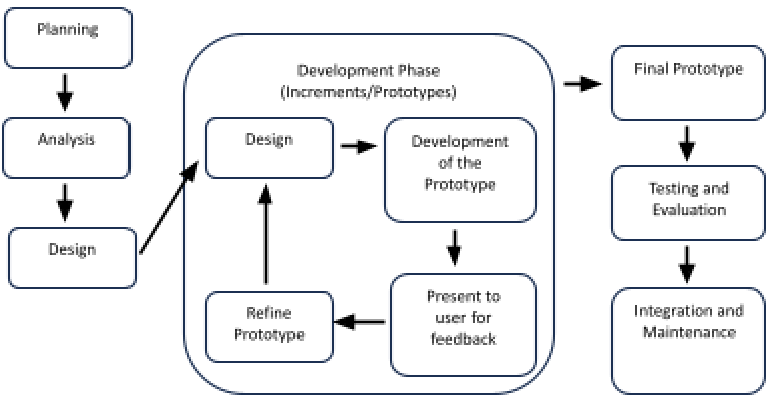


The primary objective of this study is to develop and evaluate the Guidance Office Records Management System (GORMS) by employing logistic regression to predict student retention across diverse cohorts, specifically Grade 10, Grade 12, and early college students. The research focuses on designing a robust application to effectively collect and analyze retention data, while evaluating the predictive performance of the logistic regression model through key metrics such as Accuracy, Precision, Recall, and F1 Scores. Furthermore, the study assesses the level of technology acceptance among stakeholders based on the Technology Acceptance Model (TAM) constructs of perceived usefulness, perceived ease of use, and intention to use(Davis, 1989; Davis & Venkatesh, 2004).

METHODOLOGY

This study utilized a quantitative research design to facilitate an objective, data-driven analysis of student retention outcomes and technology acceptance. The development of the GORMS followed a hybrid SDLC approach, as illustrated in Figure 2, integrating traditional sequential planning with the flexibility of incremental prototyping. This allowed major modules—profiling, prediction, and surveys—to be refined through iterative feedback loops. Complementing this, the predictive modeling component adhered to the CRISP-DM framework, using logistic regression as the classification algorithm due to its efficiency in handling binary outcomes. To ensure model stability, 10-fold cross-validation was employed.

Figure 2
Hybrid Software Development Life Cycle (SDLC) Framework Integrating Sequential Planning and Incremental Prototyping



Research Instrument. Two primary instruments were integrated into the GORMS platform. First, student retention factors as shown in Table 1, were assessed using the Institutional Integration Scales, derived from Tinto's Model to measure academic and social integration. Second, system acceptability was evaluated using an instrument grounded in the Technology Acceptance Model (TAM), measuring perceived usefulness, ease of use, and intention to use. These instruments demonstrated high internal consistency, with Cronbach's alpha values ranging from 0.90 to 0.98. To interpret these results, Table 2 translates numerical mean scores into five qualitative categories, ranging from "Very High" (4.21–5.00) to "Very Low" (1.00–1.80)

Table 1

Summary of Dataset Attributes and Predictor Variables for the Student Retention Predictive Model

Attributes	**Descriptions**	**Acceptable Values**
Age	Age of students	continuous
Sex	Sex of the Students (Male or Female)	Nominal: M – “1” F – “2”
Religion	Religion of the student	Nominal
Ethnicity/tribe	Student’s ethnicity or tribe	nominal
Average Grade	Recent Average Grade of Students Pass: Grade >=75 Failed: Grade <75	{ 75 – 100 } { 60 – 74 }
Parents’ Monthly Income	Combined income of Father and Mother or Salary of Guardian (if supported with guardian)	continuous
Institutional Integration Scales	Categorized as initial goal and institutional commitment, social integration, academic integration, and later goal and institutional commitment. There are 30 items.	Ordinal values 5-point Likert scale
Retention	*This is the target variable*. Initially, student will input whether to stay or transfer	Nominal: Retain – “1” Transfer – “0”

Table 2

Level of Technology Acceptance Interpretation Scale for Qualitative Mean Score Assessment

Mean Range	**Descriptive Scale**	**Interpretation**
4.21 – 5.00	Very High	The respondent shows very high acceptance and willingness to use technology.
3.41 – 4.20	High	The respondent is open to using technology and generally finds it useful and user friendly.
2.61 – 3.40	Moderate	The respondents neither accepts nor rejects technology, possibly unsure or indifferent.
1.81 – 2.60	Low	The respondent shows hesitation or low acceptance/confidence in using technology.
1.00 – 1.80	Very Low	The respondent rejects technology use and perceives it as unhelpful or difficult to use.

