

Research Article

Research on Forecasting of Student Grade Based on Adaptive K-Means and Deep Neural Network

Xinfen Yang,¹ Hao Zhang ,^{1,2} Renmei Chen,¹ Shasha Li,¹ Ni Zhang,¹ Biao Wang,¹ and Xin Wang³

¹Institute of Education, Guizhou Normal University, Guiyang 550025, China

²Guizhou Provincial Educational Governance Modernization Research Center, Guiyang 550025, China

³School of Business, Donghua University, Shanghai, China

Correspondence should be addressed to Hao Zhang; zh@gznu.edu.cn

Received 17 March 2022; Revised 1 April 2022; Accepted 6 April 2022; Published 22 April 2022

Academic Editor: Deepak Kumar Jain

Copyright © 2022 Xinfen Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artificial intelligence (AI) has been widely used in AIEd (AI empowers education industry). In this paper, we conduct mining and prediction of influencing factors of student performance based on education data set. First, we apply data mining and analysis methods to two public educational data sets. Next, we further analyze the results through visual analysis method and try to explain the physical meaning. Then, we screened the feature data through two computing methods based on random forest algorithm. Finally, we predicted the student performance through an improved model based on K-means and Deep Neural Network (DNN). Our results show that the proposed model feature selection Adaptive-K-means-DNN obtains the best mean of squared residuals both in two student learning data sets; the results demonstrating that the feature selection processing and the employment of the adaptive K-means have enhanced the prediction performances significantly. Furthermore, we also find that mother's education is the direct influence of the students' final grade, the difference is there is no strong correlation between absences and the students' final grade, but absences is still an important factor affecting the students' final grade, and it may affect the students' final grade through other factors; moreover, the failures also has a strong correlation between the students' final grade, and it also is a direct influence indicator for the students' final grade; in summary, the influence of mother's education on students' performance is very important. The management of students' classroom absence will effectively improve students' final performance. In addition, regular encouragement may be a good way for students to improve their learning performance.

1. Introduction

Because of the appearance of a lot of instructive information, educational data mining (EDM) has arisen in the beyond 20 years. EDM centers around creating and applying information mining techniques to distinguish designs in a lot of instructive information and to more readily get understudies and their learning surroundings. The prediction method is one of the most commonly used methods, which include classification, regression, and potential factor estimation methods. EDM can be used for a variety of tasks, such as building an intelligent tutoring system (ITS) [1], proposing appropriate exercises to students accordingly. Many of these methods are based on knowledge-tracking models and their deformation

forms. EDM has also been used in others including robotized, information-driven educational program plan, professional education arranging and educational plan suggestion, getting the effect of understudies' social way of behaving on their scholastic execution, and numerous different undertakings.

Artificial intelligence (AI) has been widely used in teaching practice and education data mining (AI-enabled education industry, AIEd), like analysis systems and HCIS system [2]. Since the debut of AIEd about 30 years, AI has been recognized as a useful method for facilitating update models of instructional design, technology development, and educational research [3, 4]. Specifically, AIEd offers new opportunities for educational innovation, such as the shift to personalized learning, the challenge of the role of

teachers, and the development of complex education systems [5, 6]. Various AIEd technologies have been used to create intelligent learning environments such as behavior classification, predictive method, and learning recommendations. [2, 7]. AIEd has become a major research focus in computing and education and has the potential to facilitate a shift in knowledge and culture [4]. Specifically, the current academic research on intelligent education is mainly divided into the following three stages:

Phase I: AI guidance, with learners as recipients. AI addresses proficient navigation and guides the learning system; while students, as beneficiaries of AI administrations, follow explicit learning ways. The hypothetical premise of the principal stage is behaviorism, which accentuates the development of painstakingly organized successions of content in the desire for a decent presentation of the student [8]. The first phase argues that learning reinforces the acquisition of knowledge through programmatic instructions that introduce new concepts in a logical, progressive manner, provide learners with immediate feedback on error responses, and maximize reinforcement of the right direction guidelines [9–11]. Learners act as recipients, react to predetermined sequences of knowledge, follow learning procedures and pathways, and perform AI-set learning activities to achieve pre-set goals [12, 13]. In phase I, the AI system inherits the characteristics of the teaching machine [11], making a logical introduction to subject knowledge that requires learners to respond [12]. Phase I is the least learner-centric phase.

Phase II: AI-enabled, learners as collaborators. That is, AI systems relinquish their control only as a supportive tool; learners work with the system as collaborators and focus on their own learning processes. The second phase is based on a cognitively recognizable view of learning, which reflects the notion that learning occurs when learners interact with people, information, and technology in a social setting [14–16]. Accordingly, in phase II, AI systems and learners should establish positive, that means, AI systems collect emerging, personalized information from learners as input to adaptively optimize decision models, while learners act as collaborators and communicate with AI systems for better and more efficient learning [17, 18]. Overall, compared to phase I, phase II has made important initiatives for learner-centered learning through interaction and continuous collaboration between learners and AI systems.

Phase III: artificial intelligence empowerment, learners are leaders. Phase III considers assistive decision-making in the learner’s learning process to be at the heart of AIEd and views AI as a tool to augment human intelligence. Phase III reflects the view of educational complexity theory, which holds that education is a complex adaptive system [19] in which collaboration between multiple entities (e.g., learners, teachers, information, and technology) is critical to enhancing learner learning outcomes. In this complex system, the design and application of AIEd need to realize that AI technology is part of a bigger system made up of humans [20]. In order to achieve collaborative cooperation in complex systems, by considering the conditions, expectations, and backgrounds of teaching participants, AI is approached from the perspective of teaching participants, and human-machine cooperation [21], human-centered AI and

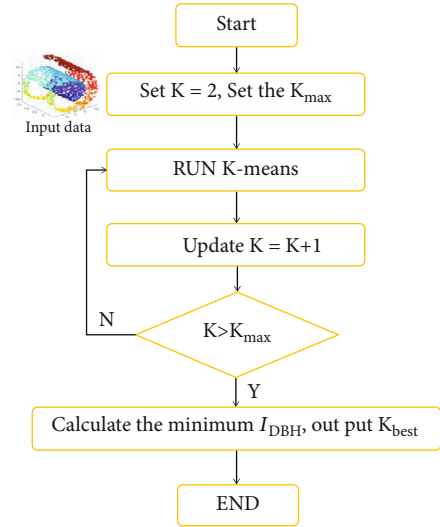


FIGURE 1: K-means process based on Davies-Bouldin index.

