

Predicting First-Year College Enrollment Using Binary Logistic Regression

Joey M. Red¹, Julian J. Meimban III²

School of Graduate Studies, New Era University
Quezon City, Philippines

Abstract - Grounded in Hossler and Gallagher's College Choice Model, this study developed and validated a binary logistic regression model to predict whether Senior High School graduates would enroll as first-year students in the same private university. Although Philippine higher education institutions routinely collect demographic and academic records, enrollment decisions are still often examined using descriptive methods rather than predictive analytics. To address this gap, the study analyzed the combined effects of gender, religion, academic strand, and General Weighted Average on enrollment outcomes. Using institutional records from 883 Senior High School graduates in Academic Year 2024–2025, the data were split into a model development sample ($n = 500$) and a validation sample ($n = 383$). Binary logistic regression using the Enter method in IBM SPSS v27 was employed, and all assumptions were satisfied. Model fit was assessed through the -2 Log Likelihood, Nagelkerke R^2 , and Hosmer–Lemeshow test, while predictive performance was evaluated using hit ratio, sensitivity, and specificity. Results showed that the model was statistically significant, $\chi^2 = 53.36$, $p < .001$, with a 64.6% hit ratio, higher sensitivity (80.9%) than specificity (47.3%), and strong validation support. Religion emerged as the strongest predictor (Wald = 42.28, $p < .001$, $\text{Exp}(B) = 3.83$, 95% CI [2.56, 5.74]). Gender, GWA, and individual academic strand were not statistically significant predictors. These findings suggest that sociocultural and curricular alignment may influence enrollment decisions more strongly than demographic or academic performance factors. The study offers a data-driven framework for enrollment prediction and provides useful evidence for recruitment and retention planning in similar private universities in the Philippines.

Keywords – Binary Logistic Regression, College Enrollment, Enrollment Prediction, Higher Education Analytics, Philippines

I. INTRODUCTION

A. Background of the Study

Student enrollment remains one of the most critical concerns in higher education institutions, particularly in private universities where sustainability depends heavily on stable student intake and retention (Varsitarian News, 2025; Saguin, 2023). In the Philippine context, many universities face increasing competition to attract and retain Senior High School (SHS) graduates for first-year college enrollment, despite having direct access to their own graduating students. Although institutions routinely collect demographic and academic information during admissions—such as gender, religion, academic strand, and General Weighted Average (GWA)—these data are often used only for descriptive reporting rather than predictive decision-making. This limits the institution's ability to anticipate enrollment behavior and design targeted interventions. The problem becomes more significant because enrollment decisions directly affect institutional planning, including faculty workload distribution, resource allocation, scholarship programming, and long-term financial sustainability. Without reliable predictive tools, universities rely heavily on assumptions or historical trends that may no longer reflect current student behavior. This creates inefficiencies in enrollment management and reduces the effectiveness of recruitment and retention strategies.

Existing studies consistently identify gender, religion, academic strand, and academic performance as influential factors in students' college choice decisions (Cruz et al., 2025; Bacon & Wang, 2024; Del Rosario et al., 2024). Gender may shape educational aspirations and program preferences, while religion influences institutional affinity, values, and cultural alignment. Academic strand reflects curricular preparedness and alignment with college programs, and GWA serves as an indicator of academic readiness and competitiveness for admission. Studies by

Manugas et al., (2022) and Allensworth & Clark, (2020) support the importance of these variables in explaining educational transitions. However, most Philippine-based studies remain descriptive or correlational, examining these predictors independently rather than within a unified predictive framework. There is limited application of Binary Logistic Regression (BLR) to estimate the probability of actual first-year college enrollment using combined demographic and academic indicators.

To address this gap, the present study proposes developing a Binary Logistic Regression model grounded in Hossler and Gallagher's College Choice Model (Hossler & Gallagher, 1987). The model estimates the likelihood that SHS graduates will enroll in the same university using four predictors: gender, religion, academic strand, and GWA. By quantifying both individual and combined effects, the study transforms routinely collected institutional data into actionable predictive insights. This work is limited to SHS graduates of one private university for the academic year 2024–2025 and focuses only on variables available in institutional records. Other potentially relevant predictors, such as socioeconomic status, parental education, financial capacity, and geographic location, were excluded due to data limitations. Despite this delimitation, the study provides a statistically grounded and institutionally relevant contribution by offering a practical enrollment forecasting tool.

The significance of this study lies in its ability to support evidence-based enrollment management. It contributes a focused and novel application of BLR in the Philippine higher education setting, strengthens local literature on predictive educational analytics, and provides university administrators with a decision-support mechanism for targeted recruitment, student retention, and strategic institutional planning.

B. Literature Review

This review synthesizes the literature on four key predictors examined in this study: gender, religion, academic strand, and academic performance (General Weighted Average, or GWA), while also situating binary logistic regression within enrollment modeling.

