Prediction of Higher Education Student Dropout based on Regularized Regression Models

Bouchra Bouihi
L2IACS Laboratory, ENSET Mohammedia, Morocco
bouihi@enset-media.ac.ma (corresponding author)

Abdelmajid Bousselham
L2IACS Laboratory, ENSET Mohammedia, Morocco
bousselham@enset-media.ac.ma

Essaadia Aoula
L2IACS Laboratory, ENSET Mohammedia, Morocco
aoula@enset-media.ac.ma

Fatna Ennibras
L2IACS Laboratory, ENSET Mohammedia, Morocco
ennibrasfatna18@gmail.com

Adel Deraoui
Casablanca-Settat Regional Center for Education and Training, Morocco
deraouiadel@yahoo.fr

Received: 6 August 2024 | Revised: 7 September 2024 | Accepted: 22 September 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.8644*

ABSTRACT

This study explores the critical topic of student dropout in higher education institutions. To allow early and precise interventions and to provide a multifaceted view of student performance, this study combined two predictive models for dropout classification and score prediction. At first, a logistic regression model was developed to predict student dropout at an early stage. Then, to enhance dropout prediction, a second-degree polynomial regression model was used to predict student results based on available academic variables (access, tests, exams, projects, and assignments) from a Moodle course. Dealing with a limited dataset is a key challenge due to the high risk of overfitting. To address this issue and achieve a balance between overfitting, data size, and model complexity, the predictive models were evaluated with L1 (Lasso) and L2 (Ridge) regularization terms. The regularization techniques of the predictive models led to an accuracy of up to 89% and an R^2 score of up to 86%.

Keywords-dropout prediction; logistic regression; polynomial regression; regularization; lasso; ridge

I. INTRODUCTION

The advent of online learning platforms in higher education has provided fertile ground for collecting information and tracking students' learning trajectories and performances [1]. Analysis of student performance data has become increasingly vital in educational data mining research [2], particularly for the early identification of at-risk students and interventions to prevent student dropout [3]. Predicting student learning failure or success remains one of the most frequently investigated areas of the Learning Analytics (LA) and Educational Data Mining (EDM) disciplines [4]. EDM deals with educational analysis to understand student behavior [4, 5]. These techniques are used to enhance learning environments, modify course structures, or predict student performance and behavior [6, 7]. LA measures, collects, analyses, and reports student data to comprehend and enhance learning experiences and environments [8]. In particular, student dropout prediction has been the focus of various studies, especially with the increased use of educational online environments. Systematic literature reviews [9, 10] have summarized studies that used various Machine Learning (ML) techniques to predict student dropout.

Several studies have used LR to predict student dropout. For instance, in [11], LR was used to predict dropout rates based on demographic and academic variables. The findings indicated that LR is effective for binary classification problems such as dropout prediction, although it often requires careful feature engineering and selection to improve performance. In [12], a predictive model was proposed based on LR using the Knowledge Discovery in Databases (KDD) model with a high accuracy, scoring 92.56% reliability. In [13], LR prediction accuracy was compared with Classification and Regression Trees (CART) and Neural Networks (NN) on data from Moodle Learning Management System (LMS) in four Finnish universities. The results showed that LMS data significantly enhanced predictive accuracy. In [14], eleven ML algorithms were compared in predicting student failure based on various metrics, such as accuracy, precision, recall, F-score, specificity, and balanced accuracy. In [15], LR and Decision Trees (DT) were used to predict student dropout at the Karlsruhe Institute of Technology, where the latter faced overfitting problems. Regularization techniques such as L1 (Lasso) and L2 (Ridge) have been increasingly applied to address overfitting and enhance model generalization [16].

In [17], a polynomial regression model was developed to predict student academic performance by testing different degrees, reaching 83.44% R^2 for degree 1. More complex architectures have been explored for performance prediction. In [18, 19], deep learning and neural networks were used, although these models often require more data and computational resources to model complex student interactions and predict academic results. In [2], the performance of various ML models was compared, highlighting the strengths of White Box (WB) over Black-Box (BB) models in terms of accuracy and understanding of results. The application of regularization in regression models, such as seen in [20] using Lasso and [6] on Ridge regression, has been instrumental in enhancing model generalization. In [21], regularized linear regression models were used to predict student grades, showing that the prediction error rate of the Ridge regularization model was the lowest among Lasso and Elastic net regularizations. Some approaches combined multilayered models [22, 23] to enhance the prediction performance of single models.