ML systems [20], human-AI collaboration [4], and people-centered AI education are proposed. In phase III, AI assists learners and facilitators in improving learning outcomes by providing high-level, highly accurate, and effective decisions [20, 22]. AI systems equip instructors with understandable, interpretable, and personalized support to facilitate learner-centered learning [3, 23]. Learners act as leaders in their learning, managing the risk of AI automated decision-making and completing more effective learning. Overall, phase III, as a development trend of AIEd, mirrors a definitive objective of AI application in the field of instruction, or at least, to improve human learning capacity and potential [24].

In this paper, our research goal will be based on the third stage of intelligent education, that is, AI empowerment, where learners are leaders. We develop a student achievement prediction system based on deep learning to provide decision-making for teaching managers. As for the prediction of students’ performance, scholars mainly study through historical data such as students’ performance in previous courses, and rarely fundamentally excavate students’ learning patterns, learning characteristics, and curriculum characteristics. Based on the individual differences of students, as well as the differences and particularity of learning content, depth, progress and students’ knowledge ability, learning style, and cognitive ability, this paper carries out students’ performance prediction and information mining. The main parts of this paper are as follows: the second section focuses on DNN model and adaptive k-means algorithm. The third section excavates the student education data set through strict statistical methods. In the fourth section, we build a random forest model, carry out feature engineering, and construct the final training data set. Then in the fifth section, we trained and predicted the model, and, the conclusion is introduced in the last section.

2. Methodology

2.1. Adaptive K-Means Algorithm. Traditional K-means is not suitable for clustering the input data set directly, because the number of clustering cannot be determined subjectively

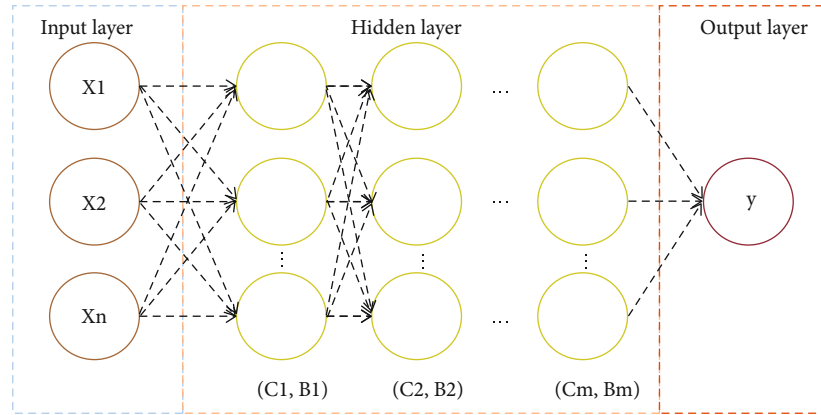


FIGURE 2: The structure of DNN.

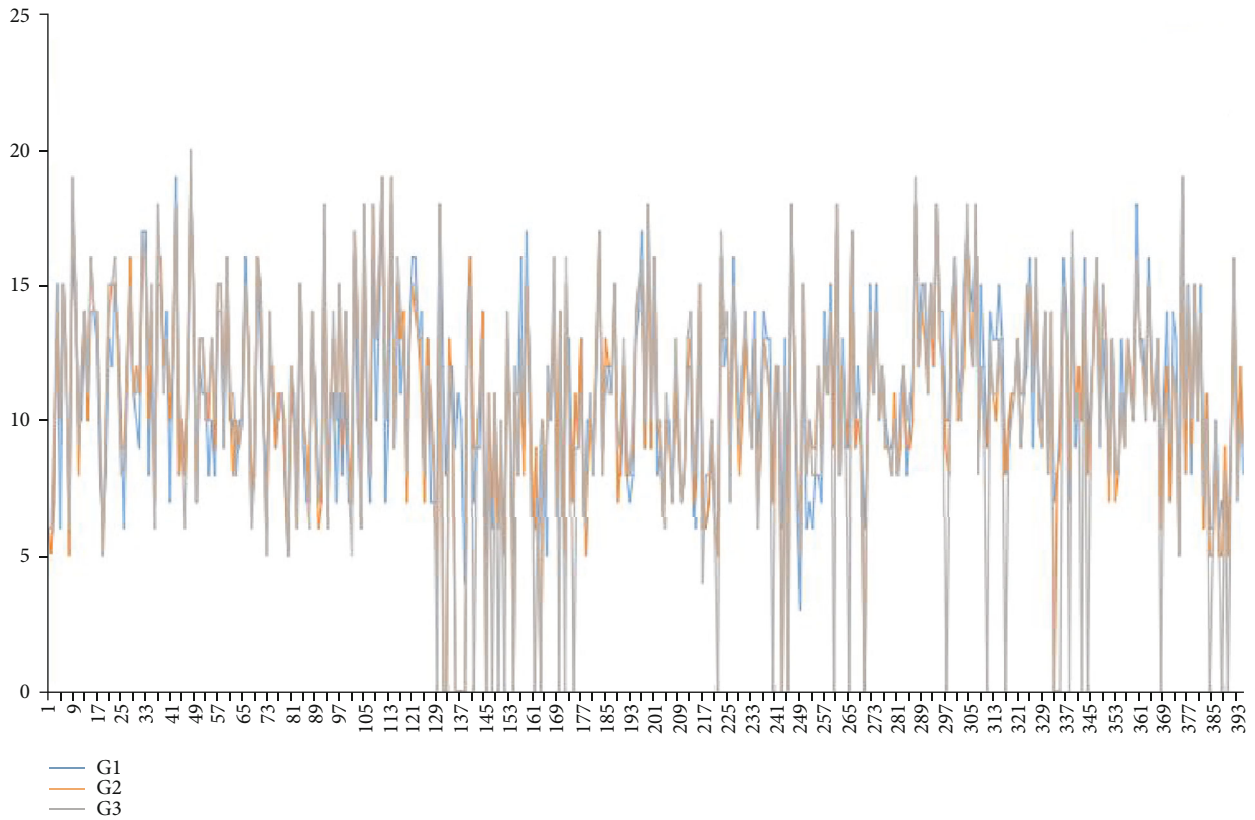


FIGURE 3: Distribution of G1, G2, and G3.

without understanding the data set. Therefore, this paper proposes an adaptive K-means, which can automatically set the number of clustering according to the input data set. Its main idea is a distance-based iterative process. The k-means algorithm steps are as follows:

- (1) k samples are randomly selected from the data set as the initial cluster centers $\{\mu_1, \mu_2, \dots, \mu_k\}$
- (2) Calculate the Euclidean distance from the remaining samples to the cluster centers and assign them to the nearest cluster center to form k clusters, and the

measure of distance is shown in equation (1).

$$D^* = (A, B) = \sqrt{\sum_{k=1}^n (A_k - B_k)^2}, \quad (1)$$

where n represents the dimensionality of space; A_k and B_k are the k -th properties of A and B , respectively.

