

Optimization of MLP-Regressor for Predicting Student's Grade Point Average (GPA)

Ihsan Fadli Tampati
Politeknik Siber dan Sandi Negara
Badan Siber dan Sandi Negara
Bogor, Indonesia
ihsan.fadli@student.poltekssn.ac.id

Hermawan Setiawan
Politeknik Siber dan Sandi Negara
Badan Siber dan Sandi Negara
Bogor, Indonesia
hermawan.setiawan@bssn.go.id

Girinoto
Politeknik Siber dan Sandi Negara
Badan Siber dan Sandi Negara
Bogor, Indonesia
girinoto@bssn.go.id

Abstract— In the digital age, vast amounts of data offer opportunities to enhance student outcomes and achieve sustainable educational objectives. Educational institutions aim to identify factors influencing academic performance and predict exam scores to provide tailored support to students in need. Educational Data Mining (EDM) techniques enable in-depth analysis of students' historical data. Prior research has identified factors such as learning methods, study environment, parental involvement, and pressure as significant influencers of academic achievement. This study takes a novel approach by training a deep learning model on historical GPA data from Politeknik Siber dan Sandi Negara, using regression to accurately predict GPA. The model is implemented in a web application, alerting educators to students at risk of declining GPA. The Multi-layer Perceptron (MLP) regressor is employed for training and evaluation, yielding optimized configurations for the number of neurons, hidden layer depth, and Adam optimizer with constant learning rate. Evaluation metrics, including MSE, RMSE, MAE, R-squared, and Breusch-Pagan Test, showcase enhanced model performance. Future research could focus on expanding the dataset to achieve greater predictive accuracy thereby empowering institutions to offer targeted support to students in need.

Keywords—educational data mining (edm), deep learning, grade point average (gpa), regression

I. INTRODUCTION

The digital age brings a lot of digital data that can be turned into useful information. This information can be analyzed in various ways to reach specific goals. To achieve sustainable education goals, we use a method called Educational Data Mining (EDM). EDM uses statistical, machine learning, and data mining algorithms to analyze educational data, helping us understand more about how students learn and their surroundings [1], [2]. Various universities have started using EDM techniques to enhance student quality, institutional quality, and institutional standards. Using EDM to improve student quality can be done by analyzing students' academic history data [1].

Researchers have analyzed that additional information such as school teaching methods, a country's educational approach, study environment, questioning intensity, participation in discussions, parent-child relationships, and parental pressure can influence students' exam scores [1], [3], [4], [5], [6], [7], [8], [9]. Analyzing the factors affecting students' exam scores and predicting students' exam scores are classification and regression problems that can be solved using machine learning algorithms [1], [3], [4], [5], [6], [7], [8], [9].

The role of machine learning is crucial in EDM techniques [1]. The use of machine learning enables the prediction of future occurrences based on analyzed data. Machine learning

algorithms commonly used for prediction in regression problems include linear regression, lasso regression, decision tree regression, random forest, k-nearest neighbors (K-NN), and multi-layer perceptron (MLP) regressor. Previous researchers have demonstrated that these algorithms can define factors related to the improvement or decline of students' exam scores [1], [3], [4], [5], [6], [7], [8], [9]. However, no researcher has proven that student exam scores, in this case grade point average (GPA), can be predicted solely based on previous GPA. Therefore, this study aims to demonstrate that deep learning algorithms can be used and optimized to predict GPA based on previous GPA. The deep learning algorithm used in this study is the Multi-Layer Perceptron regressor from the scikit-learn library. The MLP-Regressor algorithm was chosen because it has been proven effective in predicting data in regression problems and has been widely applied in various fields [10], [11], [12], [13], [14], [15]. The main idea of this research is to train the MLP regressor model using the grade point average (GPA) from Politeknik Siber dan Sandi Negara with a regression method so that it can predict students' GPA as accurately as possible, then implement it in a web application. If there are students whose GPA is predicted to decrease in the next semester, then the lecturer is obliged to pay more attention to these students. This research is expected to help universities improve the quality of students by paying attention to students who have predicted low GPA in the next semester and encourage universities to utilize educational data mining using digital technology.

