MODELING DATA MINING ALGORITHM FOR PREDICTING TIMELY STUDENT GRADUATION AT STATE UNIVERSITY SURABAYA

I Kadek Dwi Nuryana*
Information System Faculty of Engineering, State University of Surabaya, Indonesia
e-mail: dwinuryana@unesa.ac.id
*Correspondence: dwinuryana@unesa.ac.id

Abstract. Information technology aims to advance human activities. Even now it has penetrated into the field of education, as a process and industrial sector, education cannot be separated from the field of information and technology. In the implementation of the academic process, the State University of Surabaya, which has 30,000 students, faces several obstacles in predicting the students' graduation time on time. The information system owned has not been used optimally to predict the time of student graduation. It is difficult to predict because this university does not have an exact time prediction pattern to use as a basis for predicting the number of students who will graduate on time. In helping these problems, the researcher provides a solution, namely building a data mining algorithm model to predict the exact time of graduation for Surabaya State University students. The data used are Strata-1 (S1) PTI, SI and TI students from 2014-2018. The methodology in this study is FAST (Framework For The Application Of System Thinking) using the Neural Network (NN) algorithm 761 student data with the input value of the artificial neural network method 4, hidden layer 5 and output 2 providing an accuracy of 99.85%. Details of late predictions are 98.06% as many as 101 students and correct predictions are 99.85% as many as 657 students. Shows that the Neural Network can be used as a prediction of the graduation of Strata-1 (S1) PTI, SI, and IT students at the State University of Surabaya.

Keywords: data mining; neural network (NN) algorithm; prediction of student graduation.

INTRODUCTION

The development of information technology shows very rapid progress. These advances have made it easier for humans to connect with one another. This is also a fact, almost all aspects of industry and modern society are constrained by the breadth of information and technology, because facts have proven that they can provide countless values and benefits for the development of human civilization (Peng, 2021).

Information technology aims to advance human activities. Even recently penetrated into the field of education, as a field of process and industry, education cannot be separated from the field of information and technology. Even UNESCO officials and scholars believe that the development of world information technology has the greatest influence on the education sector (Júnior & dos Santos, 2015). According to (Shi et al., 2017), the peak of the application of information technology in education will be revolutionary, influencing the transformation of the teaching process from primary school to higher education or university-level schools (Hubalovsky, Hubalovska, & Musilek, 2019).

According to (Xiang, Magnini, & Fesenmaier, 2015) in the world of education, the use of information technology is needed to support the publication of institutions and the implementation of the Tri Dharma program. In today's world of education, almost all sectors have utilized information technology, including universities (Habib, Jamal, Khalil, & Khan, 2021). Related to the management of the Tri Dharma program, universities, and university accreditation, the role of information technology is needed. Information technology is one of the key factors for universities to advance institutions, publish works, compete with other universities, and facilitate the teaching and learning process, all of which are stages that cannot be avoided by educational institution administrators (Ngari et al., 2013).

In the implementation of the academic process, the State University of Surabaya, which has 30,000 students, faces several obstacles in predicting the students' graduation time on time. The information system owned has not been used optimally to predict the time of student graduation (Handayani et al., 2017). It is difficult to predict because this university does not have an exact time prediction pattern to use as a basis for predicting the number of students who will graduate on time. The academic information system owned by this university is actually website-based, but the information system has not been used optimally to predict student graduation times. Data mining is expected to be useful and valuable information for universities (Aldowah, Al-Samarraie, & Fauzy, 2019).

The research that was conducted by (Agrusti, Bonavolontà, & Mezzini, 2019) with the title "*Comparison of Data Mining Classification Methods for Predicting Student Graduation*". In this study, a comparison of data mining methods, namely the Neural Network, K-Nearest Neighbor and Decision Tree which were applied to student graduation data from the three methods was carried out, it was

found that the Neural Network method, it was found that the neural network method was the best way to solve the problem of predicting student graduation. Compared with the k-nearest neighbor method and the Decision Tree (Adeniyi, Wei, & Yongquan, 2016).

To overcome this problem, the researcher provides a solution by creating a data mining implementation model to predict the students' graduation time on time at the State University of Surabaya. The data mining used in this research is the Algorithm Neural Network for undergraduate PTI, SI and IT students. This model is expected to help Surabaya State University in particular and Indonesian universities globally in solving the problem of predicting the exact time of graduation for students.