1) Gender and College Enrollment: Gender has long been examined as a determinant of educational participation and college choice in the Philippines. Earlier studies highlight persistent gender disparities in access to higher education and in field selection, often linked to cultural expectations and socioeconomic conditions (Beltran & Arboleda, 2024; Silor & Silor, 2023; Reyes et al., 2021). However, recent literature suggests that gender disparities have narrowed in many contexts of admissions and program choice, particularly where selection criteria are more standardized, although substantial gaps remain in STEM and other highly segregated fields (Sison, 2025; Chua, 2024). For instance, evidence indicates that while female participation in higher education has increased and in some cases surpassed that of males in overall enrollment, women continue to face a "glass ceiling" in STEM fields, with persistent underrepresentation in engineering, computer science, and related technical programs (Sison, 2025; Flores et al., 2025). Recent work on mathematics education in the Philippines further suggests that equal access alone does not ensure gender equity in outcomes, as gendered classroom practices and expectations continue to shape participation and achievement in quantitative fields (Chua, 2024). Studies on technical education and certification also reveal gender-based disparities in assessment outcomes and program completion, particularly in male-dominated technical-vocational tracks (Flores et al., 2025). Together, these findings indicate that while overall gender gaps in college access have narrowed, gender continuously interacts with other variables in shaping program choice and enrolment decisions in the Philippine context (Flores et al., 2025; Sison, 2025; Torres and Narciso, 2025; Silor & Silor, 2023).

2) Religion as a Sociocultural Determinant: Religion plays a significant role in shaping educational aspirations, values, and institutional preferences. In the Philippine context, religion continues to shape educational aspirations, values, and institutional preferences, influencing both the decision to pursue higher education and the type of institution students choose (Berkley Center for Religion, Peace & World Affairs, 2025; *Perspectives of youth on religious education in the Philippines*, 2023; *Lived experiences of Muslim students in secular and Islamic educational institutions in the southern Philippines*, 2024). Oracion & Madrigal (2020) found that religiosity and spiritual well-being are associated with students' motivation and educational orientation, particularly in faith-based institutions. International research also supports the influence of religion on educational outcomes. Bacon and Wang (2024) reported that religious affiliation correlates with varying levels of educational attainment across different denominations. Similarly, longitudinal evidence indicates that students' religious environments significantly predict college participation, even after controlling for other factors. In contexts where institutions are religiously affiliated,

alignment between a student's belief system and institutional identity can enhance perceived fit and increase the likelihood of enrollment. These findings reinforce the relevance of religion within the predisposition stage of the College Choice Model, where sociocultural background and personal values shape initial educational intentions.

3) Academic Strand and Educational Alignment: The introduction of the K–12 curriculum in the Philippines has elevated the importance of the academic strand as a key factor in educational transitions. The academic strand reflects students' specialization during Senior High School and is closely linked to their preparedness for specific college programs. Empirical studies consistently show that the academic strand influences both college enrollment and academic performance. Padios et al. (2024) found that strand alignment significantly predicts whether students pursue higher education and whether their chosen courses match their SHS preparation. Similarly, Manugas et al. (2022) reported that the academic strand is a strong determinant of college GPA, indicating its long-term impact on academic success. International evidence mirrors these findings. Helbling et al. (2022) demonstrated that students in academically oriented tracks are more likely to transition to higher education than those in vocational streams. Biewen (2020) also found that early academic tracking has lasting effects on tertiary enrollment decisions. These studies highlight the role of the academic strand as a structural and preparatory factor influencing the search stage of college choice, where students evaluate institutions based on perceived academic fit.

4) Academic Performance (GWA) and Enrollment Decisions: Academic performance, commonly measured as General Weighted Average (GWA), is widely regarded as a key predictor of college readiness and success. Numerous studies have established a strong relationship between high school grades and higher education outcomes. For example, Allensworth (2020) found that high school GPA is a stronger predictor of college completion than standardized test scores. In the Philippine context, Maslang (2023) demonstrated that GWA significantly contributes to predicting enrollment using regression analyses. International studies also confirm that students with higher academic performance are more likely to enroll in and succeed in higher education programs (Wu et al., 2023). However, some studies suggest that academic performance alone may not fully explain enrollment decisions. Manugas et al., (2024) found that social, financial, and institutional factors often outweigh academic indicators in shaping college choice. This indicates that while GWA is an important measure of readiness, it must be considered alongside other contextual variables.

5) Binary Logistic Regression in Enrollment Studies: Binary Logistic Regression (BLR) is widely used in educational research to model dichotomous outcomes such as enrollment, retention, and graduation. In international contexts, BLR has been applied to examine student enrollment behavior influenced by academic, socioeconomic, and institutional factors (Othman et al., 2022; Thongsook & Somran, 2023). In the Philippines, however, its application remains limited. Daway-Ducanes et al. (2022) used regression modeling to analyze the income advantage in course choice and admission to the University of the Philippines System, demonstrating the growing relevance of predictive analytics in local educational research. Few studies integrate multiple demographic and academic variables within a single logistic regression framework to predict actual enrollment behavior. This gap highlights the need for more comprehensive, data-driven models that move beyond descriptive analysis.

