Predicting Learning Achievement through Self-Regulated Learning Strategies, Motivation, and Programming Behaviors

Pei-Xuan Wang [1] and Ting-Chia Hsu [1]

[1] *Department of Technology Application and Human Resource, National Taiwan Normal University, Taiwan*

Abstract

With the growth of digital learning, there has been an increase in research related to learning analytics. Learning analytics can be used to identify potential problems and improve the quality of education by measuring, collecting, analyzing, and reporting data about the learners and their background to understand the learner's learning situation and learning environment. In addition, further analysis of students' learning behaviors can be used to provide adaptive and personalized teaching suggestions. This study aimed to analyze operational data from the coding process on an online learning platform, as well as data obtained from students' self-regulated learning strategy scales and self-regulated learning motivation scales. The objective was to investigate the correlation of these factors with students' academic achievement and to determine whether these factors can be utilized to predict student learning outcomes. The results of the study revealed a significant correlation between programming behavior and student grades. Variances in self-regulated learning strategies and motivation levels exhibit notable differences in academic performance. When incorporating these performance-related values as features in the development of predictors, it proves effective in forecasting students' learning outcomes. However, due to the limited sample size in this study, the predictive model may experience reduced accuracy or overfitting issues when applied to larger datasets. Therefore, for future predictions with larger samples, considerations should be made to adjusting model hyperparameters or modifying the features used to improve the accuracy of the predictions.

Keywords

SRL Strategy, SRL Motivation, Random Forest, Learning Achievement

1. Introduction

In recent times, numerous online learning platforms such as Moodle, Tronclass, 1know, and Bookroll have emerged, and the outbreak of the pandemic accelerated the development and demand for these platforms. A substantial number of users engage in learning activities on these platforms, generating vast amounts of learning-related data available for scholars to conduct relevant research.

The growing body of research on self-regulated learning in recent years indicates a positive impact on academic performance (Rosen et al., 2022). Self-regulated learning is particularly crucial in online learning environments, showing correlation with academic success (Zhang, Maeda, Newby, Cheng & Xu, 2023).

With the exponential growth of technological advancements, proficiency in programming skills has become increasingly important for students. Computational thinking is recognized as a vital skill for successful adaptation to the future (Hsu, Chang & Hung, 2018) . In many Taiwanese universities, information literacy is set as a graduation requirement, with computer science fundamentals being a mandatory course.

This study aims to analyze operational data from the coding process on online learning platforms and data obtained from students' self-regulated learning strategy scales and self-regulated learning motivation scales. The objective is to understand the correlation of these factors with students' academic achievement, and to determine whether these factors can be utilized to predict student learning

LAK-WS 2024: Joint Proceedings of LAK 2024 Workshops, March 18–19, Kyoto, Japan

outcomes, providing insights for educators. Therefore, this study posed the following three research questions:

1. Is programming behavior correlated with academic performance?
2. Do students with different strategies and motivations exhibit significant differences in learning achievement?
3. Can scores from strategies and motivations, along with programming behavior, be used to predict student grades?

2. Literature Review

This section will sequentially explore the significance of self-regulated learning abilities in online learning environments, followed by an examination of the impact of self-regulated learning strategies on learning and the influence of self-regulated learning motivation on the learning process.

2.1 Self-regulated learning in online environments

Self-regulated learning (SRL) is defined as "a positive, constructive process in which learners set goals for their learning, then attempt to monitor, regulate, and control their cognitive, motivational, and behavioral processes, guided and constrained by their goals and the context" (Pintrich, 2000). Following the outbreak of COVID-19, the continuous emergence of learning platforms has made self-regulated learning a key factor in students' online learning achievements, showing correlation with academic success (Zhang et al., 2023) . Chen and Li (2021) explored the types of self-regulated learners in an asynchronous online chemistry course for university students, revealing that the high self-regulated learning group demonstrated higher academic achievement compared to the low self-regulated learning group. Previous research indicated that self-regulated learning abilities play a crucial role in achieving high performance in online learning environments. Successful outcomes in online settings often require strong self-regulated learning abilities.