II. RELATED WORKS

Various studies conducted by previous researchers aim to understand and predict students' academic performance through different approaches. A study in Tunisia [3] utilized two-stage analysis, namely the directional distance function Approach (DDF) and machine learning algorithms regression trees and random forest to identify key factors influencing students' academic performance. The results showed that factors such as school size, competition, class size, parental pressure, and interaction with females had significant effects, while school location had no significant influence. Another study [4] proposed a comprehensive student academic data processing model called AugmentedED using various data sources and features as well as machine learning and deep learning algorithms to predict academic performance with high accuracy. A comparison was made of machine learning algorithms to predict student performance based on multi-feature data, while deep learning was used to model student performance based on big data from the virtual learning environment (VLE). A study [5] attempted to identify student and school characteristics related to student exam scores using a tree-based machine learning approach. The research was conducted on students and schools from nine different

countries: Australia, Canada, France, Germany, Italy, Japan, Spain, the UK, and the USA. The results of the study showed a correlation between student and school characteristics and test scores.

Research [6] aims to identify suitable machine learning algorithms for predicting students' academic grades and categorizing final student grades. Predicting academic grades is a regression problem that the authors attempted to solve by evaluating machine learning algorithms using the mean squared error (MSE) evaluation metric, while categorizing final student grades is a classification problem that the authors attempted to solve by evaluating machine learning algorithms based on classification accuracy. The algorithms used include Decision Tree, Genetic Algorithm (GA) based Decision Tree, K-Nearest Neighbour (K-NN), and Genetic Algorithm (GA) based K-NN. The research results showed that the GA-based Decision Tree algorithm is better at categorizing final student grades compared to the other three algorithms. The obtained accuracy is 96.64%, and the GA-based Decision Tree algorithm is also better at predicting students' academic grades compared to the other three algorithms. This is evident from the smaller MSE value compared to the other three algorithms, which is 5.34.

Research [9] seeks to demonstrate the influence of using online learning resources and study duration on student performance. Performance is defined as either great or weak labels obtained from features such as learning style, family education history, non-study related work, sports, study time, use of online learning resources, and others. The relationship between these features, particularly study duration and the use of online learning resources, with the labels will be modeled using four machine learning algorithms, namely artificial neural network (ANN), naive Bayes, logistic regression, and decision trees. The results show that ANN models better than other algorithms. The best model produced by ANN interprets that there is no significant influence between study duration and the use of online learning resources on student academic performance.

Research [7] argues that now is the era where big data in education can be easily accessed, supported by increasingly advanced learning platforms, discovering problems from data with these platforms, optimizing the education environment based on the problems found, and providing recommendations for education to run smoothly. Data processing in education has been widely carried out using Educational Data Mining (EDM) techniques using machine learning algorithms. However, the use of deep learning is still rare. Researchers attempt to investigate the effectiveness of Deep Learning in modeling student performance in the form of classification. The classification labels consist of 'withdrawn-pass', 'pass-fail', 'distinction-pass', and 'distinction-fail' using datasets from VLE. The research results show that the proposed deep learning model has better accuracy compared to the logistic regression and SVM machine learning algorithms. The proposed deep learning model achieves classification accuracies from 84% to 93%, while the classification accuracies of logistic regression and SVM algorithms range from 79.82% - 85.60% and 79.95% - 89.14%, respectively.

Researcher [8] argues that in an increasingly competitive world, institutions need to be able to predict the performance of their students. This is necessary so that institutions can provide more attention to students who indeed need it. Furthermore, the success of an institution is measured by the

development of its students' performance. In the researcher's case, student performance is defined as a classification problem. To predict student performance, the author uses Educational Data Mining (EDM) techniques to analyze the relationships between features of the data and student performance using Naive Bayes, ID3, C4.5, and SVM machine learning algorithms. Subsequently, the prediction accuracy among the algorithms will be compared to obtain the best modeling. Research [10] proposes a dual-input deep learning model framework using a multi-layer perceptron (MLP) and long-short term memory (LSTM) architecture to predict grade point average (GPA) data on a 4.0 scale using a regression approach. The research results show that the model achieves the best performance compared to other models by obtaining values of 0.4142 for the mean squared error (MSE), 0.418 for the mean absolute error (MAE), and 0.4879 for R-squared.

From these various approaches, these studies contribute to a better understanding of the factors influencing students' academic performance, [2], [3], [4], [5], [6], [7], [9] the development of more accurate classification models [2], [3], [4], [5], [6], [7], [9], and the development of regression models [7]. Since the regression approach is still less used, this study will use the regression approach with deep learning.

III. IMPLEMENTATION

In this research, the implementation utilizes the MLP-Regressor from the scikit-learn library. The research methodology and implementation steps are outlined below.

A. Methodology

This research methodology utilizes the method from [16] as illustrated in the following Fig. 3.1.

