icvee 2021

by Yuni Yamasari

Predicting the Students Performance using Regularization-based Linear Regression

1st Yuni Yamasari
Department of Informatics
Universitas Negeri Surabaya
Surabaya, Indonesia
yuniyamasari@unesa.ac.id

2nd Naim Rochmawati
Department of Informatics
Universitas Negeri Surabaya
Surabaya, Indonesia
naimrochmawati @unesa.ac.id

3rd Ricky E. Putra
Department of Informatics
Universitas Negeri Surabaya
Surabaya, Indonesia
rickyeka @unesa.ac.id

4th Anita Qoiriah
Department of Informatics
Universitas Negeri Surabaya
Surabaya, Indonesia
anitaqoiriah@unesa.ac.id

5th Asmunin
Department of Informatics
Universitas Negeri Surabaya
Surabaya, Indonesia
asmunin@unesa.ac.id

6th Wiyli Yustanti
Department of Informatics
Universitas Negeri Surabaya
Surabaya, Indonesia
asmunin@unesa.ac.id

2021 Fourth International Conference on Vocational Education and Electrical Engineering (ICVEE) | 978-1-6654-0599-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICVEE54186.2021.9649704

***Abstract*—Abstract—the corona pandemic changes the education paradigm from offline- to online learning. This situation causes a crucial problem in the evaluation process of student outcome learning. A valid enough evaluation is difficult to achieve because of the lack of face-to-face between teachers and students. However, this stage is critical in the education and teaching process. Therefore, our research focuses on building a prediction model of students' performance for supporting the teacher in the evaluation process. The prediction model is created by linear regression based on Regularization. Further, we explore three regularizations, namely: ridge, lasso, and elastic net, to find the best performance of the model. The evaluation technique used is random sampling with various training set sizes. In addition, we determine the different values of alpha to discover the best model. The experimental results show that the ridge regularization model's prediction error rate is lower than the lasso and elastic net Regularization. The results of measuring values using MSE, RMSE, MAE, and R2 show that the prediction model built using ridge regularization is superior to the others.**

***Keywords*— student, performance, data mining, online, linear regression**

I. Introduction

Nowadays, almost of countries are still working hard to get out of the corona pandemic. The condition impacts all fields which are included in the education field. Relating to the situation, the Indonesia Government, through the Ministry of Education and Culture (MOEC), issued a policy to encourage online learning for students. [1] There are many obstacles in online learning in Indonesia, for example, limiting teacher abilities, parents' economic level, limited internet access, Etc. [2]

One of the critical problems of online learning is the problem relating to the evaluation process. There are two activities in this process, namely: measurement and assessment. The process is complicated because the teacher only considers the online metrics. The process is complicated because the teacher only considers the online measure. In addition, the validity of the assessment is usually in the unideal condition because students cheat during the test. This situation is caused by the lack face to face directly between the teacher and students.

Furthermore, the teacher finds obstacles to assessing student performance, such as internet connection, assessment validity, and low students enthusiasm [3]. However, this step is essential because the evaluation is one way to find out the learning outcomes. The purpose of the assessment is to assist students in knowing their learning performance. In the evaluation system, several tests and assignments are carried out periodically. The success of a students' learning is expected to be seen in tests and assignments.

The application of Data Mining techniques discovers new or hidden patterns [4],[5]. The Data Mining applied in the education area is called Educational Data Mining (EDM) [6]. The student data obtained in online learning can be mined by Data Mining methods to solve the problems in the pandemic period. One of them solve problem relating to students' performance [7], [8], [9], [10], [11], [12]. Especially in the students' performance, the researcher reviews modeling for prediction students' performance.[13] The prediction models are used to help the teacher prevent failure, improve the students' performance, Etc.

This previous research predicts the students' performance using regression using two methods: random forest and artificial neural network [14]. The following research tries to mitigate failure and to promote better achievement. It applies techniques to predict the students' academic performance. They are random forests, decision trees, support vector machines, naive Bayes, bagged trees, and boosted trees. The result shows that random forests are superior to the other classification techniques that are considered. The model built with this method achieves an accuracy level of about 95%. [15] The other research predicts students' performance to identify significant variables that influence students' performance and help stakeholders in the decision-making of plan ahead [12].