METHODS

The research framework

Entitled "Modeling Data Mining Algorithms to Determine Student Academic Achievement and Predicting the Right Time for Student Graduation at State University of Surabaya" originated from the phenomenon of academic organization. The process at the State University of Surabaya faces several obstacles in

predicting the exact time of student graduation. These problems arise because the administrative base applied is still paper-based and not yet computerized. Determination of student academic achievement is still using a manual process (Wang & Holcombe, 2010), namely by selecting student scores from the courses taken and sorting out the student grades for each department and student GPA per semester.

The procedure is carried out to predict the exact time of graduation of students at this university is still difficult. This is because the university has not made an exact pattern of graduation time as a basis for predicting the number of students who graduate on time based academic information system website, but the information system has not been implemented optimally to predict the exact time of student graduation. Data mining method is an algorithm method used by researchers to assist universities in overcoming these problems.

In helping this problem, the researcher provides a solution, namely building a data mining algorithm model to predict the exact time of graduation for students at the State University of Surabaya.

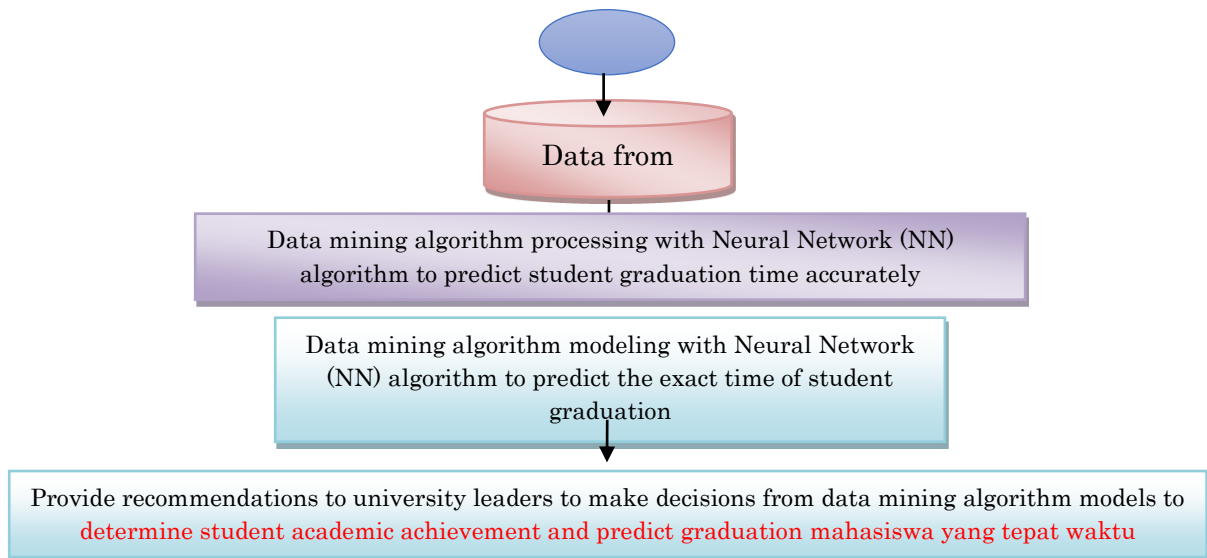The framework for this research is depicted in Figure 2.

Figure 2. Schematic Framework for Thinking

1. System Development Methodology.

The research entitled "*Data Mining Algorithm Modeling in Predicting Timely Graduation of State University Students of Surabaya*" was built using the FAST (Framework For The Application Of System Thinking) methodology (Agarina & Sutedi, 2017). The steps carried out in this study are based on the FAST methodology: 1. Initial definition range, 2. Problem analysis, 3. Needs analysis, 4. Logical design, 5. Decision analysis, 6. System testing.

2. Research Materials Research

materials are the elements needed by researchers when conducting research. The elements needed in this research process are the Neural Network (NN) algorithm described below:

a. Research Object.

The object of the research of this research is Surabaya State University.

b. Research Data

Data used to analyze and predict the exact time of graduation of students at the State University of Surabaya using the Neural Network (NN)lectures, student grade data, semester GPA data (per semester), GPA data (final), and student graduation data.

c. Data Collection Method The data

Collection method in conducting the analysis was carried out on the prediction of the exact time of graduation of students at the State University of Surabaya using the Neural Network (NN) algorithm. This will be explained below:

1) Observations

Observations were carried out by researchers to find out academic information system data to be analyzed for research

purposes. at the State University of Surabaya.

