



Forecasting Student Academic Performance at A Philippine State University Using Supervised Learning Algorithms

Abstract: Strong academic performance is a priority not only for students but also for the institution. Students with good academic standing reflect the academic reputation of universities, serving as an indicator of high-quality education. Consequently, this creates opportunities such as increased funding and improved facilities that benefit students, faculty, and the community. Predicting students' performance in the early stage of the admission process enables institutions to provide tailored support and intervention for students to successfully navigate their academics. This study intends to predict the performance of incoming first-year college students at a state university in Laguna, Philippines, using supervised learning algorithms such as Linear Regression, Random Forest, Artificial Neural Network, and k-nearest Neighbor, employing the Weka software. Furthermore, the study identified the admission criteria with the highest accuracy that can predict student academic achievement. Using a ten-fold cross-validation, results showed that Linear regression is the model that can accurately predict student academic performance in the pre-pandemic dataset with an accuracy rate of 79.63%. The algorithms, however, could not accurately forecast student performance by using the pandemic dataset. Among the criteria used, senior high school GPA (SHS_Grade) is the best predictor for student performance due to its strong correlation with the final grade (Ave_GPA) in both datasets. The findings of this study will help institutions make data-driven decisions on student admission that allow early interventions and enhance educational outcomes.

Ma. Concepcion R. Repalam *, and
Dr. Jose B. Tan, Jr

Affiliation

Institute of Graduate Studies and Research, Manuel
S. Enverga University Foundation, Lucena City, 4301
Quezon, Philippines.

Corresponding Author

Ma. Concepcion R. Repalam * -
ma.concepcion.repalam@gmail.com

Article History:

Received Date : Apr 30, 2026
Revised Date : May 25, 2026
Accepted Date : May 30, 2026
Published Date : Jun 05, 2026

Keywords

Educational Data Mining; Linear Regression;
Prediction Model; Student Academic
Performance; Weka

1. Introduction

The Philippine higher education system, which includes state universities and colleges (SUCs), plays a crucial role in the fight against poverty and in strengthening the country's competitiveness through the development of human resources. Through the Universal Access to Quality Tertiary Education Act ([Republic Act No. 10931, 2017](#)), the government provides funding to these SUCs so that they can provide accessible and quality education to hundreds of thousands of students who are unable to attend private universities due to financial constraints. Admitting students who can perform well, especially on certification and licensure exams, and graduate on time is crucial to preventing the misuse of limited public funds and maintaining a pool of skilled labor ([Laus, 2021](#)).

One of the major obstacles is the limited number of slots available. Due to the high demand for affordable, quality education, SUCs receive thousands of applications every



year. Admission slots can be very competitive, as not all students who apply can be accommodated. The acceptance rate of the University of the Philippines (UP) ranges from 10 to 19 percent ([University of the Philippines System Ranking and Overview 2024, n.d.](#)). In 2019, UP accepted only 13 percent of its examinees based on performance in the University of the Philippines College Admission Test (UPCAT) and the weighted average of final grades obtained in high school ([Solitario, 2019](#)). Another challenge that students may face is the lack of resources and facilities in some SUCs. Although SUCs are meant to offer quality education, some institutions continue to struggle due to inadequate funding, limited facilities, and outdated equipment. This can affect the quality of education students receive and may hinder their ability to compete with other institutions.

The COVID-19 pandemic has significantly impacted the global education system, with schools, colleges, and universities closing to control the virus's spread ([Tarkar, 2020](#); [Rameez et al, 2020](#); [Perez et al, 2020](#); [Dagdia and Silva, 2022](#)). Over 1.5 billion students in 165 countries, or 87% of the student population, were affected by the closure (UNESCO: Building Peace Through Education, Science and Culture, Communication and Information, n.d.). The global health crisis has had a profound impact on several stakeholders, including students, instructors, parents, and the whole of the education system, including admissions, exams, and evaluation processes ([Dagdia and Silva, 2022](#)). In the Philippines, Ateneo de Manila University (ADMU), cancelled the Ateneo College Entrance Test (ACET) from SY 2020–2021 ([Ateneo De Manila University Suspends College Entrance Exam For Upcoming School Year - The Post, 2020](#)) and 2021–2022 ([Magsambol, 2020](#)). The university's website posted that admission will rely on the applicant's past academic performance, teachers' recommendations, essays, and

information written down on the application form. Similarly, the University of Sto. Tomas (UST) suspended the UST Entrance Test (USTET) for the academic year 2021–2022 because of the prevailing quarantine restrictions in the country. The university's admission process was determined by the University of Santo Tomas Admission Rating (USTAR), as stated in its Facebook post. USTAR is a numerical assessment calculated using several criteria, mostly derived from the applicant's past academic achievements and records. ([Magsambol, 2020](#)).