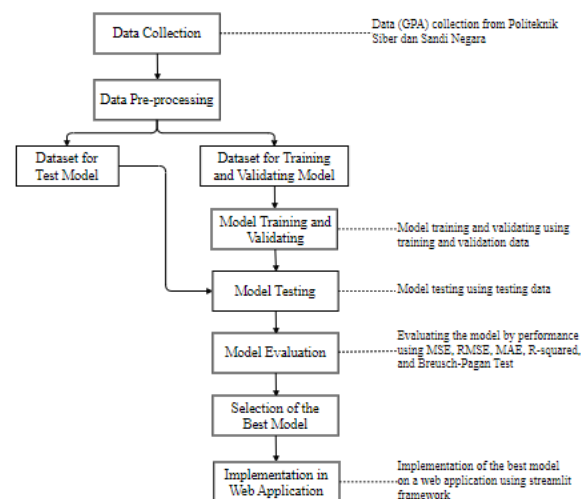


Figure III.1 Research Methodology

Based on Fig.3.1, the grade point average (GPA) data will be extracted from the GPA data of students at the Politeknik Siber dan Sandi Negara for the last nine years. The training, testing, and evaluation stages will involve hyper-parameter tuning processes for the number of neurons, layer depth within the hidden layer, various optimizers, and various learning rates. Then, the model will be evaluated using mean squared error (MSE), root-mean squared error (RMSE), mean-absolute error (MAE), coefficient of determination (R-squared), and Breusch-Pagan Test to ensure that the model is

free from heteroscedasticity and normally distributed [17]. The hyper-parameter tuning process is conducted to determine which parameters can be configured to reduce the MSE, RMSE, MAE and increase the coefficient of determination and Breusch-Pagan p -value. Furthermore, the model with the best statistical evaluation results will be implemented on the website application using the Streamlit framework.

B. Datasets

The following Fig.3.2 below provides an overview of the dataset used.

```

1 name,1,2,3,4,5,6,7,8
2 taruna-1,2.61,3.06,3.16,3.08,3.18,3.03,3.84,2.4
3 taruna-2,2.33,2.87,3.39,2.97,3.27,3.71,3.63,2.5
4 taruna-3,2.46,2.68,3.26,2.95,3.44,3.19,3.84,4.0
5 taruna-4,2.7,2.72,3.62,2.91,3.23,3.44,3.75,3.3
6 taruna-5,2.67,2.94,3.41,2.83,3.44,3.41,3.84,3.03
7 taruna-6,2.57,3.02,3.06,3.52,3.44,3.49,3.7,3.93
8 taruna-7,2.57,2.87,3.41,3.39,3.39,3.62,3.84,3.6
9 taruna-8,2.69,3.27,3.44,3.25,3.31,3.56,4.0,3.08
10 taruna-9,2.7,2.96,3.46,3.46,3.46,3.54,3.75,3.48
11 taruna-10,2.87,3.08,3.38,3.44,3.46,3.62,3.7,3.18
12 taruna-11,2.68,3.43,3.64,3.59,3.47,3.72,3.93,3.25
13 taruna-12,3.31,3.7,3.62,3.9,3.9,4.0,3.91,4.0
14 taruna-13,2.25,2.71,2.94,2.54,2.56,2.85,3.76,3.15
15 taruna-14,2.23,3.0,3.14,2.76,2.78,2.82,3.94,3.7

```

Figure III.2 GPA's Dataset

C. Neurons

The number of neurons will be configured with eight to sixteen neurons. Tables 3.1 and 3.2 below show the model evaluation results with MSE, RMSE, MAE, and coefficient of determination (R2) in the validation and testing phases of the model.

Table III.1 The results of MSE, RMSE, MAE, and R2 in the Model Validation Phase

Neurons	MSE	RMSE	MAE	R2
8	0.133032	0.364736	0.272875	-0.202540
9	0.124096	0.352273	0.254574	-0.121764
10	0.106264	0.325982	0.231283	0.039430
11	0.116069	0.340689	0.252536	-0.049198
12	0.086268	0.293714	0.214810	0.220184
13	0.109435	0.330810	0.246849	0.010762
14	0.105754	0.325199	0.243254	0.044040
15	0.107648	0.328097	0.253219	0.026923
16	0.170648	0.413096	0.310840	-0.542567