Data Mining Integration. The study uses the CRISP-DM framework to develop a predictive model to identify students at risk of transferring across Grade 10, Senior High School, and early college cohorts. This data-driven process is structured into the following phases:

- *Foundation and Preparation.* The process begins with Business Understanding, establishing a goal to provide guidance personnel with a decision-support tool for early intervention. During Data Understanding and Preparation, 1,103 student records were processed, with missing data imputed and attributes transformed to ensure consistency and quality.
- *Algorithmic Modeling.* Logistic regression was selected as the primary classification algorithm because the target variable—retention—is binary in nature (stay or transfer). The model identifies the statistical relationships between independent variables, such as student profiles and integration scales, and the likelihood of retention.
- *Performance Evaluation.* To ensure the model is reliable and generalizable, a 10-fold cross-validation technique was employed. This resulted in an average accuracy of 81.94% and a particularly high recall rate of 99.09%, which is critical for minimizing "false negatives" in student support services.
- Deployment: The finalized model was integrated into the GORMS application system, serving as a functional tool for institutional early warning and timely intervention

Research Locale. This study was conducted at a private Catholic educational institution in the western part of Mindanao, specifically in Zamboanga City. This study focused on using data to predict student retention in grade 10, SHS grade 12, and first- and second-year students. Shown in Figure 4 is the location of the institution.

Data Gathering Procedure. The procedure commenced with securing ethical clearance from the UIC-REC (Protocol No. GS-ER-10-24-0124) and obtaining administrative permission for the study. Upon receiving clearance, orientations were conducted for student cohorts to explain the voluntary nature of the study and to obtain Informed Consent Forms (ICF), including parental assent for minors. Data collection was then conducted via the online GORMS platform, where students completed their profiles and the Institutional Integration Survey. Concurrently, 22 purposively selected stakeholders—including guidance personnel, administrators, and IT experts—participated in a system orientation and user testing session. After providing informed consent, these users evaluated the system's technology acceptance using the TAM-based instrument at their respective offices.

Following collection, all data were anonymized to protect participant privacy and tabulated in MS Excel before being processed in JASP for descriptive statistics and Python for predictive validation. To ensure long-term data security, all records are stored on a password-protected internal server restricted to the researcher and the Guidance Head. In compliance with ethical recommendations, these data will be stored for five years, after which all digital files will be permanently deleted, and any physical documents will be shredded to prevent unauthorized access. A statistician reviewed the final results to ensure accurate interpretation and appropriate presentation in the study.

Statistical Tools. To address the research objectives, the study utilized a combination of descriptive statistics and predictive data mining models. The mean and standard deviation were used to analyze the descriptive results from the technology acceptance test of the GORMS application system. Simultaneously, logistic regression was utilized to develop the predictive model for student retention across the identified student cohorts, a choice justified by the binary nature of the target variable. These tools were recognized by evaluators as appropriate for the research questions, effectively facilitating the interpretation of system performance and user readiness.

RESULTS AND DISCUSSION

The development of the GORMS utilized a hybrid SDLC approach, successfully integrating traditional sequential planning with the flexibility of incremental prototyping. This methodology ensured that each functional module—student profiling, prediction, institutional integration survey, and exit interview—was refined based on continuous stakeholder feedback. As illustrated in Figure 3, the design provides a centralized platform for managing student data, moving away from reactive manual record-keeping toward a proactive digital framework. Table 3 provides further details on how the system satisfies essential functional requirements, including role-based access control and automated report generation, thereby ensuring data integrity and administrative efficiency.

Figure 3

System Architecture of the Web-Based GORMS: Highlighting the Centralized Integration of Guidance and Predictive Modules

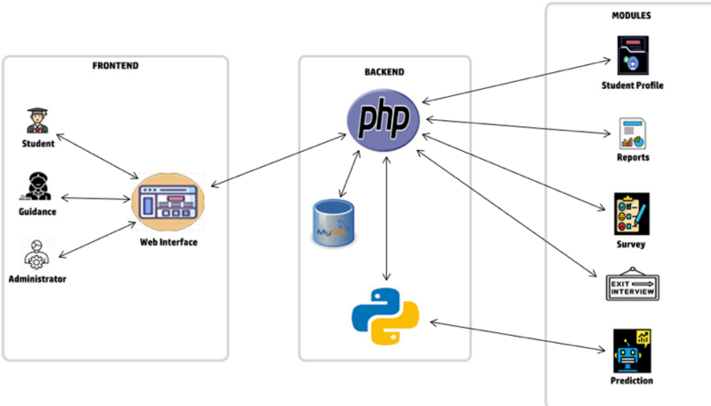


Table 3

Requirements Traceability Matrix (RTM) for the Functional and Non-Functional Specifications of GORMS