Meanwhile, dropouts in Philippine higher education are becoming a problem as families experience job loss or diminished incomes during the pandemic. The high cost of digital learning has made it too expensive for them to continue their education ([Gamboa, 2021](#); [Parreño, 2023](#)). A multinational software company, Desire 2 Learn (D2L), reported that 51 percent of Filipino students between the ages of 16 and 23 considered dropping out of school temporarily until the pandemic ends or until face-to-face classes are allowed again ([Añago, 2021](#)). This coincides with the 2017 Annual Poverty Indicators Survey (APIS) of the Philippine Statistics Authority (PSA) report, where 9% of the 39.2 million in this age group are out-of-school children and youth ([Philippine Statistics Authority, 2018](#)). The issues identified in the report included limited internet access, finding a conducive and peaceful study environment, being motivated to learn, managing other obligations and responsibilities, concerns related to mental health ([Añago, 2021](#)), marriage or family matters, and high cost of education ([Gamboa, 2021](#); [Parreño, 2023](#)), or financial concern ([Philippine Statistics Authority, 2018](#)).

Despite these challenges, SUCs remain an important option for students who wish to pursue higher education in the Philippines. SUCs provide an affordable and accessible



pathway for students to achieve their academic and career goals.

Being admitted to a university is an important milestone in a student's academic journey. It represents the recognition of one's potential and affirms the belief that they can thrive in higher education. However, success requires dedication, perseverance, and commitment. Students must be able to attend classes, participate in discussions, complete assignments, and excel on exams. Success in the classroom is measured by their final grade (Kumar *et al.*, 2021). Students who perform well academically may have more opportunities open to them and become contributing members of society.

Strong academic performance is a priority not only for students but also for the institution. Higher education is structured in a way that places a premium on students' academic achievements. The level of education that they provide is reflected in the level of achievement that their students achieve. Students who can graduate on time, achieve excellent academic standings, and succeed in licensing examinations and certifications are an advantage to the institution with which they are affiliated. Therefore, it is essential to select students who will contribute significantly to the academic community.

Every school should work toward developing an admissions process based on accurate and valid admissions criteria that choose applicants likely to succeed in its program. Also, before accepting applicants, SUCs should employ the most accurate methods for forecasting their future academic success (Moscoviz, 2022). This would assist academic institutions in making wise admissions decisions. However, most HEIs are having extreme difficulty managing the data, uncovering relevant patterns, and discovering knowledge from the huge collection of data collected through the years.

Traditional statistical techniques and database management technologies are no longer adequate for analyzing this enormous amount of data that is being collected and stored in databases.

Higher education institutions (HEI) have adopted this field of study called educational data mining (EDM) to analyze learning outcomes, engagement, and retention rates, as well as to monitor student progress and forecast future performance. EDM combines data mining, machine learning, and statistical analysis techniques to extract meaningful insights from educational data. HEIs can make data-driven decisions, optimize of student outcomes, reduce dropout rates, and improve graduation rates.

EDM is about using the data to identify patterns and build models that can be used to carry out appropriate interventions and produce better judgments. A key area of EDM is mining students' performance and enrollment data. EDM may be used to predict student performance and analyze learning data to suggest enhancements to existing educational methodologies. (Xiao *et al.*, 2022) Numerous techniques are prevalent in EDM. Data mining involves prediction, clustering, and relationship mining, while EDM emphasizes discovery with models and distillation of data for human judgment (Romero and Ventura, 2020; Bai *et al.*, 2021). Many strategies can be discovered using these techniques.

Fernando *et al.* (2019) focused on the application of the decision tree algorithm in predicting student graduation by generating rule sets that could early predict and identify students who are prone to not graduating on time.

Mengash (2020) used four data mining techniques to predict student performance at a



Saudi public university from 2016 to 2019. Pre-admission factors like high school grade point average, Scholastic Achievement Admission Test (SAAT) score, and general aptitude test score were used. The SAAT score was found to be the best predictor of future performance, with Artificial Neural Network (ANN) being the most effective approach with an accuracy rate of over 79%. Verma et al. (2022) used machine learning methods to predict future academic achievement based on variables and individual preferences.

Cruz-Jesus *et al.* (2020) used 16 demographic variables such as age, gender, citizenship, internet connection, computer ownership, and the number of courses completed to predict the academic achievement of nearly all public high school students in Portugal from 2014 to 2015. The study's findings indicate that among the methodologies used, namely ANN, SVM, DT, Random Forest (RF), k-nearest neighbors (kNN), and extremely randomized trees (ET), the RF model emerged as the most effective. The most critical factors are the number of unit courses attended, the number of failures, and gender.