However, when dealing with small educational datasets with few features, as in the case of some Moodle courses [7], the task becomes significantly more challenging. Small datasets often lead to problems such as overfitting, where models succeed in correctly predicting the training data but fail on the test data. This is a common problem in educational data mining, where collecting large amounts of data can be difficult due to limited resources, confidentiality issues, and the naturally limited populations of certain study groups or educational programs, as in the case of institutions with restricted access. The datasets commonly employed in educational data mining are typically quite small, often averaging around 200 records at the course level [4]. To address these challenges, this study proposes a hybrid approach that combines LR for binary classification with polynomial regression for continuous outcome prediction. LR is well-suited for small datasets due to its simplicity and effectiveness in binary classification tasks, such as predicting whether a student will pass or fail. On the other hand, polynomial regression can model more complex relationships within the data, providing detailed predictions of student scores.

The main purpose of combining these predictive models is to benefit from their independent strengths. LR is used to classify whether a student is at risk of dropout or not. This early prediction will highlight at-risk students who require additional support. Subsequently, polynomial regression will predict specific scores for these students, based on sufficient academic features (such as access, tests, exam, project, and assignments) that offer deeper insights into their potential performance. The performance of the proposed combined approach was evaluated using a small educational dataset [7]. Regularization techniques were explored to prevent overfitting. The focus was on developing predictive models for student dropout and performance using LR enhanced with regularization techniques, namely L1 (Lasso) and L2 (Ridge). Regularization is crucial in addressing overfitting, especially in limited or noisy datasets, by penalizing model complexity and enhancing generalization to new unseen data. Integrating these performance predictions into dropout models aims to reach a more comprehensive understanding of student trajectories, thus improving the effectiveness of early intervention strategies. The goal is to demonstrate that even with limited datasets and key features, it is possible to develop predictive models that offer significant value in an educational context by using regularization techniques.

II. METHODOLOGY

A. *Predicting Student Dropout Using Logistic Regression (LR)*

This study developed an LR model to predict student dropout based on binary classification of the target $y(i)$: 0 for non-graduate and 1 for graduate. To predict students at risk, three features were selected, based on students' interactions with course content and activities before the final exam: access, tests, and assignment. These features were extracted from the Moodle LMS platform [7]. The predictive model was developed from scratch [24]. This choice was preferred over using libraries to enhance the credibility and transparency of the research by promoting open, clear, reproducible results, and deeper and insightful analysis and discussion. The LR model is

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 \quad (1)$$

$$h_\theta(x) = \sigma(z) = \frac{1}{1+e^{-z}} \quad (2)$$

where θ_0 is the bias, and θ_1, θ_2, and θ_3 are the weights of features x_1, x_2, and x_3, respectively.

Given the limited size of the dataset, the risk of overfitting is particularly pronounced. Overfitting occurs when the model performs well on the training data but fails to generalize adequately to new data. To mitigate this issue, the Lasso (L1) and Ridge (L2) regularization terms were incorporated into the cost function of the predictive model. Lasso L1 incorporates a penalty proportional to the absolute value of the coefficients, which can also result in feature selection by setting some coefficients to zero. The Ridge L2 regularization term incorporates a penalty proportional to the square of the feature coefficients to prevent the parameters from becoming excessively large.

The cost function with the Lasso L1 regularization term is:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{m}\sum_{j=1}^{n}|\theta_j| \quad (3)$$

The cost function with the Ridge L2 regularization term is:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] + \sum_{j=1}^{n}\frac{\lambda}{2m}\theta_j^2 \quad (4)$$

The gradient descent learning algorithm was employed to minimize the cost function. The model parameters, learning rate, number of iterations, and regularization coefficient, were chosen by fine-tuning. Various values were tested for the learning rate and the number of iterations. The chosen learning rate and number of iterations were those that ensured the convergence of the gradient descent algorithm efficiently and effectively. A learning rate of $\alpha = 0.1$ and 10,000 iterations were selected. For the regularization coefficient, the value of $\lambda = 0.1$ was selected among four values (0.01, 0.1, 1, and 10).

