Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (1-9)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Machine Learning for the Model Prediction of Final Semester Assessment (FSA) using the Multiple Linear Regression Method

Fitria Rachmawati[1*], Jejen Jaenudin[2], Novita Br Ginting[3], and Panji Laksono[4]

[1,2,3]Department of Information Systems, Faculty of Science and Technology, Ibn Khaldun Bogor University
[4]Department of Information Technology, Faculty of Science and Technology, Ibn Khaldun Bogor University
[1.2,3,4]Jl. Sholeh Iskandar, Kedungbadak, District. Tanah Sereal, Bogor City, West Java 16162, Indonesia

Article:
Accepted: March 24, 2024
Revised: January 13, 2024
Issued: April 30, 2024

***Correspondence Address:**
fitria@uika-bogor.ac.id

ABSTRACT

Corona virus (COVID-19) is the reason behind the collapse of the National Assembly. The first is the Final Semester Assessment (FSA) , which is a component of the student's graduation. The aforementioned evaluation process is a crucial consideration for the teacher since it uses several intricate unsurveys and mark components. To provide suitable results for student learning, a prediction model is employed to assist teachers. The method that is used is called the multiple linear regression. This multiple linear regression algorithm yields an accuracy level of approximately 92%. The results of the analysis using the method are used as a guide to understanding of student's index. This index is a rating that appears based on the Minimum Credit Count (MCC). Therefore, the goal of this study is to determine student understanding of the FSA prediction value, which will be taken into consideration through the results of the MCC weights in the form of a range in the form of "Grade." Additionally, the research aims to determine the accuracy results from the Model obtained using multiple linear regression algorithms in predicting students' FSA.

Keywords : *covid-19; final semester assessment (FSA); multiple linear regression; minimum credit count (MCC);*

Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (1-9)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

1. INTRODUCTION

COVID-19 has become a virus spreading quickly across the world. The government of Indonesia also immediately adjusted the conditions to minimize the spread of the COVID-19 virus by implementing the health protocols 5M, namely, wearing a mask, washing hands, keeping distance, avoiding crowds, and reducing mobility [1]. The coronavirus (COVID-19) outbreak, which causes an evenly distributed impact on all sectors of human needs, is a problem that needs to be acted upon. Large-scale social restrictions (LSSR) have greatly helped medical personnel reduce the level of virus spread. However, it is unfortunate globally that the PSBB itself has resulted in other impacts that occur not only in the health sector but also in the trade sector and education.

However, with the development of the digital world, which is increasingly increasing with artificial intelligence, it is very helpful in every work carried out by every human being, even though the PSBB is still applied. Unfortunately, many changes have occurred as a result of this. In the education sector, the criteria for student graduation are no longer seen in the National Examination (NE) results. Participation and implementation of School Exams in the emergency period of the spread of COVID-19 [2]. So, the Final Semester Assessment (FSA) will be the final choice of lesson.

Therefore, teachers must be alert when responding to a circular letter to prepare and pay attention to their students when implementing the learning. The insight that is needed for teachers to know a prediction produced is very necessary because the teacher cannot just guess what value will be obtained by the student. The variables required include several daily test (DT) scores to measure student understanding, which can greatly assist teachers in knowing the results of the raised FSA.

Several algorithms have been confirmed to be able to measure a student's value, such as Radial Basis Function, Backpropagation, Neural Network Backpropagation, and Multiple Linear Regression [3]–[5]. Many other academics have already applied multiple linear regression techniques to machine learning. Based on independent variable criteria, including age, gender, body weight, number of children, smoking habits, and area, the multiple linear regression algorithm can be used to forecast health insurance prices [6]. Many of these algorithms have been widely used, so this research refers to the multiple linear regression algorithm used to implement the resulting Model. A multiple linear regression algorithm is a data mining algorithm that is learning in nature, with the teacher involving an attribute in the form of a class needed to make a model [7]. The formed Model uses a machine learning tool.

Machine learning (ML) is a machine made with artificial intelligence to be able to imitate human behavior so that the results that appear are normal, like humans who think with all their strength, but this machine makes predictions from the results of learning analysis carried out such as analyzing an emerging trend so that Machine learning has a very broad perspective compared to humans. Of course, this is very helpful for the world of analysis. Even in the future, ML will be needed because the machine can process extensive data, which is also known as Big Data. [8].