2.2 Self-regulated learning strategy

The cognitive and metacognitive aspects of SRL are considered integral "skills," comprising cognitive strategies, metacognitive strategies, and resource management strategies to support students in regulating their own learning (Pintrich, 2004). Different self-regulated learning strategies may lead to success in various programming stages and situations, with academically successful students exhibiting distinct self-regulated learning strategies from their peers (Cheng, Zou, Xie & Wang, 2024). Kalu, Wolsey and Enghiad (2023) the role of active learning strategies in fostering foundational knowledge in the taught subject, emphasizing their contribution to achieving greater success. Cheraghbeigi, Molavynejad, Rokhafroz, Elahi and Rezaei (2023) identified various student-centered strategies used during the COVID-19 pandemic to enhance digital learning, indicating the importance of self-regulated learning strategies in facilitating digital learning. An, Oh and Park (2022)pointed out a positive correlation between autonomous learning strategies, academic success, and digital learning acceptance. From past research, it is evident that self-regulated learning strategies are correlated with learning achievement.

2.3 Self-regulated learning motivation

Student motivation and achievement are core components of academic success, measured through academic self-concept and curiosity, where higher curiosity is associated with better performance (Wild & Neef, 2023) . Yang, Lian and Zhao (2023) found that different motivations lead to varied research outcomes, emphasizing the positive correlation between motivation and learning achievement. Chen, Su, Lin and Sun (2023) revealed that the motivation level of the self-regulated group surpassed that of the guided learning group, with students displaying high self-regulated learning motivation achieving better grades than their guided learning counterparts who did not display high self-regulated learning motivation. In summary, motivation is found to be correlated with achievement based on comprehensive findings from past studies.

3. Methods

This section will sequentially outline the research process, the algorithmic tools employed in the study, and the methods used to process and analyze the collected data.