The L1 regularization (Lasso) is given by:

$$\theta_j := \theta_j - \alpha\left(\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m}sign(\theta_j)\right)$$
$$for\ j = 1,2,3 \quad (5)$$

The L2 regularization (Ridge) is given by:

$$\theta_j := \theta_j - \alpha\left(\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}\theta_j\right)$$
$$for\ j = 1,2,3 \quad (6)$$

For θ_0, the update rule in gradient descent remains the same regardless of whether L1 or L2 regularization is applied.

$$\theta_0 := \theta_0 - \alpha\left(\sum_{i=1}^{m}\frac{1}{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\right) \quad (7)$$

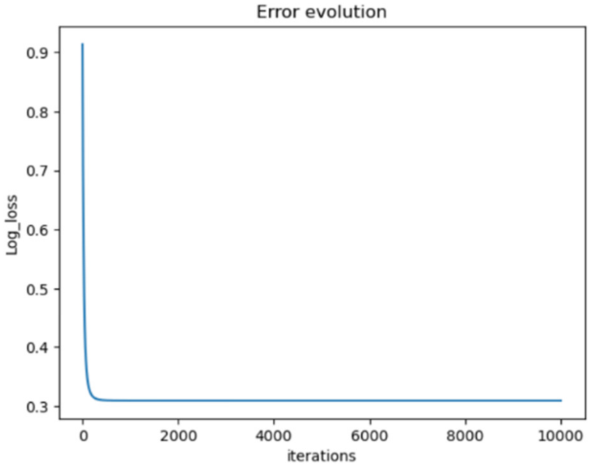

Fig. 1. Learning curve demonstrating convergence under gradient descent optimization.

B. *Predicting Student Performance Using Polynomial Regression*

After developing a classification algorithm capable of predicting early student dropout, an additional algorithm was developed to predict student scores. Specifically, a second-degree polynomial regression algorithm was developed to predict result_points based on access, tests, exam, project, and assignments. The model degree was selected based on the Bayesian Information Criterion (BIC) test, as illustrated in Figure 2. The BIC, based on Bayesian probability theory, is a criterion for selecting among a limited number of models. The BIC helps in choosing the model that optimally balances the observed data and preserves simplicity to avoid overfitting [25]. The model is given by:

$$h_\theta(x) = \theta_0 + \sum_{i=1}^{5}\theta_i x_i + \sum_{i=1}^{5}\theta_{ii}x_i^2 + \sum_{1\le i<j\le 5}\theta_{ij}x_i x_j \quad (8)$$

where θ_0 is the intercept, θ_i are the linear coefficients, θ_{ii} are the squared coefficients, and θ_{ij} are the interaction coefficients.

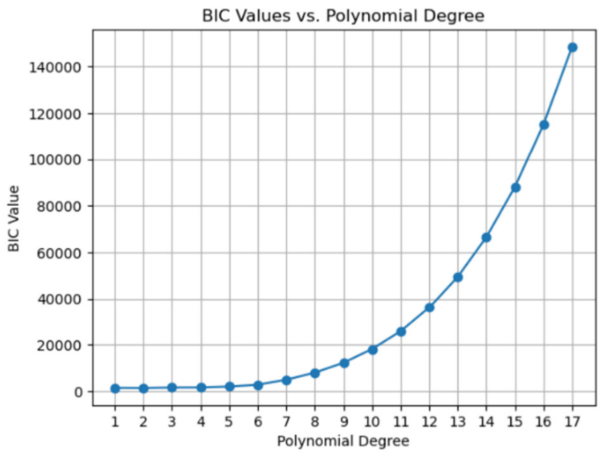

Fig. 2. Variation of BIC with model degree for polynomial regression.

