NUEJ
Naresuan University Engineering Journal

A Comparative Study of the Applicability of Regression Models in Predicting Student Academic Performance

Wenting Cui, Adisak Sangsongfar, and Noppadol Amdee*

Department of Industrial Technology Management, Faculty of Industrial Technology, Muban Chom Bueng Rajabhat University, Ratchaburi, Thailand 70150

*Corresponding author e-mail: noppadolamd@mcru.ac.th

(Received: 22 February 2024, Revised: 9 June 2024, Accepted: 27 June 2024)

Abstract

Educational Data Mining (EDM) has witnessed a surge in educational systems, as it enables the analysis and prediction of student performance, facilitating proactive measures. This paper aims to present a comparative study of regression prediction using eight mainstream models: Linear Regression, Ridge Regression, Lasso Regression, Huber Regression, Support Vector Regression (SVR), K-Nearest Neighbors Regression (KNN), Decision Tree Regression (DT), and Neural Network Regression. The applicability of these eight models is analyzed across different courses and semester GPAs, considering three distinct scenarios. Our thorough analysis underscores the substantial influence of data granularity and integrity on bolstering the precision of final CGPA predictions. In the third scenario where Semester GPAs were utilized, Lasso Regression achieved an R-value of 0.9901 with remarkably low RMSE and MAE, establishing its dominance across all scenarios. Neural Network Regression, with an R-value of 0.9832 and minimal error metrics in the same scenario, also demonstrated robust predictive capabilities. These insights highlight the imperative of tailoring regression model selection to align with specific scenario nuances and the targeted predictive precision.

Keywords: Educational Data Mining, Regression models, Student Academic Performance

1. INTRODUCTION

The emergence of big data has garnered increasing attention in education. A pivotal challenge within educational data mining is the development of student models capable of accurately predicting academic performance (Bydžovská, H., 2016). In the early 2000s, Xiushan NIE et al. (2022) pioneered association rule mining techniques to identify students requiring course remediation, paving the way for a substantial body of research focused on student performance prediction through machine learning and data mining methodologies. The query rule "student performance prediction (topic) and predicting student performance (topic) and machine learning" was entered into the "Web of Science". The results show that over half of the literature has been published in the last five years, further demonstrating the increasing interest in student academic performance prediction.

As machine learning theories and methodologies continue evolving and related technologies advance rapidly, scholars have extensively researched student performance prediction. This research can be broadly categorized into five main areas: (1) investigations into the factors influencing students' performance, (2) studies focusing on predicting students' performance in online exams, (3) research on predicting students' performance in individual courses, (4) efforts aimed at predicting students' overall GPA performance, and (5) application in real teaching scenarios, such as college alert, personalized learning, adaptive learning system, learning resource recommendation, etc. From a large amount of research literature, the current student performance prediction is still mainly based on classification problems. The prediction models are mostly single, and the corresponding classification prediction results cannot be effectively analyzed for the prediction results (Ma, Y. et al., 2000). In the process of course grade prediction, the corresponding models fit differently depending on the predicted course. Moreover, the more technical literature is not the most suitable learning resource for education and teaching administrators who lack theoretical and technical knowledge of algorithms, perhaps due to factors such as poor readability and difficulty in understanding, which becomes one of the constraints for student achievement prediction research to be truly useful in real teaching scenarios.

To address the above issues, this paper intends to conduct regression type of grade prediction and use eight mainstream regression prediction algorithms: Linear Regression, Ridge Regression, Lasso Regression, Huber Regression, Support Vector Regression (SVR), K-Nearest Neighbors Regression (KNN), Decision Tree Regression (DT), and Neural Network Regression (NN), to analyze their characteristics under different course attributes. The eight algorithms are compared in terms of prediction accuracy, error analysis, and prediction distribution, and different training samples are selected to provide a reference basis for suitable prediction algorithms for predicting college students' course grades.

2. LITERATURE REVIEW

Over the past few years, a notable surge in research has been focused on predicting student academic performance. The analyzed literature reveals that researchers often employ multiple algorithms to develop predictive models (Kumar et al., 2018). The most commonly utilized algorithms in predictive modeling encompass a variety of approaches, including Decision Trees, Bayesian Classifiers, Neural Networks, Support Vector Machines, K-Nearest Neighbors, and Logistic and Linear Regression. The selection of algorithms is contingent upon factors such as the problem type, the characteristics of the outcome to be predicted, and the variables employed in the prediction process. It is customary for researchers to experiment with multiple algorithms and assess their performance to identify the most accurate prediction methodology. Studies consistently demonstrate that the performance of algorithms varies across different datasets, with Neural Networks often outperforming other algorithms in terms of accuracy (Cavazos et al., 2017; Vijayalakshmi et al., 2019; Bravo et al., 2020; Makombe et al., 2020; Mengash, 2020; Waheed et al., 2020).

Numerous studies have undertaken comparisons of Decision Tree, Artificial Neural Networks (ANN), and Regression algorithms in their effectiveness at predicting student academic performance. These comparisons consistently demonstrate that ANN yields the best results (Mutanu & Machoka, 2019). In addition, a study by Iyanda (2018) found that Regression Neural Network outperformed MLP (Multi-Layer Perceptron). Furthermore, Hussain and Khan (2021) and Mai et al. (2022) utilized Gradient Boosting and Decision Tree algorithms to forecast students' overall performance. El Aouif et al. (2021) employed the K-Nearest Neighbor (KNN) algorithm for predicting learners' performance, while Cutad and Gerardo (2019) applied it for curriculum analysis.

