Test Turnitin

[Budi Wibawanta] Final Paper IJORER

Test Turnitin Part 1 (Moodle TT)

Test Moodle Plugins 4.5 (Moodle TT)

Universitas Pelita Harapan

Document Details

Submission ID

trn:oid:::1:3346891012

Submission Date

Sep 22, 2025, 8:04 AM GMT+7

Download Date

Sep 22, 2025, 8:25 AM GMT+7

File Name

32463_Test_Turnitin__Budi_Wibawanta__Final_Paper_IJORER_165597_1251556123.docx

File Size

181.0 KB

16 Pages

6,391 Words

39,887 Characters

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

42 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

15 Missing Quotations 2%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

8% Internet sources
5% Publications
4% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **42** Not Cited or Quoted 9%
  Matches with neither in-text citation nor quotation marks
- **15** Missing Quotations 2%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

8% Internet sources
5% Publications
4% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
Universitas Negeri Semarang - iTh		2%
2	Internet	
www.mdpi.com		<1%
3	Internet	
pmc.ncbi.nlm.nih.gov		<1%
4	Student papers	
Universitas Negeri Surabaya		<1%
5	Publication	
Lotte Dyhrberg O'Neill, Tine Nielsen. "Admission testing, pre-academic exam self-...		<1%
6	Internet	
robots.net		<1%
7	Publication	
Prio Wibisono, Akhil Deepak Vatvani, Kalis Waren, Karunia Valeriani Japar, Vivien ...		<1%
8	Student papers	
Dublin Business School		<1%
9	Publication	
Mojtaba Taheri, Fathollah Pourfayaz, Sara Hemmati. "A highly accurate model for...		<1%
10	Internet	
www.frontiersin.org		<1%

11	Internet	
bilselkongreleri.com		<1%
12	Internet	
fastercapital.com		<1%
13	Internet	
ojs.jssr.org.pk		<1%
14	Internet	
www.researchgate.net		<1%
15	Internet	
eduq.info		<1%
16	Internet	
link.springer.com		<1%
17	Internet	
gcris.ktun.edu.tr		<1%
18	Internet	
journal.ipm2kpe.or.id		<1%
19	Publication	
María José López, Maria Veronica Santelices, Carmen Maura Taveras. "Academic p...		<1%
20	Internet	
akupintar.id		<1%
21	Internet	
hdl.handle.net		<1%
22	Student papers	
Corvinus University of Budapest		<1%
23	Student papers	
Radboud Universiteit		<1%
24	Publication	
Ruqiang Yan, Jing Lin. "Equipment Intelligent Operation and Maintenance", CRC P...		<1%

25 Internet
informatticz.blogspot.com <1%

26 Internet
www2.skolverket.se <1%

27 Internet
sae.tu.kielce.pl <1%

28 Publication
Pala Ravikanth, T. Jothi Saravanan, K.I. Syed Ahmed Kabeer. "Supervised data-dri... <1%

29 Publication
Paulo Flavio Faia Nogueira, Cleber Gustavo Dias, Fabio Henrique Pereira, Francisc... <1%

30 Publication
Vardan Barsegyan, Ineke Maas. "First-generation students' educational outcome... <1%

31 Internet
core.ac.uk <1%

32 Internet
e-journal.undikma.ac.id <1%

33 Internet
ijer.inased.org <1%

34 Internet
jurnal.faperta-unras.ac.id <1%

35 Internet
ojs.unik-kediri.ac.id <1%

36 Internet
open.library.ubc.ca <1%

37 Internet
scik.org <1%

38 Internet
digitalcommons.liberty.edu <1%

Type your title in here (8 pt)

Predicting First-Year Academic Success: The Influence of Socioeconomic and Educational Backgrounds among First-Generation Scholarship Students

Budi Wibawanta[1*], Evander Banjarnahor[2], Stephanie Riady[3], Dorita Setiawan[4], HG Retno Harsanti[5]
[1]Department of Social Science Education, Faculty of Education, Universitas Pelita Harapan, Tangerang, Indonesia
[2]Department of Mathematics, Faculty of Science and Technology, Universitas Pelita Harapan, Tangerang, Indonesia
[3]Department of Educational Technology, Faculty Graduate School of Education, Universitas Pelita Harapan,
[4]Rektorat, Universitas Pelita Harapan, Tangerang, Indonesia
[5]Department of Education, National Taiwan Normal University, Taipei, Taiwan

Check for updates OPEN ACCESS

DOI : https://doi.org/10.46245/ijorer

Sections Info

Article history:
Submitted: November 23, 2022
Final Revised: January 11, 2023
Accepted: January 16, 2023
Published: January 31, 2023

Keywords:
Academic Success
Educational Background
First Generation
Linear Regression
Socioeconomic

ABSTRACT (9 pt)

Objective: *This study examines the factors influencing the academic success of first-year students who are first-generation scholarship recipients (FGCS) at Universitas X, with a focus on the impact of socioeconomic background, parental education, and entrance exam scores. Uniquely, this study combines predictive modeling with a detailed analysis of regional, school, and familial factors, providing a comprehensive early-warning system to identify at-risk students and guide targeted interventions. Using a multivariate quantitative approach, data from 511 FGCS recipients from the 2021–2024 cohorts were analyzed. Descriptive statistics reveal significant disparities in academic performance across regions, school types, and study programs. Students from Java and Sumatra achieved the highest average GPAs (3.29 and 3.21, respectively), with state school graduates performing particularly well (average GPA: 3.24). Conversely, students from private schools in Papua and Maluku faced significant academic challenges, with GPAs below the average (2.55 and 2.94). Multiple linear regression modeling, which includes entrance exam scores, predicts first-year GPA with a 6.08% error margin, providing a reliable tool for early detection of academic risk. The regression model shows that entrance exam scores in Bahasa Indonesia and TPA, as well as regional background, significantly influence academic performance. Specifically, students from Sulawesi (β = 0.203) and Java (β = 0.175) are predicted to achieve the highest GPAs, followed by students from Nusa Tenggara (β = 0.118), Sumatra (β = 0.114), Kalimantan (β = 0.101), Papua (β = 0.068), and Maluku (β = 0.059). These results indicate clear regional disparities in predicted academic outcomes. The study concludes with evidence-based policy recommendations for targeted academic support programs and interventions to enhance the academic success of FGCS, particularly those from underprivileged regions and schools.*