The reviewed literature establishes that gender, religion, academic strand, and academic performance are important factors influencing college enrollment decisions. However, existing studies—particularly in the Philippine context—are largely descriptive or examine variables in isolation. There is limited evidence on how these predictors interact when analyzed simultaneously within a predictive statistical model. Moreover, the application of binary logistic regression in modeling institutional enrollment decisions remains underdeveloped locally. This limits the ability of higher education institutions to leverage existing data for evidence-based decision-making. Addressing this gap, the present study integrates these variables into a unified binary logistic regression model to estimate enrollment probability among SHS graduates. By doing so, it contributes a focused and methodologically robust approach to understanding and predicting college enrollment behavior in the Philippine private higher education setting.

II. METHODOLOGY

A. Research Design

This study employed a quantitative predictive research design using Binary Logistic Regression (BLR) to develop and validate a model for predicting first-year college enrollment among Senior High School (SHS) graduates. Predictive research design is appropriate when the primary objective is to estimate the probability of occurrence of a dichotomous outcome based on a set of independent variables. Since the dependent variable in this study—enrollment status—was binary (enrolled = 1, not enrolled = 0), binary logistic regression was identified as the most suitable analytical approach (Hosmer et al., 2013; Menard, 2010). The study was anchored in the College Choice Model of Hossler and Gallagher, which explains college decision-making through the stages of predisposition, search, and choice. Gender and religion were aligned with the predisposition stage, academic strand with the search stage, and General Weighted Average (GWA) with the choice stage. BLR served as the empirical mechanism for quantifying the influence of these predictors on actual enrollment behavior.

B. Research Setting and Population

The study was conducted in a private university in Quezon City, Philippines, focusing on SHS graduates from the university's Basic Education Department for Academic Year 2024–2025. The population consisted of 883 graduates representing four academic strands: Science, Technology, Engineering, and Mathematics (STEM), Humanities and Social Sciences (HUMSS), Accountancy, Business, and Management (ABM), and Technical-Vocational-Livelihood (TVL). Specifically, the distribution included 524 STEM students, 168 HUMSS students, 117 ABM students, and 74 TVL students.

C. Sampling Technique and Sample Size

Although the entire population (N = 883) was included, the dataset was analytically divided into two subsets to support model development and validation. Using stratified random partitioning in *IBM SPSS Statistics Version 27*, 500 cases were assigned to the model development dataset and 383 cases to the validation dataset. Stratification ensured proportional representation across academic strands. The model development sample exceeded the minimum recommended sample size for logistic regression as recommended by Bujang et al. (2018).

D. Data Source and Variables

The study used secondary data obtained from the official institutional databases of the Registrar's Office and Admissions Office. Formal approval to access anonymized student records was secured through appropriate university channels. The dataset was encoded in Microsoft Excel. The dependent variable was Enrollment Status (ES), coded as 1 = Enrolled and 0 = Not Enrolled.

The independent variables included: Gender (GEN): coded as 1 = Male and 0 = Female; Religion (REL): coded as 1 = Majority religious affiliation and 0 = Minority religious affiliation; Academic Strand (STRAND): coded as 1 = ABM, 2 = HUMSS, 3 = STEM, and 4 = TVL; General Weighted Average (GWA): treated as a continuous variable representing academic performance during SHS. These variables were selected based on prior literature and their theoretical relevance to the College Choice Model.

E. Data Analysis Procedure

Data analysis was conducted using *IBM SPSS Statistics Version 27*, following five major stages: data preparation, exploratory analysis, assumption testing, model estimation, and model validation.

1) Preliminary Data Screening: The dataset was initially examined for completeness, consistency, and encoding accuracy. Descriptive statistics such as frequency distributions, percentages, means, and ranges were generated to assess the characteristics of the variables. Missing values, encoding errors, and out-of-range entries were checked before model estimation. No missing data were identified, and all 500 development cases were retained for analysis.

2) **Assumption Testing:** Before model estimation, the assumptions of binary logistic regression were evaluated.

- Independence of Observations
Each case represented one unique SHS graduate, ensuring independence of observations with no repeated measures or clustered cases.
- Absence of Multicollinearity
Correlation matrices and multicollinearity diagnostics were examined to ensure that predictor variables were not excessively correlated.
- Linearity of the Logit
For the continuous predictor (GWA), the assumption of linearity in the logit was assessed.
- Outliers and Influential Cases
Casewise diagnostics were examined using standardized residuals and influence statistics. No extreme outliers or influential observations were identified.

These diagnostics confirmed that the dataset met the assumptions required for valid logistic regression modeling.

3) **Model Estimation:** The binary logistic regression model was estimated using the Enter method, where all predictors were entered simultaneously to evaluate their combined and individual effects on enrollment status.

