

# Prediction Models with Machine Learning Against Student Success in Online Learning

Yennimar<sup>1)\*</sup>, Rohni Endetta Manihuruk<sup>2)</sup>, Etis Landya Br Hotang<sup>3)</sup>

<sup>1)2)3)</sup>Universitas Prima Indonesia, Indonesia

<sup>1)</sup>[yennimargucci@gmail.com](mailto:yennimargucci@gmail.com), <sup>2)</sup>[rohnimanihuruk123@gmail.com](mailto:rohnimanihuruk123@gmail.com), <sup>3)</sup>[etislandyaa@gmail.com](mailto:etislandyaa@gmail.com)

Submitted : Aug 20, 2021 | Accepted : Oct 10, 2021 | Published : Oct 10, 2021

**Abstract:** The learning system during the Covid-19 pandemic was carried out online, online learning had a negative impact and a positive impact. The impact given can affect the success of student learning. The success of learning is the main thing that must be achieved by students. From the success of learning, it can be seen that the online learning process is going well or not. To determine the success rate of online learning, testing is carried out by applying a neural network algorithm. Neural network algorithms are used because they can solve complex problems related to prediction. This research is expected to help lecturers or campus parties to create better online learning. In this study using student grade data for Academic Year 2018/2019 and Academic Year 2019/2020, data testing using Rapidminer software and operator cross validation. In testing the Academic Year 2018/2019 and Academic Year 2019/2020 value data using 700 training cycles, 0.4 momentum, 0.2 learning rate and hidden layer 2. The level of accuracy obtained in the 2018/2019 student grade data is 95, 55% and Academic Year 2019/2020 which is 93.17%. From the test results, it was found that the accuracy rate of Academic Year 2018/2019 is higher than Academic Year 2019/2020, so the success rate in Academic Year 2018/2019 before the pandemic is better than the success rate in Academic Year 2019/2020 after the pandemic.

**Keywords:** Machine Learning, Neural Network, Online Learning, Prediction, Success.

## INTRODUCTION

During the pandemic *Covid-19*, learning and teaching activities are carried out online. This decision was taken as an alternative step to break the chain of virus spread (Pratama & Mulyati, 2020). Online learning is a learning and teaching process that is carried out by utilizing technology and the internet (Prasetya & Harjanto, 2020). Several platforms used as learning media such as Google Classroom, Edmodo, Edlink, Whatsapp, E-Mail, and Zoom are used for online learning (Puspitorini, 2020). Online learning has both positive and negative impacts. The positive impact given is providing the latest experience for lecturers and students, learning can be done anywhere and anytime, being an effective choice during pandemic conditions. While the negative impacts are the internet network is not good, students find it difficult to understand the material, and the lack of facilities that support online learning (Adi, Oka, & Wati, 2021). Because the impact that occurs can affect student success in learning. The success of online learning can be known from internal factors and external factors. Internal factors that influence the success of learning are comfort, willingness, interest, participation, and student health. While external factors that affect the success of learning are support, atmosphere, and student learning facilities (Mesra, Kuntarto, & Chan, 2021).

To find out the success, prediction techniques using machine learning are needed. Several methods used to predict learning are SVM, Naïve Bayes (Supriyatna & Mustika, 2018), Fuzzy C-Means (Sismadi & Kusnadi, 2018), Back Propagation Neural Network (Kolekar, Pai, & Manohara Pai, 2017), and Artificial Neural Network (Hasibuan, Nugroho, & Santosa, 2019). One of the machine learning methods, namely Neural Network, is a collection of connected I/O modules, and each connection has an appropriate weight. Neural networks have many characteristics that make them popular in clustering. A neural network is basically a parallel and distributed processing architecture (Rahayu, Nugraha, & Shidiq, 2019). Previous research conducted a classification of student success predicates (Luvia, Hartama, Windarto, & Solikhun, 2016). This study will implement a neural network algorithm to determine student success during online learning. This research can help lecturers or campus parties in knowing students who understand the online learning that is carried out so that the online learning process can be improved even better.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## LITERATURE REVIEW

Dharshinni, et al research applied the C4.5 algorithm to find out the constraints of online learning on student interest in learning during the covid 19 pandemics from the results of research on factors that became obstacles to online learning, namely the features provided with an accuracy value of 74% (Dharshinni, Sitepu, Syuhada, Barasa, & Wijaya, 2021). Research conducted by Sismadi & Kusnadi applies Fuzzy C-Means to predict student graduation by using the parameter, namely user activity during learning, the accuracy rate obtained is 78% (Sismadi & Kusnadi, 2018). In the research conducted by Yulianto et al, they compared the KNN algorithm and the Decision Tree C4.5 algorithm to determine student performance in order to improve student achievement. From the comparison of these algorithms, a good level of accuracy is obtained in the KNN algorithm with an accuracy of 59.32%, while the accuracy of the Decision Tree is 54.80% (Yulianto, Triayudi, & Sholihati, 2020).