So, the purpose of this research is to get the value of the accuracy results from the Model obtained using multiple linear regression algorithms in predicting students' FSA and to get the results of determining student understanding regarding the FSA prediction value, which will be considered through the results of the MCC weights in the form of a range in the form of "Grade".

2. METHODS

In reviewing a basic theory of a method that will be used, the following is a linkage theory that will be explained in making a machine learning model using multiple linear regression algorithms.

2.1. Machine Learning

Machine Learning (ML) is a machine learning branch of Artificial Intelligence (artificial intelligence) where this machine can act without being explicit. ML focuses on using data algorithms to mimic how humans learn [9] and can produce the level of accuracy of the algorithms used. ML can learn from old data to use a predictive model of the results obtained [10]. Of course, producing better predictive data modeling requires many datasets because ML needs to learn from FSA data. Supervised

learning, unsupervised learning, and reinforcement learning are a few approaches to machine learning [11].

2.1.1. Tuning

Tuning is the process of maximizing model performance without overfitting or creating too high variance. Usually, in machine learning, it is referred to as a "hyperparameter" in choosing a combination of models and hyperparameters using GridSearchCV because the method used in this method is to test each combination and validate each combination to find the best average value that can be used.

2.1.2. K-Fold Cross Validation

K-Fold Cross Validation (Cross Validation) is the division of several scenarios where the training data is distributed by splitting and then looking for the average score from testing several scenarios that will be used in modeling.

2.1.3. Feature Scaling

Feature Scaling is a variable data modifier method used to normalize the range of independent variables or data features in scaling so that the data becomes more balanced. Some of the scaling that can be used include:

2.1.3.1. MinMax Scaler

MinMax Scaler works by scaling the data/adjusting the data within the range commonly used, which is 0 to 1. Both variables (x) and (y) are distributed in normalization so that when modeling, the results that appear in the data become more balanced. In searching for MinMax Scaler, it can be seen in equation (1)

$$X_{minmax} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

2.1.3.2. Standard Scaler

The Standard Scaler standardizes the features by reducing the mean and scaling to the Unit Variant. Unit variance means dividing all values by the standard deviation which can be seen in equation (2). The Standard Scaler makes the distribution mean close to 0.

$$X_{standard} = \frac{X - \mu}{\sigma} \quad (2)$$

2.1.3.3. Robust Scaler

Robust Scaler changes the feature vector by subtracting the median and then dividing it by the interquartile range, which can be seen in equation (3). Robust Scaler does not execute data into predefined intervals like MinMax Scaler because it does not meet the strict definition of scale, and the range used by Robust Scaler is more extensive than MinMax Scaler.

$$X_{robust} = \frac{X - Q_2}{Q_3 - Q_1} \quad (3)$$

2.1.4. Transform

Transform is a method that uses the mean and variance to calculate the training data to change the parameters tested against the test data. Some of the Transforms that can be used include:

2.1.4.1. Yeo-Johnson

Yeo-Johnson transforms the abnormal distribution into normal, where this type of transformation does not require the value for each input variable to be absolutely positive, 0 (zero), or negative. This means that the Yeo-Johnson method can be applied to the dataset without having to scale it first. This can be seen in equation (4) regarding the value used in the Yeo-Johnson method.

$$y_i^{(\lambda)} = \begin{cases} \frac{((y_i+1)^2-1)}{\lambda} & \text{if } \lambda \neq 0,\ y \geq 0 \\ \log(y_i+1) & \text{if } \lambda = 0,\ y \geq 0 \\ -\frac{[(-y_i+1)^{(2-\lambda)}-1]}{2-\lambda} & \text{if } \lambda \neq 2,\ y < 0 \\ -\log(-y_i+1) & \text{if } \lambda = 2,\ y < 0 \end{cases} \quad (4)$$

2.1.4.2. Box-Cox

Box-Cox is a power transformation that assumes the values of the applied input variables are really positive. That means that values 0 and negative values are not supported by Box-Cox method. It can be seen in equation (5) on the value used in the Box-Cox method.

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y_i & \text{if } \lambda = 0, \end{cases} \quad (5)$$

2.2. Multiple linear regression

The regression linear Model, or the multiple Model, illustrates the relationship between one response variable (Y) and two or more predictor variables (X1, X2,... Xn).

Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (1-9)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

Determine the response (Y) to each predictor variable (X1, X2,..., Xn) if the variables are known. This sampling also determines the degree to which variable-to-variable dependence exists [12]. Multiple Linear Regression is a data mining algorithm carried out in learning with the teacher to allow the variables used to require a class in a required attribute. [13]. In this case, the formula for a multiple linear regression can be seen in equation (6)

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \quad (6)$$

To get a search for these parameter values, a process is carried out in the search as follows:

2.2.1. Estimating parameter values β

In determining the parameter value, the data that has been normalized can be searched for the estimated value of $\beta_0, \beta_1, \beta_2$ Can be seen in the equation (7),(8),(9)

$$\beta_1 = \frac{\sum x_2^2 \sum x_1 y - \sum x_1 x_2 \sum x_2 y}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} \quad (7)$$

$$\beta_2 = \frac{\sum x_1^2 \sum x_2 y - \sum x_1 x_2 \sum x_1 y}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} \quad (8)$$

and,

$$\beta_0 = \bar{Y} - \beta_1 \overline{x_1} - \beta_2 \overline{x_2} \quad (9)$$

In equation (7), the parameter value is searched for β_1 then proceed to search for parameter values for β_2 shown in equation (8), and the last step is to search for parameter values for β_0 in equation (9).

2.2.2. Estimating error values (ε)

The purpose of determining the error value is to see the results of the error from the calculation of the prediction data with the original data so that the error value can be used as material for further analysis [14], where the resulting calculation gets the results of how much accuracy can be seen equation (10).

$$\varepsilon_n = y_n - (\beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2}) \quad (10)$$

2.2.3. Python

Python is static programming with a simple syntax for statistical calculations and website creation. The language developed by Guido van Rossum is extraordinary because of its wide appeal in the world of software development professionals, scientists, researchers, artists, and educators. This very easy-to-understand language can collaborate with machine learning, which is certainly very helpful for researchers in predicting the data to be analyzed [15].

In getting an accurate result obtained from the expected Model so that the Model can be said to be a predictive model in predicting FSA results, which will be used as teacher insight, the research procedure can be seen in Figure 1.

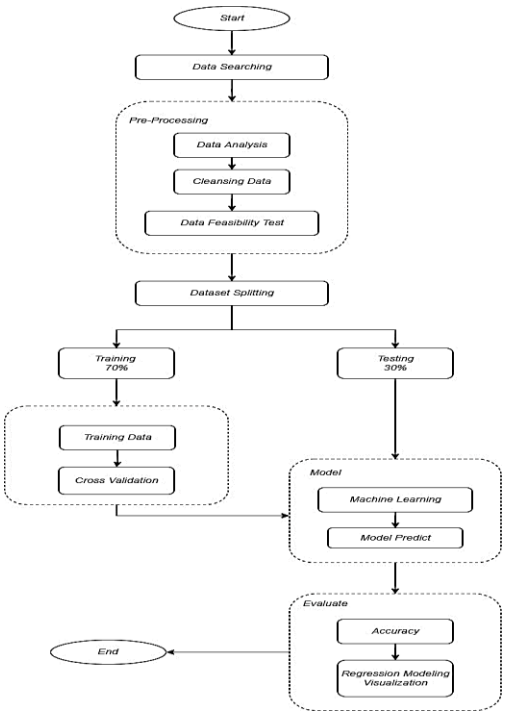

Figure 1. *Research procedure*

2.3. Data Searching

Research data is searched in the form of student scores to predict the FSA value obtained from SD Negeri Kebon Kelapa. The data that has been collected will proceed to several stages of the research process, including pre-processing, dataset splitting, modeling, and Evaluation.

2.4. Pre-Processing

Pre-processing is intended at the next stage of data processing, where the data has been collected, and carried out several stages of processing, including:

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

2.4.1. Data Analysis

Any data collected is, of course, very random, so in this case, it is necessary to analyze the data so that it can be processed properly. The data analysis includes searching for variables that will be needed in data processing and producing a model for the algorithm used.

2.4.2. Cleansing Data

At this stage, the data certainly needs to be cleaned of variables that are not used so that they do not interfere with the built Model. In this case, the variables needed in the processing consist of several Daily Test (DT) scores from each Theme and Final Semester Assessment. (FSA). In this problem, the data collected was obtained from Grade 4 students for the 2019 to 2021 period. A partial collection of the dataset can be seen in Table 1.

