Predicting Students' Mobility using Different Statistical Tools: Basis for Students' Success

PAULO V. CENAS, JENNIFER M. PARRONE, DANIEL BEZALEL A. GARCIA, FREDERICK F. PATACSIL
College of Computing,
Pangasinan State University,
Urdaneta City, Pangasinan,
PHILIPPINES

Abstract:- This paper investigates students' success at Pangasinan State University by identifying patterns and models that might be used to correctly classify and predict if a student will transfer or finish their studies. In this study, three categorical variables or attributes and one continuous variable were considered independent variables due to the availability of the data. The results from the binary logistic regression model with the high school general average and course as independent variables (Model 3), and the decision tree model with transition gain as a splitting criterion were fitted to the dataset to generate a model that possibly best describes the students' mobility in Pangasinan State University Urdaneta City Campus. The decision tree model is better than the binary logistic regression model based on accuracy, AUC, and sensitivity values. This implies that the decision tree model is better at correctly classifying observations as "transferred" than Model 3. Thus, it was concluded that the decision tree model with information gain as the splitting criterion best describes the mobility of PSU students. The results of this paper can be used for school administration involving students' mobility/success, particularly in classifying whether a student will transfer based on other.

Key-Words: - Mobility, Success, Statistical Tools, Decision Tree, Logistic Analysis

Received: March 11, 2022. Revised: September 27, 2022. Accepted: October 19, 2022. Published: November 24, 2022.

1 Introduction

Every university is responsible for preparing students for good jobs and personal growth, as well as assisting them in contributing to the betterment of society. For this, universities should improve their programs, implement updated and relevant curricula, and build personal and cultural resources. For the past five years, the Pangasinan State University has consistently done its mandates to improve the teaching and learning experience, equip the faculty members through training, seminars, and academic scholarships, provide facilities such as laboratories, and build and renovate existing edifices, all for the success of the students. The students' success plays a crucial role in every university, as it is commonly used as a performance metric for every academic institution, [1]. Students' success can be explained by the number of board or licensure passers and topnotchers, or the highest employability rates of its graduates, and of course, the number of successful graduates. Completing an academic degree is, in fact, one of the most regarded achievements of any student. Many students will enrol and start their first year of college at the Pangasinan State University Urdaneta City Campus, yet it is a constant observation that few students will finish their degree within the maximum year of residency. The reason for this is either the student transferred to another university or dropped out of the university. This implies that the university is losing a significant percentage of its enrolees every semester. Also, low completion rates among relevant attributes aside from biological sex, course, hometown, and high school general average. students have been an immense threat to the key performance indicators of the university system, [2].

This kind of student mobility is a perpetual predicament, not exclusively at PSU Urdaneta but, also in other schools and universities, and has deeper consequences for those students involved. Though transferring from one school to another may not necessarily have any significant effect on the overall academic achievement of the students, [3], changing schools is almost certain to create discord in their overall learning experience, [4]. Also, transferring from one school to another may benefit some students, but it has an overall negative impact on the possibility of getting a degree, [5]. Further, transferring may have an intense emotional effect, and social and academic problems for the students, [6]. These imply that there is a greater need for adjustment in social, emotional, and academic aspects on the part of the students, which may contribute to any difficulties when it comes to learning when maladaptation occurs.

Reasons for transferring may be linked to various factors. Academic performance has a huge impact on students' retention, and transfer, [7], [8], [9]. Transferring schools or institutions can affect access to most degree programs. This is common among students who are unsure of what they want to study in college and choose any degree program on a whim. Another reason is financial difficulties or the availability of scholarships. Students look for a college or university where they can transfer, that offers free tuition, scholarships, or student loans. It is an undeniable fact that earning a degree requires a huge amount of money, but due to free tuition at state or local colleges and universities, the expenses are reduced but still require a significant amount of money. Many other factors significantly contribute to the decision of students to transfer, and it is a challenge for every academic institution to know these in order to propose and create solutions for student mobility.