Given the limited size of the dataset, the risk of overfitting is notably high. To address this issue, Lasso (L1) and Ridge (L2) regularization terms were incorporated into the cost function, which was calculated using the mean squared error. The cost function with the Lasso L1 regularization term is

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{m}\sum_{j=1}^{n}|\theta_j| \quad (9)$$

The cost function with the Ridge L2 regularization term is

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2 \quad (10)$$

III. RESULTS AND DISCUSSION

A. *Data Source*

The dataset published in [7] was used. The dataset was processed with respect to the CRISP-DM method, which includes several key phases. Initially, data understanding involved collecting and analyzing student performance data from the Moodle LMS. This was followed by data cleaning to address noise, inconsistencies, and missing values, ensuring data quality. Feature selection was then performed to identify relevant attributes that reflect student interactions with the course: access, tests, exam, project, assignments, and results_points.

- Train dataset: The anonymized educational training dataset used comes from 261 unique students registered in the introductory database systems course at Constantine the Philosopher University in Nitra from 2016 to 2019. It was obtained from the Moodle platform, with an average dropout rate of 21.92% [7].
- Test dataset: To evaluate the performance of the models with unseen data, the test dataset was constructed from student course activities during the 2020 academic year including 60 entries [7].

B. Logistic Regression (LR) Models With and Without Regularization

After training the three models to compare the impact of regularization on LR, L1-regularized LR, and L2-regularized LR, the accuracy results shown in Table I were obtained.

TABLE I. ACCURACY RESULTS OF LR MODELS

Model	Testing accuracy	Training accuracy	Difference between testing and training accuracy
LR	0.833333	0.900383	0.06705
L2-regularized LR	0.816667	0.896552	0.079885
L1-regularized LR	0.816667	0.900383	0.083716

When applying LR, the difference between testing and training accuracy was 0.06705. After adding the regularization Ridge-L2 term, the same difference was 0.079885, and for the regularization Lasso-L1 term, the same difference was 0.083716. As regularization techniques are designed to penalize model complexity, they can help reduce overfitting. Among these, the L2-regularized LR had the smallest difference (0.079885) between the training and testing accuracy. Although the testing accuracy was slightly lower than the unregularized LR, the reduced gap suggests a better generalization capability. In [7], the following accuracies were achieved on the same dataset: Naïve Bayes (NB) 0.77, Random Forest (RF) 0.93, Neural Network (NN) 0.88, LR 0.93, Support Vector Machines (SVM) 0.92, DT 0.90. RF and LR in [7] achieved an accuracy of 0.93. In contrast, the regularized LR models in this study achieved accuracies of 0.817. Despite these results being somewhat lower than the best-performing models, the use of regularization helps address overfitting in the small-sized dataset, as reflected in the differences between test and train accuracies, ranging from 0.067 to 0.083. The accuracy of these models can be enhanced by appropriate fine-tuning of the hyperparameters. L2 penalty is applied uniformly across all coefficients, leading to preserving all model parameters with smaller weights. However, the L1 penalty can drive some coefficients to zero. Due to the simplicity of the proposed model, which includes only three features that should be retained, L2 regularization is more appropriate and achieves the best balance between avoiding overfitting and preserving good performance on both the training and testing datasets.

C. Polynomial Regression Models With and Without Regularization

After training the three models to compare the impact of regularization on multiple polynomial regression, L1-regularized polynomial regression, and L2-regularized polynomial regression, the R^2 scores shown in Table II were obtained.

TABLE II. R^2 RESULTS OF POLYNOMIAL REGRESSION MODELS

Model	Training set	Test set	Difference between Test and Train R^2
Multiple polynomial regression	0.946216	0.852597	0.093619
L1 regularization	0.908705	0.863461	0.045244
L2 regularization	0.846004	0.765425	0.080579

For unregularized multiple polynomial regression, R^2 was 0.946216 on the training set and 0.852597 on the testing set. These results indicate that the model generalizes well on the training data by capturing most of the variance within the training set. However, the drop between training and test R^2, which is about 0.094, indicates some degree of overfitting. For the L1-regularized polynomial regression, R^2 for the training set was 0.908705 and for the testing set was 0.863461. The testing performance was quite strong and was significantly higher than the testing R^2 of the unregularized multiple polynomial regression. The smaller difference of 0.045 between training and testing R^2 indicates better generalization, suggesting that L1 regularization effectively reduced overfitting. For the L2-regularized polynomial regression, the R^2 score for the training set was 0.846004 and for the testing set was 0.765425. These results were the smallest among the three models. Despite the relatively small difference between the training and test R^2 values, which was about 0.081, this indicates a reasonable generalization, but the overall fit is not as strong as on the other models.