- (3) Update the cluster centers by distance measurement to the mean of all samples belonging to the cluster

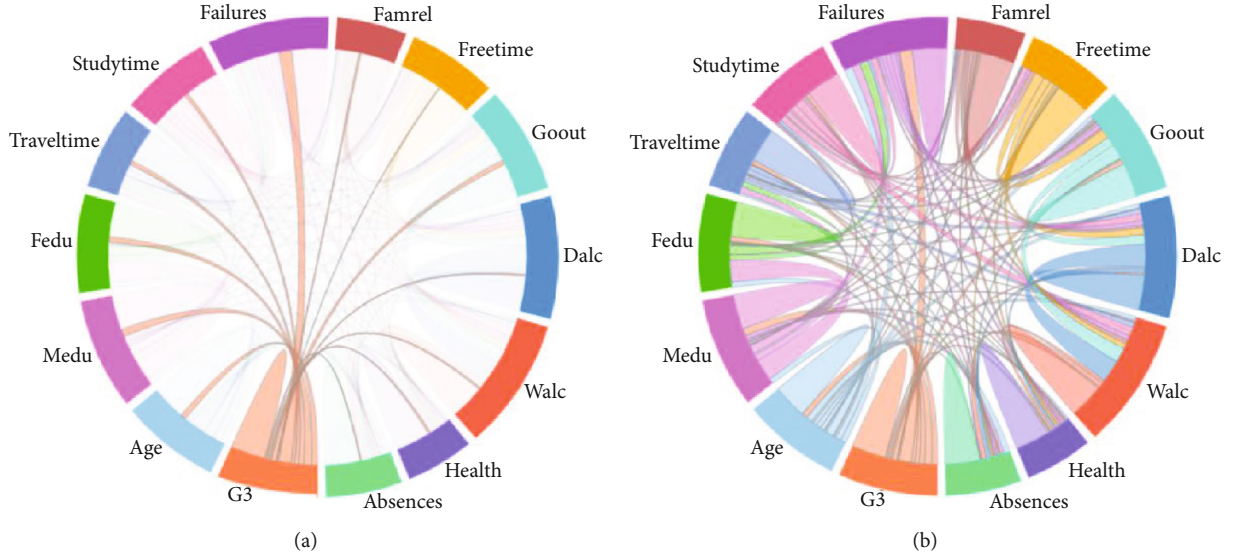


FIGURE 4: Correlation analysis between different indicators.

- (4) Repeat steps (2) and (3) until the algorithm converges

To automatically select the optimal number of clusters k , quantitative indicators are introduced to search for the best clusters for the sample. The key to the adaptive process in this paper is cluster assessment, which is associated with a large number of indicators, while the Davies-Bouldin index uses the quantities and characteristics inherent in the data set and is suitable for K-means cluster assessment. The definitions are as follows:

$$I_{DBI} = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{C_i^* + C_j^*}{D_{i,j}^*} \right), \quad (2)$$

where C_i^* and C_j^* represent the average distances from the i and j cluster samples to the centers of the corresponding clusters, respectively. $D_{i,j}^*$ represents the Euclidean distance from the center of cluster i to cluster j . The smaller the I_{DBI} , the better the cluster performance, and the best cluster book k_{best} can be obtained.

To avoid generating too many clusters, the proposed algorithm limit the number of clusters with thresholds k_{max} . Figure 1 show the processing of the adaptive clustering method..

2.2. DNN Structure. In shallow neural networks, only the number of single-hidden-layer neurons can be changed, while in DNN, the width and depth of the network can be changed. The DNN alleviates the local optimal point problem by pre-training, and the model structure is shown in Figure 2.

In Figure 2, $x_1 - x_n$ is the input data, which is a vector consisting of a sequence of loads, and the input layer uses a linear identity function as the activation function, c is the weight parameter, b is the bias parameter, and m is the number of layers of the implicit layer. The input vectors of each

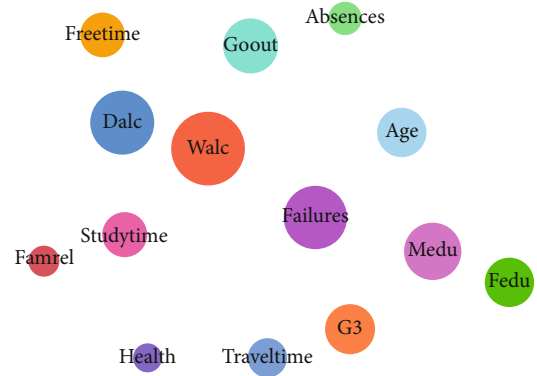


FIGURE 5: Correlation analysis of different indicators.

layer of the implied layer are taken from the previous layer and combined with the layer activation function to perform a series of nonlinear transformations, and then the obtained vectors are output to the next layer of neurons until they are passed to the output vector y .

If the input vector of the layer above each layer of the implied layer is $a^{(m-1)} = (a_1^{(m-1)}, a_2^{(m-1)}, \dots, a_n^{(m-1)})$, then the output vector of the m -th layer $a^{(m)}$ is:

$$a^{(m)} = f \left(\sum_{n=1}^a c_{m-1} a_i^{(m-1)} + b_{m-1} \right), \quad (3)$$

where a is the number of neurons in the m -th layer; f is the activation function. The activation function adopts the sigmoid function, which has the advantage of not being easy to diverge during data transfer, and the amount of computation is small, which can be expressed as:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}. \quad (4)$$