3.1 Research Process

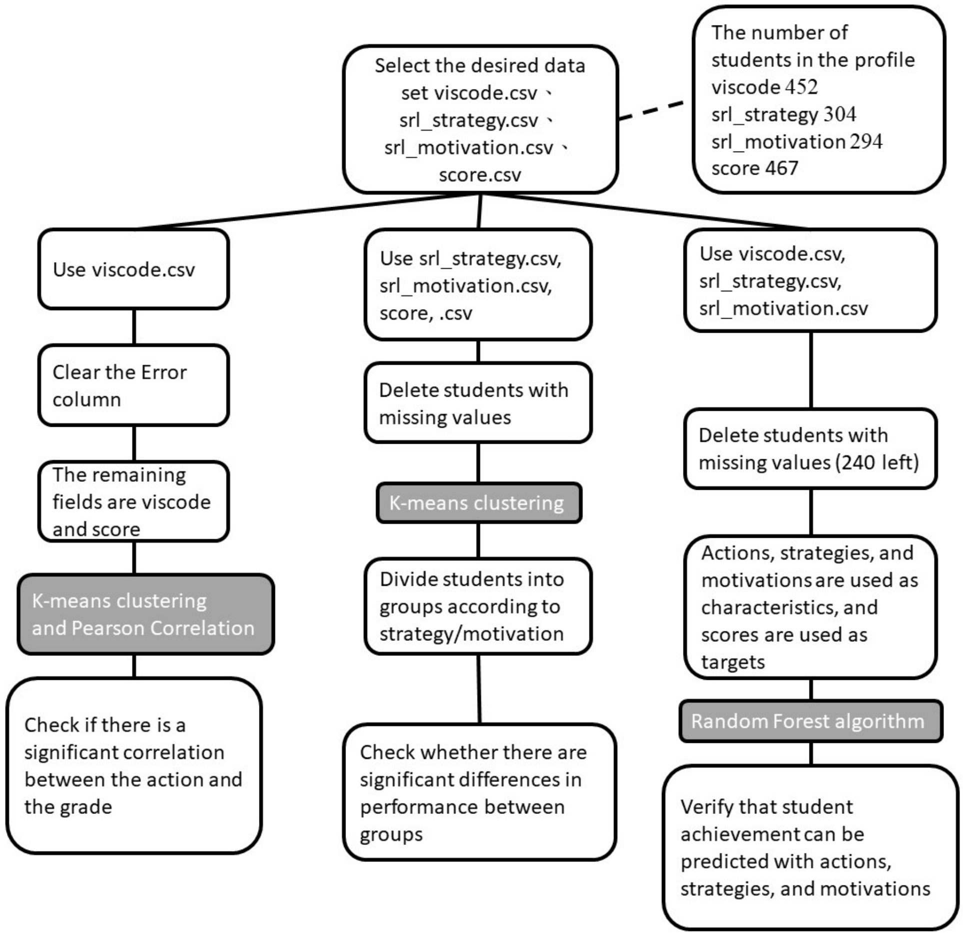

Figure 1: Research Process diagram

The study utilized datasets provided in the competition, including viscode.csv, srl_strategy.csv, srl_motivation.csv, and score.csv. The research process was divided into three distinct phases to address different research questions. For Research Question 1, only viscode.csv (excluding error-related fields) was used for correlation testing with scores. Research Question 2 involved the use of srl_strategy.csv, srl_motivation.csv, and score.csv, examining whether different strategy and motivation groups exhibited significant differences in scores. For Research Question 3, viscode.csv, srl_strategy.csv, srl_motivation.csv, and the final scores from viscode.csv were used. All strategy and motivation items, along with operational behaviors from Research Question 1, were input into a Random Forest model for training to predict student scores.

3.2 Research Tools

3.2.1 Random Forests

Random Forest is an ensemble of decision tree predictors, where each tree depends on independently sampled random vectors, and all trees in the forest share the same distribution. The generalization error of the forest converges as the number of trees in the forest increases. The generalization error of tree classifiers in the forest depends on the strength of individual trees and their correlation (Breiman, 2001).

Random Forest classifiers effectively handle high-dimensional data and multicollinearity, being both fast and insensitive to overfitting (Belgiu & Drăguţ, 2016) .

3.2.2 K-means Cluster

K-means cluster analysis (MacQueen, 1967) has found widespread application in the study of learning behaviors due to its visual interpretability and ease of use. K-means initially selects K nodes randomly as centroids, assigns each new node freely to one of the K categories, calculates the averages for each category, and then reclassifies nodes based on the nearest category, iterating until node distances are minimized and categories stabilize (Moubayed, Injadat, Shami & Lutfiyya, 2020).

3.3 Data processing and analysis

A total of 452 student records were received for viscode, with 304 valid self-regulated learning motivation scale responses and 294 valid self-regulated learning strategy scale responses. Error-related fields were removed from viscode data, leaving key operational behavior features such as code_copy, code_execution, code_paste, code_speed, notebook_open, tree_open, codeLength, Viscode-login_times, Viscode-execute_times, and Viscode-open_file_times. Pearson correlation tests were conducted between all operational features and scores to confirm positive correlations. For subsequent research questions, the optimal number of clusters determined by silhouette coefficient was used for K-means clustering. Missing values were imputed with mean values, and K-means clustering was performed. The results of strategy and motivation clustering were then merged with scores for independent sample *t* tests to analyze significant differences between groups. Once all previous research questions were addressed, all relevant features associated with scores were input into a Random Forest-based predictive model to assess its ability to predict student scores.