2) **Library Research**

Is a data collection method that is carried out by collecting data from various sources that support research with the Neural Network (NN) algorithm by searching for information through the internet or scientific journals, international

journals, papers, books, proceedings, or other articles that support research. The results of library research are in the form of models and the latest developments on concepts, data mining, data mining classification, Neural Network (NN) algorithms, association rules, and other supporting theories.

RESULTS AND DISCUSSION

In this study the research method used is the Framework for the Applications of System Technology (FAST) method with the following stages:

1. **Initial definition range,**

The data used comes from the Directorate of Information Technology, State University of Surabaya in the form of undergraduate student data (S1) PTI, TI, SI batch 2014-2018. The data obtained is the output of several existing systems, so that several stages of processing must be carried out for the accuracy of graduation to be determined based on the length of the

study. The grades range from 3.5 to 7 years. Based on the available programs, it is determined that on-time graduation occurs when a student has successfully completed his studies in 3.5 and 4 years. If it exceeds then it is considered not on time.

2. **Problem analysis,**

According to (Hla & Teru, 2015) the definition of scope, problem analysis through PIECES (Performance, Information, Economics, Control, Efficiency, and Service) can be determined in the following table:

Table 1. Analysis of System Problems using PIECES.

Parameter	Problem	Solution
Performance	Website-based academic information system, but this information system has not been implemented optimally to predict the exact graduation time.	Optimization of information systems to predict the exact time of graduation for students.
Information	Difficulty in predicting student graduation times accurately is due to the absence of an appropriate graduation	Processing data mining algorithms with Neural Network (NN) to accurately predict student graduation times

Parameter	Problem	Solution
	time pattern in predicting the number of students who will graduate on time.	
Economy	paper.	of information costscosts because the system can be accessed by related units so that information can be viewed digitally.
Control	In the current system, control of information and data is very lacking so that there are often differences in data and information.	With the information system , the data and information conveyed is more controlled and accurate.
Efficiency	With the current system, more resources are required due to manual data collection and direct inquiries by individuals.	Through the information system the required resources are less because the data and information are obtained with an integrated system.
Service	Data and information needed by units can be obtained slowly because they are waiting to get data and information.	The data and information needed by the unit can be obtained more quickly and there is no need to wait for data and information.

3. **Needs analysis,**

In the needs analysis stage, the information system to be built is determined based on the definition of the scope and problem analysis. Based on the scope of analysis of the questions above, a system of requirements is needed in the form of student data from academic Siakad. The functional requirements of the system are as follows (Kobayashi, Morisaki, Atsumi, & Yamamoto, 2016):

a. Data taken from the academic information system (SIAKAD) includes:
   1. Student grade
   2. Data, semester GPA data (per semester),
   3. GPA data (final).
b. Student graduation data.

The data used in this study is student data for semester GPA (per semester) and GPA data (end), Odd Semester Period for the 2018-2019 academic year and Even Semester 2019-2020. The data used is data from PTI, SI, and IT students from the 2014-2018 class in the Department of Informatics Engineering (JTIF). The number of data obtained is 762 data.

4. **Logical**

a. Design, Process Design of Neural Network Algorithms.

The design of the neural network process consisting of

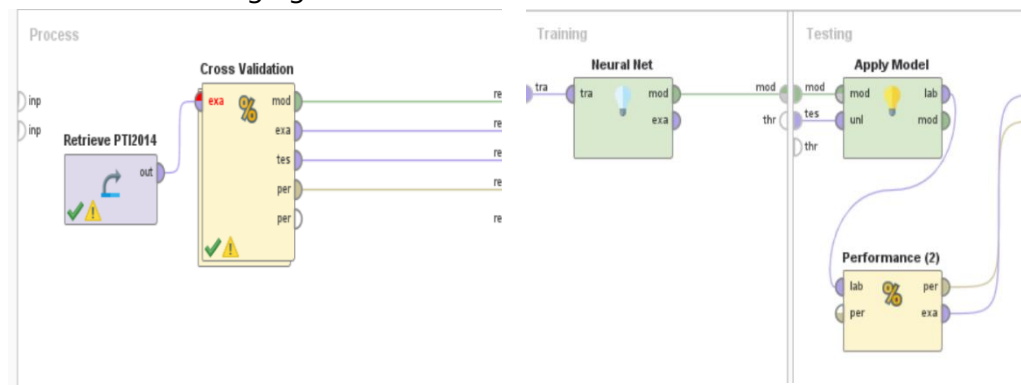processes, training, and testing can be seen in the following figure:

Figure 3. Design of Neural Network Algorithms

5. Decision analysis

a. Decision Analysis in predicting student graduation.