In the specific articles examined, Tomasevic et al. (2020) employed Artificial Neural Networks (ANN), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees to predict student performance, drawing from a combination of grades, behavioral, and demographic data. Meanwhile, Sa'ad and Mustafa (2020) utilized ELM (Extreme Learning Machines), SVM, and ANN methodologies to forecast the dropout rate of PhD students, relying primarily on behavioral data. Bujang et al. (2021) explored Naive Bayes, KNN, and Logistic Regression models to predict semester-end grades using demographic and grade data. However, these studies primarily focused on specific prediction tasks and had limitations regarding the comparison methods used and the data types considered.

In the realm of regression problems, both Single and Multiple Linear Regression have been employed for prediction tasks. Hsu Wang (2019) utilized the multiple linear regression model to forecast students’ online behavior and academic achievement. Similarly, Albalooshi et al. (2019) and Yang et al. (2018) employed this approach to assess learners’ academic performance. Meanwhile, Single Linear Regression was utilized by Omer et al. (2020) for performance evaluation and by Tuononen and Parpala (2021) to predict students’ thesis grades.

In our proposed approach, we focus on regression problems to determine the applicability of regression algorithms and emphasize the importance of selecting appropriate regression methods based on the specific scenario and desired predictive accuracy. Based on the analysis of the aforementioned related work, we have chosen eight of the most popular regression algorithms researchers use. The methodologies employed encompass a range of regression techniques, including Linear Regression, Ridge Regression, Lasso Regression, Huber Regression, SVR, KNN, DT, and Neural Network Regression, all aimed at predicting student academic performance.

3. METHODOLOGY

3.1 Proposed Approach

The proposed steps in this paper, based on eight main regression prediction methods for student performance, are as follows: First, retrieve student information from the university's academic affairs database. Second, data cleaning is performed, and the extracted data is converted into a suitable format. Third, data pre-processing on the cleaned data must be carried out to ensure its quality and relevance. Fourth, conduct course feature selection and choose different types of courses for prediction. Fifth, evaluate the application of the models, select appropriate algorithms for comparison and analysis, and recommend suitable algorithms for different courses. The specific algorithm flow is illustrated in Figure 1.

3.2 Regression algorithms used

In this paper, eight regression models were utilized following the steps of regression analysis. The models encompass Linear Regression, Ridge Regression, Lasso Regression, Huber Regression, SVR, KNN, DT, and Neural Network Regression. Each model possesses distinct characteristics and applicability. Linear regression establishes a linear relationship using the least squares method, while Ridge and Lasso regressions introduce regularization to control complexity and promote sparsity, respectively. Huber regression reduces sensitivity to outliers through the Huber loss function. SVR handles nonlinear problems and complex structures, KNN captures local patterns, DT handles nonlinear relationships and interactions, and Neural Network Regression constructs complex models for large-scale and high-dimensional data. These models were chosen to analyze regression performance and accuracy comprehensively.

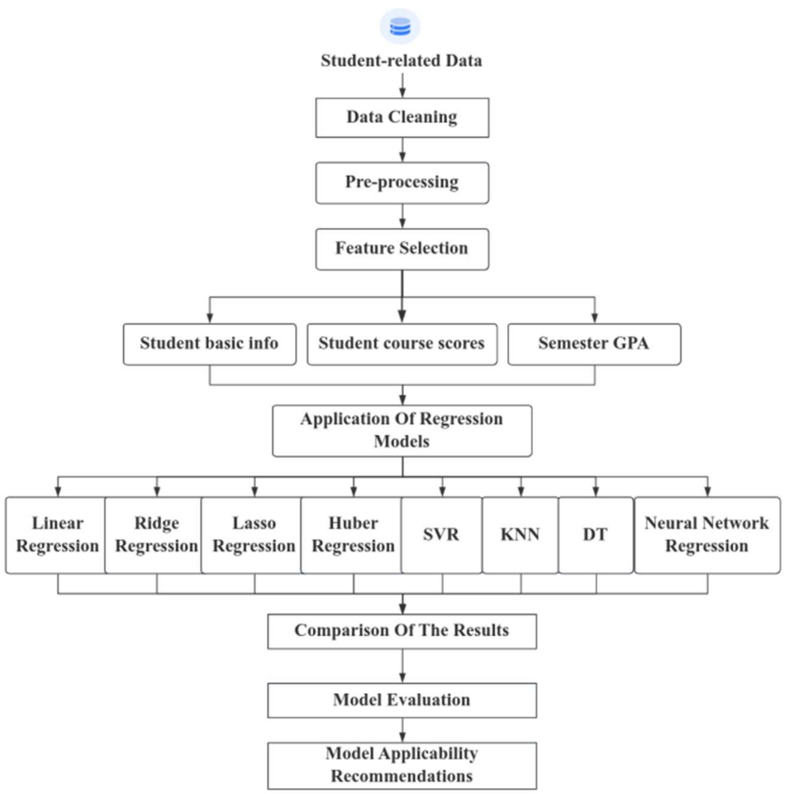


Figure 1. Predictive Comparison Flow Chart

3.3 Data description

The eight prediction algorithms were implemented for comparative analysis, and 21 courses in four grades of Computer Science and Technology under the School of Information Engineering of a university, namely 2017, 2018, 2019, and 2020, were used as experimental data. In this paper, for each student, thirty-four (34) data points were collected, including the final CGPA at graduation (Y), the values of thirty-three predictor variables (from X1 to X33), and a total of 628 students, 628 × 34 = 21,352 data points were collected. The collected data (Y, X1, X2, X3, ……, X33) were initially in different scales of measurements: X1 – X4 were encoded in different numbers, X5 – X25 varied from 0.00 to 100.00, X26 – X33 varied from 0.00 to 5.00, and Y varied from 0.00 to 5.00. The student-related attributes are given in Table 1. Before using them to establish a prediction model, the collected raw data must be pre-processed, as described in the following paragraphs.