Samsudin (2021) compared students' grade point average (GPA) before and after lockdown in face-to-face (F2F) and online learning to find the pattern of the student's academic performance. From the result, students performed better online than in F2F classes. Meanwhile, SVM was used to predict students' academic achievement based on their cumulative GPA (CGPA) and ages using the classification method. The said classification model effectively forecasted the academic performance of students by using misclassification tolerance, C , and epsilon as parameters.

Tarik *et al.* (2021) were able to predict the performance of Moroccan students during COVID-19 using Linear Regression (LR),

regression-type DT, and regression-type RF. The study adopted RF, which gave the best prediction among the three algorithms.

In Yagci's (2022) study, a new model was developed based on RF, SVM, kNN, NB, nearest neighbor (NN), and logistic regression (LoR) algorithms to predict the final exam grades of undergraduate students taking Turkish Language I course at a state university in Turkey based on midterm grades. The results showed that the proposed model achieved a classification accuracy of 70 – 75 percent. It can be said that students' midterm exam grades are important predictors to be used in predicting their final exam grades.

Locally, there have also been studies that predict student performance. Magbag and Raga (2020) developed a model that predicts the academic performance of first-year college students from three HEIs in Central Luzon. LR and Neural Network (NN) were used for modeling based on the computed first-year GPA. Multiple experiments were conducted on individual and combined datasets, evaluated for accuracy, precision, recall, and AUC to measure performance. Results showed that NN outperformed LR on all metrics.

LoR was used to forecast Philippine university students' enrollment decisions in a study by Esquivel and Esquivel (2021). The maximum accuracy was achieved using eleven variables instead of nineteen. This covers enrollment, location, gender, parent employment, senior high school type, and school choice. The model was built using the Weka data mining tool and the DSS application was designed and developed using R Weka, R Studio, and R Shiny. Results showed that LoR model was able to predict the enrollment of applicants based on the set of features indicated by Weka. It was concluded that machine learning techniques may be used as a supplement to enrollment decisions and to measure the level of



correlation between enrolment and such factors.

Laus (2021) examines how admission criteria affect student performance at a state university in Cebu City. Multiple regression was used to predict students' final grades using pre-entry grades and entrance exam results in Math, Science, APFil, English, and OLSAT. The results indicated that pre-entry grades and entrance test scores positively affect final grades. Pre-entry grade was the most significant indicator of academic achievement, while APFil score was the least. Further, scores in Science and OLSAT also do not affect their performance in school. Thus, the proposed model suggested using the applicant's pre-entry grades in the evaluation and ranking of candidates for admission, as well as assigning numerical weights to test scores in the subjects.

The synthesis of relevant research material demonstrates the significant use of data mining methods in the analysis of academic performance. Researchers have found key elements that influence academic achievement, constructed models for making predictions, and recommended educational interventions that are beneficial via the use of a variety of data mining technologies. These results give important insights for educators, legislators, and other stakeholders who are interested in improving educational processes and encouraging the success of their students.

To help improve the admissions process, the goal of this study is to utilize data mining (DM) algorithms that can accurately forecast the academic performance of incoming students at a state university in Laguna. Specifically, the study aims to identify the criteria that can accurately predict applicants' future performance, as well as generate predictive models using data collected before and during the pandemic. If academic success can be anticipated even before admission, it would

allow educational institutions to tailor their support and resources to meet the needs of each student more effectively. By understanding a student's potential and areas of strength or challenge, schools can provide targeted interventions, personalized learning plans, and enrichment opportunities. This proactive approach could lead to improved student outcomes and a more successful educational experience overall.

2. Methods

The focus of this research is to develop a model that will accurately predict the academic performance of students at Laguna State Polytechnic University in San Pablo City, Laguna (LSPU-SPCC). It involves the following steps: data collection, data preprocessing (cleaning, integration, and transformation), training and testing of the data, and model evaluation.

A correlational quantitative research approach was employed in this study to determine the predictive validity of the admission criteria used at the state university.

Data Collection

To initiate the research and acquire data from concerned offices, a letter was sent to the university officials of LSPU-SPCC. This was done to conform to the research ethics and standards that are in place. The Registrar's Office, the Information and Communications Technology Services (ICTS) Department, and the OSAS, particularly the Guidance Office, received copies of the approved letter. These letters were submitted to request a report on first-year students enrolled for school years 2019–2022. The report would include the students' profile, their grade point average (GPA) for both the first and second semesters of their first year and pre-admission records. The data collected from the three sources was cross-referenced to have a consolidated data set. This will be encoded in Microsoft Excel.



Two distinct datasets were generated as a result of the varying conditions surrounding the acquisition of the data. The first dataset, denoted as Dataset 1 in this paper, is derived from the data collected in 2019. This dataset was obtained in the pre-pandemic period, when entrance examinations were given to incoming students. The second dataset, denoted as Dataset 2, was consolidated by merging the data from the academic years 2020 to 2022. During this period, a pandemic occurred, prompting the cancellation of entrance exams. Admission is based on academic performance in junior and senior high school.