The university loses its accomplishments as students transfer, and the students who transfer or drop out sacrifice the benefit of the continuity of the services offered by the university. Low completion rates due to transferring and dropping out affect not only the student, but also the systematic changes projected by university or school reform policies [10]. Hence, early detection of student risk is necessary, and should be used for policy making, particularly in the admission of students to ensure higher completion rates within the university. Using the data from the Pangasinan State University Urdaneta City Campus, this study aims to provide a model that will describe the students' completion based on their profiles and provide recommendations based on the result of the model.

The main objective of this paper is to find a model that possibly best describes student mobility at the Pangasinan State University Urdaneta City Campus as a basis for predicting students' success. In particular, this study sought to:

1. to describe the nature and characteristics of the collected data;
2. to generate a model that possibly best describes the students' mobility in Pangasinan State University Urdaneta City Campus using:
   a. Decision Tree Model; and
   b. Binary Logistic Regression Model;
3. to compare the generated models using Decision Tree and Binary Logistic Regression based on the following criteria:
   a. Accuracy;
   b. Area Under the Curve (AUC); and
   c. Sensitivity.

2 Methodology

The classification algorithms were implemented using RapidMiner and RStudio, both of which are open-source software primarily used for data science. The decision tree model is applied to further understand patterns in students' mobility. This is named a "decision tree" because the result after using this model is a collection of nodes intended to create a decision that is akin to a tree when represented as a graph. The process of creating decision tree models depends on the purpose, whether for classification or regression. In this study, the decision tree model for classification was applied because the target attribute assigned as a label, which is the student's status (whether the student will transfer or graduate), is not numerical. Thus, the decision tree rule is utilized to separate the values belonging to different categories or classes. The criterion used for splitting in this study is the information gain criterion. The gini index criterion was also considered, but it is more applicable for larger distributions. The accuracy and gain ratio criteria were also tried, but based on the accuracy, precision in predicting the transferred class, and area under the curve (AUC) values, the information gain method for splitting is more applicable. Also, the information gain method is perfect for smaller partitions with a variety of mixed and diverse values. The application of the information gain method requires the splitting of the dataset into training and testing data sets. The rule of thumb in assigning percentages for the training and testing datasets was implemented, that is, 70% for training and 30% for testing. Stratified random sampling is used to preserve the distribution of the label (status) in both training and testing datasets,[11].

Table 1. Comparison of the four Criteria for Splitting

Splitting Method	Accuracy	Precision (Transferred Class)	AUC
Information Gain	**70.73%**	**72.70%**	**0.714**
Gain Ratio	69.40%	71.07%	0.674
Gini Index	70.55%	72.64%	0.708
Accuracy	70.55%	72.64%	0.670

Note: Values are derived based on the actual dataset. The highest values are in **boldface**. In AUC, a value closest to 1.00 is the best.

The training dataset is used to generate a decision tree model based on a maximal depth of 4 after splitting. The maximal depth parameter is used to restrict the depth of the decision tree model and depends on the size and characteristics of the dataset. This is one of the stopping criteria for decision tree models. Pruning was also allowed in this model, thus, some branches in the tree model will be replaced by leaves based on the set value for the confidence parameter. The confidence parameter prescribes the confidence level used for the pessimistic error calculation of pruning. The default confidence level value is 0.1, but in this study, the confidence level value is set to 0.3 to provide a decent and less complicated decision tree graph. All other parameters are set to default values. The model generated from the training dataset is then applied to the test dataset to predict the label. The decision tree model's accuracy and prediction performance are based on the class precision and class recall values.

Fig. 1: Methodological Process Applied in Generating Decision Tree Model

After that, the binary logistic regression model which is a classification algorithm used to predict a dichotomous variable based on a set of independent variables is employed since this study is concerned with whether a student will transfer, or graduate based on their biological sex, course enrolled in PSU, hometown municipality/province, and high school GWA. The application of binary logistic model was applied using of RStudio. Various necessary model fitting and visualization packages in R were used in this study. The list of packages and their uses can be seen in Table 2.