Table 1 Student-related attributes

Attributes	Description	Value
Student ID	The ID for a student	TEXT
Sex	The Sex of the student	1 F, 2 M
Entry year	The year the student enrolled	2017-2021
Birth place	The place where the student birth	Encoded from 1-21
Dormitory type	The university dormitory types	1,2,3,4,6
Course scores	The grade obtained by the student	0-100
Semester GPA	The student's GPA at the end of each semester	0-5
Final CGPA	The student's CGPA at semester eight(graduation) or for all the semesters passed	0-5

3.4 Data cleaning and pre-processing

Since the extracted data from the academic affairs database could not be directly applied to the prediction model, it was necessary to clean and transform the data to meet the requirements of the prediction model. The following operations were performed on the original data:

(1) Data cleaning:

a) Most student performance data with zero course grades were deleted, as these data points had limited validity and could introduce noise.

b) Selecting only courses taken consistently across all years and having consistent assessment evaluations.

c) Normalizing the performance data. The grades for each course in each data set were normalized to the range [0,1], as shown in Table 2. This normalization helped alleviate the effects of inconsistent grade distributions across courses.

(2) Missing value processing: The data set was analyzed descriptively to identify missing values. Missing data in the student's course grades could be attributed to factors such as students taking breaks from school or changing majors. However, the proportion of missing grades was very small, so the samples with missing grades were removed.

(3) Establishing a training sample: 80% of the data was allocated for training, while the remaining 20% was reserved as a test sample.

Table 2 Normalization table

Attributes	Description	Value
X1 - X4	The basic information for a student	0.00 - 1.00
X5 - X25	Course scores for a student	0.00 - 1.00
X26 - X33, Y	X26 - X33 is the Semester GPA, Y is the final CGPA	0.00 - 1.00

3.5 Model evaluation

To validate the prediction models, 10-fold cross-validation was employed. The student-related data set was randomly divided into 10 subsets, with eight subsets used for training and the remaining two for testing. This process was repeated 10 times, and the accuracy of the models was computed based on the results.

The study focused on predicting the target variable, the overall final CGPA, which is a numeric variable ranging from 0.00 to 5.00.

Consistent with most literature, this paper utilized root mean square error (RMSE) and mean absolute error (MAE) as evaluation metrics. RMSE quantifies the deviation between predicted and true values, offering insights into the overall prediction accuracy. On the other hand, MAE reflects the actual magnitude of prediction errors. The formulas for both metrics are as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2} \quad (1)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|y_i - \hat{y}_i\right| \quad (2)$$

where m is the sample size, Y_i is the true value, and $\hat{Y}_i$ is the predicted value.

3.6 Feature Selection

To conduct the experiments, a feature subset selection process was performed. Feature selection involves identifying and removing irrelevant features from the data set, specifically those that do not contribute significantly to the task at hand. This process is essential as it helps reduce the dimensionality of the data, thereby improving the efficiency of the learning algorithm, reducing execution time, and enhancing predictive accuracy.

To better understand the variables, we analyzed the correlation of all variables as the basis for feature selection in corresponding scenarios. As shown in Figure 2, X1-X4 shows a low correlation with the results, while X7, X8, X11-X13, X15-X17, X19-X22 show a high correlation with the CGPA(Y), all of which are core professional courses.

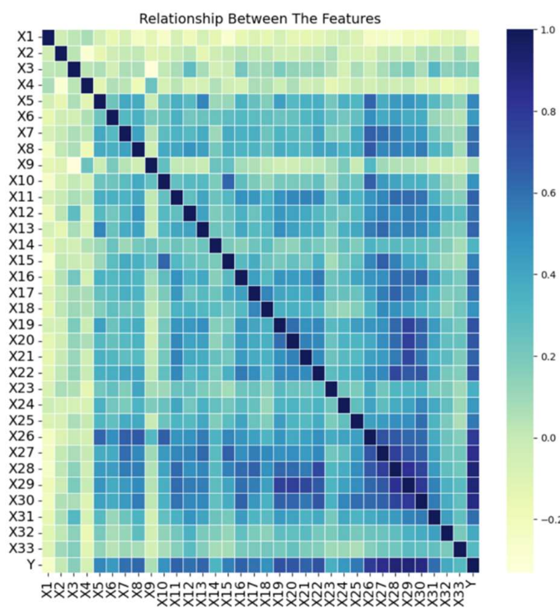


Figure 2. Relationship Between the Features

For our experiment, we considered three different combinations of predictors, all aiming to predict the final CGPA of students:

First Scenario: We utilized the students' university course scores from the first 2 years, encompassing the scores of 14 courses, as input features for predicting the final CGPA.

Second Scenario: We expanded the feature set to include the students' university course scores from the first 3 years, incorporating the scores of 21 courses to predict the final CGPA.

Third Scenario: Instead of relying on course scores, we utilized the students' Semester GPA at the end of each semester from the first 3 years of courses as input features for predicting the final CGPA.

By exploring these different scenarios, our objective was to analyze how different input features influence the accuracy of predicting the final CGPA. Through feature selection, we sought to identify the most relevant and informative predictors that would yield the best predictive performance.

