ICIC Express Letters
Part B: Applications
Volume **12**, Number **10**, October **2021**

 ISSN 2185-2766
pp. **883–890**

PREDICTING STUDENT PERFORMANCE WITH MULTI-LEVEL REPRESENTATION IN AN INTELLIGENT ACADEMIC RECOMMENDER SYSTEM USING BACKPROPAGATION NEURAL NETWORK

Dede Kurniadi[1,2], Edi Abdurachman[1]
Harco Leslie Hendric Spits Warnars[1] and Wayan Suparta[1]

[1]Doctor of Computer Science, BINUS Graduate Program
Bina Nusantara University
Jl. Kebon Jeruk Raya No. 27, Kebon Jeruk, Jakarta 11480, Indonesia
{ dede.kurniadi; edia; shendric; wayan.suparta }@binus.ac.id

[2]Department of Informatics
Sekolah Tinggi Teknologi Garut
Jl. Mayor Syamsu No. 1, Garut, West Java 44151, Indonesia
dede.kurniadi@sttgarut.ac.id

Received January 2021; accepted April 2021

Abstract. *This article aimed to develop a classification model for predicting student academic performance by representing outputs into various multi-level representations with better accuracy using supervised learning algorithms. The prediction model developed is an integral part of the intelligent academic recommendation system. The method used is a backpropagation neural network with a multi-layer perceptron architecture. The activation function uses the Tanh and SoftMax functions in each hidden layer and output layer, respectively. We train models on real-world student academic datasets using the K-fold cross-validation and the confusion matrix techniques to measure accuracy, precision, and recall results. The tested result shows that the effectiveness of the models and methods used produces the best accuracy of 80.95%, 83.20% precision, and 83.53% recall with a cross-entropy loss rate of 0.12. The Area Under Curve (AUC) Receiver Operating Characteristics (ROC) curve produces a value of 0.89, which is classified as a good result. Based on the test result, the model can be applied to the student academic early warning mechanism in the intelligent academic recommendation system to support the education process in higher education.*
Keywords: Classification, Student performance prediction, Multi-level representations, Backpropagation neural network, Intelligent recommender system

1. **Introduction.** Intelligent Recommender System (IRS) is intelligent application software that the data analysis process applies data mining methods and machine learning technology in making predictions and recommendations automatically [1]. Data mining is still an emerging trend and is used in various fields without exception in educational analytics, and student learning, often referred to as EDM (Educational Data Mining). EDM is a discipline related to developing, researching, and applying computational methods using statistics, data mining, and machine learning in studying data sourced from the world of education [2].

In Indonesia, based on statistical data from the Ministry of Research and Higher Education recorded in 2018 the number of students who failed to complete the undergraduate level as many as 245,810 people, one of the causes is students academic achievement [3]. Measuring student academic performance becomes a challenge because student academic performance depends on factors or characteristics as diverse as demographics, personal,

DOI: 10.24507/icicelb.12.10.883

educational background, psychological, academic record progression, and other environmental variables [4]. Academic performance can be drawn from the achievements of learning outcomes in each subject taken. In comparison, the representation of academic performance is influenced by several factors, including graduation status, length of study, and predicate based on the Grade-Point Average (GPA) [5].

In the education domain, research work predicting student studies success rate is needed to recommend improving student academic performance in the future. Some work has been done by researchers related to academic performance, for example, Abu Zohair's research created a performance prediction model with a dataset of 50 graduate student with variable predictions of previous academic records, methods used and research accuracy results are ANN-MLP 60.5%, Naïve Bayes 71.7%, support vector machine 76.3%, k-nearest neighbors 65.8%, and learning discriminant analysis 71.1% [6]. Mimis et al. research on intelligent academic guidance system with a focus on predicting the performance of student pass/not pass with data analysis based on academic records of the first year, socio-economic data and motivation of student with a dataset of 330 student, methods used, and accuracy results are decision tree (C45) 54.71%, Naïve Bayes 58.69% and neural network 56.10% [7]. Xu et al.'s research predicts the academic performance of the undergraduate student in an online class, in the form of predictions of the final grades of some subjects in the first year of the method, using linear regression, logistic regression, random forest, and k-Nearest Neighbors (k-NN) [8]. Guo et al.'s research predicts the academic performance of the student with various levels of representation (multi-class) by grouping output into 5 levels (outstanding, good, average, pass, fail), methods using deep learning with 3 hidden layers and using a dataset of 120,000 junior high school students with various input background, study data, academic records, and personal data with an accuracy rate of 77.2% [4]. At the same time, Al-Shehri et al.'s research presented two prediction models of support vector machine and k-nearest neighbors for estimating student performance in the possibility of obtaining final exam scores of math subjects, a dataset of 395 students with variables based on background, family, and previous history academic with the results of accuracy of both models 80% [9].