Dataset 1 has 1,726 initial records, which were consolidated from the electronic database of the university and the digital records of the Registrar's Office and the Guidance Office. It has 11 attributes, namely course enrolled, sex, hometown, SHS track, and school type. It also includes senior high school grades in English, Science, and Mathematics. The main attributes in this dataset are entrance exam rating (Exam_rating), SHS_Grade, and Ave_GPA. Dataset 2, on the other hand, has 142 records, and its attributes are sex, hometown, SHS track, school type, junior high school grades (JHSGPA), SHS_Grade, and Ave_GPA.

This paper will use the average grade of the student at the end of the first and second semesters (Ave_GPA) during their freshman year as the criterion variable and is denoted as academic performance. All collected datasets are organized in Microsoft Excel and saved in .csv format.

Data Preprocessing

All implementations of the data mining algorithms are done in Weka. Weka is an open-source tool coded in Java and used in EDM research (Mengash, 2020; Esquivel and Esquivel, 2021). It offers tools for data mining tasks such as regression, classification,

association rules mining, clustering, and visualization (Li, 2022).

A preliminary cleaning procedure was conducted on the collected dataset, whereby transferees were excluded from the analysis due to missing SHS grades. Additionally, entries with non-standard grade equivalencies or missing grades were disregarded. Finally, students who did not complete a full academic year were deleted to ensure data consistency. The cleanup of Dataset 1 reduced 1,726 instances to 419 with 11 characteristics, whereas Dataset 2 had 74 with seven attributes. The data was cleaned and converted to Weka's .arff format. A Classifier Attribute Evaluator with Ranker Search Method was used in the Attribute Selection to select the attributes that would be most relevant to the classification. In this step, the attribute 'hometown' has the lowest rank with a 0.0695 value. Verifying the relevance of this attribute, it was observed that the students live in nearby towns and provinces, which will not affect their performance. It was, therefore, removed as an attribute.

Dataset 2 underwent the same cleaning procedure. Before running Weka for preprocessing, the data was cleaned of missing and inconsistent values. The attribute 'hometown' was also removed from the dataset because it was ranked last when the dataset was subjected to the same Attribute Evaluator. Furthermore, the dataset was split into 80 percent training data and 20 percent test data using the resample method. The training set is used to build the model, whereas the test set is used to evaluate and validate the model. This step will prevent overfitting, which will lead to poor performance of the developed model. The final dataset contains 335 instances in the training data and 84 instances in the test data, with 10 attributes for Dataset 1. Dataset 2 has 59 instances in the training data and 15 instances in the test data with six attributes. All variables



included in the datasets are considered possible predictors, with Ave_GPA as the class attribute.

Model evaluation

Four data mining algorithms were employed in building the model that will best predict the academic performance of students based on admission criteria and student demographics. These are Linear Regression (LR), Artificial Neural Network (ANN), k-nearest neighbor (kNN), and Random Forest (RF).

Linear regression is a data mining method that uses supervised learning to construct a mathematical model that represents the association between a dependent variable and independent variables. This method aims to identify the best-fit line that minimizes the difference between the observed and predicted values.

Artificial neural networks are models inspired by the human brain's neural networks. They are composed of linked nodes, called neurons, which are arranged in layers. ANN can solve the nonlinear and complex relationship between different input and output variables. ANNs learn from data and adjust their weights to provide predictions or classifications. An implementation of ANN in Weka is MultiLayer Perceptron (MLP). MLP consists of multiple layers of interconnected neurons that approximate a function that best fits the data.

The k-nearest neighbor algorithm utilizes the majority class of the k-nearest data points to predict outcomes. It is a useful tool to identify patterns in similar student profiles. kNN is called IBk in Weka. IBk or Instance-Based Learner is a machine learning algorithm that provides predictions for a test instance in real time. The IBk algorithm employs a distance metric in making predictions by finding *k* adjacent instances in the training data for each test instance.

Random forest is a supervised learning method that combines multiple decision trees to reach a result. It is one of the most popular algorithms used because of its simplicity and flexibility. RF can capture complex correlations within the data and can manage noisy data.

Each model was trained using the training data and evaluated for accuracy. Subsequently, test data is fed into the model for validation, and the output is the predicted average GPA (Predicted Ave_GPA). Predicted Ave_GPA are compared with actual grades earned by students. The software displayed a summary of the performance of each model.

3. Results and Discussion

Attribute Selection

Figures 1a and 1b present the output of the Attribute Selection using the Classifier Attribute Evaluator with the Ranker Search Method. Figure 1a represents Dataset 1 and Figure 1b shows the result of Dataset 2.