The logistic regression model followed the general form:

$$\text{logit}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (\text{Equation 2-1})$$

where:

Y is a binary dependent variable
 $\text{logit}(Y)$ is the $\log_e(\text{Odds})$ or $\ln(\text{Odds})$
 β_0 is the intercept
 β_1 to β_4 are the regression coefficients
 X_1 is the Gender
 X_2 is the Religion
 X_3 is the Academic Strand
 X_4 is the GWA

In the condition that the Odds $Y = 1$, the formulas for Odds and Probability are:

$$\text{Odds}(Y) = e^{\ln[\text{odds}(Y=1)]} = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)} \quad (\text{Equation 2-2})$$

$$P(Y) = (e^{\ln[\text{odds}(Y=1)]}) \div (1 + e^{\ln[\text{odds}(Y=1)]}) \quad (\text{Equation 2-3})$$

Statistical significance of regression coefficients was tested at $\alpha = .05$ using the Wald test. Odds ratios ($\text{Exp}(B)$) and 95% confidence intervals were reported.

4) **Model Fit Assessment:** To evaluate model adequacy, the following goodness-of-fit and predictive accuracy measures were used:

Omnibus Test of Model Coefficients
 -2 Log Likelihood (-2LL)
 Nagelkerke R^2
 Hosmer-Lemeshow Goodness-of-Fit Test
 Classification Accuracy (Hit Ratio)
 Sensitivity (True Positive Rate)
 Specificity (True Negative Rate)

The default cut-off value of 0.50 was used for classification.

5) **Model Validation:** Validation was conducted using the reserved validation dataset ($n = 383$) to

assess model generalizability. The same model fit indices and classification accuracy measures were recalculated and compared with the development model. Consistency across these indices provided evidence of model stability and external validity.

III. RESULTS AND DISCUSSIONS

A. Results

Table 1 shows that, of 500 cases, 257 (51.4%) enrolled, while 243 (48.6%) did not. The nearly balanced distribution between enrolled and not enrolled students is advantageous for logistic regression modeling, as it prevents bias toward one outcome category and ensures that the model must genuinely differentiate between the two groups. Logistic regression performs best when outcome groups are reasonably balanced, as emphasized by Lever, Krzywinski, and Altman (2016), who noted that classification models require adequate representation of both classes to avoid distorted parameter estimates.

Table 1. Percentage of the Dependent Variable Enrolled vs Not Enrolled

Dependent Variable	<i>n</i>	%
Enrolled	257	51.4
Not Enrolled	243	48.6
Total	500	100

Table 2 presents the frequency distribution of the independent variables. Gender was evenly represented (47.6% male, 52.4% female), suggesting minimal risk of gender imbalance affecting the model. Religion, however, was more uneven, with the majority affiliation comprising 66.4% of the sample compared to 33.6% minority affiliation. SHS strand distribution was dominated by STEM graduates (59.4%), followed by HUMSS (19.0%), ABM (13.2%), and TVL (8.4%).

Table 2. Frequency and Percentage of Cases by Independent Variable

Independent Variables	N	%
Gender		
Male	238	47.6
Female	262	52.4
Religion		
Majority	332	66.4
Minority	168	33.6
SHS Strands		
TVL	42	8.4
ABM	66	13.2
HUMSS	95	19.0
STEM	297	59.4

Table 3 presents a case processing summary showing that all 500 cases (100%) were included in the analysis, and there were no missing cases.

Table 3. Case Processing Summary

Unweighted Cases	N	%
Selected Cases		
Included in Analysis	500	100.0
Missing Cases	0	.0
Total Cases	500	100.0
Unselected Cases	0	.0
Total	500	100.0

The baseline model without predictors in (Table 4) classified all cases as “Enrolled,” resulting in 100% correct prediction for enrolled students and 0% for not enrolled students, with an overall accuracy of 51.4%. This shows that before adding predictors, the model simply predicted the majority outcome category.

Table 4. Baseline Model of Predicted Enrollment vs Observed Enrollment

Predicted Enrollment	Not Enrolled	Enrolled	Percentage Correct
Observed Enrollment			
Not Enrolled	0	243	.0
Enrolled	0	257	100.0
Overall Percentage			51.4

Note. The cut-off value is .500

Table 5 reports the Youden Index across different cut-off values. At the SPSS default cut-off of .500, sensitivity was 80.9%, specificity was 47.3%, and the Youden Index was 28.2, yielding an overall hit ratio of 64.6%. Lower cut-off values inflated sensitivity but sacrificed specificity, while higher cut-offs reversed this pattern. The .500 threshold provided the most balanced classification, though the model was more effective at correctly identifying enrolled students than non-enrolled ones. This asymmetry suggests that the model is better suited for enrollment prediction than for identifying non-enrollment.