In the research conducted by Putra & Raharjo using the Neural Network to determine the success rate of immunotherapy and the architecture used training cycles of 500, the momentum of 0.9, and learning rate of 0.3, it obtained an accuracy rate of 80% and an AUC value of 0.72 (Putra & Raharjo, 2019). Research conducted by Al-Khowarizmi using the Evolutionary Neural Network algorithm which is a development of the Neural Network by using the Genetic algorithm to class nominal and real Rupiah values has an accuracy rate of 91.61%, has a good level of accuracy because the parameters used to generate the Neural algorithm Network is much better (Al-Khowarizmi, 2020). Neural Networks are widely used to solve complex problems related to input, prediction, and pattern recognition (Hadianto, Novitasari, & Rahmawati, 2019).

## METHOD

In this study, the research procedure was as follows:

### Data Collection The data

Used in this study were student grade data for the 2018/2019 academic year up to the 2019/2020 academic year. The data used consists of UTS value data, attendance scores, assignment scores, UAS scores, and average scores. The data collected is 3135 data.

### Pre-Processing Data

Before testing the data, the data preprocessing stage is carried out. The stages of data preprocessing carried out are:

- a. Data Normalization  
At this stage, the data for the 2018/2019 academic year and the 2019/2020 academic year are separated.
- b. Data Transformation  
At this stage, data selection is carried out for attribute selection.

### Testing

Tests are carried out using Rapidminer software and using cross validation operators. The test also applies a neural network algorithm. The neural network algorithm uses several parameters, namely:

- a. Training Cycles  
That is the number of repetitions done to get a small error.
- b. Learning Rate  
That is to determine the weight of the neuron. The weight of the neurons with large values accelerates the process but there are oscillations in the weights, while the weights of the neurons with small values slow down the process.
- c. Momentum  
Namely to increase convergence in order to speed up training time and reduce oscillations.

### Application of the Neural Network Algorithm

Neural Network was proposed by McCulloch and Pitts in 1943. Neural network cooperates with the human nervous system, namely (Rahmalia & Herlambang, 2017):

1. Signals are transmitted between neurons via connectors.
2. Connectors between neurons have weights to increase or decrease the signal.
3. To determine the output, each neuron uses an activation function that corresponds to the number of inputs received.

In NN, the activation function is used to determine the output of the neuron, and the independent variable of the activation function is a linear combination of inputs and weights.

$$\text{net} = \sum_i x_i w_i \quad (1)$$

$$f(\text{net}) = f(\sum_i x_i w_i) \quad (2)$$

In general, the activation function is continuous, differential, and non-decreasing. The activation function in this study is a binary sigmoid graph with a range (0-1).

\*name of corresponding author



$$f(net) = \frac{1}{1+e^{-net}} \quad (3)$$

$$f'(net) = f(net)(1 - f(net)) \quad (4)$$

## RESULT

### Pre-Processing Data

Before the preprocessing stage of data, the data is collected as in the table below.

Table 1  
Dataset

No	Tahun	UTS Score	Assignment scores	Attendance scores	UAS Scores	Total Value
1	2020	90	100	86	100	94,2
2	2020	60	100	4	0	28,8
3	2020	30	100	57	60	54,4
4	2018	85	100	80	80	83,5
5	2018	80	78	0	0	31,8
6	2018	80	100	90	70	80
7	2019	0	100	75	88	60,2
8	2019	50	86	80	80	71,6
9	2019	50	100	80	80	73
10	2019	0	82	100	90	64,2
...	...	...	...	...	...	...
3135	2019	67	100	68	67	70.5

At the initial stage of data preprocessing, data normalization was carried out, at this stage the data was separated into two tables, namely the student value data table for the academic year 2018/2019 and the student value data table for the 2019/2020 academic year. After the data normalization stage is complete, the data transformation stage is carried out, this stage is carried out to select the data and the selection of attributes that will be used in the test. The following is a display of the transformation table below.