Table III.2 The result of MSE, RMSE, MAE, and R2 in the Model Testing Phase

Neurons	MSE	RMSE	MAE	R2
8	0.125173	0.353798	0.280785	-0.168637
9	0.115388	0.339688	0.260061	-0.077279
10	0.110864	0.332963	0.249052	-0.035045
11	0.125576	0.354368	0.274813	-0.172403
12	0.090367	0.300612	0.220819	0.156315
13	0.096716	0.310992	0.222536	0.097042
14	0.123103	0.350861	0.267037	-0.149311
15	0.112331	0.335159	0.259050	-0.048743
16	0.150107	0.387437	0.309738	-0.401423

Based on the experiment results in Table 3.1 and Table 3.2, it was found that there are negative values in the coefficient of determination. In the validation phase, the coefficient of determination is negative when the number of neurons is configured at eight, nine, eleven, and sixteen neurons. In the testing phase, the coefficient of determination is negative when the number of neurons is configured at eight to eleven and fourteen to sixteen neurons. Based on the [18], negative values in the coefficient of determination occur when the MSE value in the testing phase is larger than the MSE value in the validation phase, indicating that the model in the testing phase has poorer prediction performance compared to the prediction performance in the validation phase.

Next, to ensure that the regression model performs well in predicting GPA values, the model needs to be examined for residual variance inequality so that it can identify models that exhibit heteroskedasticity using the Breusch-Pagan Test method. The Breusch-Pagan Test is conducted on models with non-negative coefficients of determination. Table 3.3 and Table 3.4 below show the results of the Breusch-Pagan Test in the validation and testing phases of the model.

Table III.3 The results of the Breusch-Pagan Test in the Model Validation Phase

Neurons	Breusch-Pagan Test Statistic	Breusch-Pagan p -Value
8	7.533108	0.375555
9	6.981827	0.430774
10	7.864285	0.344713
11	6.687852	0.462087
12	6.834934	0.446266
13	5.398801	0.611416
14	9.296493	0.232064
15	5.342367	0.618259
16	17.215907	0.016055

Table III.4 The results of the Breusch-Pagan Test in the Model Testing Phase

Neurons	Breusch-Pagan Test Statistic	Breusch-Pagan p -Value
8	18.565359	0.009663
9	12.725760	0.079078
10	10.588772	0.157589
11	12.20278	0.094084
12	16.744093	0.019122
13	10.933424	0.141549
14	11.695947	0.111011
15	18.247843	0.010900
16	32.611364	0.003127

Heteroskedasticity in the model is identified by high Breusch-Pagan test statistics and Breusch-Pagan p -values smaller than 0.05. Therefore, from the results of the experiments in Table 3.3 and Table 3.4, we search for models with low Breusch-Pagan test statistics and Breusch-Pagan p -

values greater than 0.05, and it is found that the model with a configuration of thirteen (13) neurons is the best model. In the next step, the configuration of thirteen (13) neurons will be used to determine the best configuration of other parameters. The main focus of configuring other parameters is to reduce the values of MSE, RMSE, MAE, and Breusch-Pagan test statistics while increasing the values of the coefficient of determination (R2) and Breusch-Pagan p-value.

D. Hidden Layer's Depth

In this stage, the depth of the layers in the hidden layer will be configured from one to nine layers deep. Tables 3.5, 3.6, 3.7, and 3.8 below present the evaluation results of the model with MSE, RMSE, MAE, and coefficient of determination (R2) during the validation and testing phases of the model.

Table III.5 The results of MSE, RMSE, MAE, and R2 in the Model Validation Phase

Hidden Layer's Depth	MSE	RMSE	MAE	R2
1	0.162070	0.402580	0.310083	-0.465027
2	0.099629	0.315640	0.235849	0.099410
3	0.109435	0.330810	0.246849	0.010762
4	0.089254	0.298751	0.217952	0.193186
5	0.109662	0.331153	0.256150	0.008712
6	0.089938	0.299898	0.222695	0.187003
7	0.085841	0.292986	0.207943	0.224045
8	0.086710	0.294466	0.215999	0.216187
9	0.097511	0.312267	0.218024	0.118555

Table III.6 The result of MSE, RMSE, MAE, and R2 in the Model Testing Phase

Hidden Layer's Depth	MSE	RMSE	MAE	R2
1	0.369468	0.294961	-0.274449	0.136507
2	0.323244	0.242227	0.024492	0.104487
3	0.310992	0.222536	0.097042	0.096716
4	0.304638	0.217673	0.133561	0.092804
5	0.317712	0.240772	0.057596	0.100941
6	0.300298	0.216751	0.158075	0.090179
7	0.296095	0.202801	0.181477	0.087672
8	0.300865	0.213214	0.154891	0.090520
9	0.369468	0.294961	-0.274449	0.136507

Based on the experiment results in Tables 3.5 and 3.6, it is evident that the configuration of the depth of the layers in the hidden layer significantly influences the decrease or increase in the values of MSE, RMSE, MAE, and coefficient of determination (R2). It can be observed that with shallower depths in the hidden layer, the model performs worse in learning the relationships among the varying dependent variables, as indicated by the decrease in the coefficient of determination (R2) values. Furthermore, to obtain a good

configuration for the depth of the layers in the hidden layer, the model needs to be further examined using the Breusch-Pagan Test. Tables 3.7 and 3.8 below present the results of the Breusch-Pagan Test.