Table 5. Youden Index at Different Cut-off Values

Cut-off Value	SN ^a %	SP ^b %	SN+SP-100 %	Hit Ratio %	Nagelkerke R ²	Hosmer-Lemeshow
.1	100	0	0	51.4	.135	.226
.2	100	0	0	51.4	.135	.226
.3	87.9	38.3	26.2	63.8	.135	.226
.4	82.5	44.0	26.5	63.8	.135	.226
.5	80.9	47.3	28.2*	64.6*	.135	.226
.6	28.4	83.1	11.5	55.0	.135	.226
.7	17.1	93.0	10.1	54.0	.135	.226
.8	1.6	100.0	1.6	49.4	.135	.226
.9	0	100.0	0	48.6	.135	.226

* Youden index with the most balanced classification. ^aSensitivity, ^bSpecificity

Table 6 summarizes the predictive accuracy measures. The -2LL was significant ($\chi^2 = 53.36$, $df = 6$, $n = 500$, $p < .001$), confirming that the model provides explanatory power beyond chance. The Nagelkerke R² value of (.135) suggests that the model accounts for approximately 13.5% of the variance in enrollment decisions.

The Hosmer-Lemeshow test ($\chi^2 = 10.59$, $df = 8$, $n = 500$, $p = .226$) indicates good model fit. Overall predictive accuracy was 64.6%, with sensitivity (80.9%) exceeding specificity (47.3%). These results demonstrates that the model is statistically valid, moderately strong, and well-calibrated, though its predictive strength is concentrated on identifying enrolled students.

Table 6. Predictive Accuracy Measures of the Estimated LR Model

Predictive Accuracy Measures	X^2	df	R^2	p	Hit Ratio	SN	SP	Y.I. ^a
Goodness-of-Fit								
-2LL	53.36	6		.001				
Cox & Snell			.101					
Nagelkerke			.135					
Hosmer-Lemeshow	10.59	8	.226					
Over-all Predictive Accuracy								
Hit Ratio					64.6			
Youden Index								28
Predictive Accuracy Actual								
Sensitivity						80.9		
Specificity							47.3	

^a.5 cut-off value, , N=500 model development sample size

Table 7 provides the classification table. The model correctly classified 208 of 257 enrolled cases (80.9%) but only 115 of 243 not enrolled cases (47.3%). The overall classification accuracy was 64.6%. This reinforces the earlier finding that the model is more effective in predicting enrollment than non-enrollment, which may reflect institutional realities where identifying likely enrollees is more straightforward than predicting attrition.

Table 7. Classification Table for the Final Estimated LR Model

Predicted Enrollment	Not Enrolled	Enrolled	Percentage Correct
Observed Enrollment			
Not Enrolled	115	128	47.3
Enrolled	49	208	80.9
Overall Percentage			64.6

Note. The cut-off value is .500

Table 8 presents the variables in the final estimated logistic regression model. Gender ($B = -0.019$, $p = .926$) and general weighted average (GWA; $B = -0.018$, $p = .551$) were not significant predictors of enrollment. Religion emerged as the lone predictor ($B = 1.343$, $p < .001$, $\text{Exp}(B) = 3.830$), indicating that students from the majority religious affiliation were nearly four times more likely to enroll compared to minority counterparts. SHS strand was also significant overall ($p = .023$), with specific strands showing borderline effects (p -values between .065 and .078), suggesting that curricular background influences enrollment decisions, though not as strongly as religion

Table 8. Logistic Regression Predicting Enrollment

Predictor	B	S.E.	Wald χ^2	p	Odds Ratio	95% C.I.
GENDER	-0.019	.200	.009	.926	.982	[0.66, 1.45]
RELIGION	1.343	.207	42.282	<.001	3.830	[2.56, 5.741]
SHS_STRAND	---	---	9.521	.023	---	
GWA	-0.018	.030	.355	.551	.982	[0.93, 1.04]

The comparison between the derived and validation binary logistic regression models in predicting first-year college enrolment is presented in Table 9. Results revealed that both models were statistically significant, as indicated by the Omnibus Test of Model Coefficients, with the derived model yielding $\chi^2 = 53.36$, $p < .001$, and the validation model yielding $\chi^2 = 42.36$, $p < .001$. These findings demonstrate that the predictor variables significantly improved the prediction of enrollment decisions compared to the constant-only model. The Nagelkerke R^2 values of .135 and .140 for the derived and validation models, respectively, indicate that both models consistently explained approximately 13% to 14% of the variance in enrollment behavior, reflecting modest but stable explanatory power. In terms of predictive performance, the validation model slightly outperformed the derived model in overall classification accuracy (66.3% vs. 64.6%) and specificity (50.8% vs. 47.3%), suggesting improved identification of non-enrolled students and evidence of good external validity with minimal overfitting. Sensitivity remained consistently high across both models (80.9% and 80.5%), indicating strong capability in correctly classifying students who were likely to enroll. Notably, religion consistently emerged as the strongest predictor in both datasets, reinforcing its significant influence on students' enrollment decisions. Overall, the similarity of findings between the derived and validation models supports the reliability, stability, and generalizability of the developed predictive model.

