Adaptive Deep Learning Technique to Predict Student's Graduation Results

Nguyen Quoc Viet[1], Vo Pham Tri Thien[2], and Nguyen Thanh Binh[3,4(✉)]

[1] Student Affairs Office, Ho Chi Minh City University of Education, Ho Chi Minh City, Vietnam
vietnqu@hcmue.edu.vn
[2] Undergraduate Studies Office, Ho Chi Minh City University of Education, Ho Chi Minh City, Vietnam
thienvpt@hcmue.edu.vn
[3] Department of Information Systems, Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam
ntbinh@hcmut.edu.vn
[4] Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

Abstract. Analyzing educational data techniques helps educational institutions predict the final graduation results of students based on the average score in the first semesters of the course. To predict student's graduation results, this research proposed the method to predict includes five steps: data collection, data preprocessing, predict student's graduation results model, evaluating model and deployment model. The proposed method is based on the average academic points data of the six main semesters out of a total of eight main semesters of students to predict their graduation result with the accuracy of about 90.53%. To objectively evaluate the effectiveness of the proposed method, the results of the proposed method, which compared with the other methods, are better than the other methods.

Keywords: Data analysis · Student's graduation result · Deep learning · Data mining

1 Introduction

Higher education is one of the most important forms of education in the country. Students will have tremendous opportunities of seeking jobs suitable to their specialized majors after the completion of the university education programs. In our country, according to statistics of the Ministry of Education and Training, there are approximately one million students enrolled by about 237 education institutions every year [1]. Many newly enrolled students will be a big concern for the training quality to which the educational institutions should pay attention. Therefore, higher education institutions need to forecast the learning situation of students to timely support students with weak academic ability to keep up with their learning progress.

P. Cong Vinh and N. Huu Nhan (Eds.): ICTCC 2021, LNICST 408, pp. 67–75, 2021.
https://doi.org/10.1007/978-3-030-92942-8_6

Today, with the advent of machine learning techniques, it is possible for users to use them to mine large data sets to find out their impact on the area of life. Educational data mining is also an area that many scientists have been researching. Using educational data in universities and colleges to analyze, predict, and support decision making has become one of the most important contributions to improve the quality of education in universities and colleges. One of the essential factors that directly affect the graduation results of students is the average score of the courses. For students with a good result, a high-grade point average (GPA) reflects their entire study procedure, and they easily achieve the desired graduation results. For students with average, good, or likely failure to graduate, we can use GPA data to give them predictions about graduation or future outcome within the following school year to change their study methods to get the results they want and graduate higher classification.

With such requirements, applying data mining techniques, especially Educational data mining to predict the students' graduation result will help the schools figuring out the opportunities as well as the risks to improve not only teaching quality but also the studying condition at the university. This research proposed a method to predict student's graduation result. The proposed model includes five steps: data collection, data preprocessing, predict student's graduation results model, evaluating model and deployment model. We use the data of the average academic points from the six main semesters out of a total of eight main semesters of students to predict their graduation result. Accuracy criteria proposed to evaluate the results between our proposed method and the recent other methods. The rest of the paper is organized as the following: the literature review is presented in Sect. 2. The proposed method for predicting outcomes for students about to graduate in 2021 is presented in Sect. 3. The result analysis is presented in Sect. 4. Finally, Sect. 5 is conclusions and future works.

2 Literature Review

Educational data mining (EDM) is a field of data mining, with superabundant data from the education industry, schools can apply data mining techniques to analyze and predict students' achievement [2–4]. Early prediction of student results can help administrators make the necessary decisions at the right time and create the right training plan to improve students' academic success rates.

Many researchers used many methods to predict how successful students will be at the time of receiving their graduation certificate or in the early years of the course [2, 5, 6]. According to previous relevant research, students' academic success is defined by factors such as academic achievement, demographics, satisfaction, competencies, acquisition of skills, attainment of learning objectives during their academic careers. Besides, the critical factors affecting students' academic success is the grade point average of the course [7], prior academic achievement and student demographics [2], which presented 69% of the total factors affecting student learning out-comes, the remaining 31% are related factors such as e-learning activities, psychological attributes, and environments.

There are many approaches to predict student learning outcomes, in which, the classification model and regression model are the most used [8–10]. The algorithms

used in the classification model are decision tree, Bayesian, artificial neural networks, rule learner's, ensemble learning, K-Nearest neighbor. These algorithms account for 95% of prior studies [2]. For forecasting student results, recent studies have used techniques such as Decision trees [12, 14, 18], Deep learning, Neural network, Naïve Bayes, Support vector machine [15]. In addition, some other models are also used for student results prediction like Generalized linear model [16]. Data mining techniques are used to predict student's performance in the first or second year in the university [17].