Table 2. R packages used for Binary Logistic Regression Modelling

R Package	Use
caret	For fitting and evaluation of the binary logistic regression model
ggplot2, visreg	For visualization of data and regression models
plotROC	For constructing the ROC curves

The first step in the binary logistic regression modelling is to plot the data to determine if the independent variables are related to the binary outcome of academic survival at PSU (graduate or transfer). Take note that the plot results are just rough estimations in determining the relationship. The next thing to do is to formulate the binary logistic regression model to be implemented. Take note that the generalized additive model is in the form.

$$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i \, ,$$

Table 3. Binary Logistic Regression Models Considered

Model	Equation
Model 1	$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_1 * High\,School\,GWA$
Model 2	$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_1 * High\,School\,GWA + \beta_2 * Sex$
Model 3	$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_1 * High\,School\,GWA + \beta_2 * Course$
Model	$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_1 * High\,School\,GWA + \beta_2 * Hometown$

It can be observed from the models considered in the study and presented in the table above that they consist of some combinations of independent variables, specifically the combination of high school GWA and any of the other independent variables. The independent variables of any binary logistic regression model can be continuous or categorical. In this study, it is the choice of the researchers to always include high school GWA since it sensible to think that this variable contributes to whether the student will transfer or graduate. The best model relative to other models presented in Table 1 was chosen based on the in-sample and out-sample accuracy, Cohen's kappa, and AUC.

Fig. 2: Methodological Process in the Binary Logistic Regression Model

Finally, the generated models from the decision tree and binary logistic model were compared based on the values of AUC, accuracy, and Cohen's Kappa.

3 Results and Discussion

The total number of transferred students from 2010 to 2021 according to the records of the campus registrar's office, is 3,002. Any entry from the data with vague or no information was removed, resulting in 2,442 entries left after the data cleaning. A random sample of 1,304 entries from the graduated class since 2000 was chosen so that it will comprise 35% of the total combined transfer-graduated data. Take note that the data considered in this study is based on the availability of data from the records of the campus registrar's office. The graph below depicts the proportion of transferred and graduated classes.

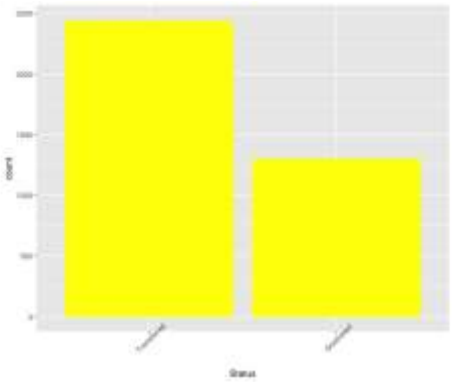

Fig. 3: The proportion of Transferred and Graduated Classes Considered in this Study

Figures 4-6 depict a visual inspection of independent variables based on the models presented in Table 3 in terms of status (transferred or graduated).

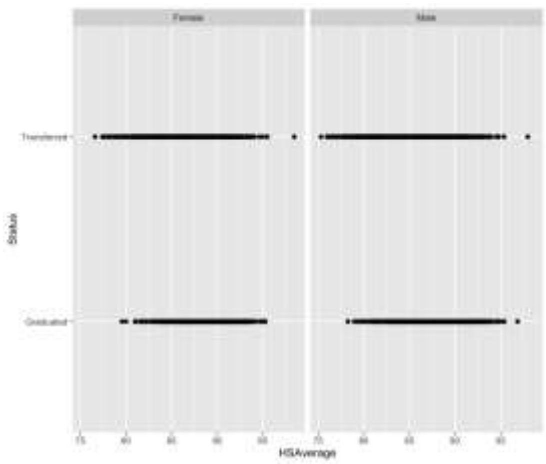

Fig. 4: Visual Inspection of High School GWA across Biological Sex in terms of Status

Figure 4 suggests that the variables high school GWA and biological sex are related to the binary outcome of status, that is, whether the student transferred or graduated. It can be shown that the range of values for the high school average for those who transferred, both male and female, is longer, particularly at the lower bound, implying that there are more transferred students whose high school average is below 80 than those who graduated. There are evidently extreme values, that is, an average greater than 97.5, under the transferred class for both males and females.