Table 2  
Transformation Result Academic Year 2018/2019

No	UTS Scores	Assignment Scores	Attendance Scores	UAS Scores	Total Value	Huruf
1	70	100	90	85	83	A
2	60	100	0	0	28	E
3	60	100	50	50	58	D
4	40	100	4	0	22,8	E
5	36	100	22	50	45,2	E
6	82	100	91	77	83,6	A
7	72	100	75	79	78,2	B
8	66	100	69	93	80,8	A
9	67	100	43	79	70,3	B
10	85	100	70	0	49,5	E
...	...	...	...	...	...	...
2403	60	94	65	85	74,4	B

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 3  
Transformation Result Academic Year 2019/2020

No	UTS Scores	Assignment Scores	Attendance Scores	UAS Scores	Total Value	Huruf
1	90	100	86	100	94,2	A
2	60	100	4	0	28,8	E
3	30	100	57	60	54,4	D
4	90	100	80	70	81	A
5	18	100	39	45	41,2	E
6	53	100	13	58	51,7	D
7	45	0	0	0	13,5	E
8	60	100	64	75	70,8	B
9	30	100	24	50	43,8	E
10	60	100	32	70	62,4	C
...	...	...	...	...	...	...
732	67	100	90	95	86,1	A

**Testing**

In the data preprocessing table, the data will be processed using Rapidminer software by implementing a neural network algorithm. The test is carried out using the Cross Validation operator, this test is carried out to find out the best level of accuracy by using student grade data for the 2018/2019 academic year and student grade data for the 2019/2020 academic year so that the results of the tests carried out can determine the level of success in online learning. Testing using Rapidminer software can be seen in the image below.

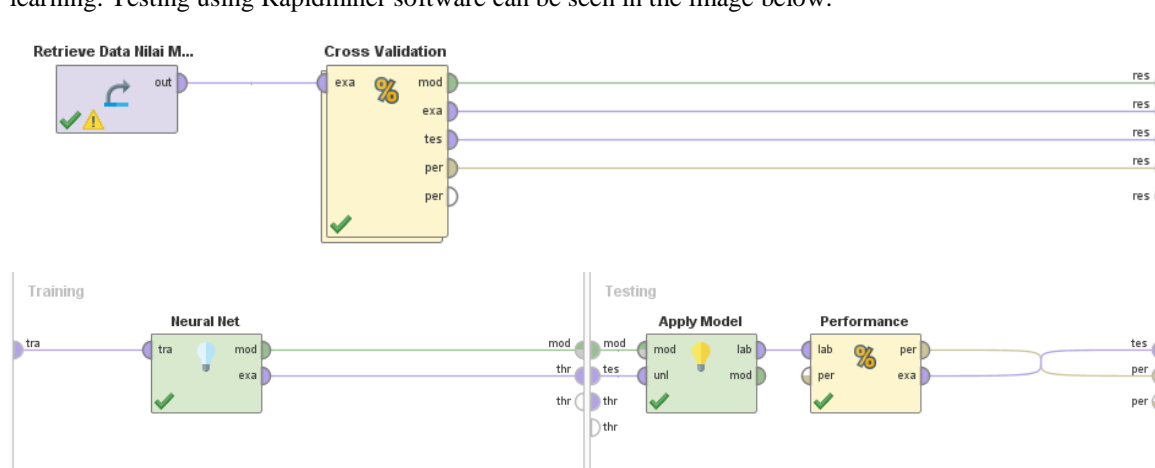


Fig. 1 Display of neural network algorithm application on Rapidminer

In the second test, the data used Rapidminer software using operator cross validation, neural network, apply model, and performance. The cross validation operator is used to make changes to the number of k-folds. Because the cross validation technique is to divide the data randomly and then the data is entered into k to find a good level of accuracy. The neural network operator uses Training Cycles 700, Learning Rate 0.2, Momentum 0.4, and for the hidden layer and neuron size 2. While the performance operators used are accuracy, absolute error, and RMSE(Root Mean Squared Error).

In the first test conducted on the 2018/2019 Academic Year value data, good accuracy was found at 7 k-fold with an accuracy of 95.55%, an absolute error was 0.129 and RMSE was 0.236. While in the 2019/2020 Academic Year value data, good accuracy is found at 8 k-fold with an accuracy of 93.17%, absolute error is 0.099 and RMSE is 0.219. If the absolute error has a small value then the error rate is also small and if the RMSE has a low value it will produce a good value. The following table of test results of the two data can be seen in the table below.