4. Results

4.1 Correlation between Programming Behaviors and Academic Performance

To verify Research Question 1, Python was utilized to conduct Pearson correlation coefficient tests between operational behaviors, excluding error-related fields, in the Viscode dataset and academic performance. The test results revealed a significant relationship between the sum of the number of programming behaviors and academic scores in Viscode ($r = 0.14$, $p < .01$). Most individual behaviors also exhibited significant correlations, as shown in Table 1. The only behavior that was not significant was codeLength ($r = 0.0862$, $p > .05$). It is hypothesized that this may be attributed to students with better programming skills employing more efficient functions, reducing unnecessary steps and resulting in shorter code.

Table 1

Correlation Coefficients of Individual Behaviors with Academic Achievement

Behavior	*r*	*p*
code_copy	0.3216^{***}	<.001
code_execution	0.3215^{***}	<.001
code_paste	0.3324^{***}	<.001
notebook_open	0.3323^{***}	<.001
tree_open	0.2226^{***}	<.001
codeLength	0.2782	>.05
Viscode-login_times	0.0862^{***}	<.001
Viscode-execute_times	0.3534^{***}	<.001
Viscode-open_file_times	0.2226^{***}	<.001

$^{***}p < .001$

4.2 Significant Differences in Academic Achievement among Students with Different Strategies and Motivations

To investigate Research Question 2, all strategy and motivation items from srl_strategy and srl_motivation were input into Python, and the optimal clustering results, determined by silhouette coefficient, indicated two distinct groups for both strategy and motivation. As illustrated in Figure 1 and Figure 2, the independent sample *t*-test results, depicted in Figure 3 and Figure 4, clearly show significant differences in academic achievement among students with different strategies (t = -4.33, p < .05) and among students with different motivations (t = 2.34, p < .001).