INTRODUCTION

First-generation students (FGS) are those whose parents do not hold a higher education degree or a university degree (Canning et al., 2024). According to recent studies, first-generation students tend to perform at lower levels compared to continuing-generation students, considering other variables such as learning outcomes, intellectual development, study skills, integration, and engagement with the academic environment (López et al., 2023). They face significant academic and non-academic challenges, including limited social capital, academic preparedness, and inadequate family support, which negatively impact their GPA and educational progress (Ives & Castillo-Montoya, 2020). However, at the same time, FGS are also known to possess positive characteristics, such as innovative thinking and perseverance (Kracht et al., 2025). These findings underscore the importance of supporting first-generation students, particularly those from low-income backgrounds, to strengthen their social capital, especially during their crucial first year (Sudbrock et al., 2024). Early detection of at-risk students, along with preventive measures, can enhance academic success (Alyahyan & Düştegör, 2020).

Empirical evidence from developing countries indicates that multiple factors contribute to FGS academic performance. Entrance exam scores in subjects such as mathematics and academic potential tests, parental education levels (particularly maternal), and access to learning resources significantly influence first-semester GPA (Munir et al., 2023). Studies also show that socioeconomic background affects educational achievement, with students from higher socioeconomic groups generally performing better. However, this effect is often mediated by family involvement, school quality, and students' capacity to adapt to the academic environment.In systematic reviews in the US and Europe, it was found that a family support-based approach can serve as an asset-based model rather than a deficit-based model to support FGCS (LeBouef & Dworkin, 2021). This approach supports the role of social and cultural capital, expands the perspective on the value of FGCS student experiences, and increases academic engagement (Barsegyan & Maas, 2024). High parental cultural background will significantly improve students' academic performance (Juhaidi, 2023), whereas parental economic background does not explain the academic achievement differences between FGS and CGS (Ives & Castillo-Montoya, 2020).

A 9-year panel study reveals that family social and cultural capital have a significant impact on the academic persistence and dropout rates of FGCS students (Barsegyan & Maas, 2024a). Family support for first-generation students is reflected in their experiences during the parental role transition, the barriers they face to success, and the knowledge systems they utilize (Spencer & Goldstein, 2025). Other findings indicate that FGCS rely on family wealth, aspirations, navigation, and resilience as sources of motivation, drive, and pride (Pancsofar et al., 2024). The findings show how the need for work, academic fatigue, and competing responsibilities present challenges, as well as how connections to culture and nature help reduce stress (Hopper et al., 2024). Previous research by Weisen et al., (2024) suggests that FGCS are at higher risk of underperforming, particularly in the Faculty of Education.

International scholarship has further underscored the role of family and institutional support in shaping FGS outcomes. Systematic reviews in the United States and Europe emphasize that family support can be understood through an asset-based rather than deficit-oriented lens, with social and cultural capital functioning as critical resources for persistence, motivation, and engagement (LeBouef & Dworkin, 2021; Barsegyan & Maas, 2024; Spencer & Goldstein, 2025; Pancsofar et al., 2024). Simultaneously, institutional interventions—such as mentoring, learning communities, and difference-education programs—have proven effective in improving GPA and retention among first-year FGS (Pressimone Beckowski, 2025). Together, these studies underscore the complex interplay of socioeconomic, cultural, and institutional factors in shaping the educational success of first-generation students.

Despite growing international research, there has been limited comprehensive analysis in Indonesia that simultaneously incorporates socioeconomic variables (occupation, income), parental education, school background, and institutional measures such as entrance exam scores in predicting the academic success of first-generation scholarship students (FGCS). Addressing this gap, the present study employs a quantitative multivariate approach using data from 511 FGCS across four cohorts (2021–2024). The novelty of this study lies in its development of a predictive regression model to examine the determinants of first-year GPA and to provide an evidence-based tool for early detection of academic risk. Beyond contributing to the academic literature, the findings

are expected to inform institutional policies and targeted interventions aimed at supporting FGCS from disadvantaged backgrounds

RESEARCH METHOD

Research Data

This study utilizes quantitative data from first-generation scholarship recipients at the Faculty of Education, Universitas X, who belong to the 2021–2024 cohorts. The study comprises 511 participants representing various educational programs. Data were collected to examine the distribution of students based on gender, region of origin, type of secondary school, and selected study program. These demographic characteristics serve as the basis for further analysis of academic achievement (first-year GPA) and the influence of educational background and geography on student performance. The dataset supporting this study is publicly accessible at go.uph.edu/datasetfirstgen

Research Method

The research design employed is a quantitative causal-comparative design with an ex post facto approach. This design was selected because the variables under investigation namely, socioeconomic background, educational background, and entrance exam scores are historical data that the researcher cannot manipulate. The goal is to identify and test the simultaneous effects of several independent variables on the first-year GPA of students classified as First-Generation College Students (FGCS). The design is non-experimental, meaning the researchers do not provide any treatment or manipulation of the subjects. The study employs multivariate statistical analysis, specifically multiple linear regression, to determine the contribution of each independent variable to the dependent variable (GPA). Data analysis is conducted simultaneously, meaning that all independent variables are tested for their effects within a single regression model, both as a whole and in partial effects. Figure 1 illustrates the proposed research design's flow.