Table III.7 The results of the Breusch-Pagan Test in the Model Validation Phase

Hidden Layer's Depth	Breusch-Pagan Test Statistic	Breusch-Pagan p-Value
1	6.487691	0.484091
2	6.766556	0.453584
3	5.398801	0.611416
4	5.422261	0.608576
5	6.032790	0.535925
6	6.348390	0.499708
7	6.850935	0.444563
8	6.024954	0.536838
9	10.714637	0.151559

Table III.8 The results of the Breusch-Pagan Test in the Model Testing Phase

Hidden Layer's Depth	Breusch-Pagan Test Statistic	Breusch-Pagan p-Value
1	13.228622	0.066730
2	17.442600	0.014754
3	10.933424	0.141549
4	15.469837	0.030426
5	19.776877	0.006072
6	15.140707	0.0342374
7	12.002047	0.100491
8	13.610676	0.058555
9	11.208736	0.129769

Based on the results of the Breusch-Pagan test in tables 3.7 and 3.8, it can be seen that the configuration with seven layers in the hidden layer is the best configuration. Next, the model with the configuration of thirteen (13) neurons and seven (7) layers in the hidden layer will be used to determine the model with the best optimizers and learning rate.

E. Optimizers

In this stage, the type of optimizers will be configured with Adam, stochastic-gradient descent (SGD), and quasi-newton (LBFGS). Tables 3.9 and Table 4.0 below show the evaluation results of the model with MSE, RMSE, MAE, and coefficient of determination (R2) in the validation and testing phases.

Table III.9 The results of MSE, RMSE, MAE, and R2 in the Model Validation Phase

Optimizers	MSE	RMSE	MAE	R2
Adam	0.085841	0.292986	0.207943	0.224045
SGD	0.111789	0.334349	0.252168	-0.010515
LBFGS	0.094406	0.307255	0.218354	0.146623

Table III.10 The results of MSE, RMSE, MAE, and R2 in the Model Testing Phase

Optimizers	MSE	RMSE	MAE	R2
Adam	0.097511	0.312267	0.218024	0.118555
SGD	0.110704	0.332722	0.250531	-0.033552
LBFGS	0.103840	0.322242	0.227131	0.030534

Based on the experiment results in tables 3.9 and 3.10, it is observed that the type of optimizer configuration significantly affects the increase or decrease in the values of MSE, RMSE, MAE, and coefficient of determination (R2). It is evident that when the model is configured with the SGD optimizer, the model has a negative value for the coefficient of determination (R2), indicating that the SGD optimizer configuration makes it difficult for the model to learn the relationship among the dependent variables. Furthermore, to obtain the best optimizer configurations, the model needs to be reviewed again using the Breusch-Pagan Test. Tables 3.11 and 3.12 below show the results of the Breusch-Pagan Test.

Table III.11 The results of the Breusch-Pagan Test in the Model Validation Phase

Optimizers	Breusch-Pagan Test Statistic	Breusch-Pagan p-Value
Adam	6.850935	0.444563
SGD	11.328200	0.124929
LBFGS	6.927729	0.436443

Table III.12 The results of the Breusch-Pagan Test in the Model Testing Phase

Optimizers	Breusch-Pagan Test Statistic	Breusch-Pagan p-Value
Adam	10.714637	0.151559
SGD	9.507533	0.218240
LBFGS	22.565600	0.002028

Based on the experiment results in Tables 3.11 and 3.12, although the SGD optimizer has the best Breusch-Pagan test statistic and Breusch-Pagan p-value, the SGD optimizer is not the best configuration. This is because in Tables 3.9 and 3.10, the SGD configuration has a negative value for the coefficient of determination (R2). Therefore, the Adam optimizer is the best optimizer in this case. Hence, in the next experiment, the model with a configuration of thirteen (13) neurons, seven (7) hidden layer depths, and the Adam optimizer will be used to determine the best learning rate configuration.