It can be seen that in Dataset 1, Course has the highest rank among all the attributes, while hometown ranked last. Notably, attributes related to grades such as SHS_Grade, English, Math, and Science have more than 50 percent correlation and were ranked on top of the demographic characteristics. The importance of grades in forecasting performance is evident in these findings. Entrance exam results, however, had a weak relationship with only 37.47 percent, while demographic information was among the attributes at the bottom of the rank, indicating that they do not show significant relevance in the process. Similarly, the output of Attribute Selection in Dataset 2 showed that grades are relevant in predicting performance because they ranked on top.

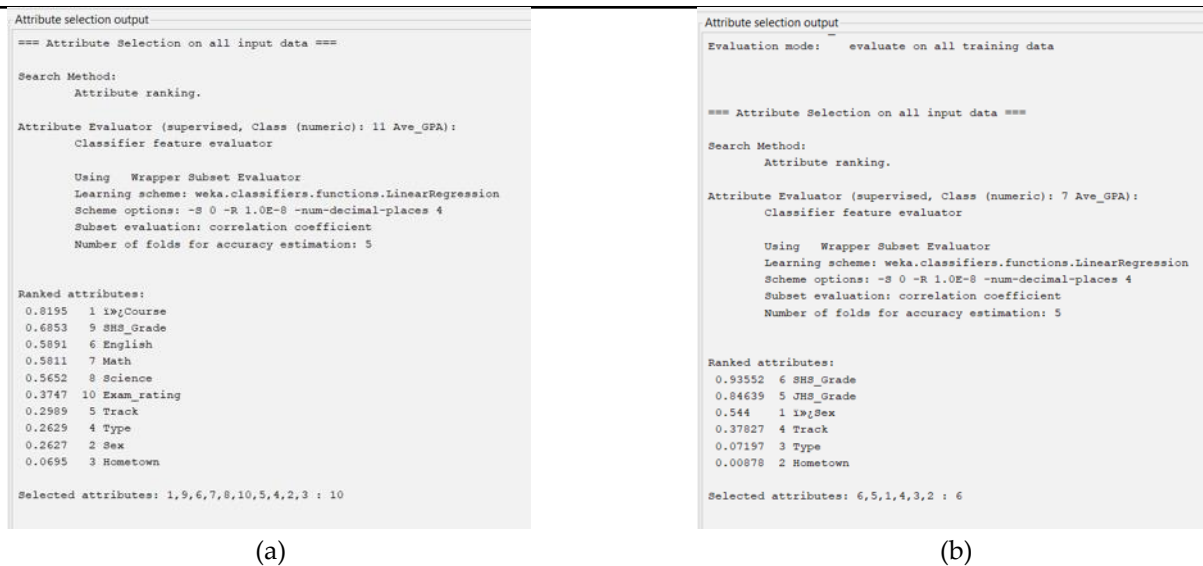


Figure 1. Screenshot of Results of Attribute Selection

Identification of the most significant criteria

LSPU, like other universities, relies heavily on high school GPA and entrance exam results for admission. Both datasets underwent regression analysis using a 10-fold cross-validation

technique to identify the most accurate admission criterion for predicting students' academic achievement. Table 1 presents a comparison of the results in Weka.

Table 1. Summary of results between admission criteria and Ave_GPA.

Dataset	Admission Criteria	Correlation Coefficient	Mean absolute error	Root Mean Squared Error
1	SHS_Grade	0.5893	0.1565	0.2098
	Exam_rating	0.2814	0.1863	0.2492
	SHS_Grade and Exam_rating	0.5910	0.1565	0.2095
2	JHS_Grade	0.5921	0.2584	0.3526
	SHS_Grade	0.6830	0.2440	0.3194
	JHS_Grade and SHS_Grade	0.6830	0.2440	0.3194

As seen in Table 1, there is a strong correlation between SHS_Grade and Ave_GPA, with values of 59 percent and 68 percent respectively, both before and during pandemic situations. This supports previous studies (Mengash, 2020; Allensworth, 2020; Al Hazaa, et al., 2021) indicating that a high school GPA is a strong foundation for college performance. Exam_rating had a 28 percent correlation, which indicates a weak relationship with Ave_GPA. This contradicts previous research by Sulphrey et al. (2018) suggesting a strong association between admission grades and academic achievement. However, Aciro (2021)

found no significant correlation between entry grades and academic performance. Other factors, such as test anxiety and stress (Javed & Abiodullah, 2021; Ayanwale, 2022), may also weaken the correlation between examinations and academic achievement.

Evaluating the two factors together strengthens their relationship with academic performance, which is also true with Sawey – Ognayon and Afalla (2022).