Recently, machine learning techniques are used to analyze data and predict students' results [3, 11–13]. Among research about academic success prediction, GPA is one of the prerequisite factors affecting the academic achievement of students [2, 14]. Deep learning is also a widely and used technique and remains an important role in predicting student final GPA [19]. However, each method has its own advantages and disadvantages.

3 Methodology

To carry out this study, this paper proposes a model to predict outcomes for students about to graduate in 2021. The proposed model includes five steps: data collection, data preprocessing, predict student's graduation results model, evaluating model and deployment model which is presented as Fig. 1. The details of the steps in the proposed model are presented below.

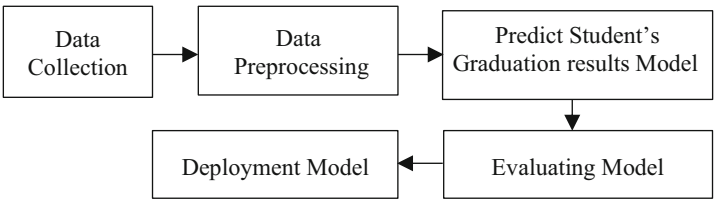

Fig. 1. The steps in the proposed model

(i) Data Collection: The data provided by our university, which is blur, includes data of graduates in 2018, 2019, and 2020 for training. Besides that, data of students, who are to graduate in 2021, are used to deploy the trained model. Our university has approximately 2.000 graduates every year. Therefore, predicting graduation results will significantly help students while studying at the university. In this research, we use the data of the average academic points from the 6 main semesters of students provided by the university. Our data has 6064 students with their data property as: Student ID, Gender, Student's GPA in semester 1, 2, 3, 4, 5, 6 and final graduation result.

(ii) Data Preprocessing: The provided data is stored in many individual datasets. It is heterogeneous in structure and properties. So, we have processed the data by selecting the necessary attributes for this study and saving them to a single dataset so that training models can be deployed.

(iii) Predict student's graduation results model: According to the related work which presented in the previous section, the neural network method has many advantages in predicting student's graduation. Therefore, we choose the method to solve this problem using neural networks.

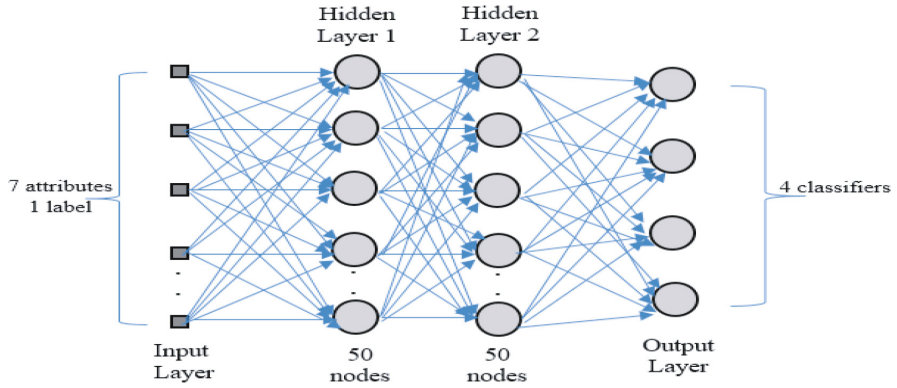

Fig. 2. Predict student's graduation results model

Based on the dataset to be evaluated, the goal of the out and the requirements of prediction of the student's graduation, the neural network model selected and presented as Fig. 2.

Table 1. The other parameters of the predict model.

* *The parameters:*
+ Loss function: Multiclass Cross Entropy (Classification)
+ Optimization method: Stochastic gradient descent
+ Learning rate: 0,005
+ Number of hidden layers: 2
+ Neuron in layers: 50
+ Epochs = 300
+ Epsilon = 1.0e-6
+ Beta1= 0,9
+ Beta2 = 0,999
* *Parameter for Hidden Layer 1 and Hidden Layer 2:*
+ Nodes = 50
+ Activation Function: Rectified Linear Unit
* *Output Layer:*
+ Nodes = 4
+ Activation Function: Softmax

As presented as Fig. 2, the input layer includes seven attributes as: Gender, Student's GPA in semester 1, 2, 3, 4, 5, 6 and final graduation result (one label). With the hidden layer 1 and 2, every layer has 50 nodes. The output layer is divided into four classifiers: Average, Good, Very Good, Excellent. Because the students do not graduate if their GPA is below average. The other parameters of the predict model selected as Table 1.

(iv) Evaluating Model: We will evaluate the performance of the model and against the other models such as: decision tree, random forest. This part will present detail in Sect. 4.

(v) Deployment Model: After conducting experiments with data of graduates in 2018, 2019, and 2020, we will deploy the above models for predicting graduation grades for graduates in 2021.