4. RESULTS

4.1 Experiment of the First Scenario

In this experiment, we evaluated the performance of various regression methods, namely Linear Regression, Ridge Regression, Lasso Regression, Huber Regression, SVR, KNN, DT, and Neural Network Regression. These methods were tested specifically for the first scenario described in section 3.6.

Table 3 Prediction result for the First Scenario

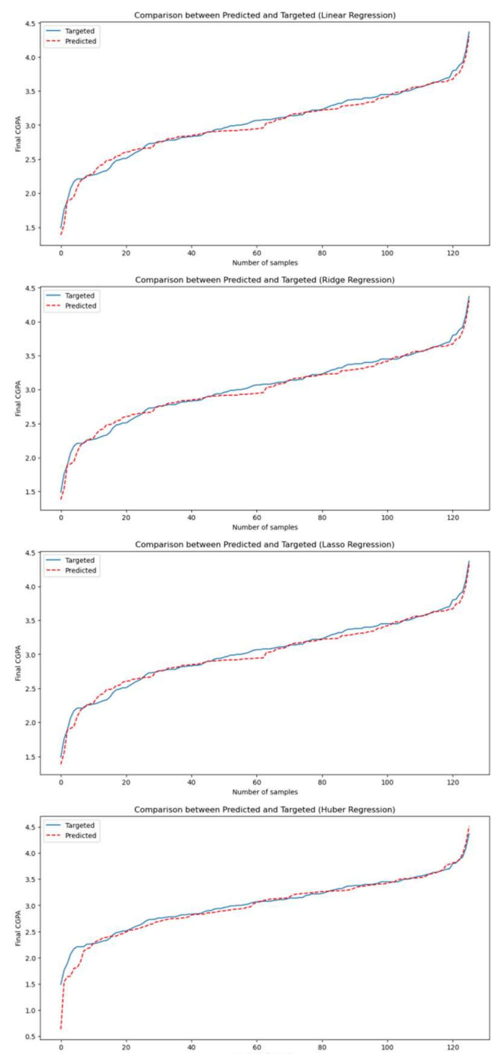
Methods	R (Correlation coefficient)	RMSE	MAE
Linear Regression	0.8769	0.4162	0.1353
Ridge Regression	0.8774	0.4158	0.1352
Lasso Regression	**0.8777**	**0.1726**	**0.1351**
Huber Regression	0.6762	0.2809	0.2089
SVR	0.8034	0.2188	0.1579
KNN	0.8230	0.2076	0.1503
DT	0.5753	0.3217	0.2390
Neural Network	**0.8752**	**0.1743**	**0.1374**

As the results indicate in Table 3, The first three linear regression algorithms showed high correlation coefficients (R), indicating a strong linear relationship between input features and the target variable. These models demonstrated strong predictive capability, with R values ranging from 0.8769 to 0.8777. They also achieve MAE values ranging from 0.1351 to 0.1353, indicating their ability to provide predictions closer, on average, to the true CGPA values. Lasso Regression and Neural Network Regression achieved the lowest mean squared error (RMSE), indicating better overall predictive accuracy and reduced squared differences between predicted and actual CGPA values. Huber Regression and Decision Tree Regression demonstrated relatively lower performance in terms of correlation coefficient and predictive accuracy metrics, suggesting they may not effectively capture the underlying relationships between input features and the final CGPA.

In addition to quantitative metrics, the prediction results were qualitatively evaluated by plotting both the predicted and target values, as shown in Figure 3.

Figure 3 shows that the SVR and KNN models have considerable deviation between the predicted and target values at lower and higher CGPA ranges. The overlap between the predicted and target values is most pronounced for Lasso Regression and Neural Network Regression, indicating their superior performance.

Overall, Lasso Regression and Neural Network Regression are promising models for predicting final CGPA using the students' university course scores from the first 2 years.



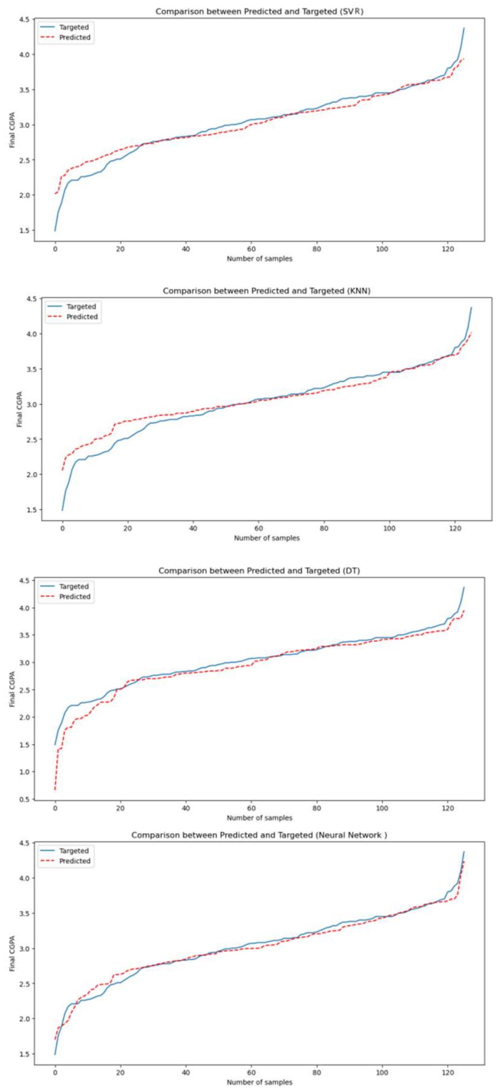


Figure 3: Prediction result of the eight methods for the First Scenario

4.2 Experiment of the Second Scenario

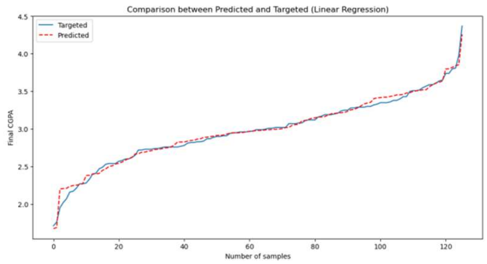
The experiments were repeated for the second scenario. Table 4 presents the results of the three prediction methods for this scenario.