The multiple polynomial regression shows signs of overfitting with a higher training R^2 compared to the test R^2. The model captures the complexities of the training data well but does not generalize as effectively to new data. The L1-Lasso regularization strikes a good balance, as evidenced by its higher test performance and a smaller gap between training and testing R^2 values. This suggests that it generalizes well while avoiding overfitting. The L2-Ridge regularization, while still effective, did not perform as well on both the training and test sets. This is because there is a complex model with 21 coefficients, as shown in (8), for a small-sized dataset. Driving some coefficients to zero, as is the case of the L1 penalty, is a more effective method to penalize model complexity. In a similar context, in [17], different polynomial regression degrees were tested until they reached an R^2 score of 83.44% for degree 1. In this case, degree 2 was the highest model performer with an R^2 score of 86.43%. In summary, the L1 regularization model demonstrates the most balanced performance, achieving a strong balance between test performance and minimizing overfitting.

IV. CONCLUSION

This study presented a hybrid approach that combined two regression models to predict student dropout and results in a higher online education context. The first model was LR to predict the probability of student dropout, whereas the second model was multiple polynomial regression to predict student results. These two models were combined to obtain a deeper and earlier identification of student attrition. This study focused on an educational context where the number of registered students was limited. In such a case, training ML algorithms with a small-sized dataset could lead to a high risk of overfitting, where the model performs well on the training dataset but fails on the test dataset. To address this problem, the regularization terms L1-Lasso and L2-Ridge were tested on both predictive models. Regularized models demonstrated valuable performance, reaching a testing accuracy of 81.66% for L2-regularized LR and an R^2 of 86.43% for L1-ridge regularized multiple polynomial regression. These regularization techniques proved their effectiveness in balancing model complexity and generalizability. Future work should explore other regularization methods or hybrid approaches to further enhance predictive accuracy. Additionally, data augmentation, feature engineering, careful model selection, hyperparameter tuning, and further exploration of advanced techniques can significantly enhance predictive performance in educational data mining.

REFERENCES

[1] M. Alsuwaiket, A. H. Blasi, and R. A. Al-Msie'deen, "Formulating Module Assessment for Improved Academic Performance Predictability in Higher Education," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4287–4291, Jun. 2019, https://doi.org/10.48084/etasr.2794.

[2] B. Alsubhi *et al.*, "Effective Feature Prediction Models for Student Performance," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11937–11944, Oct. 2023, https://doi.org/10.48084/etasr.6345.

[3] S. Kim, E. Choi, Y.-K. Jun, and S. Lee, "Student Dropout Prediction for University with High Precision and Recall," *Applied Sciences*, vol. 13, no. 10, Jan. 2023, Art. no. 6275, https://doi.org/10.3390/app13106275.

[4] W. Hämäläinen and M. Vinni, "Classifiers for Educational Data Mining," in *Handbook of Educational Data Mining*, CRC Press, 2010.

[5] L. Ji, X. Zhang, and L. Zhang, "Research on the Algorithm of Education Data Mining Based on Big Data," in *2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)*, Xinxiang, China, Jun. 2020, pp. 344–350, https://doi.org/10.1109/CSEI50228.2020.9142529.

[6] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970, https://doi.org/10.1080/00401706.1970.10488634.

[7] J. Kabathova and M. Drlik, "Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques," *Applied Sciences*, vol. 11, no. 7, Jan. 2021, Art. no. 3130, https://doi.org/10.3390/app11073130.

[8] S. Halawa, D. Greene, and J. Mitchell, "Dropout Prediction in MOOCs using Learner Activity Features," *eLearning Papers*, no. 37 (This special issue of the eLearning Papers is based on the contributions made to the EMOOCS 2014 con), 2014.

[9] F. Ennibras, E. S. Aoula, and B. Bouihi, "AI in Preventing Dropout in Distance Higher Education: A Systematic Literature Review," in *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, FEZ, Morocco, May 2024, pp. 1–7, https://doi.org/10.1109/IRASET60544.2024.10548954.