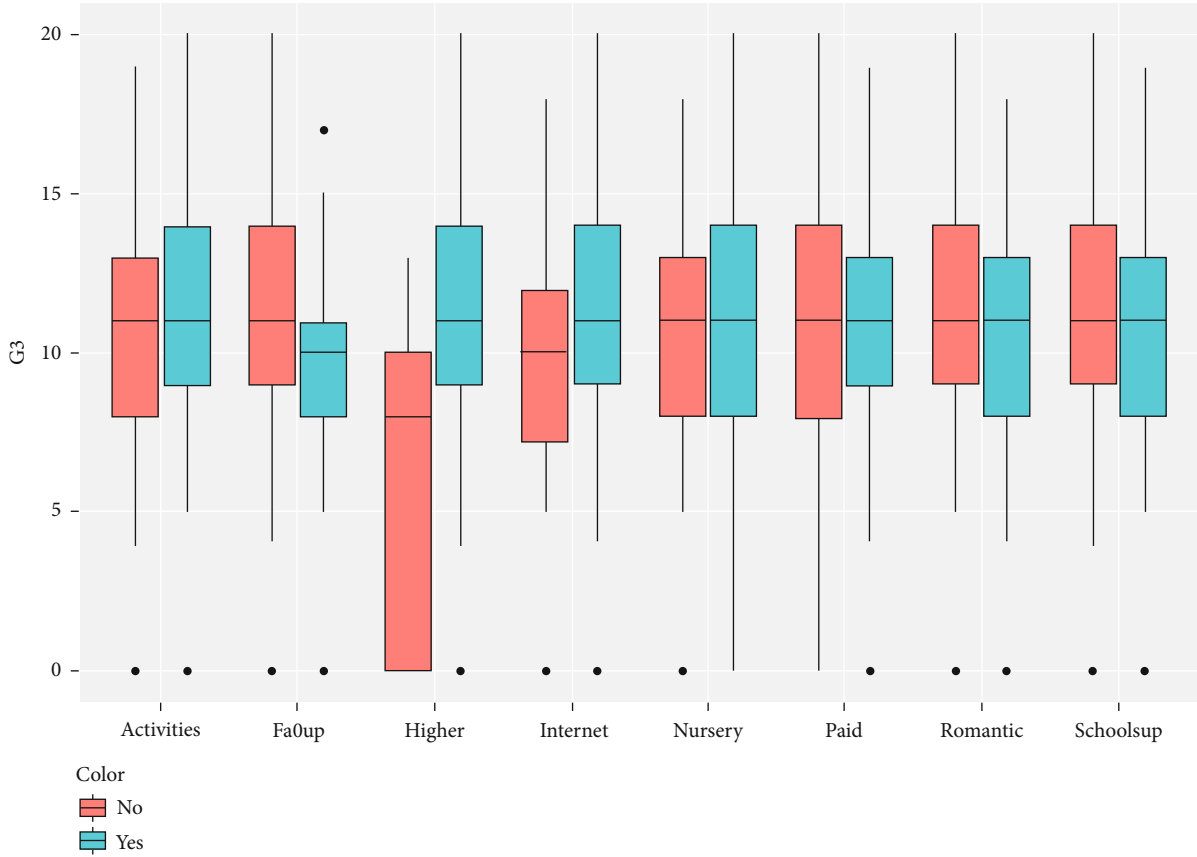


FIGURE 6: Box plot of the distribution of indicators.

The output of the network can be expressed as:

$$y = f^{m+1} \left(\sum_{n=1}^{Hd} c_{m+1} f^m + b_{m+1} \right), \quad (5)$$

where Hd is the number of neurons in the last layer of the implicit layer.

A 3-layer network structure, namely, an implicit layer, is used, on the basis of which the number of neurons in the implicit layer is adjusted to achieve the optimal prediction effect. The training is carried out in 2 steps: (1) build a single layer of neurons layer by layer with the goal of training one single-layer network each time; (2) after each layer of training is completed, tuning using the wake-sleep algorithm (wake stage: recognition process; sleep stage: generation process).

3. Date Processing and Mining

This paper uses a public student data set as the analysis object. This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, and demographic, social, and school-related features and it was collected by using school reports and questionnaires [25].

Figure 3 shows the distribution of G1, G2, and G3. The distribution of G1, G2, and G3 is very similar, which indi-

cates that after students maintain their study habits, their study performance will show some continuity unless other factors intervene. G3 means final grade (numeric: from 0 to 20, output target), and in this paper, we do not consider the G1 (first period grade) and G2 (second period grade), because the three indicators are the same type, and the prediction for final grade with G3 should ignore the previous grade.

Figure 4 shows the correlation analysis results between continuous indicators and students' final grades (G3) and correlation analysis results between continuous indicators in the data set (b), the color of the lines pointing to different indicators represents the correlation, and the thickness of the lines pointing to the correlation size. As can be seen from the figure, the correlation between G3 indicators and failures, Medu indicators is strong, and the correlation with free time and absences indicators is very small. In addition to the strong correlation of each indicator with itself, the Dalc indicator has a strong correlation with the Walc indicator and the Medu indicator has a strong correlation with the Fedu indicator.

Figure 5 shows the correlation size analysis of different indicators, and the area of the circle represents the correlation between a certain indicator and other indicators. From the graph, we can see that the area of health, absences, and famrel indicators is small, which means that the correlation between these indicators and other indicators is weak, and the area of failures and Dalc and Walc indicators is large,