The previous research is not much on the exploration of the linear regression method to build the model. Moreover, the previous research has not yet exploited the Regularization in linear regression to find the best performance of the prediction model. Therefore, in this research, we build the prediction of students' performance using linear regression based on Regularization. Here, the regularizations explored are Ridge, Lasso, and Elastic Net.

978-1-6654-0599-7/21/$31.00 ©2021 IEEE

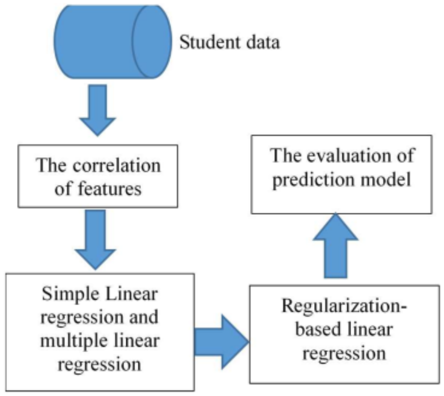

Fig. 1. The proposed method

TABLE I. STRUCTURE OF STUDENT DATA

Feature	Data Type	Description
Kehadiran	Numeric	percentage of student attendance in basic programming courses
Part	Numeric	the score relating to student participation in basic programming course, for example: asking questions, answering etc.
Tugas	Numeric	the score of assignment in basic programming course
UTS	Numeric	the score of middle test semester in basic programming course
UAS	Numeric	the score of the final test semester
NilaiAkhir	Numeric	the final score of students in basic programming course

Finally, our paper is arranged as follows: Section I is the introduction. Section II is material and method. Next, Section III is the result and discussion. Then, Section IV is the conclusion.

II. Material and Method

The proposed architecture is in this research depicted in Fig.1. Here, the architecture comprises of many steps, as follows:

Step 1: data of students

In this research, we focus on the academic performance of students. We collect data of students when they join the introductory programming course. Our research involves 115 students as participants. Then, our student data consists of 6 features as figured in TABLE I. The features are as follows: Kehadiran, Part, Tugas, UTS, UAS, and NilaiAkhir. All of them are numeric data types.

Step 2: the exploration of simple- and multiple linear regression

In this step, the data explore modeling using linear regression, namely: simple- and multiple linear regression. Linear regression assumes that the relationship between the dependent variable and independent variable can be explained linearly. The target is to find the best-estimated value for the coefficient of the equation using the values of the independent variables. The standard approach in finding coefficients is to apply the least-squares method [16]

A linear regression model with a single explanatory variable is known as simple linear regression [17],[18],[19],[20]. That is, two-dimensional sample points with one independent variable and one dependent variable are concerned. The formula of a simple linear regression is shown in equation (1).

$$y = \alpha + \beta x \qquad (1)$$

Where this formula depicts a line with slope β and y-intercept α.

The association between the dependent variable and two or more independent variables is known as multiple linear regression. The theoretical assumption is that the dependent variable changes uniformly for every one-unit change in the independent variable. The following is the model for multiple linear regression [21]:

$$y = \beta_0 + \beta_1 x_{i,1} + \ldots \beta_k x_{i,k} + \varepsilon_i \forall i \in \{1,\ldots,n\} \qquad (2)$$

Where:

k = the number of features

n = the number of samples

ε = an error term with means of zero and finite variance

Step 3: Regularization-based linear regression

The coefficient estimates are limited, regularized, or reduced towards zero in this sort of regression. Express differently; this method inhibits learning a more intricate or flexible model to minimize overfitting. There are three types in this research, namely: lasso-, ridge- and elastic net Regularization.

A loss function known as the residual sum of squares (RSS) depicted in equation (3) is used in the fitting procedure. The coefficients are chosen in such a way that the loss function is minimized.