Algorithm Back propagation neural network, the parameters displayed are training cycle (epoch), learning rate, momentum and hidden layer. Learning Rate is 0.1. The experimental results can be seen in Table 2 below.

Table 2. Experiment Value of Training Cycle (Epoch)

Training Cycle (Epoch)	Learning Rate	Momentum	Accuracy
200	0.1	0.1	%
400	0.1	0.1	99.48
600	0.1	0.1	81.73%
800	0.1	0.1	64.7 %

Based on the experimental results above, the value of the Training Cycle (Epoch) was chosen, which was 400, and the accuracy was 84.59%. This value is then used in experiments to determine the Learning Rate. Get the Learning Rate value by entering a value between 0.2 and 0.9. The experimental results can be seen in Table 3 below.

Table 3. Experiment Determination of Learning Rate

Training Cycle (Epoch)	Learning Rate	Momentum	Accuracy
400	0.2	0.1	85.12%
400	0.5	0.1	91.89%
400	0.7	0.1	89.91%
400	0.9	0.1	94, 84%

The two experiments above use a parameter value of 400 in the Training Cycle (epoch), and learning rate 0.9. The maximum accuracy of input node number 5 is 94.84%. This study uses 5 input nodes. The experimental results are as follows table 4 shows :

Table 4. Experiment Number of Input Nodes

Number of Input Nodes	Training Cycles (epoch)	Learning Rate	Momentum	accuracy (%)
2	400	0.9	0.1	94
3	400	0.9	0.1	92.6
4	400	0.9	0.1	93.35
5	400	0.9	0.1	94.84

6. System Testing,

The structure of the Neural Network (NN) network consists of 3 layers, namely the input layer, the hidden layer, and the output layer. In the input layer there are 4 nodes as input, namely (GPA 1, GPA 2, GPA 3 and GPA 4), the hidden layer is composed of 5 nodes, and the output layer is 2 nodes. If the final GPA is above 2.75 it means on time, if it is below 2.75 it means the opposite). Its structure is shown in Figure 4.
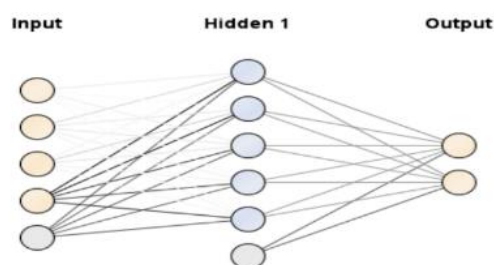

Figure 4. Structure of the Neural Network (NN)

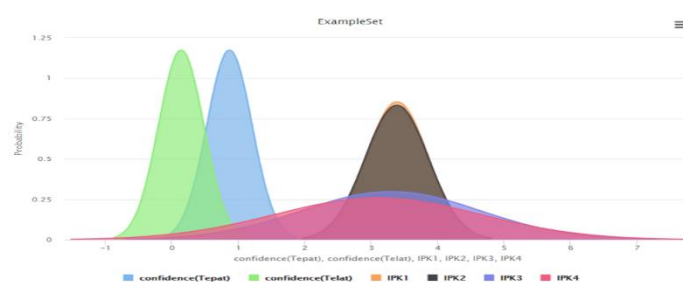Result set cross validation is shown in the graph as shown in Figure 5 below:

Figure 5. Cross Validation result set.

System testing using the Neural Network , the AUC value illustrated in the ROC graph can be seen in the following figure:
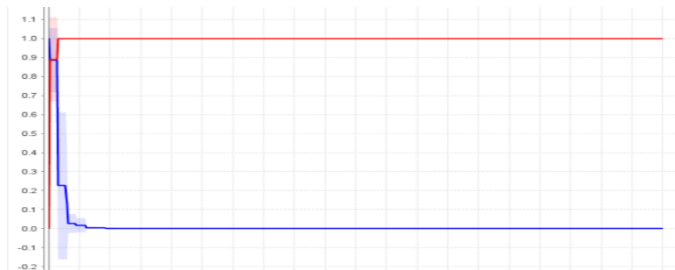

Figure 6. The AUC graph ROC

Determine the level of prediction accuracy from the error matrix obtained by processing 761 data using 75% training data, namely 571 and 25% testing data 190 to test the neural network algorithm method, as shown in Table 5 below:

Table 5. Neural Network Algorithm Method

	true Tepat	true Telat	class precision
pred. Tepat	657	1	99.85%
pred. Telat	2	101	98.06%
class recall	99.70%	99.02%	

Conducted using Neural Network showed final accuracy of 99.61%, late detail prediction 98.06% as many as 101 students, and accurate prediction accuracy 99.85% as many as 657 students.