Similarly, when predicting students' academic performance during the pandemic, SHS_Grade still has a slightly more favorable relationship

(68 percent) than junior high school grade (JHS_Grade), which has approximately 59 percent correlation with Ave_GPA. When evaluated together, the accuracy remains at 68 percent. This suggests that JHS_Grade does not have an impact when evaluated together with SHS_Grade. However, in the absence of SHS grades, admissions may consider junior high school grades when evaluating students for admission.

Other attributes in the dataset were also evaluated with Ave_GPA. This is shown in Tables 2 and 3.

Table 2. Dataset 1's summary of findings.

Attributes	Correlation Coefficient	Mean absolute error	Root mean squared error
Course	0.7159	0.1347	0.1816
Sex	0.1543	0.1921	0.2569
Type	0.1658	0.1922	0.2561
Track	0.2062	0.1961	0.2543
English	0.4923	0.1691	0.226
Math	0.4842	0.1682	0.2272
Science	0.4683	0.1709	0.2294

Table 2 shows that a student's course significantly impacts Ave_GPA, with a 71.6 percent accuracy rate. Students perform better when choosing a major that aligns with their interests and abilities (Mappadang, 2022). Grades in English, Math, and Science also slightly affect performance, with high grades in these subjects indicating continued success in college. Proficiency in English and Math is also a predictor of academic performance, as confirmed by Oducado (2020) and Balasico (2020).

However, it should be noted that the other attributes do not significantly affect academic achievement.

Table 3. Dataset 2's summary of findings.

Attributes	Correlation Coefficient	Mean absolute error	Root mean squared error
Sex	0.2696	0.3155	0.4224
Type	-0.2317	0.3232	0.4533
Track	0.0208	0.3395	0.4681

The relationship between Ave_GPA and the three attributes in Dataset 2 have weak correlations. This is consistent with the results of the same attribute in Dataset 1 but contradicts other studies where sex (Al Husaini, 2022; Mwihiya, 2020) and high school track (Lumboy, 2019) have a significant effect on student performance.

Model Performance

The study aimed to develop prediction models using LR, ANN, KNN, and RF algorithms to predict students' academic achievement. The datasets were divided into 80 percent training data and 20 percent test data using a 10-fold cross-validation, and Weka was used for classifier implementation. A 10-fold cross-validation is a method that randomly divides a dataset into ten parts, nine of those parts are dedicated to training the data, and one part is for testing. The process will then be repeated ten times, with each time having a different part for testing.

Evaluation of Results in Dataset 1

Analysis of the model shows that linear regression (LR) has the highest correlation at 79.63 percent, followed by random forest (RF), which has a 77.97 percent accuracy rate. A summary of the results displayed by the software is presented in Table 4.

From the table, the comparative performance of the models is between 72 and 80 percent, except for ANN, which has an accuracy rate of 67 percent. This shows that these algorithms perform well in the prediction of student performance. However, since LR has the highest association with Ave_GPA, we can infer

that this model can best predict the academic performance of students based on all attributes present.

Table 4. Comparison of data mining algorithms on Dataset 1 using 10-fold cross-validation

Model	Correlation coefficient	Mean absolute error	Root mean squared error
Linear Regression	0.7963	0.1204	0.1622
Random Forest	0.7797	0.1217	0.1674
K-nearest neighbor	0.7238	0.1365	0.1846
Artificial Neural Network	0.6748	0.1231	0.2474

Table 5. Comparison of data mining algorithms on Dataset 2 using 10-fold cross-validation

Model	Correlation coefficient	Mean absolute error	Root mean squared error
Linear Regression	0.5534	0.2824	0.3634
Artificial Neural Network	0.4291	0.4022	0.5484
K nearest neighbor	0.5255	0.2704	0.3637
Random Forest	0.5663	0.2657	0.3524

Evaluation of Results in Dataset 2

Looking at Table 5, results in the evaluation of Dataset 2 using 10-fold cross-validation showed that the models have an accuracy range of 42 to 57 percent. This suggests a moderately weak relationship with Ave_GPA. Random Forest has the highest accuracy at 56.63 percent, while Linear Regression has 55.34 percent accuracy. Moreover, the value of RMSE is high, ranging from 35 to 55 percent. This means that the error rate between the actual and the

model's predictions is high. It can be said that the model does not accurately predict student performance in this dataset. The classifiers chosen for the dataset were not able to predict academic performance accurately, unlike in Dataset 1.