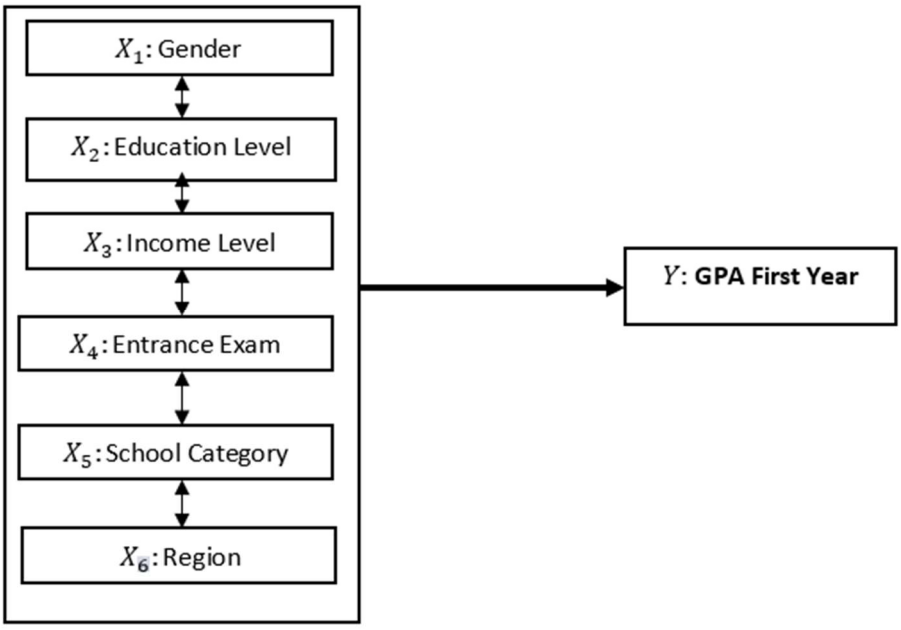

Figure 1. Design of Research

The research design is expected to provide a strong data-driven understanding of the determinants of FGCS students' academic persistence and inform the formulation of more equitable and evidence-based academic and policy interventions.

Linear Regression

This study applies linear regression algorithms to predict the academic performance of first-generation scholarship students (FGCS) in their first year of university. Regression analysis is an essential statistical inference method (He, 2025). Linear regression utilizes the linear correlation between independent variables and the target variable, enabling strong predictions, particularly when the data is linearly distributed (Iwasaki, 2020). It involves training processes such as Mean Squared Error (MSE) to find the best-fitting line that minimizes errors (Oyedeji et al., 2020). This model is well-suited for predicting continuous outcomes, such as student academic performance/GPA.

The algorithm used is multiple linear regression (MLR). MLR is more effective than simple linear regression (SLR) in handling data with various factors (Arifuddin et al., 2025). MLR enables the analysis of relationships between multiple independent variables and a single dependent variable, making it an essential method in data mining for identifying potential variables with linear relationships to the target (Howe Eng et al., 2023) (Rispandi, 2023). The general equation for multiple linear regression is expressed as follows (Phan et al., 2024).

$$y = m_1X_1 + m_2X_2 + m_3X_3 + \cdots + c \quad (1)$$

In this study, a linear regression model will be used with a machine learning approach, where the data will be divided into two parts: training data and testing data. Machine learning can be used to predict student outcomes, tailor learning experiences, and automate various administrative tasks in educational settings (Banjarnahor, Sibarani, et al., 2025). To evaluate the model's performance, commonly used evaluation metrics for measuring the performance of regression models include the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) (Ramirez-Montoya et al., 2024). The use of these metrics provides a more comprehensive view of how well the model performs in making predictions.

Additionally, to avoid overfitting and ensure the model's validity, it is simulated using the k-fold cross-validation method (Fitrianah et al., 2021). In this procedure, the dataset is partitioned into k approximately equal folds; at each iteration, the model is trained on $k-1$ folds and evaluated on the remaining fold, and performance is averaged across all folds (Banjarnahor, Belferik, et al., 2025; Farhood et al., 2024). Figure 2 illustrates the research workflow.

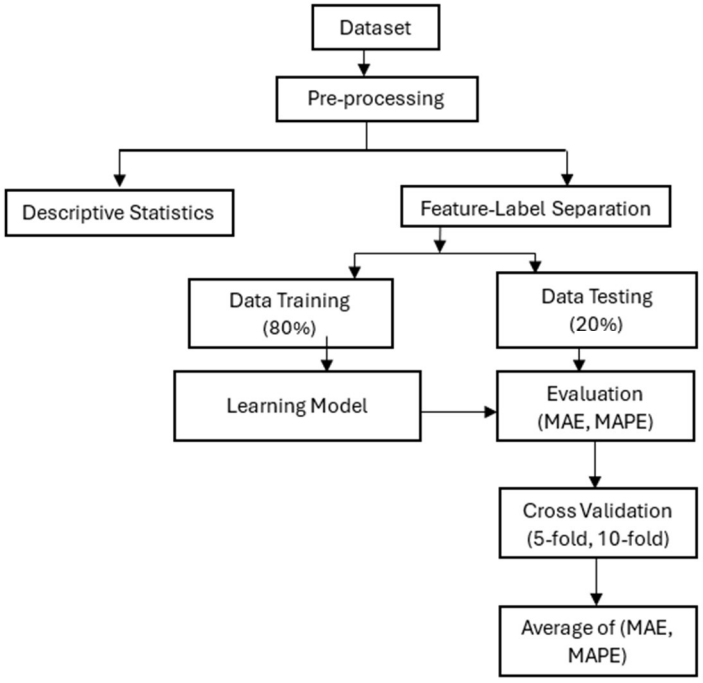

Figure 2. Flow of Research

RESULTS AND DISCUSSION

Results

Table 1 presents the demographic profile, showing that the majority of students are female (81.8%), with the largest group from Sumatra (260 students), followed by Java (103 students). The most common school category is public schools (337 students), and the study program with the highest enrollment is the Indonesian Language Education program (178 students). This diversity in regional origin and prior educational pathways provides essential context for evaluating equity of access and academic success among first-generation students in higher education.