Based on the background and previous research studies, the identified research gaps, some of which were predicting student performance, mostly focused on prediction in the first year with the output representation only depicting student performance in a limited to pass and not pass. Then the results of the classification accuracy of the method used still need to be improved.

Therefore, this study aims to propose a different approach in developing a classification model to predict student academic performance by representing output into various multi-level representations based on graduation factors, length of study, and predicate from the Grade-Point Average (GPA). The data analysis process will be based on the profile and historical data of the first-year and second-year undergraduate students academic history. The method used is backpropagation neural network with multi-layer perceptron architecture because it can work effectively in several studies for multi-class data characteristics [4,6,10] with the improvisation of modeling optimization to produce better accuracy. The model produced in this research can be applied to the academic prewarning students mechanism in the Intelligent Academic Recommender System (IARS) so that the recommendations needed to improve student academic performance in the future can be made.

The structure of the article organized is as follows. The first section is an introduction that explains the background of the problem and previous related research studies. The second section, the research method, briefly describes the research design, process flow, and methods used. The third section presents the results and discussion that explain the results, which are the study's objectives. Part 4 is the final section which is a brief description of the conclusions and future work.

2. **Research Methods.** Model development is divided into four main stages: acquisition of data from the data source, pre-processing, model construction and training, and the evaluation process. Here is the research design of the academic performance prediction process and the method used in Figure 1.