[10] J. J. da Silva and N. T. Roman, "Predicting Dropout in Higher Education: a Systematic Review," in *Simpósio Brasileiro de Informática na Educação (SBIE)*, Nov. 2021, pp. 1107–1117, https://doi.org/10.5753/sbie.2021.217437.

[11] G. Gray, C. McGuinness, and P. Owende, "An application of classification models to predict learner progression in tertiary education," in *2014 IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, Feb. 2014, pp. 549–554, https://doi.org/10.1109/IAdCC.2014.6779384.

[12] B. R. Cuji Chacha, W. L. Gavilanes López, V. X. Vicente Guerrero, and W. G. Villacis Villacis, "Student Dropout Model Based on Logistic Regression," in *Applied Technologies*, Quito, Ecuador, 2020, pp. 321–333, https://doi.org/10.1007/978-3-030-42520-3_26.

[13] M. Vaarma and H. Li, "Predicting student dropouts with machine learning: An empirical study in Finnish higher education," *Technology in Society*, vol. 76, Mar. 2024, Art. no. 102474, https://doi.org/10.1016/j.techsoc.2024.102474.

[14] A. B. Altamimi, "Big Data in Education: Students at Risk as a Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11705–11714, Oct. 2023, https://doi.org/10.48084/etasr.6190.

[15] L. Kemper, G. Vorhoff, and B. U. Wigger, "Predicting student dropout: A machine learning approach," *European Journal of Higher Education*, vol. 10, no. 1, pp. 28–47, Jan. 2020, https://doi.org/10.1080/21568235.2020.1718520.

[16] Y. Yang, "Sparse Logistic Regression with the Hybrid L1/2+1 Regularization," in *2021 6th International Conference on Mathematics and Artificial Intelligence*, Chengdu, China, Mar. 2021, pp. 8–13, https://doi.org/10.1145/3460569.3460578.

[17] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms," *Materials Today: Proceedings*, vol. 47, pp. 5260–5267, Jan. 2021, https://doi.org/10.1016/j.matpr.2021.05.646.

[18] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, "A novel methodology using RNN + LSTM + ML for predicting student's academic performance," *Education and Information Technologies*, vol. 29, no. 11, pp. 14365–14401, Aug. 2024, https://doi.org/10.1007/s10639-023-12394-0.

[19] Q. Huang and Y. Zeng, "Improving academic performance predictions with dual graph neural networks," *Complex & Intelligent Systems*, vol. 10, no. 3, pp. 3557–3575, Jun. 2024, https://doi.org/10.1007/s40747-024-01344-z.

[20] Q. Huang and Y. Zeng, "Improving academic performance predictions with dual graph neural networks," *Complex & Intelligent Systems*, vol. 10, no. 3, pp. 3557–3575, Jun. 2024, https://doi.org/10.1007/s40747-024-01344-z.

[21] Y. Yamasari, N. Rochmawati, R. E. Putra, A. Qoiriah, Asmunin, and W. Yustanti, "Predicting the Students Performance using Regularization-based Linear Regression," in *2021 Fourth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, Surabaya, Indonesia, Oct. 2021, pp. 1–5, https://doi.org/10.1109/ICVEE54186.2021.9649704.

[22] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *Journal of Applied Research in Higher Education*, vol. 10, no. 1, pp. 61–75, Jan. 2018, https://doi.org/10.1108/JARHE-09-2017-0113.

[23] E. Evangelista and B. Sy, "An approach for improved students' performance prediction using homogeneous and heterogeneous ensemble methods," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5226–5235, Oct. 2022, https://doi.org/10.11591/ijece.v12i5.pp5226-5235.

[24] U. Michelucci, "Logistic Regression from Scratch," in *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*, U. Michelucci, Ed. Berkeley, CA, USA: Apress, 2018, pp. 391–401.

[25] K. H. Pho, S. Ly, S. Ly, and T. M. Lukusa, "Comparison among Akaike Information Criterion, Bayesian Information Criterion and Vuong's test in Model Selection: A Case Study of Violated Speed Regulation in Taiwan," *Journal of Advanced Engineering and Computation*, vol. 3, no. 1, pp. 293–303, Mar. 2019, https://doi.org/10.25073/jaec.201931.220.