Types of Requirements	Requirements	Access/ Involve	Description
Functional	Student Records Management	Guidance Personnel	Add, update, and manage student profiles and cumulative records including academic performance, test results.
Functional	Prediction Module	Guidance Personnel (generate) Other end-users (view)	Allows prediction of student retention using logistic regression to create early intervention.
Functional	Institutional Integration Survey	Students (response) Guidance (manage) Other end-users (view results)	Distribute and collects survey data to assess students' sense of belonging and integration/ engagement
Functional	Exit Interview Module	Guidance Personnel (manage)	Captures feedback from students and data will be used for improvement

Functional	Report Generation	Guidance Personnel Other End-users	Generate reports for profiling, prediction outcomes, survey results and exit interview. Access varies by role.
Functional	Role-based Access Control	System Administrator	Create user accounts and what features to access based on their job roles
Functional	System Maintenance	Guidance Personnel System Administrator	Ther are other available functions that can be performed based on their accessibility like for guidance, creating academic school year. For Admin, maintaining strands, sections, & college programs.
Non-functional	Usability	All users	User-friendly interface to ensure ease of navigation even for non-technical users.
Non-functional	Security	All Users (enforced by Admin/Developer)	Protects data using role-based access, encryption, and secure login protocols.
Non-functional	Data Accuracy	Guidance Personnel, Developer	Ensures correctness of input data and prediction output
Non-functional	Performance	All Users	The system should load pages and reports within acceptable time limits

Building upon this functional foundation, the model's predictive capability, evaluated via 10-fold cross-validation on 1,103 student records, demonstrated high reliability for institutional use. As shown in Table 4, the model achieved an average accuracy of 81.94% and a remarkably high recall rate of 99.09%. In the context of student retention, this high recall is critical as it indicates the system is exceptionally sensitive in identifying students likely to be retained, thereby minimizing "false negatives"—at-risk students who might otherwise be overlooked for support. While the precision of 81.87% suggests some false positives, the institutional priority remains the sensitive identification of all potentially at-risk individuals. Although the accuracy is slightly lower than the 90.8% reported by Yaacob et al. (2020) or the 94.4% achieved by Amare and Simonova (2021), the results confirm that logistic regression is a stable and effective algorithm for identifying key retention predictors like academic standing and socioeconomic status within a localized Philippine context.

Table 4

Predictive Performance Metrics of the Logistic Regression Model for Student Retention based on 10-Fold Cross-Validation

Performance Matrix:	%
Accuracy	81.94
Recall	99.09
Precision	81.87
F1-Measure	89.60

Complementing these technical performance metrics, the evaluation of the GORMS by 22 purposively selected stakeholders revealed a "Very High" level of technology acceptance, with an overall mean score of 4.80. According to Table 5, the system received high ratings for Perceived Usefulness (4.88) and Perceived Ease of Use (4.72). These scores imply that users strongly believe the system will enhance their productivity in managing records and that the interface is intuitive enough to use with minimal mental effort. Furthermore, the high Intention to Use (4.82) reflects a motivated user base ready to adopt the GORMS into their daily workflows. These findings align with those of Davis (1989) and Buabeng-Andoh (2017), who posit that systems perceived as both useful and user-friendly are significantly more likely to be successfully adopted by an organization.

Table 5

Descriptive Analysis of Stakeholder Technology Acceptance for the GORMS based on TAM Constructs

Perceived Usefulness (PU)	**Weighted Mean**	**Standard Deviation**	**Descriptive Equivalent**
Perceived Usefulness (PU)	**4.88**	**0.275**	Very High
Perceived Ease-of-Use (PEU)	**4.72**	**0.364**	Very High
Intention to Use (UI)	**4.82**	**0.424**	Very High
Overall Mean	**4.80**	**0.272**	**Very High**

To operationalize this high level of stakeholder readiness and ensure the transition from development to practice, a strategic Integration Plan was formulated to guide the rollout across the pilot institution and other RVM schools. As detailed in Table 6, the roadmap includes key areas such as user training, data migration to internal servers, and the establishment of an intranet network to enhance data security. This structured approach is designed to foster a data-informed culture that prioritizes student success through evidence-based, early interventions.