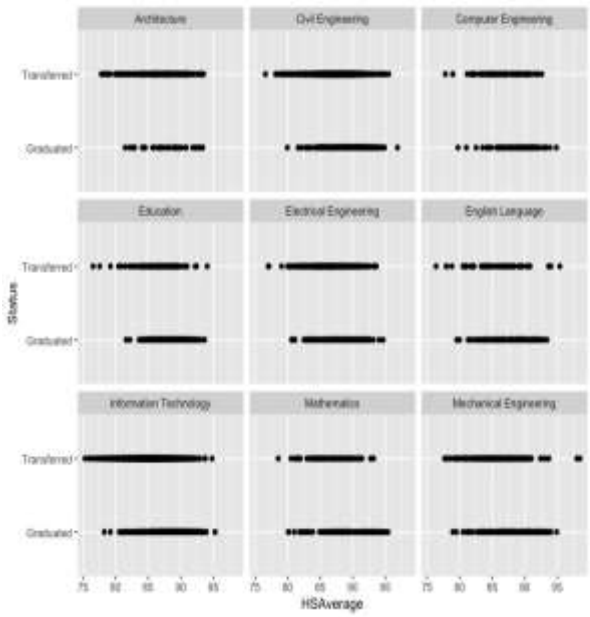

Fig. 5: Visual Inspection of High School GWA across Courses in terms of Status

Figure 5 also suggests that the variables High School GWA and Course are related to the binary outcome of Status. The graph above depicts the range of high school averages for those who transferred and graduated from the nine college majors available at the PSU Urdaneta City Campus. For instance, the range of values of the high school average in the information technology program is longer at the lower bound for those who transferred, implying that there are more transferred students whose high school average is lower than 80 than those who graduated. The same pattern can be seen in the other eight remaining courses. Further, there are extreme values, as evident in Figure 4, that are the same extreme values evident in the mechanical engineering program under the transferred class.

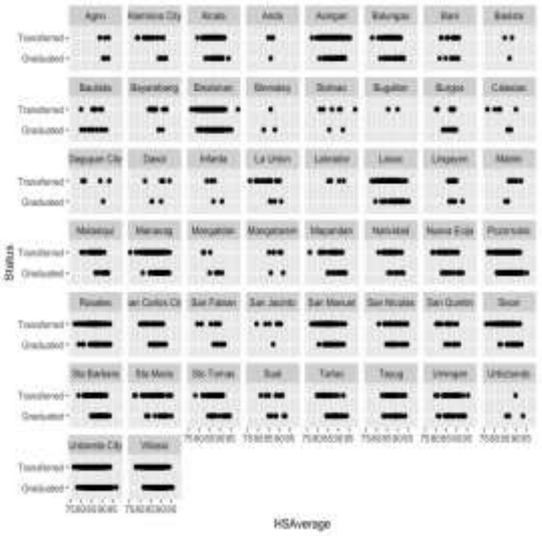

Fig. 6: Visual Inspection of High School GWA across Hometown in terms of Status

It can be observed based on Figure 6 that there are municipalities or cities within Pangasinan that have more transferred data than graduated, such as Alaminos City, Anda, Basista, Dagupan City, Dasol, Infanta, Mabini, Urbiztondo, and San Jacinto. The same is true for La Union. The possible reasons for this are either because these areas are geographically distant from Urdaneta City or they transferred to a university or college closer to their hometowns, such as Dagupan City and La Union. Some areas are geographically distant from Urdaneta City but have almost the same proportion of transfers and graduates, such as Bani and Sual in Pangasinan, Tarlac, and Nueva Ecija. There are also areas in Pangasinan with no recorded graduates from Labrador or Bugallon. Looking at Figure 7 suggests that the variables High School GWA and Hometown may or may not be related to the binary outcome of Status due to the complexity or lack of information in some areas.