Now, the coefficients will be adjusted based on training data. If the training data contains noise, the calculated coefficients will not generalize well to the following data. Noise occurs when Regularization enters the image, shrinking or setting the learned estimate to zero.

$$RSS = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}\right)^2 \qquad (3)$$

For ridge regression, this Regularization does to how the RSS is changed by adding the shrinkage quantity. By minimizing this function, the coefficients are now estimated. This is the tuning parameter that determines how much we want to penalize our model's flexibility.

Lasso (most minor absolute shrinkage and selection operator), another variation to minimize this function, is a regularized form of least squares that applies the requirement that ||β||, the L1-norm of the parameter vector, is no bigger than a specific value. [22]

The selection of variables depending on the data produces an elastic net of critique on an unstable lasso. The Regularization gives the solution with the penalties of ridge regression and lassoes to get the best result.

Step 4: evaluating the prediction model

In the last step, the prediction model built is evaluated. Here, our research measure the performance of the prediction model using metrics, namely: MSE, RMSE, MAE, and R2.

MAE is the difference between the actual and expected values obtained by averaging the absolute difference over the data set (Mean absolute error).

The MSE is the difference between the actual and anticipated values calculated by squaring the average difference over the data set (Mean Squared Error).

RMSE is the error rate multiplied by the square root of MSE (Root Mean Squared Error).

The coefficient of determination (R-squared) measures how well the values fit together compared to the original values. Percentages are calculated with values ranging from 0 to 1. The better the model, the higher the value.

The formulas of them are described in equation (4)-(7)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y')^2 \quad (4)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y')^2} \quad (5)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y'| \quad (6)$$

$$R^2 = 1 - \frac{\sum(y_i - y')^2}{\sum(y_i - \bar{y})^2} \quad (7)$$

Where,

y = actual value of y

y' = predicted value of y

$\bar{y}$ = mean value of y

N = the number of samples

III. Result and Discussion

The execution of the proposed method is represented in this section. In addition, the result is analyzed to observe the model performance. There are two subsection analyses, namely: the analysis of linear regression and the regularization-based linear regression

A. The Analysis of Linear Regression

This section explores the features shown in TABLE I to build a model using linear regression. Before we make the model using simple- and multiple linear regression, we do the correlation test to observe the relation of the features that Its result is depicted in Fig.2.

	Kehadiran	Part	Tugas	UTS	UAS	NilaiAkhir
Kehadiran	1.0000000	0.7217803	0.6328399	0.4847163	0.5232124	0.7104047
Part	0.7217803	1.0000000	0.3278195	0.6302183	0.4936028	0.6506102
Tugas	0.6328399	0.3278195	1.0000000	0.5998207	0.4876193	0.8252061
UTS	0.4847163	0.6302183	0.5998207	1.0000000	0.6744756	0.8746285
UAS	0.5232124	0.4936028	0.4876193	0.6744756	1.0000000	0.8341448
NilaiAkhir	0.7104047	0.6506102	0.8252061	0.8746285	0.8341448	1.0000000

Fig. 2. The correlation of features

The highest correlation value is about 0.8746285 showed by the correlation between UTS and NilaiAkhir. So, we build the simple linear regression for the first regression, which only involves one feature or a predictor feature first.