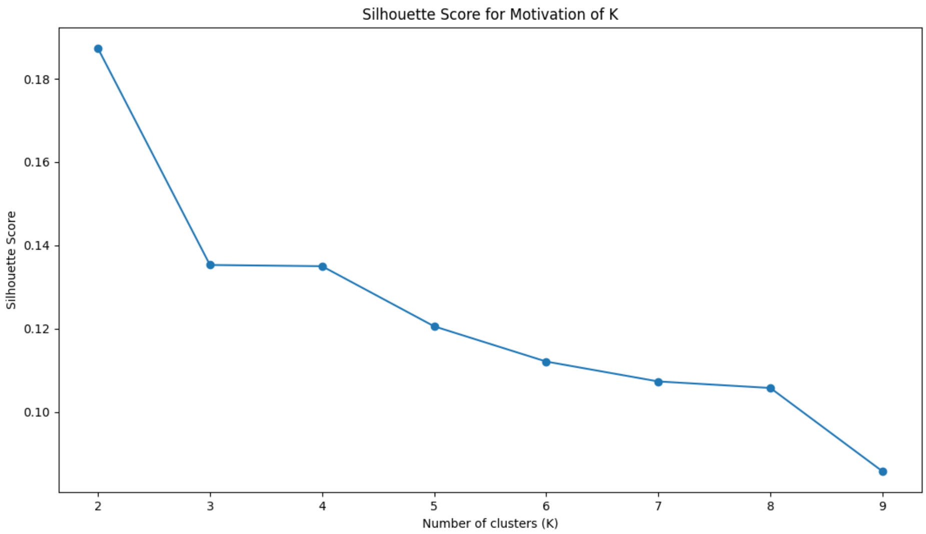

Figure 2: Silhouette Score of Motivation of K

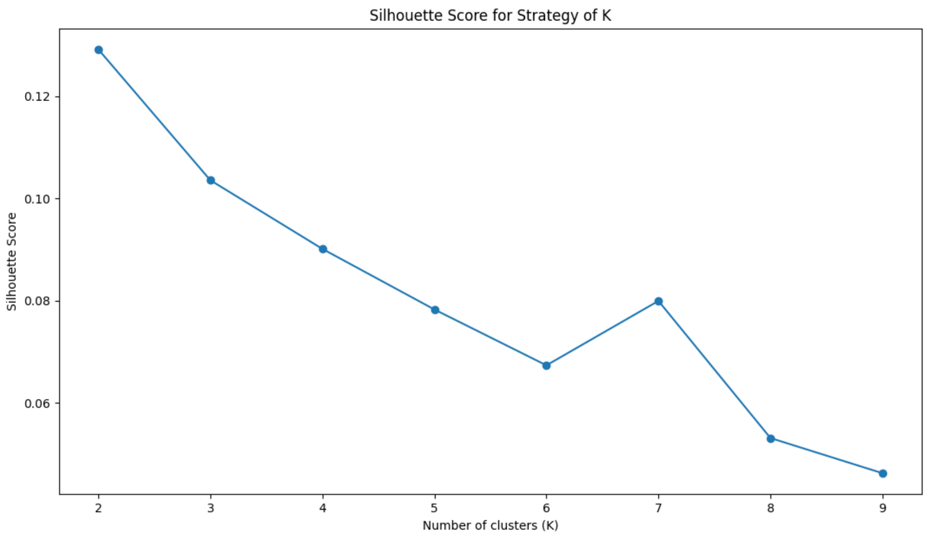

Figure 3: Silhouette Score of Strategy of K

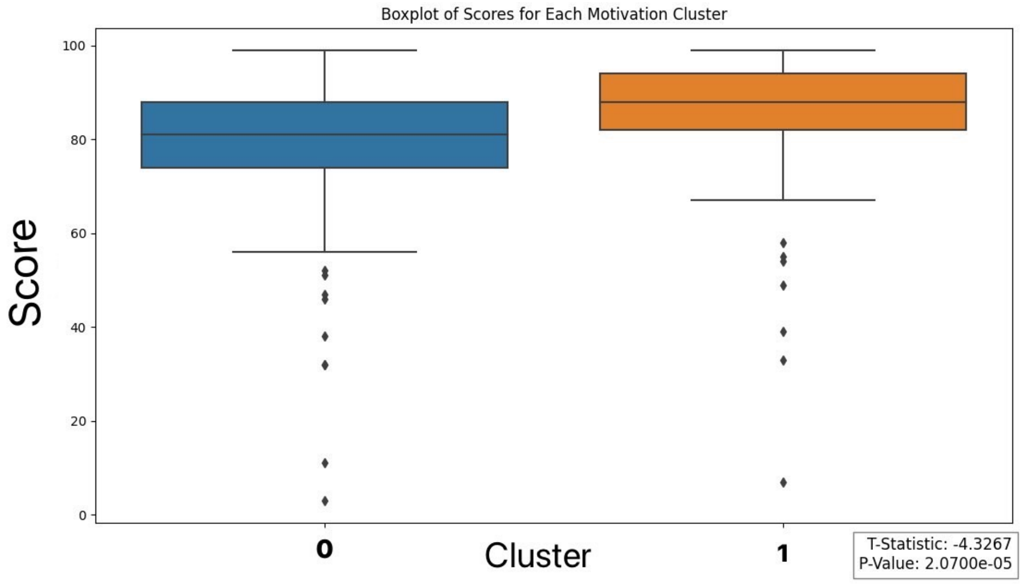

Figure 4: Motivation of *t*-test result

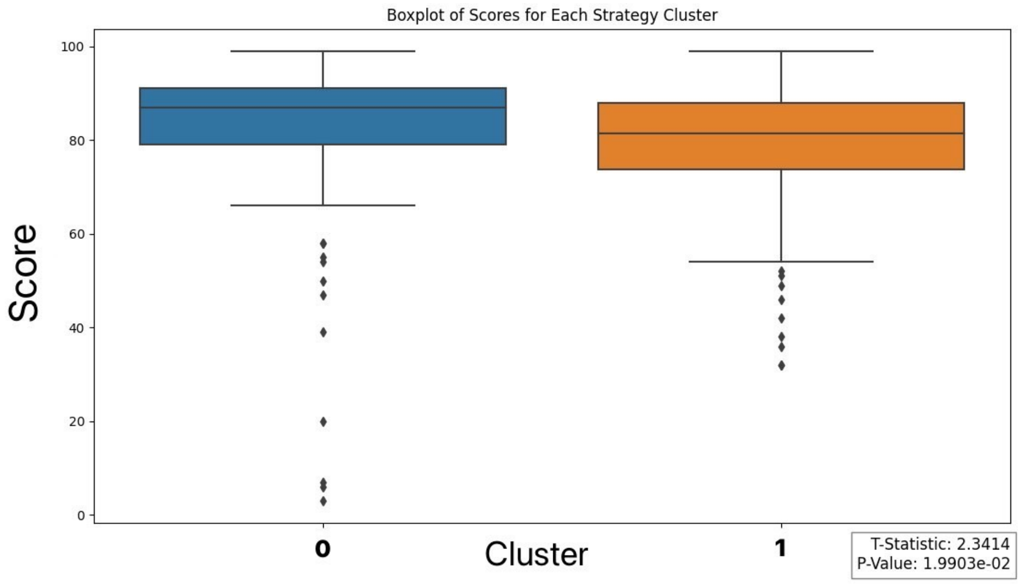

Figure 5: Strategy of *t*-test result

4.3 Predicting Student Grades using Strategy, Motivation Scores, and Programming Behaviors

To address Research Question 3, all previously used programming behavior, self-regulated learning strategy, and self-regulated learning motivation fields were input into Python. Missing values were removed, and data were merged based on user IDs, resulting in a final dataset of 240 students. The academic scores were categorized into four intervals (0~25, 26~50, 51~75, 76~100) for prediction. The Random Forest algorithm was employed as the basis for the predictive model, achieving a high accuracy of 0.795. The model demonstrates the feasibility of using programming behaviors, self-regulated learning strategies, and self-regulated learning motivations to predict students' academic performance. This approach aids in identifying high-risk students who may struggle to pass the course and require intervention from teachers. As shown in Figure 6, the decision tree from the Random Forest model provides insight into the classification process, showcasing how student behaviors contribute to predicting the potential score range, e.g., if there is a student who has a self-regulated learning strategy score less than equal to 2.5, and then further down the line there may be a code_paste behavior less than equal to 327.5, and then next to that the self-regulated learning motivation score is less than equal to 2.5, then the student's final grade is likely to fall in the 76~100 range. Therefore, Research Question 3 is confirmed: it is possible to predict student grades using strategy and motivation scores along with programming behaviors, providing valuable insights for early intervention and support.