In the next part, we implement the proposed method and results analysis.

4 Result Analysis

4.1 Material

Our experiments are developed in Python, with the computer of Intel core i5 (4 CPUs), 2.5 GHz CPU, 4 GB DDR2 memory. We collected the data which was presented in Sect. 3 (data collection and data preprocessing steps) for our experimentation. Data collection consists of two datasets: dataset A and dataset B. Detail of these dataset include:

+ Dataset A includes 4648 rows of data equivalent to 4648 students who graduated in 2018, graduated in 2019 and graduated in 2020. Every data in dataset A has nine attributes, including: Student ID, Gender, Student's GPA in semester 1, 2, 3, 4, 5, 6 and final graduation result.

+ Dataset B includes 1416 rows of data equivalent to 1416 students will graduate in 2021. Every data in dataset B has eight attributes, including: Student ID, Gender, Student's GPA in semester 1, 2, 3, 4, 5 and 6.

Based on the rules of the university, graduate classification of students depends on their GPA. Specifically, we categorize the result into four classifications as:

+ The first classification with GPA between 2.0 and 2.49 is Average.
+ The second classification with GPA between 2.5 and 3.19 is Good.
+ The third classification with GPA between 3.2 and 3.59 is Very Good.
+ The last classification with GPA between 3.6 and 4.0 is Excellent.

4.2 Experimentation and Evaluation Results

Firstly, we evaluated to predict student's graduation results model on dataset A. This dataset is divided into 60% for the training dataset and 40% for the test dataset (equivalent to 1859 students). After conducting training and testing the models, the accuracy metric proposed to evaluate results of the prediction model. With some training epochs and learning rate, the model will be fit for 300 training epochs and learning rate is 0,005. The confusion matrix of the proposed method is shown in Table 2. We see that the proposed method has good results.

Table 2. Confusion matrix for the proposed method in dataset A

Prediction level	True good	True very good	True excellent	True average
Average	1	0	0	6
Good	1092	70	0	13
Very good	62	540	26	0
Excellent	0	4	45	0

To objectively evaluate the effectiveness of the proposed method, we compared the results of the proposed method with Decision Tree method [8], Random Forest method [20] by the evaluation as accuracy criteria. Table 3 and Table 4 presented the confusion matrix of the Decision Tree method and Random Forest method with the above dataset. The accuracy of the proposed method with the other methods presented as Table 5.

Table 3. Confusion matrix for the Decision Tree method in dataset A

Prediction level	True good	True very good	True excellent	True average
Average	12	0	0	13
Good	1021	98	0	6
Very good	121	476	18	0
Excellent	1	40	53	0

Table 4. Confusion matrix for the Random Forest method in dataset A

Prediction level	True good	True very good	True excellent	True average
Average	13	1	0	8
Good	984	87	0	11
Very good	158	498	18	0
Excellent	0	28	53	0

In Table 5, the accuracy of the proposed method is 90.53% while the accuracy of the Decision Tree method and Random Forest method are 84.08% and 83%, respectively. So, the results of the proposed method are better than the other methods.

Table 5. Accuracy of the proposed method with the other methods in dataset A

Evaluation criteria	Decision tree method	Random forest method	Proposed method
Correct classified	1563	1543	1683
Wrong classified	296	316	176
Accuracy	84,08%	83,00%	90,53%
Error	15,92%	17,00%	9,47%

Secondly, we evaluated to predict student's graduation results model on dataset B. The experimentation is implemented to predict the learning outcomes for 1416 students who are preparing to graduate in 2021. The dataset B has a structure and properties quite like the dataset A. The difference in this dataset is that there is no attribute of the students' final graduation results.

With the prediction for dataset B, the results in Table 6 show that the predictive model gives relatively good results and can be applied to problems using average scores to predict graduation outcomes for students.

Table 6. Number of true and false prediction in dataset B

Prediction level	Decision tree	Random forest	Proposed method
True prediction	1194	1171	1275
False prediction	222	245	141

Table 7. Accuracy of the proposed method with the other methods in dataset B

Evaluation criteria	Decision tree	Random forest	Proposed method
Accuracy	84,32%	82,7%	90,04%
Error	15,68%	17,3%	9,96%

Table 7 presented the results of predicting our proposed method with other methods. In the dataset B with fewer data rows than the dataset A, the accuracy among the models is not significantly different. The highest accuracy belongs to the proposed method at a rate of 90.04% while the accuracy of the Decision Tree method and Random Forest method are 84.32% and 82.7%, respectively. From the experimentation on dataset A and dataset B, the results of the proposed method are better than the other methods.