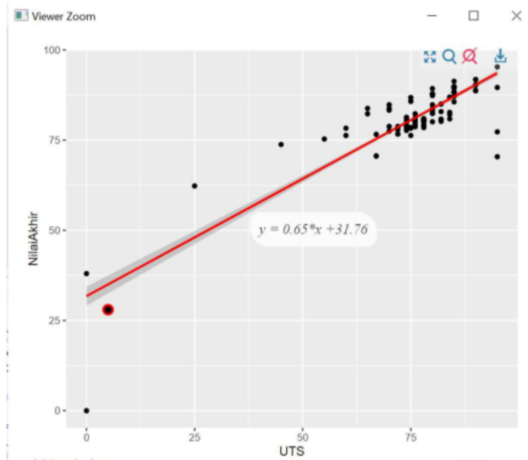

Fig. 3. The visualization of simple linear regression

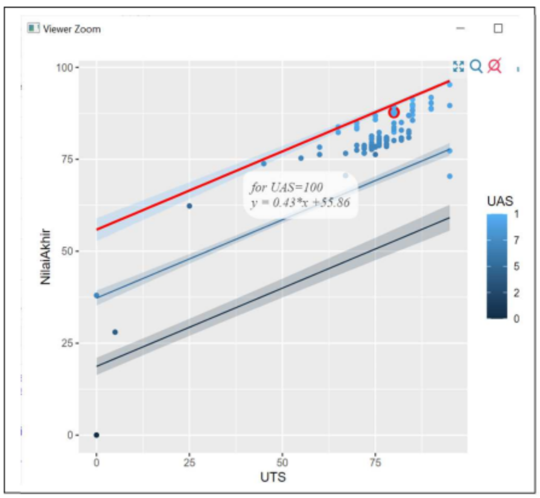

Fig. 4. The visualization of multiple linear regression

Here, the target feature or the dependent variable is NilaiAkhir, and the predictor feature or the independent variable is UTS. The experimental result of this model is depicted in Fig.3. The equation of the simple linear regression is $y = 0.65x + 31.76$

Further, we also build a model using multiple linear regression. The predictor features used to make are based on.

TABLE II. THE BEST PERFORMANCE OF PREDICTING MODEL USING LINEAR REGRESSION

Training set size	MSE	RMSE	MAE	R2
Ridge				
10	0	0	0	1
20	0	0	0	1
30	0	0	0	1
40	0	0	0	1
50	0	0	0	1
Lasso				
Training set size	MSE	RMSE	MAE	R2
10	0	0	0	1
20	0	0	0	1
30	0	0	0	1
40	0	0	0	1
50	0	0	0	1
60	0	0	0	1
Elastic net				
Training set size	MSE	RMSE	MAE	R2
10	0	0.021	0.007	1
20	0	0.002	0.001	1
30	0	0.001	0	1
40	0	0	0	1
50	0	0	0	1
60	0	0.002	0.001	1

The highest correlation value of two features that results from the correlation test. From Fig.2. , we can see that the highest correlation values of the two features are UTS and UAS when both of them are correlated with NilaiAkhir. The implementation result is visualized in Fig. 4. The equation of the multiple linear regression is $y = 0.43x + 55.86;\ for\ UAS = 100$

B. The Analysis of Regularization-Based Linear Regression

In this section, we build models using regularization-based linear regression. As explained in Section II, our research explores four regularizations: ridge, lasso, and elastic net. After models are built, they are evaluated to analyze their performance. In our research, the evaluation technique for models is random sampling with training set size: 10%,20%,30%,40%,50%, and 60%. The alpha values are 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, and 5. The measurements of performance use MSE, RMSE, MAE, and R2.

The experimental result shows that the best performance with MSE=0; RMSE=0; MAE=0; R2=1 is achieved by all regularizations depicted in TABLE II. In detail, the best performance of each Regularization is as follows:

- The ridge regularization:
  a) Alpha= 0.0005; 0.0001; 0.001 on all training set size.
  b) Alpha=0.005 on training set size: 20%, 30%, 40%, 50%, 60%
  c) Alpha=0.01;0.05 on training set size: 30%, 40%, 50%, 60%