F. Learning Rate

In this stage, the type of learning rate will be configured with constant, invscaling, and adaptive. Tables 3.13 and 3.14 below show the evaluation results of the model with MSE, RMSE, MAE, and coefficient of determination (R2) in the validation and test phases.

Table III.13 The results of MSE, RMSE, MAE, and R2 in the Model Validation Phase

Learning Rate	MSE	RMSE	MAE	R2
constant	0.085841	0.292986	0.207943	0.224045
invscaling	0.085841	0.292986	0.207943	0.224045
adaptive	0.085841	0.292986	0.207943	0.224045

Table III.14 The result of MSE, RMSE, MAE, and R2 in the Model Testing Phase

Learning Rate	MSE	RMSE	MAE	R2
constant	0.085841	0.292986	0.207943	0.224045
invscaling	0.085841	0.292986	0.207943	0.224045
adaptive	0.085841	0.292986	0.207943	0.224045

Based on the results of the experiments in Table 3.13 and Table 3.14, it can be concluded that the learning rate configuration does not significantly affect the increase or decrease in the values of MSE, RMSE, MAE, and coefficient of determination (R2). Furthermore, to obtain the best learning rate configuration, the model needs to be reviewed again using the Breusch-Pagan Test. Tables 3.15 and 3.16 below show the results of the Breusch-Pagan Test.

Table III.15 The results of the Breusch-Pagan Test in the Model Validation Phase

Learning Rate	Breusch-Pagan Test Statistic	Breusch-Pagan p-Value
constant	6.850935	0.444563
invscaling	6.850935	0.444563
adaptive	6.850935	0.444563

Table III.16 The results of the Breusch-Pagan Test in the Model Testing Phase

Learning Rate	Breusch-Pagan Test Statistic	Breusch-Pagan p-Value
constant	10.714637	0.151559
invscaling	12.002047	0.100491
adaptive	12.002047	0.100491

Based on Table 3.13, Table 3.14, Table 3.15, and Table 3.16, it is found that the best learning rate configuration is constant learning rate. Next, it is necessary to ensure that the model with the configuration of thirteen (13) neurons, seven (7) layers depth in the hidden layer, Adam optimizer, and constant learning rate is capable of learning the relationship

between independent variables and dependent variables from the dataset, thus producing normally distributed values and visually free from heteroscedasticity. Therefore, in the next step, visualization of the validation and test results of the model will be conducted using Q-Q plots, density plots, and residual plots.

G. Visualization of Residual Plot, Q-Q Plot, and Density Plot

Figure 3.3, Figure 3.4, and Figure 3.5 below are visualizations of the residual plot, Q-Q plot, and density plot in the validation and testing phases of the model.

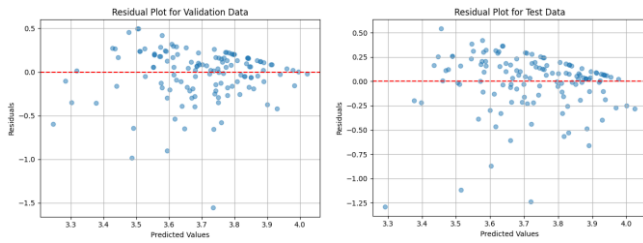


Figure III.3 Comparison of Residual Plots

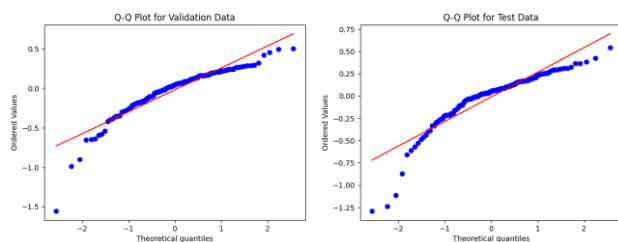


Figure III.4 Comparison of Q-Q Plots

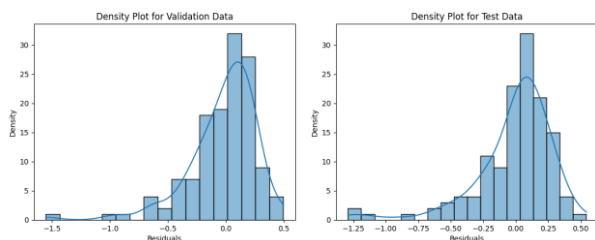


Figure III.5 Comparison of Density Plots

Based on the visualization in Figure 3.3, Figure 3.4, and Figure 3.5, it can be seen that the model in both the validation and testing phases is free from heteroskedasticity, normally distributed, and exhibits similar characteristics. Based on these facts, it can be concluded that the model is capable of generalizing new GPA values well, even though the coefficient of determination (R^2) and the Breusch-Pagan p-value of the model are still far from 1.