Table 1. Demographic Data of First-Generation Students by Cohort

	Cohort				
	2021 (N=89)	**2022 (N=136)**	**2023 (N=142)**	**2024 (N=144)**	**Total (N=511)**
Biologicalsex					
Laki-laki	17	30	27	29	103
Perempuan	72	106	115	115	408
Region					
Bali	1	1	2		4
Jawa	28	23	29	23	103
Kalimantan	2	5	8	8	23
Maluku	4	2	4	4	14
Nusa Tenggara	8	12	11	15	46
Papua	1	1		5	7
Sulawesi	8	18	14	14	54
Sumatera	37	74	74	75	260

Type your title in here (8 pt)

Scholl Origin Category					
Sekolah Katolik	2	4	4	1	11
Sekolah Kristen	19	36	28	34	117
Sekolah Negeri	58	85	96	98	337
Sekolah Swasta Umum	10	11	14	11	46
Study Program					
Pend. Agama Kristen	9	14	9	13	45
Pend. Bahasa Indonesia	4	22	8	12	46
Pend. Bahasa Inggris			7	7	14
Pend. Biologi	8	9	18	11	46
Pend. Ekonomi	11	11			22
Pend. Fisika	4	6			10
Pend. IPS	9	9	19	17	54
Pend. Kimia	5	8			13
Pend. Matematika	8	7	15	20	50
PGSD_BI	25	45	58	50	178
PGSD_EN	6	5	8	14	33

Table 2 presents descriptive statistics in the form of the average first-year Grade Point Average (GPA) of First-Generation scholarship recipients at the Faculty of Education, grouped by region of origin, school category, and study program. This statistic aims to provide an initial overview of the distribution of academic achievement among students based on their geographical and institutional backgrounds. Through this analysis, common patterns can be identified that show differences in academic performance across groups, which will serve as the basis for further inferential analysis.

Table 2. Distribution of GPA by Region, School Category, and Study Program

	Sumatera (N= 260)	Jawa (N= 103)	Bali (N= 4)	Kalimantan (N= 23)	Sulawesi (N= 54)	Maluku (N= 14)	Nusa Tenggara (N= 46)	Papua (N= 7)	Grand Total
School Category									
Sekolah Katolik	3,14	3,28							3,20
Sekolah Kristen	3,20	3,28		3,18	3,17	3,15	3,23	3,15	3,20
Sekolah Negeri	3,21	3,32	3,15	3,24	3,29	3,18	3,19	3,19	3,24
Sekolah Swasta Umum	3,25	3,21	3,09	3,18	3,47	2,94	3,10	2,55	3,22
Grand Total	3,21	3,29	3,13	3,21	3,26	3,15	3,20	3,08	3,23
Study Program									
Pend. Agama Kristen	3,19	3,20		3,26	3,29	3,23	3,24	2,55	3,21
Pend. Bahasa Indonesia	3,24	3,31		3,45	3,45		3,32	3,22	3,29
Pend. Bahasa Inggris	3,31	3,23		3,36			3,06		3,26
Pend. Biologi	3,20	3,26		3,12	3,21		3,18	2,83	3,20
Pend. Ekonomi	3,37	3,60	3,16		3,48	3,25	3,61		3,43
Pend. Fisika	3,37	3,26			3,71	3,26	3,29		3,36
Pend. IPS	3,27	3,29	3,14	3,03	3,30	3,20	3,12	3,30	3,24

Type your title in here (8 pt)

Pend. Kimia	3,28	3,31				2,94			3,26
Pend. Matematika	3,16	3,39	3,09	3,28	3,25	3,25	3,31	3,16	3,24
PGSD_BI	3,15	3,26		3,25	3,10	3,07	3,15		3,17
PGSD_EN	3,43	3,27		2,98	3,35	2,99	3,13	3,26	3,28
Grand Total	3,13	3,29	3,21	3,15	3,20	3,08	3,26	3,21	3,23
Father Income Level									
Kelas Atas	3,22	3,23		3,25	3,32				3,24
Kelas Menengah	3,24	3,30		3,24	3,41	3,24	3,28	3,24	3,28
Menuju Kelas Menengah	3,24	3,36	3,13	3,23	3,29	3,10	3,03	2,69	3,24
Rentan Miskin	3,13	3,26	3,14	3,39	3,25	3,46	3,31		3,19
Miskin	3,17	3,20		3,15	3,17	2,94	3,19		3,17
Tidak Berpenghasilan	3,28	3,25		2,99	3,18	3,10	3,15		3,24
Grand Total	3,21	3,29	3,13	3,21	3,26	3,15	3,20	3,08	3,23
Mother Income Level									
Kelas Atas	2,99	3,14			3,64				3,31
Kelas Menengah	3,32	3,33		3,18	3,22	3,25	3,09		3,29
Menuju Kelas Menengah	3,24	3,27	3,13	3,24	3,11		3,02		3,21
Rentan Miskin	3,23	3,23	2,94	3,27	3,34	3,26	3,10		3,23
Miskin	3,14	3,16		3,19	3,56	3,14	3,18		3,15
Tidak Berpenghasilan	3,21	3,32	3,34	3,21	3,25	3,14	3,23	3,08	3,24
Grand Total	3,21	3,29	3,13	3,21	3,26	3,15	3,20	3,08	3,23

Based on data from 511 First-Generation scholarship recipients from the 2021–2024 cohorts, there are significant differences in first-year academic performance across regions, school categories, and study programs. In general, students from Java and Sumatra recorded the highest average GPAs (3.29 and 3.21, respectively), with a dominant contribution from graduates of public schools (national average GPA: 3.24). The public school category consistently showed the best academic performance across nearly all regions. In contrast, private schools exhibited fluctuating results very high in Sulawesi (3.47), but very low in Papua (2.55) and Maluku (2.94). High performance in Java is also reflected in study programs such as Economics Education (3.60), Mathematics (3.39), and Indonesian Language Education (3.31), indicating better academic preparedness and possibly the influence of more equitable secondary education quality in the region.

Conversely, the Papua and Maluku regions face greater academic challenges, both in terms of average GPA and performance across various study programs. In Papua, for example, students in the Christian Religious Education and Biology programs recorded GPAs of 2.55 and 2.83, respectively, well below the national average. This highlights structural gaps in academic preparedness carried over from prior educational levels, especially among graduates of private schools in the 3T (underdeveloped) regions. These findings underscore the importance of data-driven academic intervention strategies such as matriculation programs, early academic mentoring, and basic skills reinforcement specifically designed for students from more vulnerable educational and geographical backgrounds.