7. Implementation

Implementation Prediction of student graduation for each batch from the predictions of the 2014 – 2018 batch as follows:

a. Graduation Prediction for the 2014 batch.

Determining the level of accuracy of the 2014 graduation prediction from the error matrix obtained in 81 data with the neural network algorithm, as shown in Table 6 below:

Table 6. Accuracy Value of Graduation Prediction Class of 2014

	true Tepat	true Telat	class precision
pred. Tepat	70	1	98.59%
pred. Telat	1	9	90.00%
class recall	98.59%	90.00%	

Tests conducted using the Neural Network showed final accuracy of 97.50%, late detail prediction of 90% of 9 students, and accurate prediction of 98% accuracy of 70 students. The AUC value illustrated in the ROC graph can be seen in the following figure:
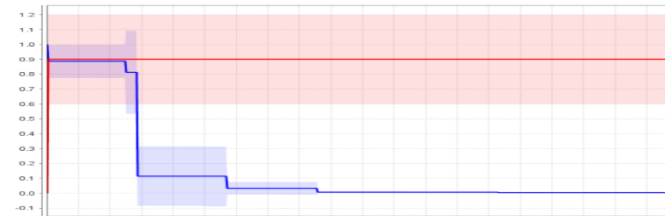

Figure 7. The AUC Graph ROC predicts graduation class 2014

b. Prediction Graduation batch 2015.

Determines the level of prediction accuracy for class 2015 graduation from the error matrix obtained in 165 data with neural network algorithms, such as It can be seen in Table 7 below:

Table 7. Accuracy Value of Graduation Prediction Class of 2015

	true Tepat	true Telat	class precision
pred. Tepat	89	7	92.71%
pred. Telat	20	49	71.01%
class recall	81.65%	87.50%	

Tests conducted using the Neural Network showed final accuracy of 83.79%, late detail prediction 71.01% as many as 49 students, and accurate prediction accuracy 92.71% as many as 89 students. The AUC value illustrated in the ROC graph can be seen in the following figure:
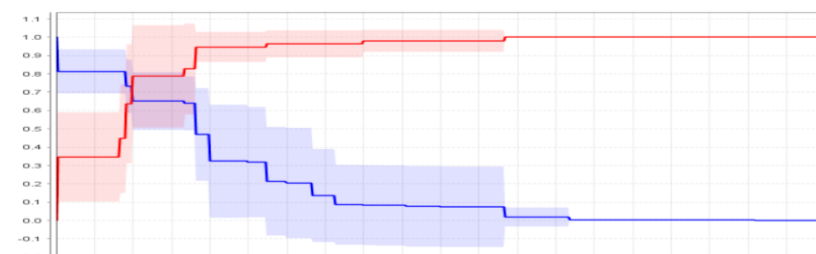

Figure 8. The AUC Graph ROC for 2015 graduation

c. Predictions 2016 Graduation Predictions

Determines the accuracy level of 2016 graduation predictions from the error matrix obtained in 128

data with neural network algorithms, as shown in Table 8 below:

Table 8. Accuracy Value of Graduation Prediction for Class of 2016

	true Tepat	true Telat	class precision
pred. Tepat	126	0	100.00%
pred. Telat	0	2	100.00%
class recall	100.00%	100.00%	

Tests conducted using the Neural Network showed final accuracy of 100%, prediction of 100% late details for 2 students, and ccurate prediction accuracy of 100% with 126 students. The AUC value illustrated in the ROC graph can be seen in the following figure:
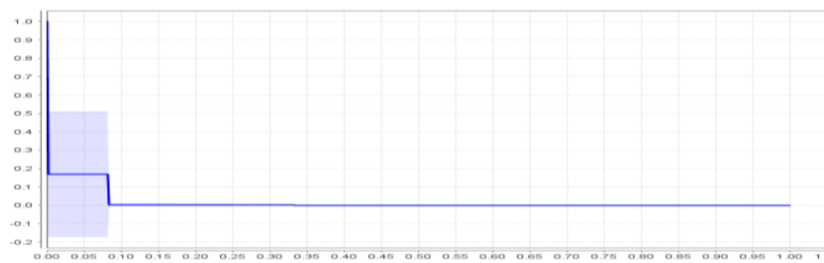