Table 6

Strategic Integration Plan and Sustainability Roadmap for the Institutional Deployment of GORMS

Key Results Area	Objectives	Strategies/ Activities	Budget	Time/ Duration	Persons Involved	Monitoring and Evaluation	Performance Indicators
Institutional Integration	Adopt the application within the participating institution and RVM schools	1. Conduct a Meeting with RVM EMC 2. Create policies for system usage 3. Establish a Memorandum of Agreement (MOA) with RVM schools	50,000	1 month	Researcher, RVM EMC, Participating school, Other RVM schools	Review of MOA and policy compliance	Agreed and sent communication to RVM schools
System Deployment	Ensure successful implementation of the system in both participating institution and RVM schools	1. Conduct initial pilot testing at the participating institution 2. Assess system performance and address technical issues 3. Install and configure the system in identified RVM schools.	200,000	2 months	IT Team, Guidance Personnel, School Administrations IT Officers	System performance evaluation and user feedback	System successfully installed and functional in target schools
User Training	Train users for smooth system adoption	1. Conduct training for all end-users 2. Provide hands-on sessions 3. Develop user manuals and training modules	50,000	1 month	Guidance Staff, IT Team, RVM representatives	Post-training assessment and feedback	At least 90% of trainees demonstrate proficiency in system usage
Data Transfer	Conduct data transfer	For the participating institution, transfer student records from the online portal to internal server of the institution.	20,000	1 month	IT Team and ITRC personnel	Checking of student records	All records are successfully transferred at the internal server of the participating institution

Key Results Area	Objectives	Strategies/ Activities	Budget	Time/ Duration	Persons Involved	Monitoring and Evaluation	Performance Indicators
Data encoding	Encode manual records to the application system	For RVM schools without system automation at the guidance office: 1. assign staff to encode data into the application system 2. schedule for data encoding 3. encode the records of students into the system		3 months	Guidance personnel	System performance evaluation and user feedback	100% of the records are encoded in to the sys-tem
Security	Ensure security and data privacy of students' records	Implement security protocols to safeguard data integrity	c/o ITRC Budget (Php. 100,000)	Monthly /as needed	ITRC Head and Staff, Guidance Staff	Security Audit	Secured students' records
Internal Network	To ensure data privacy and avoid from intrusion	Create an internal network (intranet) to implement the GORMS in each RVM school.		1 month	ITRC Head and Staff	System performance evaluation and user feedback	GORMS runs within the internal network of the institution
Technical Support and Maintenance	Ensure continuous technical support and updates	1. Appoint and establish a technical support for system troubleshooting 2. Conduct periodic maintenance checks	ITRC Budget (1,500 monthly)	Start: September 2025 Monthly / as needed	School President, ITRC Head, ITRC Staff	Review of issue logs and system maintenance reports	Issues are address with issue resolution within a given time
Evaluation and Continuous Improvement	Evaluate and refine the system regularly	1.Conduct annual evaluation 2. Implement refinements based on user feedback	c/o Guidance Budget (Php. 3,000)	Year-end / as needed	School President, ITRC and Guidance Heads and Staff, Users of the system	Evaluation Reports and action plans	Continues improvement based on the feedback

CONCLUSION

The study successfully developed the Guidance Office Records Management System (GORMS) using a hybrid SDLC approach, providing a centralized platform for efficient data management and proactive student support. Logistic regression proved an effective predictive tool across Grade 10, Senior High School, and early college cohorts, yielding an exceptional 99.09% recall rate, critical for identifying at-risk students and minimizing dropout risk. Furthermore, the system achieved a "Very High" level of technology acceptance among stakeholders, confirming its institutional viability and potential to foster a data-informed culture within RVM schools. Collectively, these results demonstrate that integrating predictive analytics into guidance services enables timely, targeted interventions that significantly enhance both student success and institutional effectiveness.

To ensure the continued efficiency and security of the GORMS, the institution should conduct regular usability testing and technical reviews to maintain alignment with evolving guidance needs. For data privacy, the system should be implemented within an internal intranet to protect sensitive records from external hacking and unauthorized access. Future research should expand beyond user acceptance to incorporate "effectiveness indicators" that measure the actual impact of system-triggered interventions on student retention. Additionally, subsequent studies should conduct comparative analyses using advanced algorithms, such as random forests or neural networks, on larger, multi-year datasets. They should utilize mixed-method approaches to gain deeper qualitative insights into the complex factors driving student attrition. Finally, for journal publication, technical descriptions should be condensed to meet length requirements while explicitly reporting data storage duration and disposal protocols.