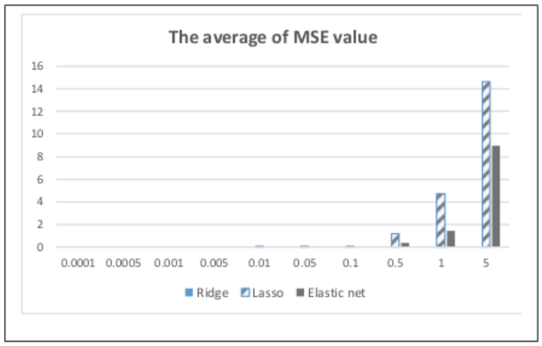

Fig. 5. The average of MSE value on all Regularization and all scenarios

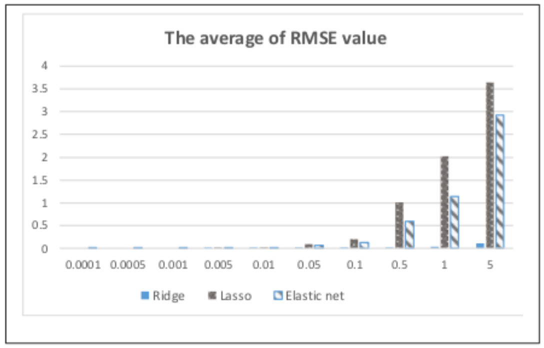

Fig. 6. The average of RMSE value on all Regularization and all scenarios

- The lasso regularization:
  a) Alpha= 0.0005; 0.0001; 0.001 on all training set size.
  b) Alpha=0.005 on training set size: 20%, 30%, 40%, 50%, 60%
  c) The elastic net Regularization
  Alpha= 0.001 on training set size: 40% and 50%.

Further, the measurement result values of the model's performance are computed in the average for each metric. The results are visualized in the bar graph in Fig.5-8.

The first measurement of the model performance is the average MSE value, as depicted in Fig.5. Ridge dominates from the other Regularization. The MSE value average of the ridge is closer to 0 more than additional Regularization on all scenarios. On the contrary, the model performance of linear regression based on lasso is not much.

The second measurement of the model performance is the average RMSE value, as depicted in Fig.6. Ridge also dominates from the other Regularization. In detail, the measurement result uses RMSE almost similar to MSE.

The third measurement of the model performance is the average MAE value, as described in Fig.7. Then, it is followed by the elastic net and then a ridge. Ridge is also superior to the other Regularization. The MAE value average of the ridge is closer to 0 more than additional Regularization. On the contrary, the model performance of linear regression based on lasso is not much. This means that the level of error prediction of linear regression based on lasso is high enough.

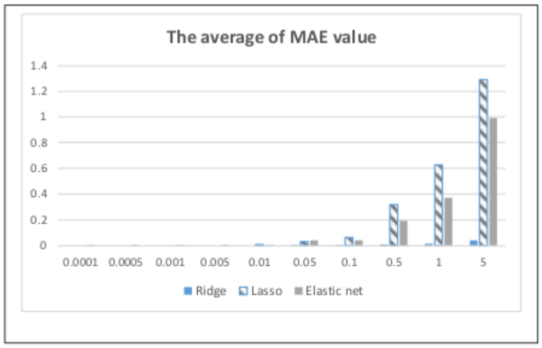

Fig. 7. The average of MAE value on all Regularization and all scenarios.

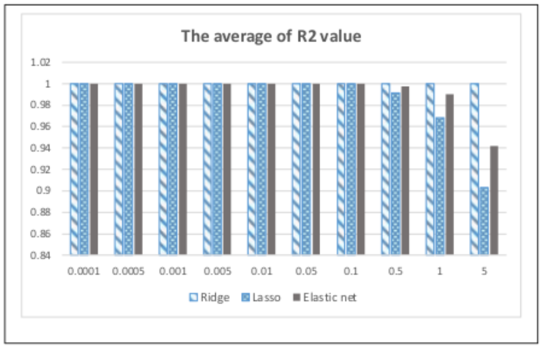

Fig. 8. The average of R2 value on all Regularization and all scenarios

The last measurement of the model performance is the average R2 value, as visualized in Fig.8. The R2 value average of the ridge is closer to 1 more than other Regularization. On the contrary, the model performance of linear regression based on lasso is not much. This means that the level of error prediction of linear regression based on lasso is high enough. Then, it is followed by the elastic net and then lasso.

IV. Conclusion

The linear regression based on Regularization is a method that can be used to build the prediction model on students' performance. The exploration results on three regularizations, namely: ridge, lasso, elastic net, and good tuning, result in better predictive model performance. In our research, the ridge regularization shows domination over the others, which is based on four metrics, namely: MSE, RMSE, MAE, and R2..

References

[1] A. Abidah, H. N. Hidaayatullaah, R. M. Simamora, D. Fehabutar, and L. Mutakinati, "The Impact of Covid-19 to Indonesian Education and Its Relation to the Philosophy of 'Merdeka Belajar,'" *Stud. Philos. Sci. Educ.*, vol. 1, no. 1, pp. 38–49, 2020, Accessed: Aug. 26, 2021. [Online]. Available: http://scie-journal.com/index.php/SiPoSE.