Table 1. *Partial collection of data set*

No	Nama Siswa	DT1	DT2	DT3	DT4	FSA
1	Adi Satria	80	88	78	78	79
2	Agung Terry	96	95	86	86	86
3	Agus Indrawan	84	92	88	88	84
...	...	...	...	...	...	...
247	Zeevana N	65	76	86	70	76

2.4.3. Data Feasibility Test

When the data has been cleaned of variables that are not needed, then the data will be carried out a data feasibility test process to find out whether the data is feasible enough to use a Linear Regression model or not by looking at the correlation relationship between the variables used.

2.5. Dataset Splitting

Data that has been cleaned and the variables ready to be used can be divided into the data, which is 80% of the data for training and 20% for testing. The contribution of the distribution of the dataset can be seen in Table 2.

Table 2. *Contribution of dataset sharing*

Jumlah Siswa	Training (80%)	Testing (20%)
247 data	197 data random	50 data random

2.6. Validate

Validation is carried out on the distribution of training data, which will then be carried out in several stages, including:

2.6.1. Training Data

Machine learning will carry out a learning stage on the data sharing results from the collected training data. This stage will also make some changes or additions to parameters such as scaling and transform, which aims to make the scenario results appear more balanced so that they are very good to use in the next test scenario.

2.6.2. Cross Validation

Cross-validation is the stage of dividing the test into several scenarios. Usually, the final validation stage is carried out to get the prediction results evenly so that validation is carried out in each different scenario, and the average value of the division's results is then searched for. Then, the next stage will be to make a prediction model through machine learning.

2.7. Model

The modeling of data that has been validated, as well as data that is divided into testing data, will be continued into the model stage, where there are several stages of the process, including:

2.7.1. Machine Learning

The validated data in the form of the best average value will be used as a model that will be built on machine learning to create a prediction model.

2.7.2. Model Predict

This prediction model will be used to implement value testing in predicting students' FSA, which can be done by searching for accurate results outlined in the evaluation stage.

2.8. Evaluate

The evaluation of each research obtained will certainly be a new insight into research development, where the evaluation can be raised in the form of accuracy results and visualization of the algorithm model used.

Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (1-9)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

3. RESULTS AND DISCUSSION

3.1. Model Making

From the results of pre-processing the data analysis that has been carried out on the feasibility of the data in the use of a linear regression model, in making the Model, several parameter changes (hyperparameters) will be made where changes in these parameters will affect the modeling of the learning process that will be carried out. The parameter scenarios can be seen in Table 3.

Table 3. *Parameter scenario*

No	Parameter	Keterangan
1	*Variable Data Uses*	*4 data*
2	*Variable Data Target*	*1 data*
3	*(X_train, y_train)*	*(197, 4) (197,)*
4	*(X_test, y_test)*	*(50, 4) (50,)*
5	*Tuning*	*"GridSearchCV" params*
6	*Cross Validation*	*3 scenario*
7	*Scaling*	*"STANDARD"*
8	*Transform*	*"YEO-JOHNSON"*
9	*Plot Model*	*"LinearRegression"*
10	*Polynomial*	*2 degree*

In Table 3. the data variables used are 4 data, including the value of the Daily Test 1, the value of the Daily Test 2, the value of the Daily Test 3, and the value of the Daily Test 4. Some of the values of the Daily Test will be made into a model based on odd semesters. Namely DT1 (Score on Theme 1), DT2 (Score on Theme 2), DT3 (Score on Theme 3), DT4 (Score on Theme 4), and models based on even semesters, namely DT5 (Score on Theme 5), DT6 (Score on Theme 5) on Theme 6), DT7 (Score on Theme 7), DT8 (Score on Theme 8). The target variable only requires 1 data set, namely the Final Semester Assessment (FSA), which will be predicted based on the Odd Semester or Even Semester FSA. So, in distributing the splitting dataset, divided into training and testing data, there are 197 training data and 50 testing data.

Tuning is done using GridSearchCV params to search for DT over the parameter value specified as an estimator. Where the validation will be divided into 3 scenarios by multiplying some of the interception results (slices) obtained from tuning in producing the best average value that will be used as material for evaluating the search for the accuracy value of a model to be built, the implementation of this method produces several average values. The average has been scaled using "STANDARD" and transformed using the "Yeo-Johnson" model because the distribution of the required variables does not pay attention to negative or positive values in transforming variables.