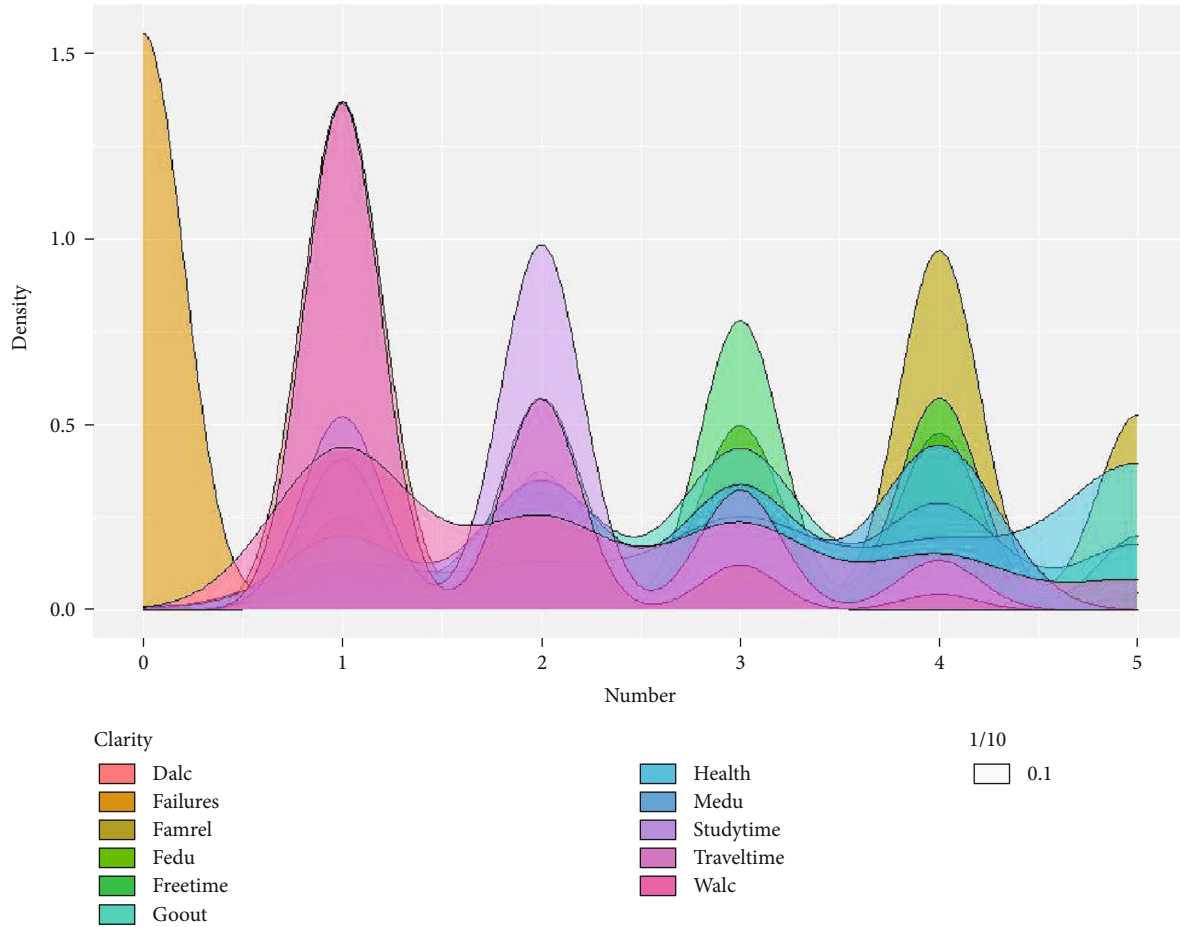


FIGURE 7: Density distribution of the indicators.

which means that these three indicators have a strong correlation with other indicators.

Figure 6 shows the box plots of the distribution of the 01 indicator, depicting the distribution states of the indicators activities, fa0up, higher, internet, nursery, paid, romantic, and schoolsup, respectively. Among them, the final grade with no extra educational support (schoolsup) is better than yes, it may be because the extra educational support is special for the students who are not good in any aspect. Higher (wants to take higher education) is very important for student grade, and it is the most obvious factor between yes and no in this data set; this shows that the internal factors of students are the most important in some situations, and the internal influence and motivation will be the important factors affecting students' performance, that means, based on their own motivation and initiative to make progress, it will be very helpful to improve their academic performance.

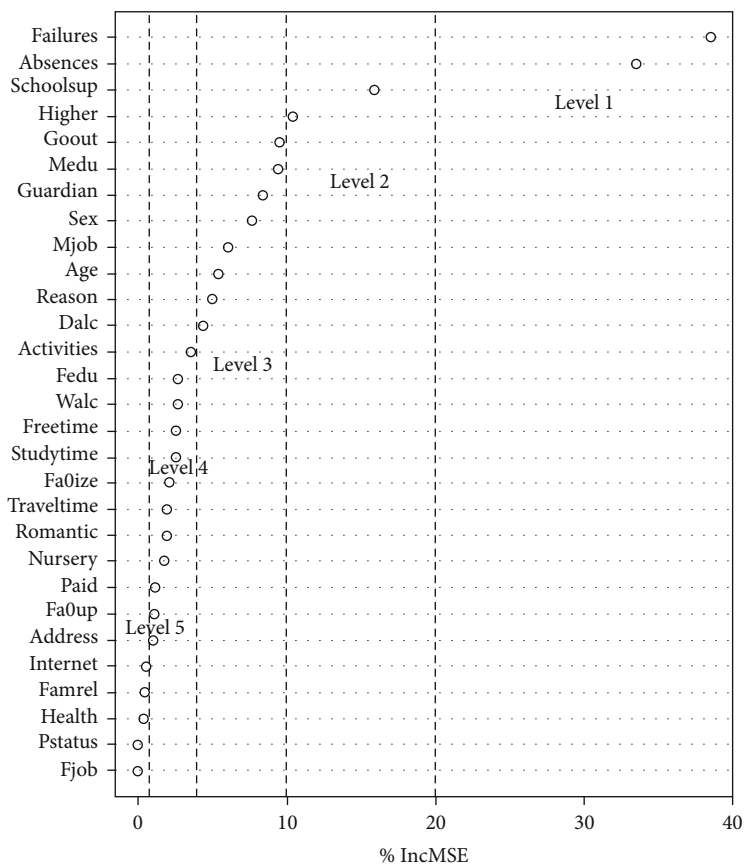
Figure 7 shows the density distribution of different continuous indicators. failures (number of past class failures) is mainly between 0 and 1. famrel (quality of family relationships) is mainly between 4 and 5, which means that most people have good family relationships. studytime (weekly study time) is mainly between 1 and 2, which means that most students spend no more than five hours a week study-

ing. traveltime (home to school travel time) is mainly around 1, which means that most students spend no more than 30 minutes traveling.

4. Feature Selection

Figure 8 shows the index importance ranking chart calculated by two different ways according to the random forest model. We roughly divide it into five levels according to the importance value (the greater the value, the higher the importance) and distribution trend of different indicators. We mark it with dotted lines and notes in the figure, and failures, goout, Medu, and absences are the most important indicators to the students' final grade; the value of Pstatus and internet is very small; it means that parent's cohabitation status and Internet access at home are not very important for the student performance.

Figure 9 shows the variation of the error of the random forest model with the increase of the number of trees in the random forest model. It can be seen from the figure that the error of the regression model decreases rapidly at the beginning. When the number of trees is 30, the error of the model gradually stabilizes, and then gradually converges to 15. This shows that our model is effective in the regression of students' achievement.



(a)

FIGURE 8: Continued.