\*name of corresponding author



Table 4  
Data Value Test Result Academic Year 2018/2019

Sampling Type: Linear Sampling			
K-Fold	Accuracy	Absolute Errorr	RMSE
2	95,21%	0,133	0,248
3	95,26%	0,138	0,250
4	90,97%	0,141	0,260
5	95,59%	0,130	0,238
6	93,13%	0,133	0,250
7	95,55%	0,129	0,236
8	93,63%	0,138	0,249
9	93,63%	0,139	0,244
10	93,92%	0,148	0,257

Table 5  
Data Value Test Result Academic Year 2019/2020

Sampling Type: Linear Sampling			
K-Fold	Accuracy	Absolute Errorr	RMSE
2	92,62%	0,150	0,281
3	81,56%	0,158	0,280
4	93,17%	0,100	0,231
5	81,81%	0,197	0,321
6	93,17%	0,121	0,238
7	86,52%	0,135	0,252
8	93,17%	0,099	0,219
9	92,88%	0,105	0,229
10	93,17%	0,117	0,226

### DISCUSSIONS

Based on the test results on the 2018/2019 academic year student score data using 7 k-fold because it has an accuracy rate of 95.55% with an absolute error of 0.129 and the RMSE obtained is 0.236 while 2019/2020 student grade data uses 8 k-fold because it has The level of accuracy is 93.17% with an absolute error of 0.009 and RMSE of 0.219. In the second test, the student value data produced 6 input layers, 3 hidden layers, and 5 output layers. So judging by the comparison of Academic Year 2018/2019 before the pandemic and Academic Year 2019/2020 after the pandemic, it can be seen that the success rate of students was greater in Academic Year 2018/2019 before the pandemic because it was easier for students to understand the material provided, while during a pandemic, students had difficulty. in understanding the material due to several factors experienced. The following shows the performance vector and neural network architecture of the two student value data.

Table 6  
Performance Vektor In Testing Data Value Academic Year 2018/2019

Accuracy: 95,55%					
Absolute Error: 0,129					
RMSE: 0,236					
True	A	E	D	B	C
A	1069	0	0	2	0
E	0	166	62	0	2
D	0	0	0	0	0
B	2	0	0	831	24
C	0	1	13	1	230

\*name of corresponding author



Table 7  
Performance Vector In Testing Data Value Academic Year 2019/2020

Accuracy: 93,17%					
Absolute Error: 0,099					
RMSE: 0,219					
True	A	E	D	B	C
A	469	0	0	6	0
E	0	6	2	0	4
D	0	0	0	0	0
B	7	0	0	171	4
C	0	17	10	0	36

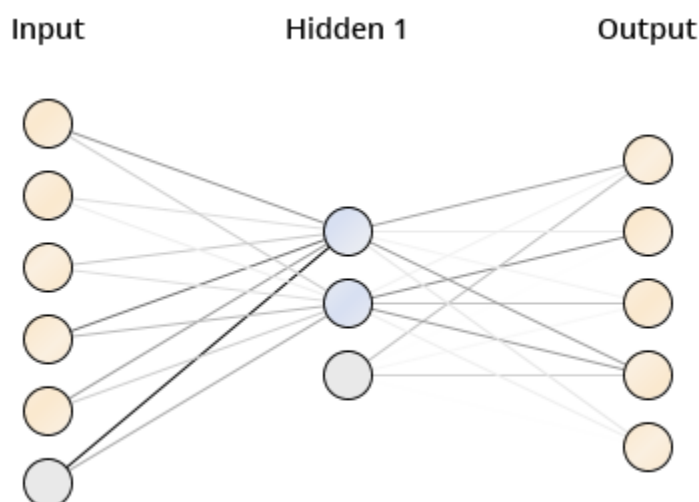


Fig. 2 Neural Network Architecture

The input layer is a layer consisting of input nodes that receive input from outside such as student grade data consisting of UTS, UAS, assignments, attendance, total values, and range values so that the input data in the image above has 6 layers. The value of the input layer is normalized first and then it will be sent from the input layer to the hidden layer. The hidden layer serves to improve the network's ability to solve problems. In the picture above there are 3 neurons and each neuron in the input layer and output layer will be connected to the hidden layer through weights and activation functions. The output weight of the hidden layer will be forwarded to the output layer which has 3 outputs. The output layer functions to distribute the output signals resulting from network processing.

### CONCLUSION

The success of learning can be seen in the rise and fall of student scores during the pandemic based on the scores obtained by students. In this study, the data attributes of student scores such as UAS, UTS, Assignments, Attendance scores data were used. Based on the test results in this study, it is known that the success of learning in the 2018/2019 Academic Year has an accuracy rate of 95.55% while the learning success in 2019/2020 Academic Year has an accuracy of 93.17%. The results show that the success rate of students in Academic Year 2018/2019 has a higher accuracy rate than Academic Year 2019/2020. Therefore it is necessary to make improvements in the online learning process in order to increase the success of students in participating in online learning.