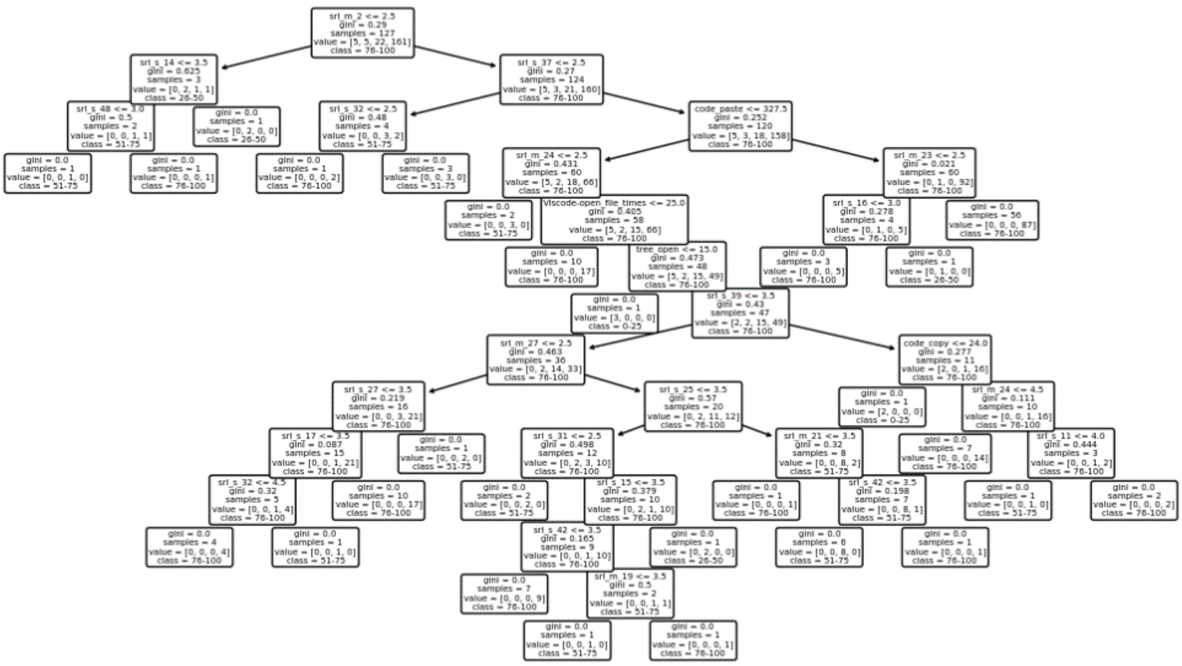

Figure 6: A Random Forest tree

5. Discussion and Conclusion

Based on the research results, we confirmed significant differences in programming behaviors among students with different grades in coding activities, with the only non-significant behavior being "codeLength." This lack of significance may be attributed to more proficient students using functions more effectively, resulting in shorter, more efficient code.

Both the level of self-regulated learning strategy and motivation significantly influenced academic performance. Significant differences in grades were observed for different strategy and motivation scores. Higher levels of self-regulated learning strategy and motivation correspond to better learning outcomes. The potential interplay between motivation and strategy, whether they mutually influence each other, remains a topic for further investigation.

In conclusion, the predictor model, incorporating variables significantly correlated with grades, effectively forecasts the range in which students' academic performance is likely to fall. Early identification of students who may encounter learning difficulties and require intervention is possible. However, due to the limited sample size in this study (240 students meeting the criteria), it is acknowledged that the predictive model's accuracy might decrease with a larger sample size. Therefore, future endeavors should consider adjusting model hyperparameters or modifying features when employing a larger sample size. Including error types encountered by students in the feature set for model training could be explored as a potential enhancement, which might increase the accuracy of the predictions. Random forest is also an easy-to-use prediction algorithm, and it should be very feasible to use this method in other disciplines, and it can be adapted to different disciplines by putting the platform operation behavior of the discipline as a feature for training.

6. Acknowledgements

This paper is partially supported by the research project of National Science and Technology Council in Taiwan under the contract number NSTC 112-2628-H-003-007-.

7. References

[1] J. An, J. Oh, K. Park, Self-regulated learning strategies for nursing students: A pilot randomized controlled trial, International Journal of Environmental Research and Public Health 19(15) 9058 (2022).

[2] M. Belgiu, L. Drăguţ, Random forest in remote sensing: A review of applications and future directions, ISPRS journal of photogrammetry and remote sensing 114 (2016) 24-31. doi: 10.1016/j.isprsjprs.2016.01.011.