Forming a model is based on variable needs so that students can use linear regression modeling in predicting the FSA, where the variable x is the data variable and the variable y is the target variable. The polynomial (many terms) used is 2 degrees, meaning that in this case, the scope of the resulting straight-line data will look for points within the scope of 2 degrees, whether it be on training data or testing data to get more coefficient results. So that the output in searching for the accuracy level of the algorithm in predicting students' FSA can be seen in Figure 2.

3.2. Model Test Results

The Model will be tested based on the discussion of the model results, including parameter tuning, accuracy results, and dataset visualization.

3.2.1. Tuning Parameter

In the tuning parameter experiment, the data used is one of the results of dataset splitting, namely training data; because machine learning requires learning from the data, it has to be able to validate the Model by dividing it into 3 alternative scenarios of testing in determining the search for the best average value. The experimental results of the analysis of the 3 scenarios can be seen in Table 4.

Table 4. *Cross validation*

No	*Split 0 Test*	*Split 1 Test*	*Split 2 Test*	*Mean Test Score*	*Rank Score*
1	0.76721	0.909698	0.854394	0.843767	1
2	0.76692	0.909689	0.854515	0.843711	2
3	0.76692	0.909689	0.854515	0.843711	2
4	0.76690	0.909674	0.854515	0.843697	4
5	0.76589	0.909689	0.854515	0.843366	5
6	0.76589	0.909689	0.854515	0.843366	5

Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (1-9)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

Table 4 continued…

No	*Split 0 Test*	*Split 1 Test*	*Split 2 Test*	*Mean Test Score*	*Rank Score*
7	0.76589	0.909689	0.854515	0.843366	7
8	0.76589	0.909689	0.854515	0.843366	7
9	0.76696	0.910074	0.852520	0.843187	9
10	0.76685	0.910087	0.852520	0.843154	10
11	0.76201	0.881712	0.857733	0.833820	11
12	0.39800	0.680342	0.856360	0.644901	12

The conclusion in Table 4 is the tuning parameter experiments obtained from the results of the division of 3 scenarios. These results are obtained in the split test scores of the 3 scenarios, which are then accumulated into the mean test score to get the best average results that can be used as multiple linear regression modeling algorithms.

So the scenario results obtained from the training data above are that in test 1 it produces a mean test score of 0.843767, in test 2 it produces a mean test score of 0.843711, in test 3 it produces a mean test score of 0.843711, in test 4 produces a mean test score of 0.843697, in test 5 it produces a mean test score of 0.843366, in test 6 it produces a mean test score of 0.843366, in test 7 it produces a mean test score of 0.843366, in test 8 it produces a mean test score of 0.843366, in test 9 produces a mean test score of 0.843187, in test 10 it produces a mean test score of 0.843154, in test 11 it produces a mean test score of 0.833820, and in test 12 it produces a mean test score of 0.644901.

From the results above, it was concluded that the best validation experiment was obtained from test 1, which stated the mean test score resulted in a value of 0.843767 so that the results of this training validation (train-val) modeling can be carried out in the next stage, namely testing the process that is entered into machine learning modeling so that the data ready to be processed to produce an evaluation of the learning that has been obtained.

3.3. Accuracy

In the test results of the train-val, the search stage is carried out to obtain the accuracy value of the multiple linear regression algorithm in predicting the final semester assessment by testing 4 variable data and 1 target variable data by changing parameters (hyperparameters) and transforming the data that has been calculated. Tuning using the "YEO-JOHNSON" transform, the main function of the transform is to distribute data with negative or positive values so that the managed data can be completed (compared) appropriately. The accuracy results obtained can be seen in Table 5, and the model accuracy results can be visualized in Figure 2.

Table 5. *Accuracy results*

Model Score (X_train, y_train)	**Model Best Score**	**Model Score (X_test, y_test)**
0.8698398924412 651 (87%)	0.8437671166060 108	0.9242165795609 549 (92%)

In Table 5. the results obtained from the search for modeling the accuracy level using training data are 0.870 (87%), and the accuracy level using testing data is 0.924 (92%). The results of the best score model obtained from modeling the distribution of 3 scenarios are described in the conclusion of Table 3.