[2] B. Azhari and I. Fajri, "Distance learning during the COVID-19 pandemic: School closure in Indonesia," *https://doi.org/10.1080/0020739X.2021.1875072*, 2021, DOI: 10.1080/0020739X.2021.1875072.

[3] D. Yulianto and N. M. Mujtahid, "Online Assessment during Covid-19 Pandemic: EFL Teachers' Perspectives and Their Practices," *JET (Journal English Teaching)*, vol. 7, no. 2, pp. 229–242, Jun. 2021, DOI: 10.33541/JET.V7I2.2770.

[4] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Pearson Addison Wesley, 2005.

[5] J. Han, J., Kamber, M., & Pei, *Data Mining Concepts and Techniques*. USA: Elsevier, 2012.

[6] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37. Elsevier Ltd, pp. 13–49, Apr. 01, 2019, DOI: 10.1016/j.tele.2019.01.007.

[7] R. Bansal, A. Mishra, and S. N. Singh, "Mining of educational data for analyzing students' overall performance," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Jan. 2017, pp. 495–497, DOI: 10.1109/CONFLUENCE.2017.7943202.

[8] Y. Ayalew, E. Tshukudu, and M. Lefoane, "Factors Affecting Programming Performance of First-Year Students at a University in Botswana," *African J. Res. Math. Sci. Technol. Educ.*, vol. 22, no. 3, pp. 363–373, Sep. 2018, DOI: 10.1080/18117295.2018.1540169.

[9] S. M. Merchan Rubiano and J. A. Duarte Garcia, "Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance," *IEEE Lat. Am. Trans.*, vol. 14, no. 6, 2016, DOI: 10.1109/TLA.2016.7555255.

[10] Y. Altujjar, W. Altamimi, I. Al-Turaiki, and M. Al-Razgan, "Predicting Critical Courses Affecting Students Performance: A Case Study," *Procedia Comput. Sci.*, vol. 82, pp. 65–71, Jan. 2016, DOI: 10.1016/J.PROCS.2016.04.010.

[11] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Electr. Eng.*, vol. 66, pp. 541–556, Feb. 2018, doi: 10.1016/j.compeleceng.2017.03.005.

[12] A. Al Mazidi and E. Abusham, "Study of general education diploma students' performance and prediction in Sultanate of Oman, based on data mining approaches," *Int. J. Eng. Bus. Manag.*, vol. 10, p. 184797901880702, Jan. 2018, DOI: 10.1177/1847979018807020.

[13] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Students' Performance Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, Jan. 2015, DOI: 10.1016/J.PROCS.2015.12.157.

[14] L. M. Crivei, V.-S. Ionescu, and G. Czibula, "An Analysis of Supervised Learning Methods for Predicting Students' Performance in Academic Environments," *ICIC Express Lett.*, vol. 13, no. 3, pp. 181–189, 2019, DOI: 10.24507/icicel.13.03.181.

[15] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, "Early segmentation of students according to their academic performance: A predictive modeling approach," *Decis. Support Syst.*, vol. 115, pp. 36–51, Nov. 2018, DOI: 10.1016/j.dss.2018.09.001.

[16] H. Chen and P. A. . Ward, "Predicting student performance using data from an auto-grading system," 2019, Accessed: Aug. 25, 2021. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3370272.3370297.

[17] H. J. Seltman, "Experimental Design and Analysis," 2018.

[18] D. M. Lane, D. Scott, M. Hebl, R. Guerra, D. Osherson, and H. Zimmer, "Introduction to Statistics."

[19] Z. KH, T. K, and S. SG, "Correlation and simple linear regression," *Radiology*, vol. 227, no. 3, pp. 617–622, Jun. 2003, DOI: 10.1148/RADIOL.2273011499.

[20] A. N and K. M, "Simple linear regression," *Nat. Methods*, vol. 12, no. 11, pp. 999–1000, Nov. 2015, DOI: 10.1038/NMETH.3627.