3.1 Decision Tree Model

The generated decision tree model using information gain as a splitting criterion is shown in the figure below.

Fig. 7: The Decision Tree Model of the Student Mobility Dataset with Information Gain as the Splitting Criterion (maximal depth = 4)

It can be observed from Figure 7 that some of the evident rules present here are those enrolled in the Information Technology program with a high school general average of below 86.537 who later transferred, and those enrolled in Civil Engineering program with a high school average of less than 88.434 or greater than 88.434 but less than 92.035 who later transferred. Based on the graph, the most striking case is in the part of the architecture where no splitting occurred. The model showed that when a student is enrolled in the Architecture program, there is a huge possibility that he/she will transfer, regardless of their biological sex, high school general average, or hometown. Another difference is that biological sex is only used in splitting in the English language program. This means that in the English Language program, there is a huge possibility that a student will transfer if the student has a high school general average of less than 84.779 and is female when compared to a male. Overall, it is sensible to think that the high school general average has the biggest factor that might affect students' mobility. The branches of the decision tree graph in text form below.

Course = Architecture: Transferred {Graduated=19, Transferred=163}
Course = Civil Engineering
| HSAverage > 88.434
| | HSAverage > 92.035: Graduated {Graduated=37, Transferred=17}
| | HSAverage ≤ 92.035: Transferred {Graduated=80, Transferred=129}
| HSAverage ≤ 88.434: Transferred {Graduated=68, Transferred=274}
Course = Computer Engineering
| HSAverage > 87.404: Graduated {Graduated=36, Transferred=17}
| HSAverage ≤ 87.404: Transferred {Graduated=13, Transferred=48}
Course = Education
| HSAverage > 89.585: Graduated {Graduated=44, Transferred=4}
| HSAverage ≤ 89.585
| | HSAverage > 84.800: Graduated {Graduated=90, Transferred=47}
| | HSAverage ≤ 84.800: Transferred {Graduated=8, Transferred=19}
Course = Electrical Engineering
| HSAverage > 87.370
| | HSAverage > 91.393: Graduated {Graduated=16, Transferred=9}
| | HSAverage ≤ 91.393: Transferred {Graduated=47, Transferred=61}
| HSAverage ≤ 87.370: Transferred {Graduated=35, Transferred=115}
Course = English Language
| HSAverage > 84.779
| | HSAverage > 93.565: Transferred {Graduated=0, Transferred=3}
| | HSAverage ≤ 93.565: Graduated {Graduated=61, Transferred=25}
| HSAverage ≤ 84.779
| | Sex = Female: Transferred {Graduated=5, Transferred=19}
| | Sex = Male: Graduated {Graduated=5, Transferred=4}
Course = Information Technology
| HSAverage > 86.537
| | HSAverage > 90.517: Graduated {Graduated=45, Transferred=26}
| | HSAverage ≤ 90.517: Transferred {Graduated=114, Transferred=161}
| HSAverage ≤ 86.537: Transferred {Graduated=45, Transferred=375}
Course = Mathematics
| HSAverage > 85.705: Graduated {Graduated=51, Transferred=25}
| HSAverage ≤ 85.705: Transferred {Graduated=7, Transferred=28}
Course = Mechanical Engineering
| HSAverage > 90.082
| | HSAverage > 96.423: Transferred {Graduated=0, Transferred=2}
| | HSAverage ≤ 96.423: Graduated {Graduated=33, Transferred=7}
| HSAverage ≤ 90.082: Transferred {Graduated=54, Transferred=131}