Analysis based on father's income indicates that the "transitioning to middle class" group in the Papua region recorded the lowest first-year GPA nationally (2.69), suggesting that lower to middle economic status in 3T regions does not necessarily

guarantee academic readiness. Conversely, in Maluku, the group with fathers in the "at-risk of poverty" category showed the highest GPA (3.46), which may reflect a compensatory effect or strong motivation from families with economic limitations but a strong educational aspiration. On the mother's side, the "upper class" group showed lower GPAs than the national average, as seen in Sumatra (2.99) and Bali (2.94) in the "at-risk of poverty" income group. This indicates that high economic status from the mother does not always correlate with academic achievement, which may be influenced by factors such as active involvement in learning support or the family's educational value orientation. Further analysis also compares the educational level and occupation of parents with the first-year GPA of First Generation students. Table 3 shows a comparison of student GPAs based on the educational level and occupation of the parents.

Table 3. Comparison of GPA Based on Parental Occupation and Education Level

	Ocupation					
	Pegawai pemerintah/swasta	**Pendeta**	**Petani, buruh, nelayan**	**Tidak Bekerja**	**Wirausaha**	**Grand Total**
Father Edu Level						
SD/Sederajat		2,86	3,15	3,29	3,15	3,16
SMA/SMK/Sederajat	3,29	3,13	3,25	3,23	3,24	3,25
SMP/Sederajat	3,07	3,00	3,16	3,31	3,37	3,19
Tidak Sekolah			3,21	3,13		3,17
Grand Total	3,28	3,05	3,21	3,24	3,24	3,23
Mother EduLevel						
SD/Sederajat	3,63	3,00	3,19	3,21	3,00	3,20
SMA/SMK/Sederajat	3,28	3,17	3,20	3,26	3,32	3,25
SMP/Sederajat	3,32		3,22	3,20	3,10	3,20
Tidak Sekolah			3,03	2,95	3,11	3,00
Grand Total	3,32	3,11	3,20	3,24	3,26	3,23

The descriptive analysis of the socio-economic aspects shows that First-Generation students with parents who have lower educational backgrounds and occupations tend to have first-year GPAs below the average, with the father's education level being no schooling (3.17) and primary school (3.16). Specifically, students with fathers who are pastors and have only completed primary school recorded the lowest GPA, at 2.86. Additionally, students whose mothers do not work and have never received formal education also showed lower GPAs, ranging from 2.95 to 3.00, which is lower than that of other groups.

These findings suggest a significant impact of parental education and socioeconomic status on the academic preparedness and performance of students. Students with parents who have higher education or more stable occupations (such as civil servants or entrepreneurs) tend to show higher and more consistent GPAs. Therefore, students from the most socio-economically vulnerable backgrounds may require additional support, in the form of academic mentoring, psychosocial guidance, and affirmative interventions, to ensure they can thrive in higher education environments.

Predictive Analysis

In this study, a predictive modeling simulation was performed using a machine learning approach to predict the academic performance of First-Generation scholarship recipients, specifically the first-year GPA at the Faculty of Education. The model used is

linear regression, a supervised learning method commonly applied in numerical predictive analysis. Before the model training process, encoding was performed on categorical variables such as gender, parental education level (father and mother), income level, school category, and region of origin, so that they could be interpreted numerically in the regression model.

The dataset was divided into two parts: 80% for training and 20% for testing. After the model training process, the final regression coefficients were obtained, as shown in Table 4 below:

Table 4. Regression Coefficients for Each Variable

Category	Feature	Coefficient
Intercept	intercept	2.523.257
Gender	gender	-0.101532
Education Level	father_edu_level	0.003530
	mother_edu_level	-0.009798
Income Level	father_income_level	-0.006533
	mother_income_level	-0.000007
Entrance exam	math	0.000090
	ing	0.001761
	ind	0.0059
	tpa	0.002363
School Category	school_category_Sekolah Kristen	0.070847
	school_category_Sekolah Negeri	0.086010
	school_category_Sekolah Swasta	0.096788
	Umum	
Region	region_Jawa	0.175396
	region_Kalimantan	0.100515
	region_Maluku	0.058526
	region_Nusa Tenggara	0.118260
	region_Papua	0.068014
	region_Sulawesi	0.202953
	region_Sumatera	0.114351

Based on the training results, the linear regression equation for predicting the first-year GPA of First-Generation students at the Faculty of Education can be formulated as follows:

$$GPA\ First\ Year = -0.1015 * gender + 0.0035 * father\_edu\_level - 0.009 * mother\_edu\_level + \cdots + 0.1143\ region\_sumatera + 2.523$$

This equation indicates that the initial GPA prediction value is 2.523, which is then adjusted based on the values of each predictor feature. The regression coefficients provide insight into the direction and strength of the influence of each variable on the GPA prediction. For instance, the Gender variable, with a negative coefficient (-0.102), suggests that specific categories (e.g., male = 1) tend to have a lower GPA compared to the reference group (e.g., female = 0). Similarly, the father's education level has a positive contribution to the GPA prediction (β = 0.0035). In contrast, the mother's education level shows a negative direction (β = -0.0098), which may indicate that the mother's education effect in this model is complex and possibly influenced by other uncontrolled variables.

The feature with the highest positive coefficient is Region_Sulawesi (β = 0.2029), followed by Region_Jawa (β = 0.1754), indicating that students from these regions are predicted to have a higher GPA compared to the reference group (likely Bali or the first

category in one-hot encoding). Additionally, the school category also contributes positively to the GPA prediction, with students from private general schools (β = 0.0968) and public schools (β = 0.0860) having higher predicted GPAs compared to the reference group (likely Catholic schools). Meanwhile, entrance exam variables such as Bahasa Indonesia (β = 0.0059) and TPA (β = 0.0024) also provide positive contributions, though relatively small. Overall, this model indicates that a combination of demographic factors, region of origin, school category, and entrance exam results can be used to predict first-year GPA outcomes. Although this model is linear and straightforward, its results provide an initial foundation for developing an early academic risk detection system, which can assist in designing more targeted academic support strategies and interventions for First-Generation students.

After the training process is complete, the model's performance is evaluated using the testing set to assess how accurately the linear regression model predicts the GPA of first-generation students. Table 5 presents the evaluation of the simulated linear regression model.