[21] O. O. and C. P., "Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression," *Int. J. Comput. Appl.*, vol. 157, no. 4, pp. 37–44, Jan. 2017, DOI: 10.5120/IJCA2017912671.

[22] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58. WileyRoyal Statistical Society, pp. 267–288, 1996, DOI: 10.2307/2346178.

icvee 2021

ORIGINALITY REPORT

17%	12%	10%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Y N Widhitama, A Lukito, S Khabibah. "Problem Solving-based Learning Materials on Fraction for Training Creativity of Elementary School Students", Journal of Physics: Conference Series, 2018 Publication	2%
2	link.springer.com Internet Source	1%
3	Submitted to University of Ulster Student Paper	1%
4	rain.atmos.colostate.edu Internet Source	1%
5	Submitted to University of Leicester Student Paper	1%
6	www.hindawi.com Internet Source	1%
7	dokumen.pub Internet Source	1%

8	Dasol Kim, Mintae Kim, Wooju Kim. "Wafer Edge Yield Prediction Using a Combined Long Short-Term Memory and Feed- Forward Neural Network Model for Semiconductor Manufacturing", IEEE Access, 2020 Publication	1 %
9	Submitted to National Institute of Technology, Rourkela Student Paper	1 %
10	www.insightsociety.org Internet Source	1 %
11	www.gia.edu Internet Source	1 %
12	ejournal.akakom.ac.id Internet Source	<1 %
13	kuva.mis.hiroshima-u.ac.jp Internet Source	<1 %
14	Submitted to Franklin University Student Paper	<1 %
15	Wenty Dwi Yuniarti, Edi Winarko, Aina Musdholifah. "Data Mining for Student Assessment in e-Learning: A Survey", 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020 Publication	<1 %
	digitalcommons.wayne.edu	

16	Internet Source	<1%
17	www.sciencedirect.com Internet Source	<1%
18	uir.unisa.ac.za Internet Source	<1%
19	Kdhaiya Sulaiman Al Shibli, Amal Sulaiman Sayed Al Abri, Linitha Sunny, Nandakishore Ishwar, Sherimon Puliprathu Cherian. "Model for Prediction of Student Grades using Data Mining Algorithms", European Journal of Information Technologies and Computer Science, 2022 Publication	<1%
20	is.muni.cz Internet Source	<1%
21	jurnal.ideaspublishing.co.id Internet Source	<1%
22	www.atlantis-press.com Internet Source	<1%
23	www.biorxiv.org Internet Source	<1%
24	mafiadoc.com Internet Source	<1%
25	www.ijstr.org Internet Source	

<1%

26 www.springerprofessional.de <1%
Internet Source

27 Luxhoj, J.T.. "An artificial neural network for nonlinear estimation of the turbine flow-meter coefficient", Engineering Applications of Artificial Intelligence, 199812 <1%
Publication

28 Muhammad Adnan, Asad Habib, Jawad Ashraf, Shafaq Mussadiq, Arsalan Ali Raza, Muhammad Abid, Maryam Bashir, Sana Ullah Khan. "Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models", IEEE Access, 2021 <1%
Publication

29 Zhao Wang, Quande Liu, Qi Dou. "Contrastive Cross-site Learning with Redesigned Net for COVID-19 CT Classification", IEEE Journal of Biomedical and Health Informatics, 2020 <1%
Publication

30 Hanan Abdullah Mengash. "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems", IEEE Access, 2020 <1%
Publication

31	Marcos Wander Rodrigues, Luiz Enrique Zárate, Seiji Isotani. "Educational Data Mining: A review of evaluation process in the e-learning", Telematics and Informatics, 2018 Publication	<1%
32	Nurafifah Mohammad Suhaimi, Shuzlina Abdul-Rahman, Sofianita Mutalib, Nurzeatul Hamimah Abdul Hamid, Abdul Hamid. "Review on Predicting Students' Graduation Time Using Machine Learning Algorithms", International Journal of Modern Education and Computer Science, 2019 Publication	<1%
33	V.L. Miguéis, Ana Freitas, Paulo J.V. Garcia, André Silva. "Early segmentation of students according to their academic performance: A predictive modelling approach", Decision Support Systems, 2018 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		