4. Conclusions and Recommendations

Selecting students who will perform well academically throughout their university journey is an advantage to the university. Knowing which applicants to admit every school year is crucial. This study aimed to develop a model that can effectively forecast the academic achievement of incoming college students based on the university's admission criteria. Further, it determined the criteria that can predict performance. Lastly, this research was able to evaluate the accuracy of the models given different circumstances. The results revealed that most of the models chosen can predict student academic performance correctly, but Linear Regression exhibited the best level of accuracy in Dataset 1. However, the models showed a weak relationship with academic performance in Dataset 2. Furthermore, it was identified that the SHS grade point average is the criteria that has the strongest correlation with the average grade of the student in both pre-pandemic and pandemic periods. But course enrolled and grades in English, Math and Science should also be considered when assessing an applicant's credentials. Examinations did not have a significant impact on academic performance, but since it is a major part of the criteria, universities should look into formulating a more comprehensive strategy that will reflect the student's abilities. To improve the accuracy and performance of the predictive models, a follow-up study is in place, as this is a continuing research project. A more extensive dataset may be used to normalize the training sample, which would further improve the accuracy rate and predict all of the grades more precisely. Future



research can include other parameters and other algorithms in the process.

Conflict of Interest.

There is none among the authors

Funding.

No external funding, private or government financial support or grant was received before, during and after this study.

References

- [1] Aciro, R., Onen, D., Malinga, G. M., Ezati, B. A., & Openjuru, G. L. (2021). Entry grades and the academic performance of university students: A review of literature. *Education Quarterly Reviews*, 4(1).
- [2] Al Hazaa, K., Abdel-Salam, A. S. G., Ismail, R., Johnson, C., Al-Tameemi, R. A. N., Romanowski, M. H., ... & Elatawneh, A. (2021). The effects of attendance and high school GPA on student performance in first-year undergraduate courses. *Cogent Education*, 8(1), 1956857.
- [3] Al Husaini, Y. N. S., & Shukor, N. S. A. (2022). Factors affecting students' academic performance: A review. *RES MILITARIS*, 12(6), 284–294.
- [4] 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.
- [5] Añago, B. A. (2021, June 23). Half of PHL students consider putting off school amid pandemic. *BusinessWorld* Online. <https://www.bworldonline.com/the-nation/2021/06/23/377499/half-of-phl-students-consider-putting-off-school-amid-pandemic/>
- [6] ATENEO DE MANILA UNIVERSITY SUSPENDS COLLEGE ENTRANCE EXAM FOR UPCOMING SCHOOL YEAR. (2020, September 2). *The POST*. <https://thepost.net.ph/news/campus/ateneo-de-manila-university-suspends-college-entrance-exam-for-upcoming-school-year/>
- [7] Ayanwale, M. A. (2022). Nexus between test-anxiety and response aberrance patterns in mathematics test items.
- [8] Bai, X., Zhang, F., Li, J., Guo, T., Aziz, A., Jin, A., & Xia, F. (2021). Educational big data: Predictions, applications and challenges. *Big Data Research*, 26, 100270. <https://doi.org/10.1016/j.bdr.2021.100270>
- [9] Balasico, C. L., & Tan, D. A. (2020). Predictors of performance of Central Mindanao University Laboratory High School students. *PEOPLE: International Journal of Social Sciences*, 6(2), 1–21.
- [10] Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R. M. N., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon*, 6(6), e04081. <https://doi.org/10.1016/j.heliyon.2020.e04081>
- [11] Dagdia, Z. C., & Silva, A. C. S. E. (2022). Effects of COVID-19 pandemic on education and society. *STEM Education*, 2(3), 197. <https://doi.org/10.3934/steme.2022013>
- [12] Esquivel, J. A., & Esquivel, J. A. (2021). A machine learning based DSS in predicting undergraduate freshmen enrolment in a Philippine university. *International Journal of Computer Trends and Technology*, 69(5), 50–54. <https://doi.org/10.14445/22312803/ijctt-v69i5p107>
- [13] Fernando, M. C. G., Goh, M. L. I., Lagman, A. C., Jayar, P. L., Ortega, J. H. J. C., & Solomo, M. V. S. (2019). Development of a predictive decision support system for student graduation using a decision tree algorithm. *International Journal of Simulation--Systems, Science & Technology*, 20.
- [14] Gamboa, R. (2021, May 27). Coping with higher college dropouts. *Philstar.com*. <https://www.philstar.com/business/2021/05/27/2101015/coping-higher-college-dropouts>
- [15] Javed, E., & Abiodullah, M. (2021). Effects of test anxiety on academic achievement at secondary school level in Lahore. *Bulletin of Education and Research*, 43(3), 67–80.
- [16] Kumar, S., Agarwal, M., & Agarwal, N. (2021). Defining and measuring academic performance of HEI students: A critical review. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(6), 3091–3105.
- [17] Laus, M. A. (2021). Admission profiles as predictors of academic performance. *Asia Pacific Journal of Management and Sustainable Development*, 9(1), 39–47.
- [18] Lumboy, M. (2019). Senior high school strand choice: Its implication to college academic performance. *Ascendens Asia Journal of Multidisciplinary Research Abstracts*, 3(7).
- [19] Magbag, A., & Raga, R. (2020). Prediction of college academic performance of senior high school graduates using classification techniques. *International Journal of Scientific & Technology Research*, 9(4), 2104–2109.
- [20] Magsambol, B. (2020, September 1). No college entrance exams in Ateneo, UST for academic year 2021–2022 due to pandemic. *RAPPLER*. <https://www.rappler.com/nation/college-entrance-exams-philippines-2020>
- [21] Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision-making in university admission systems. *IEEE Access*, 8, 55462–55470. <https://doi.org/10.1109/ACCESS.2020.2981905>