Figure 9. The AUC Graph ROC predicts graduation class 2016.

d. Prediction Graduation class 2017

Determines the level of prediction accuracy for class 2017 graduation from the error matrix obtained in processing 128 data with neural network algorithms, as shown in Table 9 below:

Table 9. Accuracy Value of Graduation Prediction for Class of 2017

	true Tepat	true Telat	class precision
pred. Tepat	187	2	98.94%
pred. Telat	0	0	0.00%
class recall	100.00%	0.00%	

Tests conducted using the Neural Network showed final accuracy of 98.95%, late detail prediction 0%, and accurate prediction accuracy of 98.94% as many as 187 students. The AUC value illustrated in the ROC graph can be seen in the following figure:
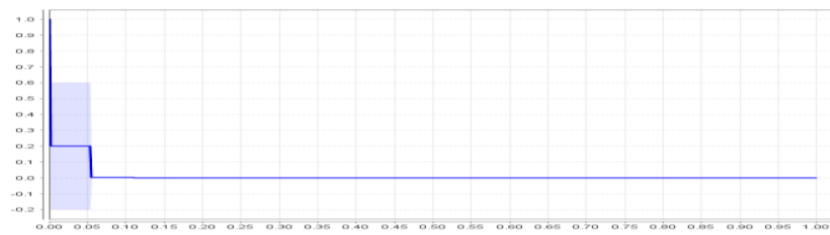
Figure 10. The AUC Graph ROC predicts graduation class 2017

e. Prediction Graduation class 2018

Determines the level of prediction accuracy for class 2018 graduation from the error matrix obtained in processing 198 data with a neural network algorithm, as shown in Table 10 below:

Table 10. Accuracy Value of Graduation Prediction for class of 2018

	true Tepat	true Telat	class precision
pred. Tepat	184	0	100.00%
pred. Telat	4	10	71.43%
class recall	97.87%	100.00%	

Tests conducted using the Neural Network showed final accuracy of 98%, late detail prediction of 71.43% for 10 students and accurate prediction accuracy of 100% for 184 students.AUC illustrated in the ROC graph can be seen in the following figure:
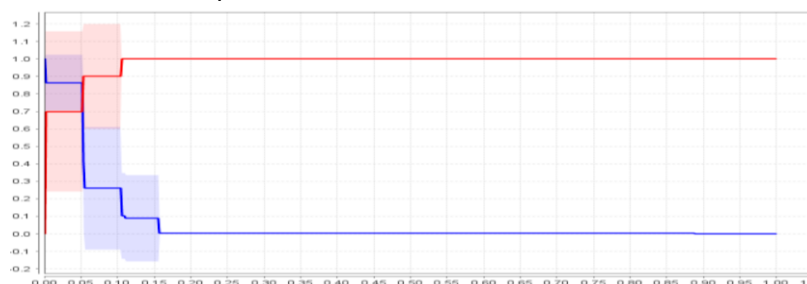

Figure 11. The AUC Graph ROC predicts graduation for class 2018.

CONCLUSIONS

In predicting the graduation of undergraduate students (S1) PTI, SI, and IT, State University of Surabaya from the 2014 to 2018 batch, an algorithm is used neural networks. Algorithm Neural Network \ used is a Neural Network, the input value for the artificial neural network method is 4, hidden layer 5 and output is 2, the accuracy is 99.85%. Detailed information on predictions of 98.06% late for 101 students, and accurate predictions for 99.85% for 761

students. Predictions for graduation for each batch are as follows:

1. Prediction of PTI Graduation 2014 batch of final accuracy 97.50%, Prediction of late details 90% as many as 9 students, and 98% correct prediction as many as 70 students.

2. Prediction of Graduation PTI, SI, TI 2015 final accuracy 83.79%, late detail prediction 71.01% as many as 49 students, and accurate predictions 92.71% as many as 89 students.