Table 5. Performance Results of the Linear Regression Model

MAE	0.189902686572911
MAPE	0.06082533556860566

The initial evaluation results show that the linear regression model produced a Mean Absolute Error (MAE) of 0.1899 and a Mean Absolute Percentage Error (MAPE) of 0.0608 or 6.08%. The MAE indicates that, on average, the model's predictions deviate by approximately 0.19 points from the actual GPA values. At the same time, the MAPE shows that the prediction error relative to the actual GPA is approximately 6%. In the context of higher education, this level of prediction error is considered relatively low and still within practical tolerance limits, particularly for mapping academic risks and determining early interventions for students.

To test the model's validity and stability, and to minimize bias and mitigate the risk of overfitting due to the limited sample size (N = 511), cross-validation testing was conducted using two data splitting schemes: 5-fold and 10-fold cross-validation. The performance evaluation results of the linear regression model with cross-validation are presented in Table 6.

Table 6. Performance Results of the Linear Regression Model with Cross-Validation

	5-fold	**10-fold**
MAE	0.190209	0.190894
MAPE	0.060175 (6,01%)	0.060388 (6,038%)

Table 6 shows the stability of the linear regression model's performance with cross-validation, with MAE values of 0.1902 (5-fold) and 0.1909 (10-fold), and MAPE values of 6.01% and 6.038%, respectively. The consistency of the error values in both cross-validation schemes suggests that the linear regression model possesses good generalization capabilities for new data and can serve as a baseline for the development of more complex predictive models in subsequent stages.

Discussion

Regional Background

The study's results indicate that first-generation students from the Java and Sulawesi regions have the highest predicted GPAs, while students from Papua and Maluku tend to have lower GPAs. These findings suggest that regional origin plays a significant role in the academic readiness and achievement of first-generation students during their first year of college. Previous research has highlighted that students from rural and underdeveloped areas frequently encounter challenges in transitioning to higher education due to a lack of both academic and social capital. More urbanized areas typically offer higher-quality secondary education, more experienced teachers, and better learning facilities, which enhance students' academic readiness (Hodge, 2022).

Furthermore, the study shows that students from marginalized regions experience more academic stress and a lack of social support, which directly affects their educational engagement and performance (Castillo-Montoya & Ives, 2021) (Conefrey, 2021). This aligns with systematic studies that found limited social networks in rural communities impact the academic adaptation process of first-generation students (FGS).

However, some research has found that geographical influence does not have a direct effect. If the quality of schools and entrance exam scores are controlled for, the regional influence on GPA can significantly decrease (Barsegyan & Maas, 2024b) This suggests that regions often serve as a proxy for other systemic inequalities, such as disparities in the quality of education in schools. From the school category perspective, students from public schools consistently recorded higher GPAs. This is likely because the national curriculum and readiness for entrance exams are more standardized in public schools compared to small or religiously based private schools in 3T areas. Triwahyuni et al. (2021) and Roy et al. (2025) show that school background influences students' academic literacy and academic expectations.

Parental Background & Socioeconomic Status

The study found that the students with highly educated parents or stable occupations (such as employees or entrepreneurs) tend to have higher GPAs. Conversely, students from families with lower education levels or informal occupations (such as farmers, laborers, or pastors) tend to have lower GPAs. These findings strengthen the concept of parental educational capital proposed by Bourdieu, which suggests that parents' education serves as a source of cultural capital that influences children's aspirations and learning patterns (Khan Eusafzai, 2024). Regression studies examining how socio-economic status predicts academic outcomes, while controlling for factors such as family involvement and school resources, demonstrate that socio-economic status has a significant influence on academic performance. Students with higher socio-economic status tend to perform better academically. However, parental involvement and school resources can buffer the relationship between socio-economic status and educational achievement (Munir et al., 2023b). Learning communities and first-year mentoring programs that facilitate relationships between students and faculty, along with essential academic support, enhance student success in higher education (Mulyadin & Rahyasih, 2021).

Entrance Exam & Academic Preparedness

The regression results indicate that entrance exam scores in Bahasa Indonesia, Academic Potential Test, and Mathematics have a positive impact on the first-year GPA of First-Generation students. Although the coefficients are small, these findings are consistent with the academic preparedness theory, which states that entrance exam scores reflect fundamental cognitive competencies crucial for success in higher education (Payne et al., 2024).

The impact of entrance exams on student retention in higher education is not yet well understood, but several studies focusing on this topic have emerged in recent decades (Nielsen et al., 2017) (Nielsen et al., 2022).Studies show that pre-entry academic scores, such as the SAT/ACT in the U.S. or the TPA in Indonesia, contribute to predicting academic success, but only when combined with other psychological variables, including grit, self-efficacy, and social support (Weisen et al., 2024b). Entrance exams are designed to test traditional academic attributes, such as subject knowledge, text comprehension, or reasoning skills. On the other hand, there is considerable criticism of the reliability of entrance exams as the sole indicator of readiness (Tierney & Sablan, 2014). Mentoring-based interventions and early learning experiences have a more significant impact on academic retention compared to entrance exam scores alone (Watts et al., 2023).

Beyond Descriptive Differences: Key Findings

The findings of this study reveal deeper structural mechanisms underlying the academic outcomes of first-generation scholarship students. Regional disparities, for instance, appear to significantly shape student performance. First-generation students from Java and Sulawesi consistently outperform those from Papua and Maluku, reflecting not only geographic variation but also systemic inequalities in access to quality education. Students from more developed regions benefit from standardized curricula, better teaching resources, and enriched academic environments, which collectively enhance preparedness for higher education. Conversely, students from 3T (underdeveloped, frontier, and outermost) areas often enter university with cumulative disadvantages, making regional origin a proxy for broader educational inequities.

School category further contributes to these differences. Graduates of public schools demonstrate stronger academic readiness, likely due to nationally standardized curricula and examination systems that provide more consistent academic foundations. In contrast, students from small private or religious schools in under-resourced areas face challenges in bridging academic gaps, suggesting the need for interventions such as bridging programs, remedial courses, or early mentoring, particularly for students from eastern Indonesia.