Table 5. Parameter Estimates for the Binary Logistic Regression Model 3

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	22.55775	1.15766	19.486	< 2e-16***
HSAverage	-0.23313	0.01286	-18.13	< 2e-16***
Course = Civil Engineering	-1.19828	0.21621	-5.542	0.0000000299***
Course = Computer Engineering	-1.89583	0.26334	-7.199	0.000000000000606***
Course = Education	-2.69151	0.23737	-11.339	< 2e-16***
Course = Electrical Engineering	-1.60395	0.23069	-6.953	0.00000000000358***
Course = English Language	-2.73283	0.26374	-10.362	< 2e-16***
Course = Information Technology	-1.36175	0.21512	-6.33	0.000000000245***
Course = Mathematics	-2.27361	0.26685	-8.52	< 2e-16***
Course = Mechanical Engineering	-1.8733	0.2367	-7.914	0.00000000000000249***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3.2 Binary Logistic Regression Model Fitting

The results of performance measures of the binary logistic regression model fitting of the four models considered in this study are in Table 6.

Table 6. Performance Measures of the Four Binary Logistic Regression Models

	In-Sample Accuracy	AUC	Out-Sample Accuracy	Kappa	Sensitivity	Specificity
Model 1	0.6839295	0.6924294	0.6839554	0.2139176	0.3013109	**0.8882875**
Model 2	0.6855312	0.6924047	0.6827809	0.2108405	0.2992361	0.8876336
Model 3	**0.7178324**	**0.7487583**	**0.7156181**	**0.3187784**	**0.4209419**	0.8736256
Model 4	0.692205	0.7111978	0.6781635	0.2147846	0.3302894	0.8661349

Note: Values are derived based on the actual dataset. The highest values are in **boldface**.

It is important to look at the following performance measures to identify the best binary logistic regression model compared to other models in comparison. The simplest among these performance indicators is the in-sample accuracy, which is defined as the proportion of correct classifications a model makes. Based on the results, Model 3 exhibits the largest in-sample accuracy value of 71.78%. Take note that this is called "in-sample accuracy" because the interpretation of this is limited to our dataset. The generalized version of this is the out-of-sample accuracy. Model 4 achieves the highest out-of-sample accuracy among all models considered in this study. Another performance metric is the area under the receiver operating characteristic (ROC) curve, also known as "area under the curve." An AUC value close to 0.5 indicates that the classification is based on random guessing, while an AUC value equal to 1.0 indicates perfect classification. Thus, a value closer to 1.0 is better.

Based on Table 5, the Model 3 exhibits the highest AUC value. Another performance measure is the Kappa, which measures the inter-rater reliability and is commonly used to measure the level of agreement between the model's predictions and the actual data. A kappa value of at least 0.60 is considered substantial. Based on the results, all kappa values are fair, nevertheless, Model 3 exhibits the largest kappa values among the other models in consideration. Lastly, the sensitivity, which measures the capability of the model to correctly classify an observation as "transferred,", and specificity, which measures the capability of the model to correctly classify an observation as "graduated," are other performance measures that might help us to further identify the best model relative to other models in this study. All models in comparison have high specificity values, which indicate these models are good at correctly classifying observations as "graduated.". Since our concern is the ability of the model to correctly classify those who have been transferred, Model 3 is probably the best model based on the sensitivity value. Overall, Model 3 is the best model compared to the other binary logistic regression models considered in this study. The estimated coefficient values for Model 3 are shown in the next table.
In the interpretation of the estimates, it is important to remember that each coefficient represents an additive linear contribution on the log-odds scale. For the case of a categorical variable just like the course in Model 3, if the observation belongs to one of these 9 courses, then its value is equal to 1 for that particular course and 0 for the other 8 courses, making it the baseline of the model in terms of the Course variable. In this model, each 1-unit increase in the high school general average decreases the log odds of transferring by 0.23313, and if the student is enrolled in the English Language Program, the log odds of transferring will further decrease by 2.73283. Another way to interpret this is by exponentiating the coefficients. For instance, suggests that the odds of transferring change by a factor of approximately 0.7921 for each 1-unit increase in the high school GWA. Moreover, observe that the decrease in log-odds of transferring depends on the per-unit increase in the high school average and the enrolled degree program, except for architecture. This means that if a student is enrolled in the Architecture program, the decrease in the log odds of transferring solely depends on the per-unit increase in his/her high school general average. The Wald test was used to test the significance of the individual regression coefficients in Model 3. All regression coefficients, aside from that for architecture as a course, were found to be significant based on the results.