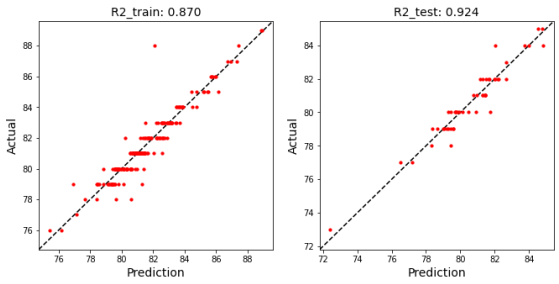

Figure 2. *Plot Visualization of Accuracy Results*

In Figure 2, the points mentioned above are data derived from training data and testing data to obtain accurate results from linear regression modeling. At the same time, for straight lines, they are prediction results from machine learning modeling using linear regression models obtained to predict the existence of new data to be tested, where if the point approaches a straight line, then the new data to be predicted can produce optimal and predictable values.

Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (1-9)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.28652

3.4. Dataset Visualization

In looking at the results obtained from predictions on the data, the dataset modeling needs to be considered. The visualization of the dataset modeling can be seen in Figure 3.

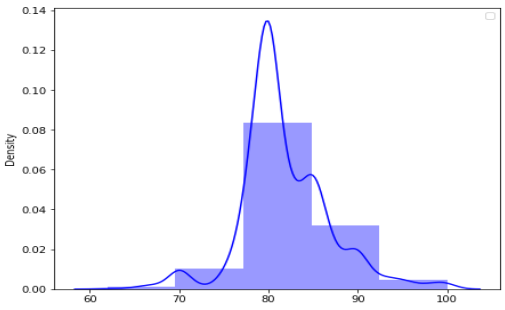

Figure 3. *Dataset modeling visualization*

In Figure 3. the visualized dataset is quite good in processing learning data on machine learning, but the data still needs to be considered for normal distribution so that the data generated in the search for accuracy levels for the linear regression model is not too underfit (referring to a model that cannot model training data or generalize to new data.) or overfit (good performance on training data, poor generalization to other data.). If the data is too underfitted or overfit, then this will result in poor algorithm performance against machine learning.

3.5. Prediction Model Testing

The results of modeling and the results of model testing in the search for accuracy values, after which visualization of the training data modeling and testing data are carried out to obtain a model that can be implemented on the new data to be tested so that the prediction model test can be seen in the following stages:

3.5.1. Prediction Results

In testing the Model that has been created, the new data will later be tested through the machine learning model in a prediction that produces an output in the form of the student's FSA value. The test can be seen in the sample, which can be seen in Figure 4.

Out[14]:

	UH1	UH2	UH3	UH4	Prediksi PAS	Grade
0	100	90	77	100	87.483586	Baik

Figure 4. *New data prediction results*

In Figure 4 above, it can be concluded that testing new data by entering several variables$x_1 = 100,\ x_2 = 90,\ x_3 = 77,\ x_4 = 100$, then it produces a y value (prediction result) of 87.483 with a Grade description that is "Good".

3.5.2. Prediction Result Model Regression Visualization

Visualization was done to determine the accumulated predictive value generated from the test data, where one of the data for class 1 students was used, as seen in Figure 5.

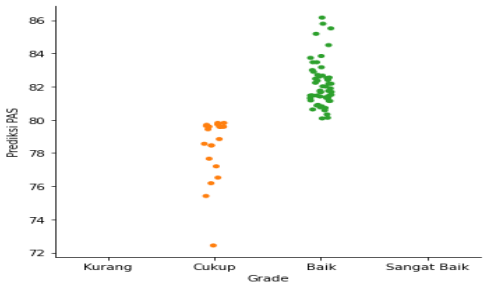

Figure 5. *Visualization of the predicted result accumulation model*

In Figure 5. it can be concluded that the plot of the Model is the result of the accumulation of visualized predictions so that from a collection of classes, it can be concluded that in these subjects, there is a lot of data collected in the "Good" and "Enough" Grades so that the teacher can conclude that students can understand the learning that has been done.

CONCLUSION

Based on the results of testing data analysis and the application of multiple linear regression algorithms in the prediction of students' final semester assessment (FSA), several conclusions can be drawn, including (1) The results obtained from the accuracy of the Model obtained using multiple linear regression algorithms in predicting students' FSA obtained results from the search for modeling the accuracy level using training data of 0.870 (87%). The accuracy level using testing data is 0.924 (92%). Meanwhile, the best score model results are obtained from modeling the distribution of 3 scenarios. (2) The results obtained in testing the analysis of an FSA prediction using new data as an example in the student score index in determining student understanding generated through "Grade".