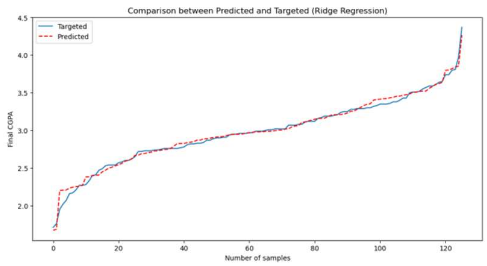
Table 4 Prediction result for the Second Scenario

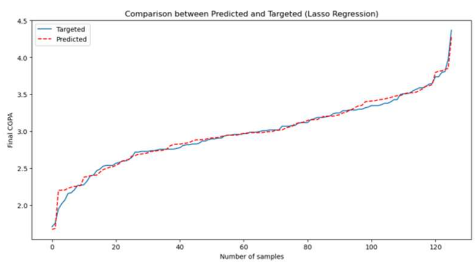
Methods	R (Correlation coefficient)	RMSE	MAE
Linear Regression	0.9436	0.3279	0.0855
Ridge Regression	0.9436	0.3279	0.0854
Lasso Regression	**0.9442**	**0.1070**	**0.0849**
Huber Regression	**0.9396**	**0.1114**	**0.0882**
SVR	0.8002	0.2025	0.1263
KNN	0.8373	0.1827	0.1374
DT	0.5247	0.3123	0.2497
Neural Network	**0.9248**	**0.1243**	**0.0986**

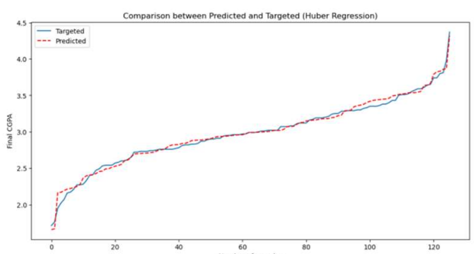
The prediction results for the second scenario are summarized in Table 4. Among the evaluated regression methods, Lasso Regression attained the highest correlation coefficient (R) of 0.9442, denoting a robust linear relationship between the input features and the target variable. Furthermore, it showcased the lowest RMSE value of 0.1070 and the lowest MAE value of 0.0849, suggesting superior predictive accuracy with minimal squared differences and closer predictions to the true values. Ridge Regression performed similarly to Lasso Regression, with a correlation coefficient (R) of 0.9436 and slightly higher RMSE and MAE values. Huber Regression also showed good performance with a correlation coefficient (R) of 0.9396 and relatively low RMSE and MAE values. On the other hand, SVR, KNN, and DT exhibited lower correlation coefficients and higher RMSE and MAE values, indicating less accurate predictions. Neural Network Regression achieved a correlation coefficient (R) of 0.9248, demonstrating a strong linear relationship, and showcased competitive performance with low RMSE and MAE values. Overall, Lasso Regression, Ridge Regression, and Neural Network Regression emerged as the top-performing methods for the second scenario, providing accurate predictions and capturing the underlying relationships between the input features and the target variable.

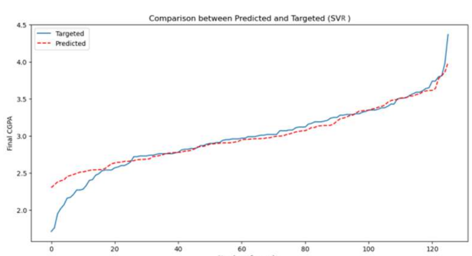
The prediction results evaluated by plotting both the predicted and target values also show that the three models have the best performance, as depicted in Figure 4.

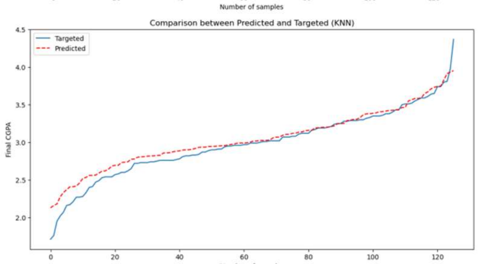


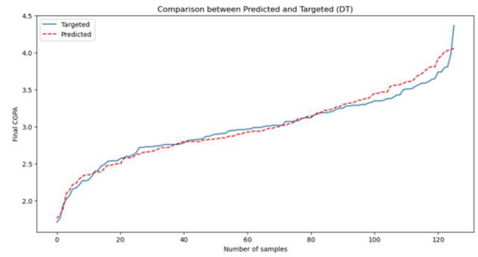












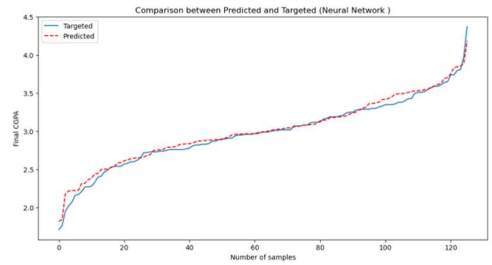


Figure 4: Prediction result of the eight methods for the Second Scenario

4.3 Experiment of the Third Scenario

The prediction results for the third scenario are summarized in Table 5.