[3] L. Breiman, Random forests, Machine learning 45 (2001) 5-32. doi:10.1023/A:1010933404324

[4] C. Y. Chen, S. W. Su, Y. Z. Lin, C. T. Sun, The Effect of Time Management and Help-Seeking in Self-Regulation-Based Computational Thinking Learning in Taiwanese Primary School Students, Sustainability 15(16) 12494 (2023). doi:10.3390/su151612494

[5] K. Z. Chen, S. C. Li, Sequential, typological, and academic dynamics of self-regulated learners: Learning analytics of an undergraduate chemistry online course, Computers and Education: Artificial Intelligence 2 100024 (2021). doi:10.1016/j.caeai.2021.100024

[6] G. Cheng, D. Zou, H. Xie, F. L. Wang, Exploring differences in self-regulated learning strategy use between high-and low-performing students in introductory programming: An analysis of eye-tracking and retrospective think-aloud data from program comprehension, Computers & Education 208 104948 (2024). doi:10.1016/j.compedu.2023.104948

[7] N. Cheraghbeigi, S. Molavynejad, D. Rokhafroz, N. Elahi, E. Rezaei, Nursing students' learning strategies for e-learning during the Covid-19 pandemic in Iran: a qualitative study, BMC Medical Education 23.1 (2023) 1-11. doi: 10.1186/s12909-023-04270-9

[8] Flanagan, B., Ogata, H. (2018). Learning Analytics Platform in Higher Education in Japan, Knowledge Management & E-Learning (KM&EL), Vol.10, No.4, pp.469-484.

[9] T. C. Hsu, S. C. Chang, Y. T. Hung, How to learn and how to teach computational thinking: Suggestions based on a review of the literature, Computers & Education 126 (2018) 296-310. doi:10.1016/j.compedu.2018.07.004

[10] F. Kalu, C. Wolsey, P. Enghiad, Undergraduate nursing students' perceptions of active learning strategies: A focus group study, Nurse Education Today 131 105986 (2023). doi: 10.1016/j.nedt.2023.105986.

[11] Lu, O.H., Huang, A.Y., Flanagan, B., Ogata, H., & Yang, S.J. (2022). A Quality Data Set for Data Challenge: Featuring 160 Students' Learning Behaviors and Learning Strategies in a Programming Course. 30th International Conference on Computers in Education Conference Proceedings, pp. 64-73.

[12] J. MacQueen, Some methods for classification and analysis of multivariate observations, In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability Vol.1, No.14, pp.281-297.

[13] A. Moubayed, M. Injadat, A. Shami, H. Lutfiyya, Student engagement level in an e-learning environment: Clustering using k-means, American Journal of Distance Education 34.2 (2020) 137-156. doi:10.1080/08923647.2020.1696140

[14] P. R. Pintrich, A conceptual framework for assessing motivation and self-regulated learning in college students, Educational psychology review 16 (2004) 385-407. doi:10.1007/s10648-004-0006-x

[15] P. R. Pintrich, The role of goal orientation in self-regulated learning, Handbook of self-regulation, Academic Press, (2000) 451-502. doi:10.1016/B978-012109890-2/50043-3

[16] J. A. Rosen, Y. Hong, S. Rutledge, M. Cannata, S. Rotermund, K. Leu, Improving SEL outcomes for US 9th graders: Results from a District Wide School Reform, Studies in Educational Evaluation 75 101212 (2022) doi:10.1016/j.stueduc.2022.101212
[17] S. Wild, N. Christoph, Analyzing the associations between motivation and academic performance via the mediator variables of specific mathematic cognitive learning strategies in different subject domains of higher education, International Journal of STEM Education 10(1) 32 (2023). doi:10.1186/s40594-023-00423-w
[18] Q. F. Yang, L. W. Lian, J. H. Zhao, Developing a gamified artificial intelligence educational robot to promote learning effectiveness and behavior in laboratory safety courses for undergraduate students, International Journal of Educational Technology in Higher Education 20(1) 18 (2023). doi:10.1186/s41239-023-00391-9
[19] Z. Zhang, Y. Maeda, T. Newby, Z. Cheng, Q. Xu, The effect of preservice teachers' ICT integration self-efficacy beliefs on their ICT competencies: The mediating role of online self-regulated learning strategies, Computers & Education 193 104673 (2023). doi:10.1016/j.compedu.2022.104673