- [22] Moscoviz, L., & Evans, D. K. (2022). Learning loss and student dropouts during the COVID-19 pandemic: A review of the evidence two years after schools shut down.
- [23] Mwihiha, C. (2020). Gender difference in academic achievement of students in Kinangop Sub County, Nyandarua County, Kenya.
- [24] Oducado, R. M. F., Sotelo, M. G., Ramirez, L. M. M., Habaña, M. P., & Belo-Delariarte, R. G. (2020). English language proficiency and its relationship with academic performance and the nurse licensure examination.
- [25] Parreño, S. J. (2023). School dropouts in the Philippines: Causes, changes and statistics. *Sapienza: International Journal of Interdisciplinary Studies*, 4(1), e23002–e23002.
- [26] Perez, M., Williams, C., & Vapiwala, N. (2020). Holistic admissions during a pandemic. *Academic Medicine*.
<https://doi.org/10.1097/ACM.0000000000003613>
- [27] Philippine Statistics Authority. (2018). Nine percent of Filipinos aged 6 to 24 years are out of school (Results from the 2017 Annual Poverty Indicators Survey). <https://psa.gov.ph/content/nine-percent-filipinos-aged-6-24-years-are-out-school-results-2017-annual-poverty-indicators>
- [28] Rameez, A., Fowsar, M. A. M., & Lumna, N. (2020). Impact of COVID-19 on higher education sectors in Sri Lanka: A study based on South Eastern University of Sri Lanka.
- [29] Republic Act No. 10931. (2017, August 3). Official Gazette of the Republic of the Philippines. <https://www.officialgazette.gov.ph/2017/08/03/republic-act-no-10931/>
- [30] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3). <https://doi.org/10.1002/widm.1355>
- [31] Samsudin, E. A. N. A. M. (2021). Modeling student's academic performance during COVID-19 based on classification in support vector machine. *Turkish Journal of Computer and Mathematics Education*, 12(5), 1798–1804.
<https://doi.org/10.17762/turcomat.v12i5.2190>
- [32] Sawey-Ognayon, G. B., & Afalla, B. T. (2022). Admission criteria as predictors of licensure performances among graduates of a state university in the Philippine Cordillera Administrative Region. *Academic Journal of Interdisciplinary Studies*, 11(6), 147. <https://doi.org/10.36941/ajis-2022-0157>
- [33] Solitario, J. M. (2019, October 3). [INFOGRAPHIC] Tips for UPCAT 2020 examinees. University of the Philippines. <https://up.edu.ph/infographic-tips-for-upcat-2020-examinees/>
- [34] Sulphey, M. M., Al-Kahtani, N. S., & Syed, A. M. (2018). Relationship between admission grades and academic achievement. *Entrepreneurship and Sustainability Issues*, 5(3), 648–658.
- [35] Tarik, A., Aissa, H., & Yousef, F. (2021). Artificial intelligence and machine learning to predict student performance during the COVID-19. *Procedia Computer Science*, 184, 835–840.
- [36] Tarkar, P. (2020). Impact of COVID-19 pandemic on education system. *International Journal of Advanced Science and Technology*, 29(9), 3812–3814.
- [37] UNESCO. (n.d.). Building peace through education, science and culture, communication and information. <https://en.unesco.org/news/unesco-rallies-international-organizations-civil-society-and-private-sector-partners-broad>
- [38] University of the Philippines System ranking & overview 2024. (n.d.). <https://www.4icu.org/reviews/9097.htm>
- [39] Verma, U., Garg, C., Bhushan, M., Samant, P., Kumar, A., & Negi, A. (2022). Prediction of students' academic performance using machine learning techniques. 2022 International Mobile and Embedded Technology Conference (MECON). <https://doi.org/10.1109/MECON53876.2022.9751956>
- [40] Xiao, W., Ji, P., & Hu, J. (2021). A survey on educational data mining methods used for predicting students' performance. *Engineering Reports*, 4(5). <https://doi.org/10.1002/eng2.12482>
- [41] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-022-00192-z>