H. Implementation in Web Application

Here is the web application for GPA prediction using the MLP-Regressor model.

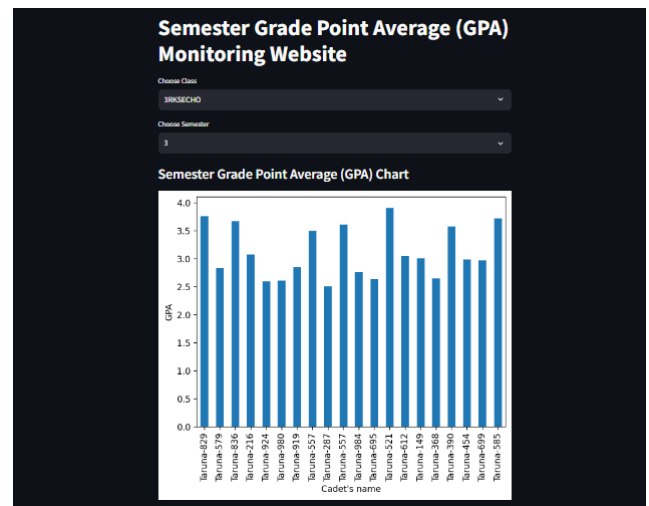


Figure III.6 GPA List Page

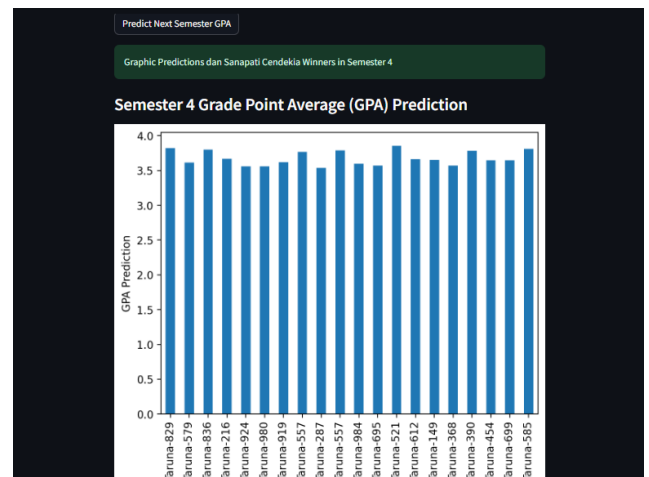


Figure III.7 GPA Prediction Page

Based on Fig.6 and Fig.7, users can select the student level and semester they want to predict. Then, the system will predict the chosen semester's GPA. If there are predicted GPA values for students that decrease in the next semester, the website will display a separate list of those students.

IV. CONCLUSION

Based on the experiment results, it is proven that MLP-Regressor can be used to predict GPA based on historical GPA data. Additionally, it can be concluded that the configuration of the learning rate parameter does not significantly affect the performance of the MLP-Regressor model. Unlike the learning rate, the configuration of the number of neurons, the depth of the hidden layer, and the type of optimizer affect the improvement or deterioration of the regression model's performance. Although statistically, the model performs well, the values of the coefficient of determination (R -squared) and Breusch-Pagan p-value are still far from 1. Therefore, further research can be developed by increasing the data in the dataset to improve the coefficient of determination and Breusch-Pagan p-value towards 1 so that the resulting MLP-Regressor model can accurately predict student GPA and assist educational institutions in providing more attention to students predicted to experience a decrease in GPA.