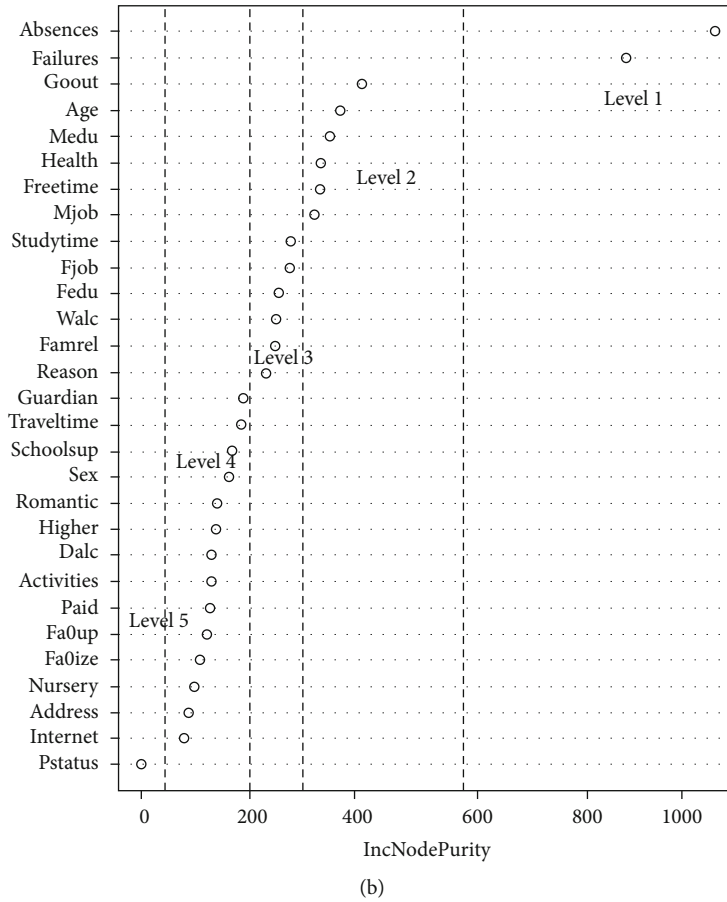


FIGURE 8: Index importance ranking.

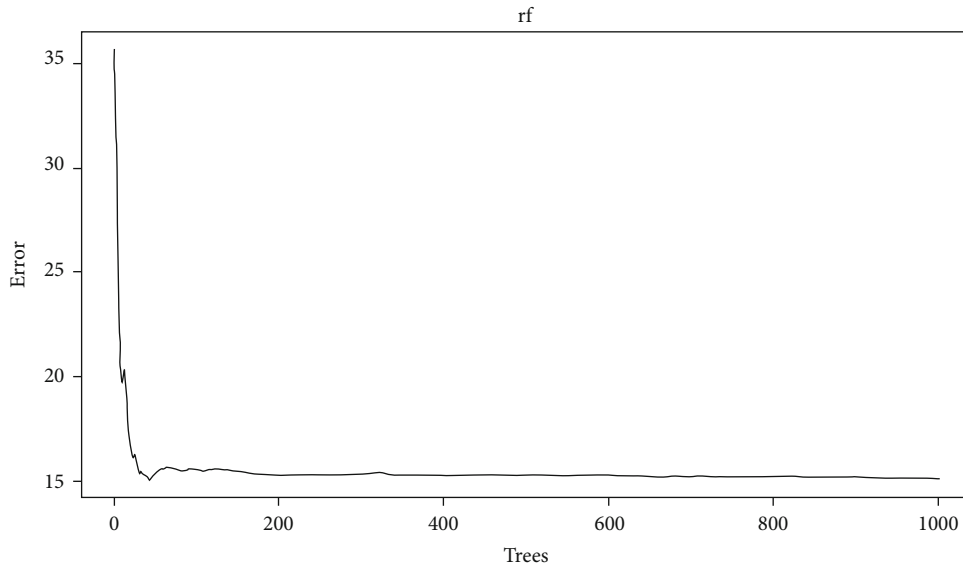


FIGURE 9: The error of random forest model.

5. Result and Analysis

The final input of our model will depend on the screening of indicators by the random forest model, and the importance

of indicators will be an important reference basis for the selection of indicators in this paper. There are some differences in the importance of indicators in the two different calculation methods, for example, some indicators are

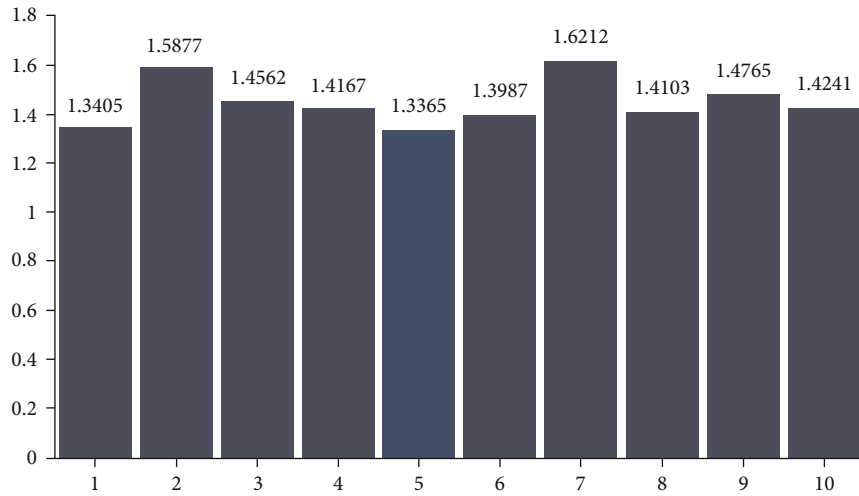


FIGURE 10: Davies-Bouldin index for different clustering schemes in the Portuguese student learning data set.