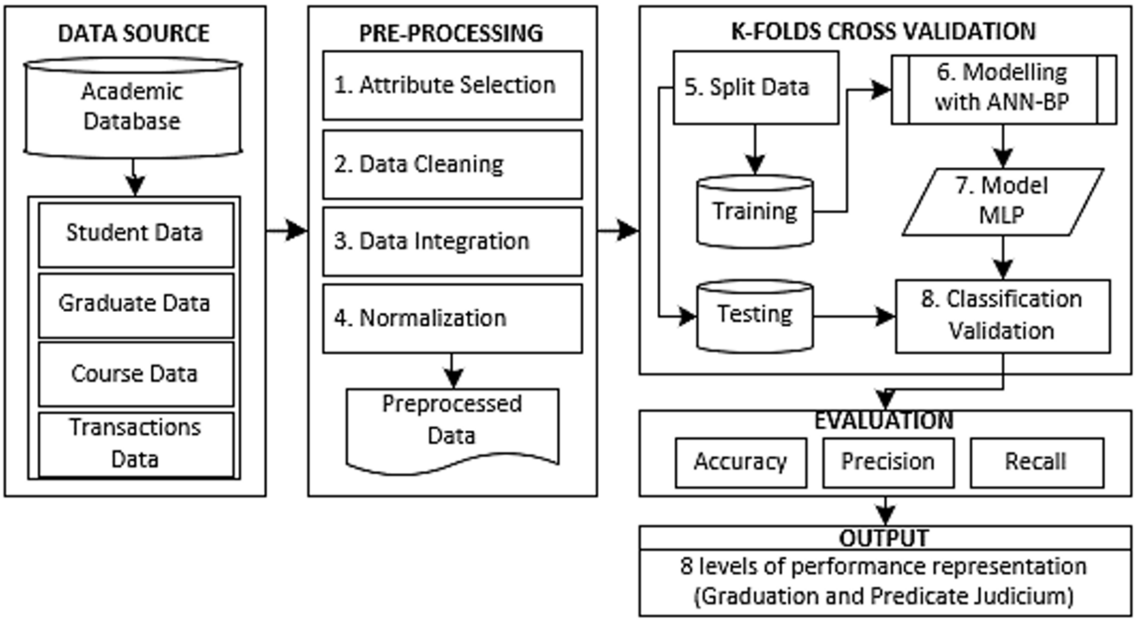


Figure 1. Research design for student performance prediction

The explanation is based on Figure 1, and the first stage begins with data acquisition from the academic database to prepare data that will be used in the modeling and data analysis process. The second stage of pre-processing consists of the process of selecting the input and output attributes/variables to be used; data cleaning, namely the process of detecting and correcting/deleting incomplete or empty data from a set of records, tables in the database; data integration, namely the process of combining relevant attributes to the needs of data modeling and analysis; and the data normalization process to equalize the minimum and maximum weights on the data that will be used at a later stage. The third stage is model construction and training using the backpropagation neural network algorithm with a multi-layer perceptron architecture. Furthermore, to validate the model's feasibility using the K-fold cross validation technique to validate the accuracy of a model built on a specific data set by dividing the data into two segments, namely training data and testing data. Training data is used in the model building process, while testing data is used to validate the model. In cross-validation, the training and validation sets must be crossover consecutively so that each data has a chance to be validated. The fourth stage is the evaluation process to measure performance based on true or false objects using the confusion matrix performance measurement parameters, namely Accuracy, Precision, and Recall. While the representation of the output data model predicts student performance into 8 level groups representing student learning achievement based on graduation, length of study, and judicium predicate (multi-level representations), which will be detailedly explained in the result and discussion section.

3. Result and Discussion.

3.1. **Data preparation.** The total dataset collected from acquisitions in undergraduate student academic database amounted to 626 people from 3 engineering majors with 41,492 records. Identification of input and output variables is carried out through the research literature study process, academic handbook, dataset characteristics, and interview with the study program manager based on the problems to be researched. The model training process variable inputs are gender, majors, age, school origin, Grade-Point Average (GPA) for each semester (semester 1 to 4), and Semester Credit Units (SCU) for each semester (semester 1 to 4).

The classing of the output level of undergraduate student performance is influenced by several factors, including graduation status, length of study, and judicium predicate based on Grade-Point Average (GPA) acquisition on a scale of 0-4 as in Table 1.

Table 1. Multi-level representation of student performance

Performance level	**Graduation status**	**Length study (Semester)**	**Judicium predicate (GPA)**
Level 1	On-Time Graduation, with Praise	<= 8 semesters	Praise ($3.51 \leq$ GPA ≤ 4)
Level 2	On-Time Graduation, Very Satisfying	<= 8 semesters	Very Satisfying ($3.01 \leq$ GPA ≤ 3.50)
Level 3	On-Time Graduation, Satisfactory	<= 8 semesters	Satisfactory ($2.76 \leq$ GPA ≤ 3.00)
Level 4	On-Time Graduation, Not Predicate	<= 8 semesters	Not Predicate ($2.00 \leq$ GPA ≤ 2.75)
Level 5	Not On-Time Graduation, Very Satisfying	> 8 to 14 semesters	Very Satisfying ($3.01 \leq$ GPA ≤ 3.50)
Level 6	Not On-Time Graduation, Satisfactory	> 8 to 14 semesters	Satisfactory ($2.76 \leq$ GPA ≤ 3.00)
Level 7	Not On-Time Graduation, Not Predicate	> 8 to 14 semesters	Not Predicate ($2.00 \leq$ GPA ≤ 2.75)
Level 8	Failed/*Dropout*	> 14 semesters	Failed/*Dropout*

The determination of variable output in Table 1 is based on a multi-level class of student academic performance representing student performance and success in higher education.

3.2. **Model construction.** Before conducting model training using a neural network method with the backpropagation algorithm, the first thing to do is to determine the Multi-Layer Perceptron network (MLP). The determination of the MLP network architecture used in the training process aims to obtain a better accuracy degree. The proposed MLP network architecture as in Figure 2 with layer input consists of 12 input neurons, 3 hidden layers, and an output layer consisting of 8 outputs, while the number of neurons/nodes in the hidden layer will be determined based on the results of experiments to obtain models with the best accuracy.

The activation function in the hidden layer using *Tanh* function Equation (1) because it is suitable for multi-layer architecture with data normalization has a range of interval $(-1, 1)$ [11].

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{1}$$

where x is the input value on the layer, the output layer uses the equation of *SoftMax* function (2) because the function activation is very suitable for classification whose output representation is multi-label/multi-class [4].

$$f_i(x) = \frac{e^x i}{\sum_{j=1}^{k} e^x j} \text{ for } i = 1, 2, 3, \dots, k \tag{2}$$

where x is the input value on the output layer. To calculate the value of a loss, the activation function used Equation (3) cross-entropy [12]:

$$J_{bce} = \frac{-1}{M} \sum_{m=1}^{M} [y_m x \log(h_\theta(x_m))] + (1 - y_m) x \log(1 - h_\theta(x_m)) \tag{3}$$

where M is the number of examples of training, y_m is the target label for the training example, x_m is input from the examples of training and h_θ is a model with neural network weights θ. The next stage is to conduct a model training process using backpropagation neural network, which consists of 3 phases, namely forward, backward, and update using a predetermined formula [13]. For training, testing and evaluation, we use the Python Programming Language as a tool for analyzing experimental results. Pseudocode for model training is shown on page 6.