REFERENCES

Al-Dossary, S. A. (2008). A study of the factors affecting student retention at King Saud University, Saudi Arabia: Structural equation modeling and qualitative methods [Doctoral dissertation, University of Stirling]. https://core.ac.uk/download/pdf/9047402.pdf

Amare, M. Y., & Šimonová, S. (2021). Global challenges of student dropout: A prediction model development using machine learning algorithms on higher education datasets. SHS Web of Conferences, 129, 09001. https://doi.org/10.1051/shsconf/202112909001

Armstead, K. (2023, November 21). Student retention: Definition, importance & effects. Study.com. https://bit.ly/3vKMt8r

ASO Staff Writers. (2023, October 17). How to prevent students from dropping out. Accredited Schools Online. https://www.accreditedschoolsonline.org/resources/preventing-students-dropping-out/

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. Journal of College Student Personnel, 25(4), 297–308. (Added from general knowledge)

Barshay, J., & Barshay, J. (2021, April 8). Federal data shows 3.9 million students dropped out of college with debt in 2015 and 2016. The Hechinger Report. https://hechingerreport.org/federal-data-shows-3-9-million-students-dropped-college-debt-2015-2016/

Biescas, J. P. A., Colao, J. M., & Peniones, V. S. (2022). Retention prediction of computer students in the Institute of Computer Studies using logistic regression. https://www.studocu.com/ph/document/western-mindanao-state-university/computer-science/retention-prediction-1-v5-update-aug1/44133563

Bravo, E. D. (2023). Trends and causes of student dropouts in a public higher education institution in northern Philippines: A data visualization approach. Journal for Educators, Teachers, and Trainers. https://doi.org/10.47750/jett.2023.14.03.081

Buabeng-Andoh, C. (2018). Predicting students' intention to adopt mobile Bustillo, E., & Aguilos, M. (2022). The challenges of modular learning in the wake of COVID-19: A digital divide in the Philippine countryside revealed. Education Sciences, 12(7), 449. https://doi.org/10.3390/educsci12070449

Dalangin, J. J. G. (2018). Factors affecting student retention at De La Salle Araneta University. Journal of Global Business. https://www.researchgate.net/publication/340333513_Factors_Affectin_Student_Retention

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Davis, F. D., & Venkatesh, V. (2004). Toward preprototype user acceptance testing of new information systems: Implications for software project management. IEEE Transactions on Engineering Management, 51(1), 31–46. https://doi.org/10.1109/TEM.2003.822468

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley. (Added from general knowledge)

Gabana, J. J., & Madrigal, D. V. (2021). A glimpse into the lives of transferring private school students to a state college. Technium Social Sciences Journal, 20(1), 313–322. https://doi.org/10.47577/tssj.v20i1.3610

Morris, L. V. (2016). Mining data for student success. Innovative Higher Education, 41(3), 183–185. https://doi.org/10.1007/s10755-016-9367-6

Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. The Journal of Higher Education, 51(1), 60–75. https://doi.org/10.2307/1981125 (Added from general knowledge)

Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: A systematic literature review. IEEE Access, 10, 72492–72539. https://doi.org/10.1109/ACCESS.2022.3188767

Solo, A. M. (2021). Exploring job satisfaction and employee retention in a private higher education institution (HEI) in Zamboanga City. International Social Science Review, 3, 1–25. https://uz.edu.ph/wp-content/uploads/2022/05/ISSR-VOL3-doc.pdf

Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (2nd ed.). University of Chicago Press.

Villegas-Ch, W., Govea, J., & Revelo-Tapia, S. (2023). Improving student retention in institutions of higher education through machine learning: A sustainable approach. Sustainability, 15(19), 14512. https://doi.org/10.3390/su151914512

Watermark Insights. (2022, April 29). Top 5 challenges for student retention. https://www.watermarkinsights.com/resources/blog/top-5-challenges-for-student-retention

Yaacob, W. F. W., Sobri, N. M., Nasir, S. A. M., Norshahidi, N. D., & Husin, W. Z. W. (2020). Predicting student drop-out in higher institutions using data mining techniques. Journal of Physics: Conference Series, 1496(1), 012005. https://doi.org/10.1088/1742-6596/1496/1/012005Acceptance Model (TAM) to Predict University Students' Intentions to Use Metaverse-Based Learning Platforms. Education and Information Technologies. https://di.org/10.1007/s10639-023-11816-3