Table 9. Comparison of the Derived and Validation Models

Model Evaluation Indicator	Derived Model (n = 500)	Validation Model (n=383)	Comparative Interpretation
Omnibus Test of Model Coefficients	$\chi^2 = 53.36$, $p < .001$	$\chi^2 = 42.358$, $p < .001$	Both models were statistically significant, indicating that the predictors improved enrollment prediction over the constant-only model.
Nagelkerke R^2	.135	.140	Both models explained approximately 13–14% of the variance in enrollment decisions, indicating moderate but consistent explanatory capability.
Overall Classification Accuracy	64.6%	66.3%	The validation model slightly improved overall predictive accuracy, suggesting good external validity and minimal overfitting.
Sensitivity (Correctly Classified Enrolled Students)	80.9%	80.5%	Sensitivity remained consistently high across both models, confirming strong ability to identify students likely to enroll.
Specificity (Correctly Classified Non-Enrolled Students)	47.3%	50.8%	Specificity slightly improved in the validation model, indicating better identification of non-enrolled students.
Strongest Predictor	Religion	Religion	Religion consistently emerged as the most influential predictor of enrollment in both datasets.

B. Discussion of Findings

The principal findings derived from the data analysis are summarized below:

The logistic regression model demonstrated moderate predictive accuracy. At the cut-off value of .500, the model achieved an overall hit ratio of 64.6%, correctly classified 208 of 257 enrolled cases 80.9% sensitivity and 115 of 243 not enrolled cases 47.3% specificity (Table 5; Table 6). The Nagelkerke R^2 of .135 indicates that the model explains approximately 13.5% of the variance in enrollment decisions, while the Hosmer-Lemeshow test ($\chi^2 = 0.226$, $p = .11$) confirmed good model fit. These results suggested that the model was statistically valid and practically useful, with particular strength in predicting enrollment outcomes, though it was less effective in identifying non-enrollment.

The model's strength thrived in predicting enrollment with high sensitivity, but it struggled to predict non-enrollment, as reflected in its lower specificity. From an institutional perspective, this bias may be acceptable, as universities were primarily concerned with identifying and securing enrollees. However, for retention and outreach planning, improving specificity would be beneficial. The moderate explanatory power (Nagelkerke $R^2 = .135$) and good fit (Hosmer- Lemeshow $p = .11$) confirmed that the model was both statistically sound and moderately useful for enrollment forecasting.

Among the independent variables tested, Religion emerged as a statistically significant predictor ($B = 1.34$, $SE = 0.21$, $Wald = 42.28$, $p < .001$, $Exp(B) = 3.83$, 95% CI [2.56, 5.74]). Students from the majority religion were approximately 4 times as likely to enroll, suggesting that sociocultural alignment may influence institutional choice. This finding was consistent with studies highlighting the role of cultural congruence and institutional fit in shaping student decisions (Oracion & Madrigal, 2020). This finding also underscored the sociocultural dimension of enrollment decisions, suggesting that institutional affiliation and cultural identity played a decisive role in shaping student choices, complementing the findings of Bacon and Wang (2024).

The SHS Academic Strand was also found to be a significant predictor of enrollment. The academic strand was statistically significant overall, $Wald \chi^2 (3) = 9.52$, $p = .023$, ($p < .05$), indicating that prior academic preparation strands influenced enrollment decisions when considered as a group. This supported literature emphasizes the importance of academic alignment and preparedness in college choice (Padios et al., 2024; Manugas et al., 2022).

Gender was found not to be a significant predictor, $B = -0.02$, $SE = 0.20$, $Wald = 0.01$, $p = .926$, ($p > .05$). This finding contrasts with several previous studies that identified gender as an influential factor in higher education participation and institutional choice (Flores et al., 2025; Sison, 2025; Torres and Narciso, 2025; Silor & Silor, 2023). Moreover, the result may reflect changing societal attitudes toward gender equality in education, particularly in urbanized and modern educational environments where institutional policies promote inclusivity and equal access. The finding also implies that both male and female students may now perceive similar educational opportunities and institutional support within the university, thereby reducing the predictive influence of gender on enrollment outcomes (Sison, 2025).

General Weighted Average (GWA) was not a statistically significant predictor of enrollment ($B = -0.02$, $SE = 0.03$, $Wald = 0.36$, $p = .551$; $p > .05$). This suggests that academic performance alone does not determine enrollment decisions. This finding aligned with studies indicating that social, financial, and institutional factors often outweighed academic indicators in influencing college choice (Manugas et al., 2024). Furthermore, the findings imply that while GWA may remain important for measuring academic preparedness and potential college performance, it may not independently determine actual enrollment behavior in all institutional settings. Consequently, the present study contributes to the literature by demonstrating that, within the examined university context, GWA did not significantly predict enrollment despite the strong predictive relationships reported in previous local and international studies.

C. Theoretical Implications

The findings of this study provides important theoretical implications for understanding college enrollment behavior within the context of higher education decision-making. Anchored on the College Choice Model developed by David Hossler and Karen Gallagher (1987), the study empirically validated the relevance of sociocultural and academic factors in shaping students' enrollment decisions. The significant influence of religion and academic strand supported the model's assertion that students' predisposition, educational preparation, and institutional alignment collectively affected college choice behavior.