REFERENCES

- [1] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, "Classification and prediction of student performance data using various machine learning algorithms," *Mater. Today Proc.*, vol. 80, pp. 3782–3785, 2023, doi: <https://doi.org/10.1016/j.matpr.2021.07.382>.
- [2] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 40, no. 6, pp. 601–618, 2010, doi: [10.1109/TSMCC.2010.2053532](https://doi.org/10.1109/TSMCC.2010.2053532).
- [3] S. Rebai, F. Ben Yahia, and H. Essid, "A graphically based machine learning approach to predict secondary schools performance in Tunisia," *Socioecon. Plann. Sci.*, vol. 70, p. 100724, 2020, doi: <https://doi.org/10.1016/j.seps.2019.06.009>.
- [4] L. Zhao et al., "Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data," *IEEE Access*, vol. 9, pp. 5453–5465, 2021, doi: [10.1109/ACCESS.2020.3002791](https://doi.org/10.1109/ACCESS.2020.3002791).
- [5] C. Masci, G. Johnes, and T. Agasisti, "Student and school performance across countries: A machine learning approach," *Eur. J. Oper. Res.*, vol. 269, no. 3, pp. 1072–1085, 2018, doi: <https://doi.org/10.1016/j.ejor.2018.02.031>.
- [6] S. Hussain and M. Q. Khan, "Student-Performer: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning," *Ann. Data Sci.*, vol. 10, no. 3, pp. 637–655, 2023, doi: [10.1007/s40745-021-00341-0](https://doi.org/10.1007/s40745-021-00341-0).
- [7] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Human Behav.*, vol. 104, p. 106189, 2020, doi: <https://doi.org/10.1016/j.chb.2019.106189>.
- [8] H. Prabowo, A. A. Hidayat, T. W. Cenggoro, R. Rahutomo, K. Purwandari, and B. Pardamean, "Aggregating Time Series and Tabular Data in Deep Learning Model for University Students' GPA Prediction," *IEEE Access*, vol. 9, pp. 87370–87377, 2021, doi: [10.1109/ACCESS.2021.3088152](https://doi.org/10.1109/ACCESS.2021.3088152).
- [9] "Predicting Students' Performance Using Machine Learning Techniques," *J. Univ. BABYLON Pure Appl. Sci.*, vol. 27, no. 1 SE-Articles, pp. 194–205, Apr. 2019, doi: [10.29196/jubpas.v27i1.2108](https://doi.org/10.29196/jubpas.v27i1.2108).
- [10] M. I. Dutt and W. Saadeh, "A Multilayer Perceptron Neural Network Approach for Predicting Student Performance," *NEWCAS (NEWCAS)*, 2022, pp. 251–255. doi: [10.1109/NEWCAS52662.2022.9842242](https://doi.org/10.1109/NEWCAS52662.2022.9842242).
- [11] H. Shimanaka, T. Kajiwara, and M. Komachi, "{RUSE}: Regressor Using Sentence Embeddings for Automatic Machine Translation Evaluation," in *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, C. Monz, M. Negri, A. Névéol, M. Neves, M. Post, L. Specia, M. Turchi, and K. Verspoor, Eds., Belgium, Brussels: Association for Computational Linguistics, 2018, pp. 751–758. doi: [10.18653/v1/W18-6456](https://doi.org/10.18653/v1/W18-6456).
- [12] X. Xu, S. L. Keoh, C. K. Seow, Q. Cao, and S. K. B. A. Rahim, "Towards Prediction of Bus Arrival Time using Multi-layer Perceptron (MLP) and MLP Regressor," in *2023 8th International Conference on Business and Industrial Research (ICBIR)*, 2023, pp. 669–674. doi: [10.1109/ICBIR57571.2023.10147614](https://doi.org/10.1109/ICBIR57571.2023.10147614).
- [13] E. Agirre-Basurko, G. Ibarra-Berastegi, and I. Madariaga, "Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area," *Environ. Model. Softw.*, vol. 21, no. 4, pp. 430–446, 2006, doi: <https://doi.org/10.1016/j.envsoft.2004.07.008>.
- [14] X. Feng, G. Ma, S.-F. Su, C. Huang, M. K. Boswell, and P. Xue, "A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan," *Ocean Eng.*, vol. 211, p. 107526, 2020, doi: <https://doi.org/10.1016/j.oceaneng.2020.107526>.
- [15] DoreSwamy, H. K S, Y. KM, and I. Gad, "Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models," *Procedia Comput. Sci.*, vol. 171, pp. 2057–2066, 2020, doi: <https://doi.org/10.1016/j.procs.2020.04.221>.
- [16] J. Kamiri and G. Mariga, "Research Methods in Machine Learning: A Content Analysis," *Int. J. Comput. Inf. Technol.*, vol. 10, no. 2, 2021, doi: [10.24203/ijcit.v10i2.79](https://doi.org/10.24203/ijcit.v10i2.79).
- [17] T. S. Breusch and A. R. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica.*, vol. 47, no. 6, pp. 1287–1294, 1979, doi: <https://doi.org/10.2307/1911963>.
- [18] A. Chen, "Predictive Modeling of Webpage Aesthetics," *Missouri University of Science and Technology.*, pp. 108–114, 2019, doi: https://scholarsmine.mst.edu/masters_theses/7913.