5 Conclusions and Future Works

Analyzing educational data techniques helps educational institutions predict the final graduation results of students based on the average score in the first semesters of the course. The experiment shows that using the average scores in the first six semesters of the course will help educational institutions to categorize students. From that, they can intervene in the students' study procedure, encourage students to improve their study results and to desired academic success. The proposed method used in this study gives pretty good output in terms of performance. Along with training and testing the models, the research has also deployed models to predict the grade results for students graduating in 2021 with the high prediction accuracy. In the future work, besides using the average score and the grade classification, we need to use more data from the results of each subject to increase the results predicted.

Acknowledgment. We would like to thank HCMC University of Education for the data in this research, so this research is able to be done and hope can bring benefit for all their students. We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

References

1. Ministry of Education and Training. https://moet.gov.vn/thong-ke/Pages/thong-ko-giao-duc-dai-hoc.aspx?ItemID=7389. Accessed 4 July 2021
2. Alyahyan, E., Düştegör, D.: Predicting academic success in higher education: literature review and best practices. Int. J. Educ. Technol. High. Educ. **17**, 1–21 (2020)
3. Romero, C., Ventura, S.: Data mining in education. Wiley Interdiscipl. Rev. Data Min. Knowl. Discov. **3**(1), 11–27 (2013)
4. Martins, M.P.G., Miguéis, V.L., Fonseca, D.S.B., Alves, A.: A Data mining approach for predicting academic success – a case study. In: Rocha, Á., Ferrás, C., Paredes, M. (eds.) ICITS 2019. AISC, vol. 918, pp. 45–56. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11890-7_5
5. Acharya, A., Sinha, D.: Early prediction of students performance using machine learning techniques. Int. J. Comput. Appl. **107**(1), 37–43 (2014). https://doi.org/10.5120/18717-9939
6. Tanuar, E., Heryadi, Y., Abbas, B.S., Gaol, F.L.: Using machine learning techniques to earlier predict student's performance. In: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 85–89. IEEE (2018)
7. York, T.T., Gibson, C., Rankin, S.: Defining and measuring academic success. Pract. Assess. Res. Eval. **20**, 5 (2015). https://doi.org/10.7275/hz5x-tx03
8. Adekitan, A.I., Odunayo, S.: The impact of engineering students' performance in the first three years on their graduation result using educational data mining. Heliyon **5**(2), e01250 (2019). https://doi.org/10.1012/j.heliyon.2019.e01250
9. Almarabeh, H.: Analysis of students' performance by using different data mining classifiers. Int. J. Mod. Educ. Comput. Sci. **9**, 9–15 (2017). https://doi.org/10.5815/ijmecs.2017.08.02
10. Asif, R., Merceron, A., Ali, S.A., Haider, N.G.: Analyzing undergraduate students' performance using educational data mining. Comput. Educ. **113**, 177–194 (2017)
11. Garg, R.: Predicting student performance of different regions of Punjab using classification techniques. Int. J. Adv. Res. Comput. Sci. **9**, 236–241 (2018)
12. Hamoud, A., Hashim, A.S., Awadh. W.A.: Predicting student performance in higher education institutions using decision tree analysis. Int. J. Interact. Multim. Artif. Intell. **5**, 26–31 (2018)
13. Xu, J., Moon, K.H., Van Der Schaar, M.: A machine learning approach for tracking and predicting student performance in degree programs. IEEE J. Select. Top. Sig. Process. 11, 742–753 (2017)
14. Al-Barrak, M.A., Al-Razgan, M.: Predicting students final GPA using decision trees: a case study. Int. J. Inf. Educ. Technol. **6**(7), 528–533 (2016). https://doi.org/10.7763/IJIET.2016.V6.745
15. Babić, I.Đ.: Machine learning methods in predicting the student academic motivation. Croat. Oper. Res. Rev. **8**, 443–461 (2017)
16. Kumar, A., Naughton, J., Patel, J.M.: Learning generalized linear models over normalized data. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (2015). https://doi.org/10.1145/2723372.2723713

17. Adekitan, A.I., Noma-Osaghae, E.: Data mining approach to predicting the performance of first year student in a university using the admission requirements. Educ. Inf. Technol. **24**(2), 1527–1543 (2018). https://doi.org/10.1007/s10639-018-9839-7
18. Pandey, M., Sharma, V.K.: A decision tree algorithm pertaining to the student performance analysis and prediction. Int. J. Comput. Appl. **61**(13) (2013)
19. Sultana, J., Usha, M., Farquad, M.A.H.: Student's performance prediction using deep learning and data mining methods. Int. J. Recent Technol. Eng. **8**(1S4) (2019). ISSN: 2277-3878
20. Jayaprakash, S., Krishnan, S., Jaiganesh, V.: Predicting students academic performance using an improved random forest classifier. 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 238–243. IEEE (2020)