Table 5 Prediction results for the Third Scenario

Methods	R (Correlation coefficient)	RMSE	MAE
Linear Regression	0.9905	0.2311	0.0413
Ridge Regression	0.9900	0.2341	0.0423
Lasso Regression	**0.9901**	**0.0546**	**0.0428**
Huber Regression	**0.9899**	**0.0550**	**0.0438**
SVR	**0.9882**	**0.0595**	**0.0459**
KNN	0.8549	0.2085	0.1462
DT	0.9076	0.1665	0.1262
Neural Network	**0.9832**	**0.0710**	**0.0555**

In the third scenario, the shift to Semester GPA as the predictive feature marked a significant advancement in modeling accuracy, as evidenced by the strikingly higher R values across all models. This transition from individual course scores to a cumulative measure of academic performance over three semesters encapsulates a more integrated view of student learning. The Lasso Regression and Huber Regression models, in particular, excelled with R values of 0.9901 and 0.9899, respectively, indicating an exceptionally strong linear relationship between the Semester GPA and the final CGPA. Moreover, these models' RMSE and MAE values were notably lower, suggesting a more precise prediction of student outcomes. The superior performance in the third scenario, as depicted in Figure 5, underscores the predictive power of aggregated academic metrics and the effectiveness of regularization techniques in regression analysis.

When synthesized, the comparative analysis of the three scenarios reveals a clear trend: the predictive models' accuracy escalates with the use of more comprehensive and aggregated academic data. The third scenario, focusing on Semester GPA, provides a more stable and generalized prediction and aligns with the educational systems' objective to assess students' overall performance rather than isolated course achievements. The consistent dominance of Lasso Regression, as observed in Figure 5, reinforces its utility in educational data mining, where the ability to handle complex datasets and multicollinearity is paramount. Additionally, Neural Network Regression has consistently performed consistently throughout all scenarios, suggesting its capacity to model complex, nonlinear relationships effectively. These insights are invaluable for institutions seeking to optimize their predictive models, suggesting that the strategic employment of aggregated GPA data can substantially enhance the forecasting of student success.

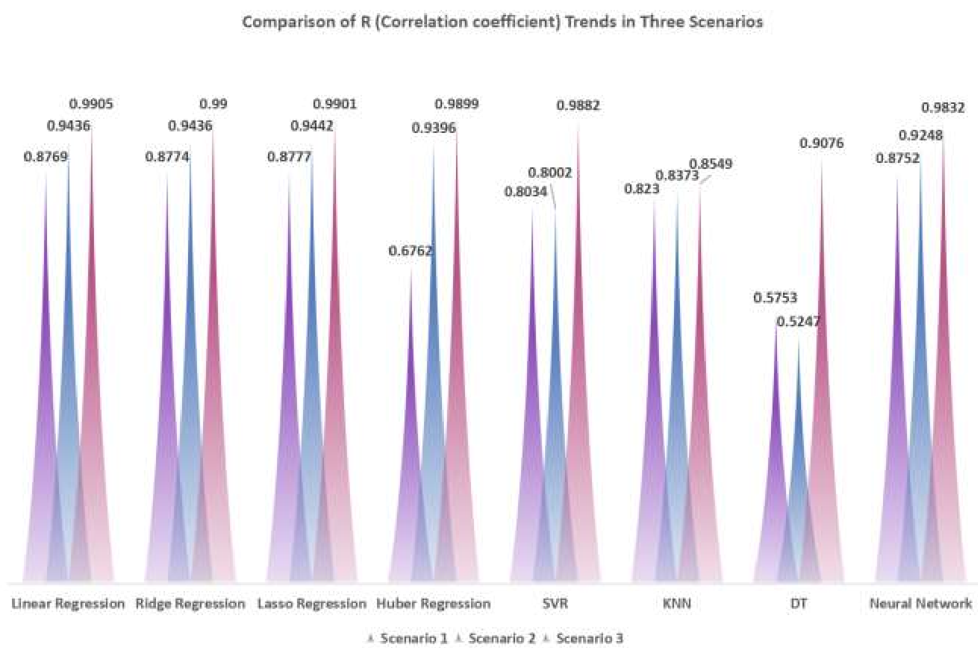


Figure 5: Comparison of Results Trends in Three Scenarios

4.4 Experimental Analysis

The experimental outcomes across three scenarios underscore the pivotal role of data quantity and type in enhancing predictive model accuracy. Initially, employing 14-course scores from the first two academic years, Lasso Regression distinguished itself with a high correlation coefficient and minimized error metrics, showcasing its adeptness at discerning patterns within a moderate feature set. This trend intensified in the subsequent scenario, expanding to 21-course scores across three years. Lasso Regression's performance peaked with an even higher correlation coefficient and significantly reduced RMSE and MAE, underscoring the value of comprehensive data in refining predictive accuracy.

The transition to utilizing Semester GPA as an aggregated performance metric in the culminating scenario engendered a significant leap in the predictive efficacy of the models, particularly Lasso Regression and Huber Regression, which showcased exceptional correlation values and minimized error metrics. This advancement illustrates a heightened sensitivity and precision when models are based on comprehensive academic data. The incremental refinement from discrete course scores to consolidated GPA metrics has enhanced the models' predictive capabilities. The consistent outperformance of Lasso Regression is attributed to its advanced feature selection and regularization techniques, which together advocate for the strategic use of aggregated and representative data in academic forecasting. These findings underscore educational institutions' need to adopt data-driven methodologies, providing a robust framework for evidence-based decision-making and strategic planning to augment student success.