Socioeconomic status and parental education play a critical role in shaping academic outcomes. The transmission of cultural capital from parents to children, including academic aspirations, values, and learning behaviors, explains why students from families with higher education levels or stable occupations tend to perform better. Students from families with lower educational attainment or informal occupations often lack these resources, which can weaken academic persistence. Interestingly, maternal education exhibited a complex, sometimes negative association with GPA, indicating that education alone does not guarantee active learning support. This highlights the importance of exploring family dynamics and parental involvement more deeply.

Entrance exam scores, particularly in Bahasa Indonesia and the Academic Potential Test (TPA), positively predict first-year GPA, although the coefficients are modest. These results suggest that while standardized exams reflect baseline cognitive competencies, they capture only a portion of the skills required for success in higher education. Consequently, universities should consider entrance exam scores as complementary indicators, integrating them with psychosocial measures such as self-efficacy, resilience, and social support.

Taken together, these findings emphasize that supporting first-generation scholarship students requires multidimensional strategies that address structural, institutional, and familial factors. Regional-based matriculation programs, school-targeted interventions for under-resourced graduates, and family engagement initiatives are essential for promoting academic equity. The predictive model developed in this study, with low error margins, provides an evidence-based tool for early identification of at-risk students, offering both theoretical insights and practical guidance for universities aiming to enhance persistence and success among first-generation students.

CONCLUSION

This study offers valuable insights into the factors influencing the academic success of first-generation scholarship students (FGCS) in their first year at Universitas X. The analysis reveals significant disparities in educational performance based on region, school type, and socioeconomic factors. Students from Java and Sumatra had the highest average GPAs, 3.29 and 3.21, respectively, while students from private schools in Papua and Maluku faced significant academic challenges, with GPAs of 2.55 and 2.94, respectively. State school graduates from Java and Sumatra consistently performed better, with an average GPA of 3.24. These findings highlight the clear academic readiness gap among students from different regions and school backgrounds.

Socioeconomic factors, including parental education and income levels, were also found to have a significant impact on academic success. Students with parents who had higher education levels or stable employment were more likely to achieve higher GPAs. Conversely, students whose parents had lower education levels or worked in informal sectors, such as farming or religious professions, tended to have lower GPAs. This highlights the significance of family support and social capital in promoting the academic success of first-generation students.

In predictive analysis using multiple linear regression, it was found that variables such as entrance exam scores in Bahasa Indonesia and TPA, as well as regional background, play an essential role in predicting first-year GPA. The regression model revealed that students from Sulawesi ($\beta = 0.2029$) and Java ($\beta = 0.1754$) are expected to perform better than students from other regions. The model also indicated that entrance exam scores, particularly in Bahasa Indonesia and TPA, made a positive contribution to the prediction of GPA. With a Mean Absolute Error (MAE) of 0.1899 and a Mean Absolute Percentage Error (MAPE) of 6.08%, the model demonstrates reliability in early detection of academic risk. Furthermore, cross-validation tests with 5-fold and 10-fold showed good stability, with MAPE values of 6.01% and 6.038%, respectively. These findings support the importance of using predictive models to design more targeted academic interventions for FGCS.

However, this study has several limitations. The regression model primarily considers cognitive and demographic variables, without accounting for non-cognitive factors such as grit, self-efficacy, motivation, or social support, which may also influence academic

outcomes. Additionally, the study is based on data from a single university, which may limit the generalizability of the findings to other contexts. Future research should incorporate non-cognitive variables and longitudinal data to develop more comprehensive predictive models of academic success. Investigating intervention strategies tailored to students from underrepresented regions and schools could further inform policy and institutional practices aimed at supporting first-generation students.

REFERENCES

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), 3. https://doi.org/10.1186/s41239-020-0177-7

Arifuddin, D., Kusrini, K., & Kusnawi, K. (2025). Perbandingan Performansi Algoritma Multiple Linear Regression dan Multi Layer Perceptron Neural Network dalam Memprediksi Penjualan Obat. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(2), 722–737. https://doi.org/10.57152/malcom.v5i2.1952

Banjarnahor, E., Belferik, R., Cendana, W., Adi, Y., & Abraham, S. (2025). *Analisis Implementasi Support Vector Machine dan Random Forest untuk Prediksi Kategori Indeks Kualitas Udara Jakarta under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)*. *10*(1), 2541–1179. https://journal.uin-alauddin.ac.id/index.php/instek

Banjarnahor, E., Sibarani, D. P., Wibawanta, B., Sihotang, D. A. G., & Abraham, Y. A. S. (2025). A Machine Learning Approach to Predicting Student Success Through Data Mining of LMS Moodle Activity Data. *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, 233–238. https://doi.org/10.1109/ICERA66156.2025.11086633

Barsegyan, V., & Maas, I. (2024). First-generation students' educational outcomes: The role of parental educational, cultural, and economic capital – A 9-years panel study. *Research in Social Stratification and Mobility*, *91*, 100939. https://doi.org/10.1016/j.rssm.2024.100939

Canning, E. A., White, M., & Davis, W. B. (2024). Growth Mindset Messages from Instructors Improve Academic Performance Among First-Generation College Students. *CBE—Life Sciences Education*, *23*(2). https://doi.org/10.1187/cbe.23-07-0131

Farhood, H., Joudah, I., Beheshti, A., & Muller, S. (2024). Evaluating and Enhancing Artificial Intelligence Models for Predicting Student Learning Outcomes. *Informatics*, *11*(3), 46. https://doi.org/10.3390/informatics11030046

Fitrianah, D., Dwiasnati, S., H, H. H., & Baihaqi, K. A. (2021). Penerapan Metode Machine Learning untuk Prediksi Nasabah Potensial menggunakan Algoritma Klasifikasi Naïve Bayes. *Faktor Exacta*, *14*(2), 92. https://doi.org/10.30998/faktorexacta.v14i2.9297

He, M. (2025). A linear regression modeling study of university language education and students' expressive skills. *Journal of Combinatorial Mathematics and Combinatorial Computing*, *127a*. https://doi.org/10.61091/jcmcc127a-316

Hopper, L. N., Yada, F. N., Abdulkareem, O., Jerger Butts, S., & Lacci-Reilly, K. R. (2024). A Picture of Perfect Mental Health: Exploring Barriers and Facilitators to First-Generation College Student Wellness Through Photovoice. *Journal of First-Generation Student Success*, 1–23. https://doi.org/10.1080/26906015.2024.2415091