The findings further extend the theoretical application of the College Choice Model by demonstrating that enrollment decisions could be quantitatively modeled using Binary Logistic Regression. While the original framework primarily explains the stages of the college choice conceptually, the present study operationalizes these constructs into measurable variables capable of predicting enrollment probability. This integration bridges the gap between descriptive educational theory and predictive statistical modeling, thereby contributing to the growing field of educational data analytics and predictive enrollment research.

Moreover, the non-significant effects of gender and General Weighted Average (GWA) suggests that enrollment behavior in the institutional context of the study may have been influenced more strongly by sociocultural affiliation and curricular alignment than by demographic characteristics or academic performance alone. This finding contributed to the ongoing discourse in enrollment theory by highlighting the contextual nature of college choice determinants, particularly within faith-oriented educational institutions in the Philippine setting.

Overall, the study strengthened the theoretical proposition that college enrollment decisions are multidimensional and context-dependent, while also demonstrating the value of predictive analytics in advancing enrollment theory from explanatory understanding toward empirical prediction and institutional application.

D. Conclusions

The study developed and validated a Binary Logistic Regression (BLR) model capable of predicting the likelihood of Senior High School (SHS) graduates enrolling as first-year college students in the same university. The model demonstrated acceptable goodness-of-fit and predictive accuracy, indicating its suitability for institutional enrollment forecasting and decision-making.

The findings support the applicability of the College Choice Model of Hossler and Gallagher (1987) in explaining enrollment behavior and further demonstrate that educational decision-making can be effectively analyzed through predictive statistical modeling. The study contributes to the growing body of Philippine-based educational predictive analytics research by demonstrating the practical use of institutional data in forecasting enrollment outcomes. The developed model may serve as a valuable evidence-based tool for enrollment management, strategic planning, and policy formulation in higher education institutions.

E. Recommendations

Future researchers are encouraged to extend the present study by incorporating additional predictors such as socioeconomic status, parental education, financial capability, geographic location, and institutional perception to improve the predictive accuracy of the model. Replication of the study across multiple institutions and longitudinal cohorts is also recommended to strengthen generalizability and contribute further to the growing field of educational predictive analytics in the Philippine context.

ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, 49(3), 198–211. <https://doi.org/10.3102/0013189X20902110>
- [2] Bacon, R. J., & Wang, L. (2024). Selection into higher education and subsequent religious decline in a United States cohort. *Social Science Research*, 124, Article 103067. <https://doi.org/10.1016/j.ssresearch.2024.103067>
- [3] Beltran, J. K. D., & Arboleda, E. R. (2024). Investigating gender disparities in electronics engineering program enrollments at a state university in the Philippines. *International Journal of Multidisciplinary: Applied Business and Education Research*, 5(1), 21–28. <https://doi.org/10.11594/ijmaber.05.01.03>
- [4] Biewen, M., & Tapalaga, M. (2020). Early tracking, academic vs. vocational training, and the value of 'second- chance' options. *Labour Economics*, 66, Article 101900. <https://doi.org/10.1016/j.labeco.2020.101900>
- [5] Bujang, M. A., Sa'at, N., Sidik, T. M. I. T. A. B., & Joo, L. C. (2018). Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. *The Malaysian journal of medical sciences : MJMS*, 25(4), 122–130. <https://pubmed.ncbi.nlm.nih.gov/30914854/>
- [6] Camral, K. A. (2025). Lived Experiences of Muslim Students in Secular and Islamic Educational Institutions in the Southern Philippines. *Psychology and Education: A Multidisciplinary Journal*, 45(7), 870-879. <https://doi.org/10.70838/pemj.450707>