4.5 Model Applicability Suggestions

Based on the findings from the experiments, certain regression models demonstrate better applicability for specific scenarios. Lasso Regression consistently performs well across all three scenarios, indicating its robustness and effectiveness in capturing the underlying relationships between input features and the target variable. Neural Network Regression is a competitive model, showing strong linear relationships and relatively accurate predictions. These two models, Lasso Regression and Neural Network Regression, can be recommended for various applications where accurate prediction and capturing complex interactions are crucial. However, it's crucial to consider each scenario's unique requirements and choose the regression method accordingly. This is because different models may exhibit varying performance levels based on the characteristics of the data and the relationships between the variables.

5. CONCLUSION AND DISCUSSION

5.1 Conclusion

The experimental analysis underscores the efficacy of Lasso Regression and Neural Network Regression in predicting student grades across varying scenarios, emphasizing their ability to capture complex relationships with high accuracy. Lasso Regression, in particular, has proven to be a robust model, consistently outperforming others with its high correlation coefficients and minimized RMSE values. The progression from the First to the Third Scenario illustrates the substantial impact of data quality and aggregation on predictive accuracy, with the use of Semester GPA in the Third Scenario yielding the most significant improvements. These findings underscore the importance of model selection tailored to specific scenarios and highlight Lasso Regression as a top performer, particularly suitable for applications requiring precise predictions and handling complex interactions. As we advance, further exploration into additional regression techniques and hybrid models could potentially enhance predictive accuracy and broaden the scope of regression analysis in diverse fields.

In conclusion, our comprehensive evaluation has demonstrated that the depth and quality of academic data significantly enhance the predictive accuracy of students' final CGPA. Lasso Regression has emerged as a consistently superior model, with its predictive power notably increasing as the data set expanded from 14-course scores to a more holistic measure of Semester GPA. The results highlight the profound impact of representative and aggregated data features on model efficacy, advocating for the strategic use of such data in educational modeling and decision-making processes.

5.2 Discussion

Our study contributes to the body of research on applying regression models in predicting student academic performance. Through a comparative analysis of eight regression models, we have identified Lasso Regression and Neural Network Regression as particularly effective in predicting final CGPA, especially when using Semester GPAs as input data. Our findings resonate with those of Lyu (2023), who found Random Forest to be the best model for predicting student grades based on past scores, highlighting the potential of machine-learning techniques in educational data mining.

Moreover, our results align with the study by Kumar et al. (2020), which applied regression analysis to accurately predict student performance, underscoring the importance of past academic records as a strong predictor of future performance. The high R values and low error metrics observed in our study are consistent with the effectiveness of regression models in capturing the nuances of student performance, as demonstrated in the work of Lyu (2023) and Kumar et al. (2020).

Our study extends the existing literature by examining the impact of data granularity and integrity on the precision of CGPA predictions. The dominance of Lasso Regression in our third scenario, with an R-value of 0.9901 and minimal error metrics, is particularly noteworthy. This aligns with the findings of Rusli and Ibrahim (2007), who reported high accuracy in predicting academic performance using artificial neural networks, suggesting that advanced regression techniques can outperform traditional methods.

Furthermore, our comparative study adds depth to the work of Arsad et al. (2014), who compared neural networks and linear regression for predicting academic achievement by providing a more comprehensive analysis of various regression models. Our results indicate that while Neural Network Regression showed robust predictive capabilities, Lasso Regression was superior in prediction accuracy.

In conclusion, our study emphasizes the importance of selecting the appropriate regression model tailored to the specific characteristics of the data and the predictive precision required. The high predictive accuracy of Lasso Regression and Neural Network Regression in our study, along with the support from the literature, validates the effectiveness of these models in predicting student academic performance. Future research should continue to explore the integration of various regression models and their potential for enhancing student performance prediction in educational settings.

6. ACKNOWLEDGMENT

Data from the School of Information and Engineering, Wenzhou Business College, supported this work. The authors thank Wenzhou Business College and Muban Chom Bueng Rajabhat University for their support.

7. References

Albalooshi, F., Alobaidy, H., & Ghanim, A. (2019). Mining students outcomes: An empirical study. International Journal of Computing and Digital Systems, 8(3), 229–241. https://doi.org/10.12785/ ijcds/080303

Arsad, P. M., & Buniyamin, N. (2014, April). Neural Network and Linear Regression methods for prediction of students' academic achievement. In 2014 IEEE Global Engineering Education Conference (EDUCON) (pp. 916-921). IEEE. https://doi.org/10.1109/EDUCON.2014.6826206

Bravo, L. E. C., Molano, J. I. R., & Trujillo, E. R. (2020). Exploration of a system to determine the academic performance of engineering students through machine learning. International Journal of Advanced Science and Technology, 29(7), 11894–11905. http://sersc.org/journals/index.php/IJAST/article/view/27865

Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass prediction model for student grade prediction using machine learning. Ieee Access: Practical Innovations, Open Solutions, 9, 95608–95621. https://doi.org/10.1109/ACCESS.2021.3093563

Bydžovská, H. (2016). A Comparative Analysis of Techniques for Predicting Student Performance. International Educational Data Mining Society. https://doi.org/10.1109/FIE.2007.4417993