Howe Eng, T., Eng, L. S., Ying, L. C., Daneil, I. L. A., & Ting, G. L. C. (2023). An Application of Multiple Linear Regression Analysis to Establish the Relationship between Higher Order Thinking Skills and Affective Domain. *International Journal of Academic Research in Business and Social Sciences*, *13*(5). https://doi.org/10.6007/IJARBSS/v13-i5/16907

Ives, J., & Castillo-Montoya, M. (2020). First-Generation College Students as Academic Learners: A Systematic Review. *Review of Educational Research*, *90*(2), 139–178. https://doi.org/10.3102/0034654319899707

Iwasaki, M. (2020). *Multiple Regression Analysis from Data Science Perspective* (pp. 131–140). https://doi.org/10.1007/978-981-15-2700-5_8

Juhaidi, A. (2023). Gender, Latar Belakang Sosio-ekonomi Keluarga, dan Prestasi Akademik Mahasiswa (Kasus pada Universitas Islam Negeri Terbesar di Kalimantan). *Metta : Jurnal Ilmu Multidisiplin*, *3*(4), 393–404. https://doi.org/10.37329/metta.v3i4.2853

Kracht, A. G., Doherr, M. G., & Jensen, K. C. (2025). Disparities in veterinary education: a survey comparing first-generation and continuing-generation students in Germany. *Frontiers in Veterinary Science*, *12*. https://doi.org/10.3389/fvets.2025.1595643

LeBouef, S., & Dworkin, J. (2021). First-Generation College Students and Family Support: A Critical Review of Empirical Research Literature. *Education Sciences*, *11*(6), 294. https://doi.org/10.3390/educsci11060294

López, M. J., Santelices, M. V., & Taveras, C. M. (2023). Academic performance and adjustment of first-generation students to higher education: A systematic review. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2023.2209484

Munir, J., Faiza, M., Jamal, B., Daud, S., & Iqbal, K. (2023). The Impact of Socio-economic Status on Academic Achievement. *Journal of Social Sciences Review*, *3*(2), 695–705. https://doi.org/10.54183/jssr.v3i2.308

Oyedeji, A. O., Salami, A. M., Folorunsho, O., & Abolade, O. R. (2020). Analysis and Prediction of Student Academic Performance Using Machine Learning. *JITCE (Journal of Information Technology and Computer Engineering)*, *4*(01), 10–15. https://doi.org/10.25077/jitce.4.01.10-15.2020

Pancsofar, N., Hall, M., Gouin, K., & Espinal, A. C. (2024). Motivation, Drive, and Pride: Exploring Assets of First-Generation College Students Majoring in Education. *Journal of First-Generation Student Success*, 1–18. https://doi.org/10.1080/26906015.2024.2432873

Pressimone Beckowski, C. (2025). Navigators in Training: First-Generation Graduate Students in College Student Affairs. *Journal of First-Generation Student Success*, *5*(2), 123–139. https://doi.org/10.1080/26906015.2024.2376761

Ramirez-Montoya, M. S., Morales-Menendez, R., Tworek, M., Escobar, C. A., Tariq, R., & Tenorio-Sepulveda, G. C. (2024). Complex competencies for leader education: artificial intelligence analysis in student achievement profiling. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2378508

Rispandi. (2023). Implementasi Metode K-Medoids Untuk Clustring Penerima Bantuan Berdasarkan Normalisasi Data Masyarakat Miskin Dengan Metode Desimal Scaling. *Journal of Computing and Informatics Research*, *3*(1), 141–152. https://doi.org/10.47065/comforch.v3i1.1063

Spencer, G. A., & Goldstein, S. E. (2025). "We Don't Know These Answers. We Don't Know Where to Go": The Experience of First-Generation College Student Mothers.

Journal of First-Generation Student Success, 1–20. https://doi.org/10.1080/26906015.2025.2515485

Sudbrock, C., Uribe-Zarain, X., Glaessgen, T. A., Sottile, J., & Finch, K. K. (2024). Examining Social Capital and First-Generation Student Status: Evidence from a Midwest University. *Journal of Research in Technical Careers*, *8*(2), 32. https://doi.org/10.9741/2578-2118.1155

Weisen, S., Do, T., Peczuh, M. C., Hufnagle, A. S., & Maruyama, G. (2024). How are first-generation students doing throughout their college years? An examination of academic success, retention, and completion rates. *Analyses of Social Issues and Public Policy*, *24*(3), 1274–1287. https://doi.org/10.1111/asap.12413

***Budi Wibawanta (Corresponding Author)**
Department of Teacher Profesional Education, Faculty of Education, Universitas Pelita Harapan, Tangerang, Indonesia
Jalan M.H. Thamrin Boulevard No.1100, Klp. Dua, Kec. Klp. Dua, Kabupaten Tangerang Email: budi.wibawanta@uph.edu

Evander Banjarnahor
Department of Mathematics, Faculty of Science and Technology, Universitas Pelita Harapan, Tangerang, Indonesia
Jalan M.H. Thamrin Boulevard No.1100, Klp. Dua, Kec. Klp. Dua, Kabupaten Tangerang Email: evander.banjarnahor@uph.edu

Stephanie Riady
Department of Educational Technology, Faculty Graduate School of Education, Universitas Pelita Harapan, Tangerang, Indonesia
Jalan M.H. Thamrin Boulevard No.1100, Klp. Dua, Kec. Klp. Dua, Kabupaten Tangerang Email: sriady@uph.edu

Dorita Setiawan
Universitas Pelita Harapan, Tangerang, Indonesia
Jalan M.H. Thamrin Boulevard No.1100, Klp. Dua, Kec. Klp. Dua, Kabupaten Tangerang Email: dorita.setiawan@uph.edu

HG Retno Harsanti
Department of Education, National Taiwan Normal University, Taipei, Taiwan
No. 162號, Section 1, Heping E Rd, Da'an District, Taipei City, Taiwan 106308
Email: 81200011e@ntnu.edu.tw