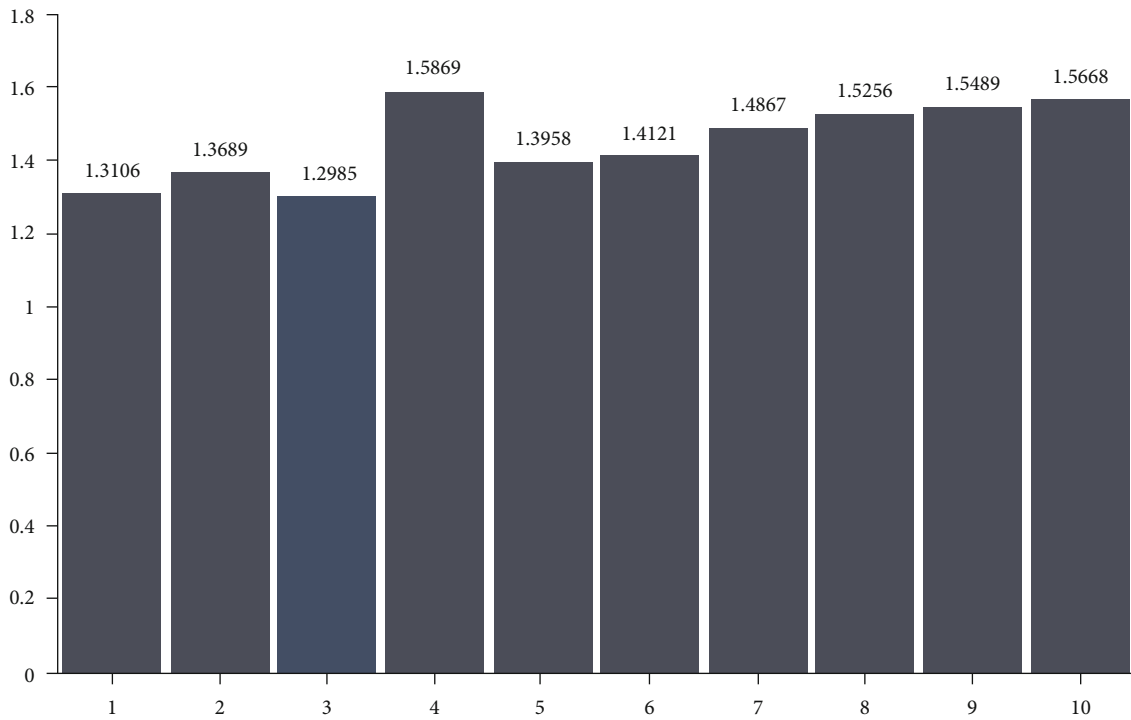


FIGURE 11: Davies-Bouldin index of different clustering schemes in the mathematics student learning data set.

TABLE 1: The result of Adaptive-K-means in Portuguese student learning data set.

Portuguese student learning data set	Mean of squared residuals	Adaptive-K-means ($K=5$) Total mean of squared residuals:15.41
	15.42 (clustered set 1)	
	15.88 (clustered set 2)	
	15.01 (clustered set 3)	
	15.18 (clustered set 4)	
	15.24 (clustered set 5)	
	15.28	Adaptive-K-means ($K=1$)

TABLE 2: The result of Adaptive-K-means in mathematics student learning data set.

	Mean of squared residuals	
Mathematics student learning data set	15.32 (clustered set 1)	Adaptive-K-means ($K=3$) Total mean of squared residuals:15.2
	15.1 (clustered set 2)	
	15.12 (clustered set 3)	
	15.61	Adaptive-K-means ($K=1$)

TABLE 3: The result of Ablation experiment in Portuguese student learning data set.

Method	Mean of squared residuals
DNN	15.61
Feature selection-DNN	15.58
Adaptive-K-means-DNN	15.43
Feature selection-Adaptive-K-means-DNN	15.41
RNN	15.71
RF	15.82

TABLE 4: The result of Ablation experiment in mathematics student learning data set.

Method	Mean of squared residuals
DNN	15.79
Feature selection-DNN	15.77
Adaptive-K-means-DNN	15.54
Feature selection-Adaptive-K-means-DNN	15.2
RNN	15.45
RF	15.53

ranked higher in importance in the first calculation method, but not very high in importance in the second calculation method, but in general, most indicators show roughly the same distribution in the two different indicator importance scores in both calculation methods. Therefore, we set all indicators within LEVEL3, LEVEL4, and LEVEL5 as the final input indicators of our model (including the different indicators that appear in both calculation methods in index importance ranking with random forest feature selection) in order to ensure the comprehensiveness of our experiments regarding the selection of indicators. In this experiment, a K-means algorithm based on the Davies-Bouldin index was used to cluster the two educational data sets in order to build a performance prediction model for two different educational data sets, the Portuguese student learning data set and the mathematics student learning data set, respectively. The data sets are the Portuguese language student learning data set and the mathematics student learning data set. The results of the Davies-Bouldin indices for the different clustering schemes are detailed in Figure 10 (Portuguese student learning data set) and Figure 11 (mathematics student learning data set).

From Figure 10, we can see that the minimum Davies-Bouldin index for clustering in the Portuguese student learning data set is 1.3365, which indicates that the best clustering number is when the clustering category is 5. The Davies-Bouldin index is 1.3405, which is very close to the minimum Davies-Bouldin index, so we use both clustering results as input to the model for the final comparison experiments.

From Figure 11, we can see that the Davies-Bouldin index with the smallest clustering in the mathematics student learning data set is 1.2985, which indicates that the best clustering number is when the clustering category is 3. Therefore, in this paper, we choose a clustering number of 3 as the data input for the prediction model constructed from the mathematics student learning data set.

Tables 1 and 2 show the result of Adaptive-K-means in Portuguese and mathematics student learning data set, respectively. In this experiment, a K-means algorithm based on the Davies-Bouldin index was used to cluster the two educational data sets in order to build a performance prediction model for two different educational data sets. In order to further determine the number of selected clusters, we took several clustering results with smaller Davies-Bouldin index as the final input data of the model for training and prediction. Detailed results are listed in Tables 1 and 2, from which we can find that, for mathematics student learning data set, when the number of clusters is three, the result is better.

Table 3 shows the result of Ablation experiment in Portuguese student learning data set; in order to verify the effectiveness of the proposed model in this paper, different benchmark models were selected for comparison, including common DNN model, standard RNN model, and standard RF model. In addition, to verify the effectiveness of the feature selection method, this paper also added the standard DNN model after feature selection to compare the original DNN model, and the standard Adaptive-K-means-DNN model and the Adaptive-K-means-DNN model with feature selection. From the experiment results shown in the table, we can obtain the following findings: (1) The mean of squared residuals for DNN-based model performs better than others (15.61 for DNN, 15.71 for RNN, and 15.82 for RF). (2) The proposed feature selection Adaptive-K-means-DNN method performs best under all the based model. (3) The result of the feature selection DNN (15.58) is better than single DNN (15.61), and the result of the feature selection Adaptive-K-means-DNN (15.41) is also better than Adaptive-K-means-DNN (15.43), indicating that the feature selection processing could be employed as a useful method for Portuguese student grade prediction.

Table 4 shows the result of Ablation experiment in mathematics student learning data set; from the mean of

squared residuals results for mathematics displayed in Table 4, we get the similar conclusions with Table 3; the proposed model feature selection Adaptive-K-means-DNN obtains the best mean of squared residuals both in mathematics and Portuguese student learning data sets, and the feature selection processing is important and effective; the results demonstrating that the feature selection processing has enhanced the prediction performances significantly. In addition, both in Table 3 and Table 4, among the benchmark methods for mean of squared residuals results, RNN (15.45 of mathematics and 15.71 of Portuguese) performs better than RF (15.53 of mathematics and 15.82 of Portuguese).

6. Conclusion

Two public student learning data sets as mathematics and Portuguese are collected in this paper. We conduct the proposed students' grade prediction by integrating DNN and K-means algorithm. First, we make a feature selection through two computing methods based on random forest algorithm, and then, we predicted the student performance through an improved model based on K-means and Deep Neural Network. The proposed model Adaptive-K-means-DNN obtains the best mean of squared residuals both in two student learning data sets and outperforms other benchmark methods; the results demonstrating that the feature selection processing and the employment of the adaptive K-means have enhanced the prediction performances significantly.