- [7] Chua, V. C. G. (2024). Engendering gendered mathematics education in the Philippines: Is equality of access enough? *Power and Education*, 16(2), 139–149. <https://doi.org/10.1177/17577438231174132>
- [8] Cruz, A. U., Salamatin, A. M., & Espinosa, K. M. (2025). Factors influencing school preference: Mapping the student market. *International Journal of Multidisciplinary: Applied Business and Education Research*, 6(12), 6079–6097. <https://doi.org/10.11594/ijmaber.06.12.18>
- [9] Daway- Ducanes, S. L. S., Pernia, E. E., & Ramos, V. J. R. (2022). On the income advantage in course choices and admissions: Evidence from the University of the Philippines. *International Journal of Educational Development*, 91, Article 102578. <https://doi.org/10.1016/j.ijedudev.2022.102578>
- [10] del Castillo, F. A., & Cacho, R. (2023). Perspectives of youth on religious education in the Philippines. *Religious Education*, 119(1), 43–58. <https://doi.org/10.1080/00344087.2023.2289264>
- [11] Del Rosario, R. A., Cruz, A. R., Bartolome, C. N., Manipon, N., Dela Cruz, A. M., & Villarama, J. (2024). Understanding how senior high school students choose a college degree program: A phenomenological study. *Journal of Interdisciplinary Perspectives*, 2(7), 118–127. <https://doi.org/10.69569/jip.2024.0151>
- [12] Flores, J. P., Sambaan, R. M., Dullente, J. H., Jamilio, J. A., & Mahinay, R. D. (2025). Gender- based disparities in technical education assessment and certification outcomes across Philippine regions: A regional perspective. *Journal of Interdisciplinary Perspectives*, 3(10), 63–69. <https://doi.org/10.69569/jip.2025.561>
- [13] Helbling, L. A., Tomasik, M. J., & Mosler, U. (2022). Putting an explanatory understanding into a predictive perspective: An exemplary study on school track enrollment. *Frontiers in Education*, 6, Article 793447. <https://doi.org/10.3389/feduc.2021.793447>
- [14] Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: A three- phase model and the implications for policymakers. *College and University*, 62(3), 207–221.
- [15] Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- [16] Lever, J., Krzywinski, M., & Altman, N. (2016). *Logistic regression*. *Nature Methods*, 13, 541–542. <https://doi.org/10.1038/nmeth.3904>
- [17] Manugas, S. B., Pepito, M. T., Fernandez, J. J. S., & Canque, M. A. (2022). Senior high school tract as determinant for college GPA: A correlational study. *International Journal of Science and Management Studies*, 5(3), 230–234. <https://doi.org/10.51386/25815946/ijsms-v5i3p126>
- [18] Maslang, K. L. (2023). Predictors of Academic Performance Among Pre-Service Teachers: The Case of a Private Higher Educational Institution in Northern Philippines. [Online]. Available: <https://ssrn.com/abstract=4325530> or <http://dx.doi.org/10.2139/ssrn.4325530>
- [19] Menard, S. (2020). *Logistic regression: From introductory to advanced concepts and applications*. Sage Publications.
- [20] Oracion, E. G. & Madrigal, D. V. (2020). Religiosity and Spiritual Well-Being of Senior High School Students of a Catholic College in the Philippines. [Online]. Available: <http://serp-p.pids.gov.ph/publication/public/view?slug=religiosity-and-spiritual-well-being-of-senior-high-school-students-of-a-catholic-college-in-the-phi>
- [21] Othman, W. R. W., Wahab, S. F. A., Ngah, N. H., Halim, I. I. A., & Motalib, S. S. S. A. (2024). Determining significant factors for selection of private higher education institutions in Malaysia using binary logistic regression. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 37(2), 58–67. <https://doi.org/10.37934/araset.37.2.5867>
- [22] Padios, Jr, A. ,Lejano, R. ., Gorospe, S. A. ., & De Asis, V. . (2021). Strand and Statehood Predictors of Senior High School Graduates: A Tracer Study. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 55(1), 211-224. [Online]. Available: <https://www.gssrr.org/JournalOfBasicAndApplied/article/view/12087>
- [23] Reyes, W., Bayten, E., & Mercado, F. (2021). Differences in gender perceptions of higher education institutions (HEI) students and teachers on science technology engineering agriculture mathematics (STEAM) education. *The Normal Lights*, 15(2). <https://doi.org/10.56278/tnl.v15i2.1750>
- [24] Saguin, K. (2023). The politics of de-privatisation: Philippine higher education in transition. *Journal of Contemporary Asia*, 53(3), 471–493. <https://doi.org/10.1080/00472336.2022.2035424>
- [25] Sison, M. (2025). Filipinas stare through the glass ceiling in STEM fields. UNESCO. [Online]. Available: <https://www.unesco.org/en/articles/filipinas-stare-through-glass-ceiling-stem-fields>
- [26] Silor, F., & Silor, A. (2023). Gender inequality in STEM education: Basis for gender sensitization program. In *The Barcelona Conference on Education 2023: Official Conference Proceedings* (pp. 517–530). <https://doi.org/10.22492/issn.2435-9467.2023.40>
- [27] Thongsook, S., & Somran, S. (2022). Binary logistic regression analysis for the relationship of an undergraduate study choice at Phibunsong Kham Rajabhat University with various factors. *Life Sciences an Environment Journal*, 23(2), 306–323. <https://doi.org/10.14456/lsej.2022.24>

-
- [28] Torres, C. S., & Narciso, P. G. (2025). From imbalance to inclusion: A quantitative study on gender participation trends in engineering and architecture education (AY 2020–2025). *International Journal of Multidisciplinary: Applied Business and Education Research*, 6(7), 3335–3346. <https://doi.org/10.11594/ijmaber.06.07.10>
- [29] Varsitarian News. (2025, February 2). Private sectarian schools suffer 8% drop in enrollment as students flock to SUCs, LUCs. The Varsitarian. [Online]. Available: <https://varsitarian.net/private-sectarian-schools-suffer-8-drop-in-enrollment-as-students-flock-to-sucs-lucs/>
- [30] Wu, J.-P., Lin, L., Lin, L., & Tsai, C.-L. (2023). A predictive model that aligns admission offers with student enrollment probability. *Education Sciences*, 13(5), 440. <https://doi.org/10.3390/educsci13050440>
- [31] World Faiths Development Dialogue. (2025, August 11). Religion and education in the Philippines. Berkley Center for Religion, Peace, and World Affairs. [Online]. Available: <https://berkeleycenter.georgetown.edu/publications/religion-and-education-in-the-philippines>