REFERENCES

[1] Arrahman Kaffi et. al, The Development of Interactive Games for Covid-19 Prevention Using Indonesian Health Protocols, Jurnal Teknik Informatika Vol. 15 No. 2, (101-109), 2022.

[2] KEMENDIKBUD, "SURAT EDARAN TENTANG PENIADAAN UJIAN NASIONAL DAN UJIAN KESETARAAN SERTA PELAKSANAAN UJIAN SEKOLAH," 2021. https://lldikti13.kemdikbud.go.id/2021/02/04/surat-edaran-menteri-pendidikan-dan-kebudayaan-nomor-1-tahun-2021-tentang-peniadaan-ujian-nasional-dan-ujian-kesetaraan-serta-pelaksanaan-ujian-sekolah-dalam-masa-darurat-penyebaraan-corona-virus-dise/

[3] K. Farouq Mauladi and Masruroh, "PERBANDINGAN METODE REGRESI LINEAR DAN NEURAL NETWORK BACKPROPAGATION DALAM PREDIKSI NILAI UJIAN NASIONAL SISWA SMP MENGGUNAKAN SOFTWARE R," *JOUTICA*, vol. 5, no. 1, pp. 331–336, 2020.

[4] S. Kumala Sari and J. Manurung, "Implementasi Jaringan Syaraf Tiruan Untuk Memprediksi Tingkat Pemahaman Siswa Pada Mata Pelajaran Ujian Akhir Sekolah (UAS) di SD Mis An Nur Sukamandi Menggunakan Metode Backpropragation," *JIKOMSI : Jurna Ilmu Komputer dan Sistem Informasi*, vol. 3, no. 3, pp. 283–292, 2021.

[5] R. Afriani, S. Hotlan Sitorus, and U. Ristian, "Aplikasi Prediksi Nilai Siswa Sekolah Dasar Menggunakan Metode Radial Basis Function Berbasis Web (Studi Kasus : SDN 19 Sungai Raya)," *Coding: Jurnal Komputer dan Aplikasi*, vol. 07, no. 03, pp. 85–96, 2019.

[6] Jesi Pebralia, Rainfall Analysis using Machine Learning-Multiple Linear Regression Method Based on Python and Jupyter Notebook, JIFP (Jurnal Ilmu Fisika dan Pembelajarannya) Vol. 6, No. 2, 23-30, Desember 2022.

[7] G. MARDIATMOKO, "PENTINGNYA UJI ASUMSI KLASIK PADA ANALISIS REGRESI LINIER BERGANDA," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 14, no. 3, pp. 333–342, Oct. 2020, doi: 10.30598/barekengvol14iss3pp333-342.

[8] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python," *BINA INSANI ICT JOURNAL*, vol. 7, no. 2, pp. 156–165, 2020.

[9] A. Ahmad, "Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning," *Yayasan Cahaya Islam, Jurnal Teknologi Indonesia*, 2017.

[10] S. Kumar, D. Shah dan S. Patel, "Heart Disease Prediction using Machine Learning Techniques," SN Computer Science, vol. I, no. 1, p. 345, Oct. 2020.

[11] W. Budiharto, *Machine Learning & Computational Intelligence*, VI. Yogyakarta: ANDI OFFSET, 2016.

[12] R. Z. Nainggolan, K. Ibnutama, and M. G. Suryanata, "Implementasi Data Mining Dengan Metode Regresi Linier BergandaDalam Estimasi Mahasiswa Baru Pada SekolahTinggi Agama Islam Raudhatul Akmal BatangKuis," Jurnal CyberTech, vol. 1, no. 1, pp. 13–20, 2021

[13] M. S. Prof. Dr. Suyono, "Analisis Regresi untuk Penelitian," I., Yogyakarta: DEEPUBLISH (CV BUDI UTAMA), 2015.

[14] M. S. Prof. Dr. Suyono, "Analisis Regresi untuk Penelitian," I., Yogyakarta: DEEPUBLISH (CV BUDI UTAMA), 2015.

[15] P. Wentworth, J. Elkner, A. B. Downey, and C. Meyers, "How to Think Like a Computer Scientist: Learning with Python 3 Documentation Release 3rd Edition," 2012.