In this paper, our research is not limited to the traditional prediction of student performance; we have mined different aspects based on a publicly available student data set, and have done our best to visualize our results; we have also tried to explain the physical aspects of our mining findings. Our data mining experiments show that, after students maintain their study habits, their performance will show some continuity unless other factors intervene; failures, goout, Medu, and absences are the most important indicators to the students' final grade. Among it, mother's education is the direct influence of the students' final grade; the difference is there is no strong correlation between absences and the students' final grade, but absences is still an important factor affecting the students' final grade, and it may affect the students' final grade through other factors; moreover, the failures also has a strong correlation between the students' final grade, and it also is a direct influence indicator for the students' final grade; in summary, we can find that (1) the influence of mother's education on students' performance is very important. (2) The management of students' classroom absence will effectively improve students' final performance. (3) Regular encouragement may be a good way for students to improve their learning performance. (4) The internal influence and based on their own motivation and initiative to make progress will be the important factors affecting students' performance.

In the future, we can continue to mine the student data set through different mining methods. For the continuous distribution of student grades, we can also convert the continuous values of student grades into a 01-distribution data set by manually setting the pass rate, merit rate, and other conditions, and then perform a dichotomous study of student pass rates. On the

other hand, we will continue our research by extending it to additional data sets, including some online course student data sets, or some private data sets that we will collect and build in the future, and expect to find more findings.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Acknowledgments

This work is supported by the National Natural Science Foundation of China "Research on the Construction and Application Strategy of New Training Models for Rural Teachers Supported by Artificial Intelligence"[Grant Number: 72164004], the University-level Doctor Starting Fund project "Construction and Application of Internet Smart Campus based on Big Data" (Grant Number: 0517041/11904), and the Philosophy and Social Science Planning Topic of Guizhou Province "Construction Research on Rural Teachers Team of Guizhou Province" [Grant Number: 21GZYB54].

References

- [1] J. R. Anderson, C. F. Boyle, A. T. Corbett, and M. W. Lewis, "Cognitive modeling and intelligent tutoring," *Artificial Intelligence*, vol. 42, no. 1, pp. 7–49, 1990.
- [2] X. Chen, H. Xie, D. Zou, and G. J. Hwang, "Application and Theory Gaps During the Rise of Artificial Intelligence in Education," *Computers and Education: Artificial Intelligence*, vol. 1, article 100002, 2020.
- [3] W. Holmes, M. Bialik, and C. Fadel, *Encyclopedia of Education and Information Technologies*, Springer International Publishing, Cham, 2019.
- [4] G.-J. Hwang, H. Xie, B. W. Wah, and D. Gašević, "Vision, challenges, roles and research issues of Artificial Intelligence in Education," *Computers and Education: Artificial Intelligence*, vol. 1, article 100001, 2020.
- [5] W. Holmes, S. Anastopoulou, H. Schaumburg, and M. Mavrikis, "Technology-enhanced personalised learning: Untangling the evidence," in Robert Bosch Stiftung GmbH, 2018.
- [6] A. I. Starčič, "Human learning and learning analytics in the age of artificial intelligence," *British Journal of Educational Technology*, vol. 50, no. 6, pp. 2974–2976, 2019.
- [7] M. Rowe, "Shaping our algorithms before they shape us," in *Artificial Intelligence and Inclusive Education*, pp. 151–163, Springer, Singapore, 2019.
- [8] B. F. Skinner, *Science and Human Behavior*, Macmillan, 1953.
- [9] J. G. Greeno, A. M. Collins, and L. B. Resnick, "Cognition and learning," in *Handbook of Educational Psychology*, D. Berliner and R. Calfee, Eds., New York, MacMillan, 1996.

- [10] M. Schommer, "Effects of beliefs about the nature of knowledge on comprehension," *Journal of Educational Psychology*, vol. 82, no. 3, pp. 498–504, 1990.
- [11] B. F. Skinner, "Teaching machines," *Science*, vol. 128, no. 3330, pp. 969–977, 1958.
- [12] J. K. Burton, D. M. Moore, and S. G. Magliaro, "Behaviorism and instructional technology," in *Handbook of Research on Educational Communications & Technology*, Lawrence Erlbaum Associates Publishers, 2004.
- [13] T. Koschmann, "Paradigm shifts and instructional technology: An introduction," *CACL: Theory and practice of an emerging paradigm*, vol. 116, pp. 1–23, 1996.
- [14] A. Bandura and N. Cliffs, *Social Foundations of Thought and Action: Cognitive Theory*, Prentice-Hall, 1987.
- [15] C. H. Liu and R. Matthews, "Vygotsky's philosophy: constructivism and its criticisms examined," *International Education Journal*, vol. 6, no. 3, pp. 386–399, 2005.
- [16] L. S. Vygotsky and M. Cole, "Mind in society: Development of higher psychological processes," in Harvard University Press, 1978.
- [17] B. D. Boulay, "Can we learn from ITSs," in vol. 1839 of *Lecture Notes in Computer Science*, pp. 9–17, Springer, 2000.
- [18] C. P. Rosé, E. A. McLaughlin, R. Liu, and K. R. Koedinger, "Explanatory learner models: why machine learning (alone) is not the answer," *British Journal of Educational Technology*, vol. 50, no. 6, pp. 2943–2958, 2019.
- [19] M. Mason, "What is complexity theory and what are its implications for educational change," in *Complexity Theory and the Philosophy of Education*, pp. 32–45, Wiley-Blackwell, Oxford, UK, 2008.
- [20] M. O. Riedl, "Human-centered artificial intelligence and machine learning," *Human Behavior & Emerging Technologies*, vol. 1, no. 1, pp. 33–36, 2019.
- [21] H. Jean-Michel, "From human – machine interaction to human – machine cooperation," *Ergonomics*, vol. 43, no. 7, pp. 833–843, 2000.
- [22] S. Yang, H. Ogata, T. Matsui, and N. S. Chen, "Human-centered artificial intelligence in education: seeing the invisible through the visible," *Computers and Education: Artificial Intelligence*, vol. 2, article 100008, 2021.
- [23] I. Roll and R. Wylie, "Evolution and revolution in artificial intelligence in education," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 582–599, 2016.
- [24] S. Russell, "Artificial intelligence: the future is superintelligent," *Nature*, vol. 548, no. 7669, pp. 520–521, 2017.
- [25] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in *Proceedings of 5th Annual Future Business Technology Conference, Porto, 2008*, pp. 5–12, Porto, Portugal, 2008.