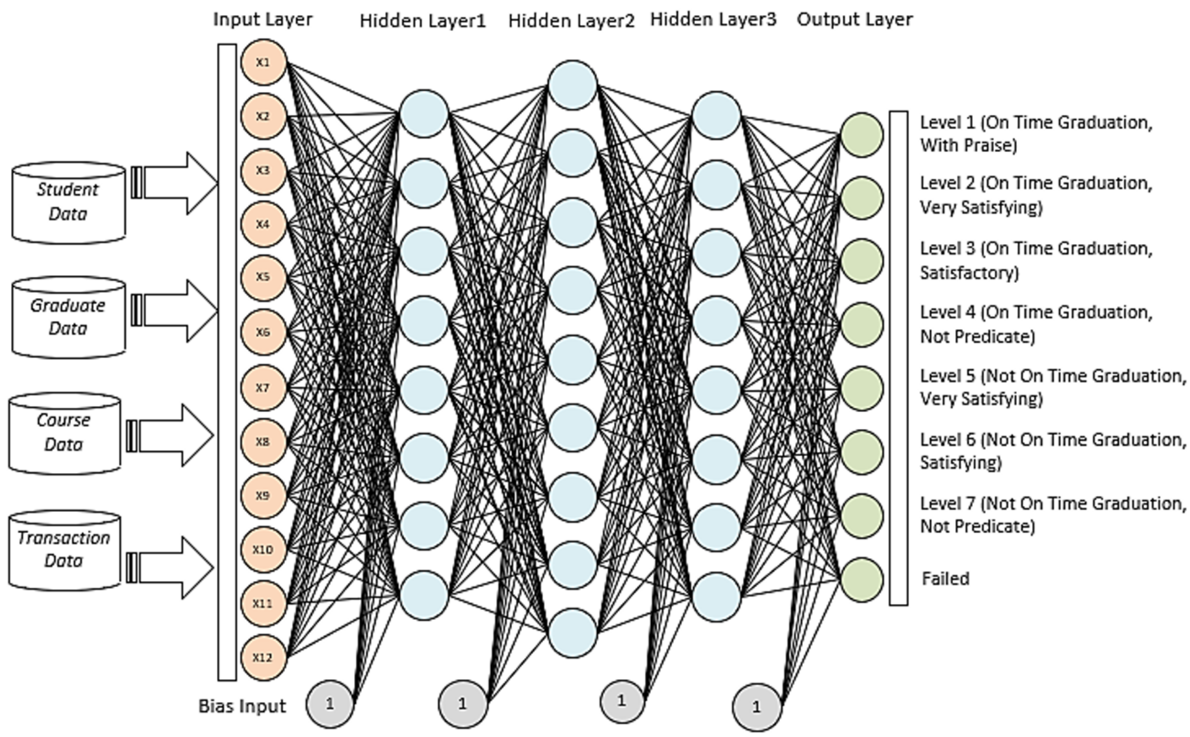


Figure 2. The proposed MLP network model architecture

3.3. **Testing and evaluation.** The training experiment was conducted with 4 MLP network architecture modeling with hidden layers, with each model conducting five times the experiment so that there are a total of 20 experiments. Each experiment was conducted 1000 Epoch with the number of neurons/nodes in the hidden layer, and the learning rate values vary to get the model with the best accuracy. For accuracy validation test conduct with 10-fold cross-validation trial technique by comparing prediction results and actual data. Figure 3 shows a visualization of the experimental results with the best accuracy.

The best accuracy obtained was 99.82% for data training and 80.95% for data testing with optimum value in the 951st epoch. Figure 3(a) shows that the accuracy between training and testing is also better and not far compared to other experiments. In Figure 3(b) it can be seen Area Under Curve (AUC) in the ROC curve micro-average value of ROC 0.89, and then this model is categorized as Good Classification [14].

For more details, here is the overall resume of test results from 20 experiments that have been done, as shown in Table 2.