3. Prediction of Graduation PTI, SI, TI 2016 final accuracy 100%, 100% late detail predictions as many as 2 students, and 100% accurate predictions as many as 126 students.

4. Prediction of Graduation PTI, SI, TI class 2017 final accuracy 98.95%, late detail prediction 0%, and predictions right 98.94% as many as 187 students.

5. Predictions for Graduation PTI, SI, TI 2018 final accuracy 98%, late detail prediction 71.43% as many as 10 students, and exactly 100% as many as 184 students.

6. Based on these results, it can be concluded that the Neural Network can be used to predict the graduation of undergraduate (S1) PTI, TI, and SI at the State University of Surabaya from 2014 to 2018.

REFERENCES

Adeniyi, David Adedayo, Wei, Zhaoqiang, & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, *2*(12), 90–108. https://doi.org/10.1016/j.aci.2014.10.001

Agarina, Melda, & Sutedi, Sutedi. (2017). Information System Design Of Goods Stock Using Framework For The Application Of System Thinking (FAST) Method (Case Study CV. Aneka Mandiri Lestari). *International Conference on Information Technology and Business (ICITB)*, 125–128.

Agrusti, Francesco, Bonavolontà, Gianmarco, & Mezzini, Mauro. (2019). University dropout prediction through educational data mining techniques: A systematic review. *Journal of E-Learning and Knowledge Society*, *15*(3), 161–182. https://doi.org/10.20368/1971-8829/1135017

Aldowah, Hanan, Al-Samarraie, Hosam, & Fauzy, Wan Mohamad. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis.

Telematics and Informatics, *3*(7), 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Habib, Muhammad Nauman, Jamal, Waseef, Khalil, Uzma, & Khan, Zunnoorain. (2021). Transforming universities in interactive digital platform: case of city university of science and information technology. *Education and Information Technologies*, *6*(1), 517–541.

Handayani, Putu Wuri, Hidayanto, Achmad Nizar, Pinem, Ave Adriana, Hapsari, Ika Chandra, Sandhyaduhita, Puspa Indahati, & Budi, Indra. (2017). Acceptance model of a hospital information system. *International Journal of Medical Informatics*, 9(9), 11–28. https://doi.org/10.1016/j.ijmedinf.2016.12.004

Hla, Daw, & Teru, Susan Peter. (2015). Efficiency of accounting information system and performance measures. *International Journal of Multidisciplinary and Current Research*, 3(2), 976–984.

Hubalovsky, S., Hubalovska, Marie, & Musilek, Michal. (2019). Assessment of the influence of adaptive E-learning on learning effectiveness of primary school pupils. *Computers in Human Behavior*, 9(2), 691–705. https://doi.org/10.1016/j.chb.2018.05.033

Júnior, José Policarpo, & dos Santos, Claudemir Inacio. (2015). " Learning to live together in peace and harmony": um olhar comprometido com a formação humana. *Conjectura: Filosofia e Educação*, 20(2), 94–111.

Kobayashi, Nobuhide, Morisaki, Shuji, Atsumi, Noritoshi, & Yamamoto, Shuichiro. (2016). Quantitative Non Functional Requirements evaluation using softgoal weight. *J. Internet Serv. Inf. Secur.*, 6(1), 37–46.

Ngari, S. M., Ndungu, A., Mwonya, R., Ngumi, O., Mumiukha, C., Chepchieng, M., & Kariuki, M. (2013). Levels of stress among secondary school administrators and its implication in education management in Kenya. *Educational Research and Reviews*, 8(11), 677–681. https://doi.org/10.5897/ERR10.021

Peng, Bingzhuan. (2021). Intercultural Communicative Competence Teaching and Assessment Based on Modern Information Technology. *International Journal of Emerging Technologies in Learning*, 16(7).

Shi, Victor Guang, Baines, Tim, Baldwin, James, Ridgway, Keith, Petridis, Panagiotis, Bigdeli, Ali Ziaee, Uren, Victoria, & Andrews, Daniel. (2017). Using gamification to transform the adoption of servitization. *Industrial Marketing Management*, 6(3), 82–91. https://doi.org/10.1016/j.indmarman.2016.12.005

Wang, Ming Te, & Holcombe, Rebecca. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, 4(3), 633–662. https://doi.org/10.3102/0002831209361209

Xiang, Zheng, Magnini, Vincent P., & Fesenmaier, Daniel R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, 2(2), 244–249. https://doi.org/10.1016/j.jretconser.2014.08.005