Cavazos, R., & Garza, S. E. (2017). Learning models for student performance prediction. Paper presented at the Mexican International Conference on Artifcial Intelligence. https://doi.org/10.1007/978-3-030-02840-4_14

Cutad, R. E. E., & Gerardo, B. D. (2019). A prediction-based curriculum analysis using the modifed artifcial Bee Colony Algorithm. International Journal of Advanced Computer Science and Applications, 10(10), 117–123. https://doi.org/10.14569/ijacsa.2019.0101017

El Aouifi, H., El Hajji, M., Es-Saady, Y., & Douzi, H. (2021). Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining. Education and Information Technologies, 26(5), 5799-5814. https://doi.org/10.1007/s10639-021-10512-4

Hsu Wang, F. (2019). On prediction of online behaviors and achievement using self-regulated learning awareness in fipped classrooms. International Journal of Information and Education Technology, 9(12), 874–879. https://doi.org/10.18178/ijiet.2019.9.12.1320

Hussain, S., & Khan, M. Q. (2021). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of Data Science. https://doi.org/10.1007/s40745-021-00341-0

Ibrahim, Z., & Rusli, D. (2007, September). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In 21st Annual SAS Malaysia Forum, 5th September. https://www.researchgate.net/publication/228894873

Iyanda, A. R., Ninan, O. D., Ajayi, A. O., & Anyabolu, O. G. (2018). Predicting student academic performance in computer science courses: A comparison of neural network models. International Journal of Modern Education & Computer Science, 10(6). https://dx.doi.org/10.5815/ijmecs.2018.06.01

Lyu, S. (2023, November). Student's Academic Performance Prediction Based on Machine Learning Regression Models. In 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023) (pp. 293-299). Atlantis Press. https://doi.org/10.2991/978-94-6463-300-9_29

Kumar, A., Eldhose, K. K., Sridharan, R., & Panicker, V. V. (2020, July). Students' Academic Performance Prediction using Regression: A Case Study. In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-6). IEEE. https://doi.org/10.1109/ICSCAN49426.2020.9262346

Kumar, A., Selvam, R., & Kumar, K. (2018). Review on prediction algorithms in educational data mining. International Journal of Pure and Applied Mathematics, 118, 531–536. https://www.acadpubl.eu/jsi/2018-118-7-9/articles/8/77.pdf

Ma, Y., Liu, B., Wong, C. K., Yu, P. S., & Lee, S. M. (2000, August). Targeting the right students using data mining. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 457-464). http://scholarbank.nus.edu.sg/handle/10635/39963

Mai, T. T., Bezbradica, M., & Crane, M. (2022). Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data. Future Generation Computer Systems, 127, 42–55. https://doi.org/10.1016/j.future.2021.08.026

Makombe, F., & Lall, M. (2020). A predictive model for the determination of academic performance in private higher education institutions. International Journal of Advanced Computer Science and Applications (IJACSA), 11(9). http://dx.doi.org/10.14569/IJACSA.2020.0110949

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

Mutanu, L., & Machoka, P. (2019). Enhancing computer students' academic performance through predictive modelling-a proactive approach. In Paper presented at the 2019 14th international conference on Computer Science & Education (ICCSE). https://doi.org/10.1109/ICCSE.2019.8845452

Omer, U., Farooq, M. S., & Abid, A. (2020). Cognitive learning analytics using assessment data and concept map: A framework-based approach for sustainability of programming courses. Sustainability (Switzerland), 12(17). https://doi.org/10.3390/su12176990

Sa'ad, M. I., & Mustafa, M. S. (2020, October). Student Prediction of Drop Out Using Extreme Learning Machine (ELM) Algorithm. In 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1–6). IEEE. https://doi.org/10.1109/ICORIS50180.2020.9320831

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & Education, 143, 103676. https://doi.org/10.1016/j.compedu.2019.103676

Tuononen, T., & Parpala, A. (2021). The role of academic competences and learning processes in predicting Bachelor's and Master's thesis grades. Studies in Educational Evaluation, 70. https://doi.org/10.1016/j.stueduc.2021.101001

Vijayalakshmi, V., & Venkatachalapathy, K. (2019). Comparison of predicting student's performance using machine learning algorithms. International Journal of Intelligent Systems and Applications, 11(12), 34. https://doi.org/10.5815/ijisa.2019.12.04

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189. https://doi.org/10.1016/j.chb.2019.106189

Xiushan NIE, Yuling MA, Huiyan QIAO, Jie GUO, Chaoran CUI, Zhiyun YU, Xingbo LIU, & Yilong YIN. (2022). Surver on student academic performance prediction from the perspective of task granularity. Journal of Shandong University (Engineering Science), 52(2): 1-14. https://doi.org/10.6040/j.issn.1672-3961.0.2021.489

Yang, S. J. H., Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. Journal of Information Processing, 26, 170–176. https://doi.org/10.2197/ipsjjip.26.170

7. BIOGRAPHIES (OPTIONAL)

Miss Wenting Sui (Ph.D. Student)

Department of Industrial Technology Management, Faculty of Industrial Technology, Muban Chom Bueng Rajabhat University, Ratchaburi, Thailand

Dr.Adisak Sangsongfar

Department of Industrial Technology Management, Faculty of Industrial Technology, Muban Chom Bueng Rajabhat University, Ratchaburi, Thailand

Assistance Professor Dr.Noppadol Amdee

Department of Industrial Technology Management, Faculty of Industrial Technology, Muban Chom Bueng Rajabhat University, Ratchaburi, Thailand