Based on Table 2, the best test results obtained on MLP models with a pattern of 12-18.18.16-8 (12 nodes on input layer, 18 nodes on hidden layer 1, 18 nodes on hidden layer 2, 16 nodes on hidden layer 3, and 8 nodes on the output layer) with learning rate 0.003. The best accuracy obtained is 80.95%, precision is 83.20%, and recall is 83.53%.

Initialize the number of inputs, hidden layer, and output neurons.
Initialize all learning rates, bias, weights with random values.
Initialization epoch $\leftarrow n$
While (epoch $\leq n$)
For each layer in the network
  For each neuron layer (Z)
    1. Calculate all hidden layer output *(Forward)*

$$f(x) = W_{j0} + \sum_{i=1}^{n} x_i W_{ij}$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

    2. Calculate all networks at the output layer

$$f(x) = W_{j0} + \sum_{i=1}^{n} x_i W_{ij}$$

$$f_i(x) = \frac{e^x i}{\sum_{j=1}^{k} e^x j} \text{ for } i = 1, 2, 3, \ldots, k$$

    3. Calculate all factor *(Backward)*

$$\delta = (t - y_{out}) y_{out} (1 - y_{out})$$

$$\Delta w_{kj} = \alpha \delta z_{out}$$

    4. Calculate factor in hidden layer

$$\delta_{net} = \sum_{k=1}^{m} \delta_k w_{kj}$$

$$\delta_j = \delta_{net} = f'(z_{net}) = \delta_{net} z(1 - z)$$

$$\Delta w_{ji} = \alpha \delta_j x_i$$

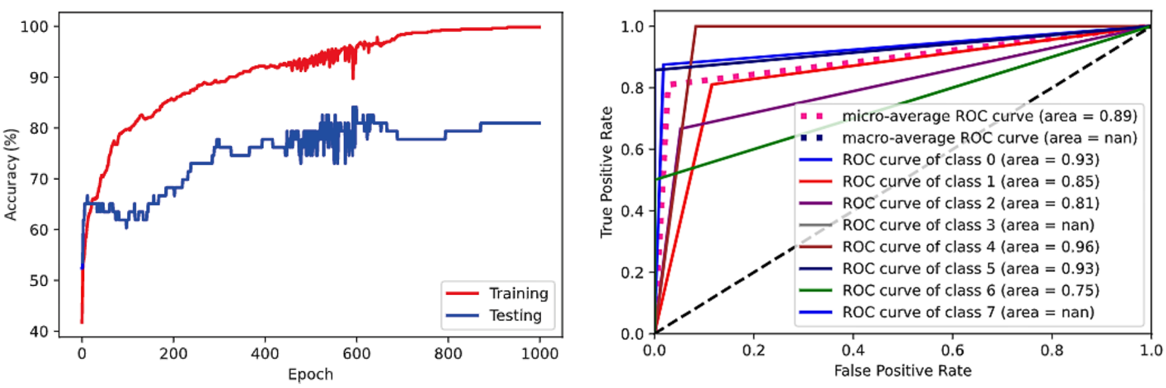
    5. Update weight values

$$w_{kj}(new) = w_{kj}(now) + \Delta w_{kj}$$

  End
End
  For each neuron do output layer
  Calculate output error value against target

$$J_{bce} = \frac{-1}{M} \sum_{m=1}^{M} [y_m x \log(h_\theta(x_m))] + (1 - y_m) x \log(1 - h_\theta(x_m))$$

  End
End While



(a) Accuracy (b) Curve ROC & AUC

FIGURE 3. (color online) Experiment result diagram with best accuracy

TABLE 2. Summary of testing student performance prediction models

MLP (Node layer)	L. rate	Loss	Accuracy	Precision	Recall	AUC
Model 1 12-18.20.10-8	0.001	0.20	69.84	56.77	67.52	0.83
	0.003	0.21	73.02	66.29	77.68	0.85
	0.005	0.31	71.43	70.50	40.24	0.84
	0.007	0.17	77.42	61.22	59.08	0.87
	0.009	0.21	71.43	67.35	65.16	0.84
Model 2 12-10.20.10-8	0.001	0.14	74.19	79.75	68.71	0.85
	0.003	0.24	68.25	60.16	57.56	0.82
	0.005	0.19	74.60	70.82	59.50	0.85
	0.007	0.24	73.02	62.54	53.66	0.85
	0.009	0.27	69.84	58.09	49.25	0.83
Model 3 12-16.20.16-8	0.001	0.25	74.19	74.75	61.85	0.85
	0.003	0.28	74.60	70.29	47.25	0.85
	0.005	0.22	71.43	67.81	68.68	0.84
	0.007	0.17	70.97	70.09	65.69	0.83
	0.009	0.24	73.02	65.44	61.32	0.85
Model 4 12-18.18.16-8	0.001	0.19	73.02	77.92	69.40	0.85
	0.003	0.12	80.95	83.20	83.53	0.89
	0.005	0.24	72.58	55.97	48.35	0.85
	0.007	0.27	72.58	71.89	48.45	0.84
	0.009	0.37	68.25	63.29	52.49	0.82