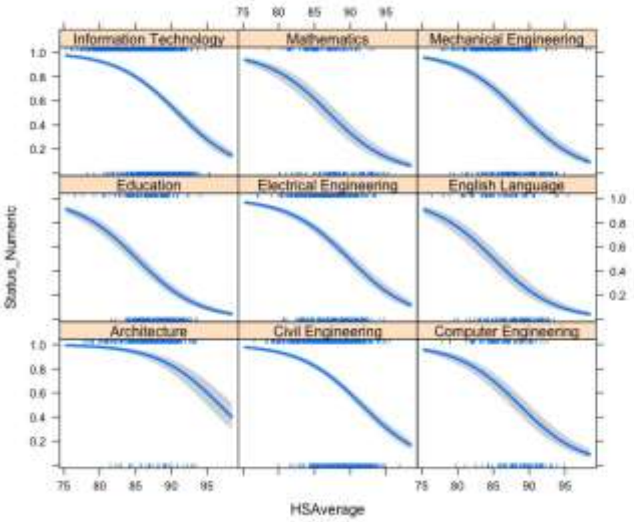

Fig. 8: The Visualization of Model 3

Figure 8 shows the visualization of Model 3. This was done by plotting the relationship between the independent variable high school GWA and the probability of transferring on the y-axis per course.

3.3 Model Comparison

The decision tree model with information gain splitting criterion and the binary logistic regression model with high school GWA and course as predictors are compared based on the performance measures presented in Table 7.

Table 7. Comparison between the Decision Tree and Binary Logistic Regression Models based on Accuracy, AUC, and Sensitivity

Model	Accuracy	AUC	Sensitivity
Decision Tree Model	0.7073	**0.7490**	**0.8827**
Binary Logistic Regression: Model 3	**0.7178**	0.7488	0.4209

Note: Values are derived based on the actual dataset. The highest values are in **boldface**.

It can be seen that, though Model 3 of the binary logistic regression model has a higher in-sample accuracy than the decision tree model, the latter exhibits a slightly higher AUC and significantly higher sensitivity values. This means that the decision tree model is better at correctly classifying observations as "transferred" than model 3. This supports many studies and literatures that suggest decision tree models are usually superior to binary logistic models. In this study, the decision tree model with information gain as the splitting criterion best describes the mobility of PSU students.

Table 8. Confusion Matrix of the Decision Tree Model

	true Graduated	true Transferred	class precision
pred. Graduated	148	86	63.25%
pred. Transferred	243	647	72.70%
class recall/sensitivity	37.85%	**88.27%**	

Table 8 above shows the performance of the Decision Tree Model with information gain splitting criterion using the testing dataset.