### REFERENCES

- Adi, N. N. S., Oka, D. N., & Wati, N. M. S. (2021). Dampak Positif dan Negatif Pembelajaran Jarak Jauh di Masa Pandemi Covid-19. *Journal Ilmiah Pendidikan Dan Pembelajaran*, 5(1), 1–13.
- Al-Khowarizmi. (2020). Model Classification Of Nominal Value And The Original Of IDR Money By Applying Evolutionary Neural Network. *Journal of Informatics and Telecommunication Engineering*, 3(2), 258–265.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- <https://doi.org/10.31289/jite.v3i2.3284>
- Dharshinni, N. P., Sitepu, A. H., Syuhada, R. Y., Barasa, D., & Wijaya, A. C. (2021). Moodle Web-Based Learning Constraints toward Student Learning Interest Using C4.5 Algorithm during Covid-19 Pandemic. *JITE (Journal of Informatics and Telecommunication Engineering)*, 5(1), 132–141. <https://doi.org/10.31289/jite.v5i1.5301>
- Hadianto, N., Novitasari, H. B., & Rahmawati, A. (2019). Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network. *Jurnal Pilar Nusa Mandiri*, 15(2), 163–170. <https://doi.org/10.33480/pilar.v15i2.658>
- Hasibuan, M. S., Nugroho, L. E., & Santosa, P. I. (2019). Model detecting learning styles with artificial neural network. *Journal of Technology and Science Education*, 9(1), 85–95. <https://doi.org/10.3926/jotse.540>
- Kolekar, S. V., Pai, R. M., & Manohara Pai, M. M. (2017). Prediction of Learner's Profile Based on Learning Styles in Adaptive E-learning System. *International Journal of Emerging Technologies in Learning*, 12(6), 31–51. <https://doi.org/10.3991/ijet.v12i06.6579>
- Luvia, Y. S., Hartama, D., Windarto, A. P., & Solikhun. (2016). Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa. *JURASIK (Jurnal Riset Sistem Informasi & Teknik Informatika)*, 1(1), 75–79. <https://doi.org/jurasik.v1i1.12.g10>
- Mesra, P., Kuntarto, E., & Chan, F. (2021). Faktor-Faktor Yang Mempengaruhi Minat Belajar Siswa di Masa Pandemi. *Jurnal Ilmiah Wahana Pendidikan*, 7(3), 177–183. <https://doi.org/10.5281/zenodo.5037881>
- Prasetya, T. A., & Harjanto, C. T. (2020). Pengaruh Mutu Pembelajaran Online Dan Tingkat Kepuasan Mahasiswa Terhadap Hasil Belajar Saat Pandemi. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(2), 188–197. <https://doi.org/10.23887/jptk-undiksha.v17i2.25286>
- Pratama, R. E., & Mulyati, S. (2020). Pembelajaran Daring dan Luring pada Masa Pandemi Covid-19. *Gagasan Pendidikan Indonesia*, 1(2), 49. <https://doi.org/10.30870/gpi.v1i2.9405>
- Puspitorini, F. (2020). Strategi Pembelajaran Di Perguruan Tinggi Pada Masa Pandemi Covid-19. *Jurnal Kajian Ilmiah*, 1(1), 99–106. <https://doi.org/10.31599/jki.v1i1.274>
- Putra, J. L., & Raharjo, M. (2019). Penerapan Neural Network Dalam Menentukan Tingkat Keberhasilan Immunotherapy. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 4(2), 132–136. <https://doi.org/10.31294/ijcit.v4i2.6242>
- Rahayu, S., Nugraha, F. S., & Shidiq, M. J. (2019). Analisis Tingkat Keberhasilan Cryoterapy Menggunakan Neural Network. *Jurnal Pilar Nusa Mandiri*, 15(2), 141–148. <https://doi.org/10.33480/pilar.v15i2.599>
- Rahmalia, D., & Herlambang, T. (2017). Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSO-NN). *Seminar Nasional Matematika Dan Aplikasinya (SNMA)*, 41–48.
- Sismadi, & Kusnadi, Y. (2018). Prediksi Tingkat Kelulusan Siswa Elearning Berbasis Algoritma Fuzzy C-Means. *Jurnal TECHNO Nusa Mandiri*, 15(1), 1–6.
- Supriyatna, A., & Mustika, W. P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kulit. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 152. <https://doi.org/10.30645/j-sakti.v2i2.78>
- Yulianto, L. D., Triayudi, A., & Sholihati, I. D. (2020). Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4.5. *Jurnal Mantik*, 4(1), 441–451.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.