Besides that, a loss rate of 0.12 is the smallest among other experiments with the Area Under Curve (AUC) producing a value of 0.89, classified as a good result.

4. **Conclusions.** The use of the backpropagation neural network method with the application of the Multi-Layer Perceptron (MLP) network architecture and the use of the Tanh activation function in the hidden layer and SoftMax on the output layer are able to handle classifications whose output representations are multi-label/multi-class. Determining the number of hidden layer nodes in each MLP architectural model with different learning rate values can increase the accuracy value of each model tested by obtaining the best accuracy of 80.95% so that it can be categorized as Good Classification. Thus, the model with the best accuracy can be applied to predicting student performance in the student academic pre-warning mechanism. In the future, we will explore the association rule method to generate automatic recommendation patterns to improve student academic performance based on the predictive results that have been made to be applied to an intelligent academic recommender system.

REFERENCES

[1] J. Aguilar, P. Valdiviezo-Díaz and G. Riofrio, A general framework for intelligent recommender systems, *Appl. Comput. Informatics*, vol.13, no.2, pp.147-160, DOI: 10.1016/j.aci.2016.08.002, 2017.
[2] Y. Yamasari, S. M. S. Nugroho, K. Yoshimoto, H. Takahashi and M. H. Purnomo, Expanding tree-based classifiers using meta-algorithm approach: An application for identifying students' cognitive level, *International Journal of Innovative Computing, Information and Control*, vol.15, no.6, pp.2085-2107, 2019.
[3] Kemristekdikti, Higher education statistical 2018, in *Higher Education Statistical Year Book*, 2018.
[4] B. Guo, R. Zhang, G. Xu, C. Shi and L. Yang, Predicting students performance in educational data mining, *2015 International Symposium on Educational Technology (ISET)*, pp.125-128, DOI: 10.1109/ISET.2015.33, 2015.
[5] Ministry of Research Technology and Higher Education of Republic of Indonesia, *National Standards for Higher Education*, 2015.

[6] L. M. Abu Zohair, Prediction of student's performance by modelling small dataset size, *Int. J. Educ. Technol. High. Educ.*, vol.16, no.1, DOI: 10.1186/s41239-019-0160-3, 2019.

[7] M. Mimis, M. El Hajji and Y. Es-Saady, A framework for smart academic guidance using educational data mining, *Educ. Inf. Technol.*, DOI: 10/1007/s10639-018-9838-8, 2018.

[8] J. Xu, K. H. Moon and M. Van Der Schaar, A machine learning approach for tracking and predicting student performance in degree programs, *IEEE J. Sel. Top. Signal Process.*, vol.11, no.5, pp.742-753, DOI: 10.1109/JSTSP.2017.2692560, 2017.

[9] H. Al-Shehri et al., Student performance prediction using support vector machine and k-nearest neighbor, *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp.1-4, 2017.

[10] S. A. Naser, I. Zaqout, M. A. Ghosh, R. Atallah and E. Alajrami, Predicting student performance using artificial neural network: In the Faculty of Engineering and Information Technology, *Int. J. Hybrid Inf. Technol.*, vol.8, no.2, pp.221-228, DOI: 10.14257/ijhit.2015.8.2.20, 2015.

[11] L. Zhang, T. Luo, F. Zhang and Y. Wu, A recommendation model based on deep neural network, *IEEE Access*, vol.6, pp.9454-9463, DOI: 10.1109/ACCESS.2018.2789866, 2018.

[12] Y. Ho and S. Wookey, The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling, *IEEE Access*, vol.8, pp.4806-4813, DOI: 10.1109/ACCESS.2019.2962617, 2020.

[13] R. Hecht-Nielsen, Theory of the backpropagation neural network, *Neural Networks*, vol.1, p.445, 1988.

[14] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Springer Science & Business Media, Berlin, Heidelberg, 2011.