Fig. 9: The AUC (optimistic) graph of the Decision Tree Model

4 Conclusion

This paper investigates students' success at Pangasinan State University by identifying patterns and models that might be used to correctly classify and predict if a student will transfer or finish their studies. In this study, three categorical variables or attributes and one continuous variable were considered independent variables due to the availability of the data. These are the biological sex, hometown, course, and the high school general average. The objectives of this study were accomplished by applying decision tree models, and binary logistic regression models. The results from the binary logistic regression model with the high school general average and course as independent variables (Model 3), and the decision tree model with transition gain as a splitting criterion were fitted to the dataset to generate a model that possibly best describes the students' mobility in Pangasinan State University Urdaneta City Campus. The decision tree model is better than the binary logistic regression model based on accuracy, AUC, and sensitivity values. This implies that the decision tree model is better at correctly classifying observations as "Transferred" than Model 3. Thus, it was concluded that the decision tree model best described the mobility of the students using information gain as the splitting criterion. The decision tree model shows some significant findings. First, it can be concluded that there is a high chance that when a student is enrolled in an architecture program, the student will transfer regardless of biological sex, high school general average, or hometown. There is no splitting present under the architecture program when compared to other degree programs present in the model. One possible reason for this is that it might be the case that the independent variables considered as factors in modeling the students' mobility in this study are not applicable under the Architecture program. This means that other factors aside from those considered in this study, such as the number of absences or college GWA might be used instead. Another possible reason is that the degree program itself is very challenging because it requires strong analytical skills while focusing on detail with a huge academic workload and pressure. A testament to its difficulty is that the degree program requires 5 years of regular residency to finish and 2 years of apprenticeship in architectural firms or industry experience before taking the licensure. Second, the results from the decision tree model can be used as the basis for admission. It can be seen that for a student to have a high chance of survival enrolled in any of the following majors: civil engineering, computer engineering, education, electrical engineering, English language, information technology, mathematics, or mechanical engineering, they must have at least a high school general average of 92.035, 87.404, 84.800, 91.393, 84.779, 90.517, 85.705, or 90.082, respectively. Generally, a student must have at least a high school general average of 85.000 for non-engineering courses and at least 87.000 for engineering courses. Lastly, the results of this paper can be used for future research involving students' mobility, particularly in classifying whether a student will transfer based on other relevant attributes aside from biological sex, course, hometown, and high school general average.

References:

[1] Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17(1). doi:10.1186/s41239-020-0177-7.

[2] Priyadarshini, M., Gurnam, K., Sian Hoon, T., Geethanjali, N., & Yuen Fook, C. (2022). Key Factors Influencing Graduation on Time Among Postgraduate Students: A PLS-SEM Approach. *Asian Journal Of University Education*, *18*(1), 51-64. doi:10.24191/ajue.v18i1.17169.

[3] Wing, Michael D. (2008). *Student Transfer: The Effect of Timing on Academic Achievement*. *Electronic Theses and Dissertations*. 415. https://digitalcommons.library.umaine.edu/etd/415.

[4] Alexander, K.L., Entwisle, D.R., & Dauber, S.L. (1994). *Children in motion: Schools transfers and elementary school performance*. Paper presented at the American Sociological Association Annual Meeting, Los Angeles, CA.

[5] Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

[6] Grais, Benjamin M. (2011). *High School Transfer Student Transitions and Changes: Risk, Success, Failure, and the Vital Role of the Counseling Curriculum*. *Dissertations*. 66. https://ecommons.luc.edu/luc_diss/66

[7] Allen, J., Robbins, S. B., Casillas, A., & Oh, I.-S. (2008). *Third-year College Retention and Transfer: Effects of Academic Performance, Motivation, and Social Connectedness. Research in Higher Education, 49(7), 647–664.* doi:10.1007/s11162-008-9098-3

[8] Colobong, R. & Cenas, P. (2011). *Student Mobility in Pangasinan State University*. Pangasinan State University Urdaneta City Campus Research Journal s. 2011-2012. pp. 13-20

[9] Dela Cruz, R. O. (2015). Persistence and retention towards degree xcompletion of BS agriculture students in selected State Universities in Region IV-A, Philippines. African Journal of Agricultural Research, 10*9130*: pp. 1543 – 1556 Retrieved from: https://academicjournals.org/journal/AJAR/article-full-text-pdf/6AA4CD351984. Retrieved on: June 7, 2022.

[10] Kerbow, D. (1996). *Patterns of Urban Student Mobility and Local School Reform. Journal of Education for Students Placed at Risk (JESPAR), 1(2), 147–169.* doi:10.1207/s15327671espr0102_5

[11] Hazra, A. (2019). *Top 7 Cross-Validation Techniques with Python Code*, Analytics Vidhay, Retrieved : https://www.analyticsvidhya.com/blog/2021/11/top-7-cross-validation-techniques-with-python-code/