Employing Tree-based Algorithms to Predict Students' Self-Efficacy in PISA 2018

Bin Tan
University of Alberta
btan4@ualberta.ca

Maria Cutumisu
University of Alberta
cutumisu@ualberta.ca

ABSTRACT

Self-efficacy is a critical psychological construct that has a substantial impact on students' learning experience and global well-being. Thus, the early identification of low self-efficacious learners is an important task for educators and researchers. This study uses machine learning (ML) approaches to model the self-efficacy of over 520,000 students based on their test performance and responses to survey questions in the Programme for International Student Assessment (PISA) 2018. Two tree-based ensemble learning models (random forest and XGBoost) were built using 64 predictors and evaluated using nested cross-validation with a grid search method. The results showed that, although both algorithms predicted self-efficacy accurately, XGBoost slightly outperformed Random Forest (RF). The findings also revealed that students' non-cognitive constructs such as meaning in life and the motivation for mastering tasks were the most important predictors. Theoretical contributions include the expansion of the body of literature on ML applications that predict students' self-efficacy and the potential advancement of theoretical models of self-efficacy. Practical contributions include the applications of tree-based algorithms to identify low self-efficacious individuals at scale, in a large international assessment. Implications include the development of systems that use ML algorithms to detect low self-efficacious learners and provide support for early interventions.

Keywords

Self-efficacy, large-scale assessment, PISA 2018, students' traits

1. INTRODUCTION

Self-efficacy represents individuals' general beliefs about their competencies of performing specific tasks or achieving goals [4]. Students' self-efficacy has been consistently associated with their learning achievement [3, 11, 14]. Students with higher self-efficacy, at all levels of competency, are more successful in school activities and use more effective learning strategies [21]. In addition, various empirical findings showed that self-efficacy is associated with academic engagement [25]. High self-efficacious students tend to report a higher level of academic aspirations, spend more time on homework, and gain more positive learning experiences [6]. Those students are more gratified and satisfied with their accomplishments [27]. Moreover, self-efficacy is highly linked to students' global well-being and life outcomes [11]. It has been found that students with low self-efficacy are more likely to drop out of school, which jeopardizes their future employment prospects [5]. In addition, low self-efficacious students tend to suffer from many mental and behavioral problems such as depression [2], suicidal ideation and attempts [25], social avoidance [24], and addictive behaviors [23]. Thus, students' self-efficacy is an important topic for psychological and educational research. If students' self-efficacy can be screened and predicted, practitioners may be able to deliver early intervention to help low self-efficacious students improve their learning experience, global well-being, and life outcomes.

In this present study, two decision tree-based algorithms (RF and XGboost) were trained based on over 520,000 students' test performance and responses to survey questions in PISA 2018. The proposed research questions (RQ) were:

(1) Is it possible to use the RF and XGBoost algorithms to predict students' self-efficacy with a small error rate?

(2) What are the most important predictors of self-efficacy in these models?

2. RELATED WORK

Self-efficacy has been increasingly used as a predictor in ML models. However, to date, despite the significance of self-efficacy for students' learning and life, there are very few studies that treated self-efficacy as the focal variable to be predicted. The first such study [17] used Naive Bayes and decision tree algorithms to generate two sets of classification models of self-efficacy (high vs. low). The first set of models were built based on the demographic factors of the students, whereas the other set of models added additional predictors that were obtained when students were exposed to an intelligent problem-solving tutoring system including biofeedback signals and recorded log data. The classification accuracy of the models ranged from 82.1% to 87.3%.

Later, a K-medoids clustering algorithm was employed to group similar students based on their gender, survey-reported self-efficacy, and collected natural language utterances during dialogue in an intelligent tutorial dialogue system [9]. Results revealed differences in the use of utterances between students with high and low self-efficacy. For example, students with high self-efficacy tend to use more confident utterances to express their understanding of the knowledge, compared to students with low self-efficacy who usually make less confident utterances.

Recent efforts have examined domain-specific self-efficacy. A study trained a K-nearest neighbor algorithm to classify 127 students' responses as low, middle, or high using a 21-item self-efficacy survey [1]. The optimal results of the model performance based on validation-set approach reached 92.3%. Another study applied a decision tree classifier to a dataset containing 1894

B. Tan and M. Cutumisu. Employing tree-based algorithms to predict students' self-efficacy in PISA 2018. In A. Mitrovic and N. Bosch, editors, *Proceedings of the 15th International Conference on Educational Data Mining*, pages 634–639, Durham, United Kingdom, July 2022. International Educational Data Mining Society.

https://doi.org/10.5281/zenodo.6852970

undergraduate students' survey data, obtaining the highest accuracy score of 82.58% [26].

3. METHODS

3.1 Data Source

The international large-scale dataset used in this study contained students' self-reported survey data and their test results of the OECD's PISA 2018. The dataset was publicly accessible at [20]. All students participating in PISA 2018 were included in this study regardless of their country of citizenship or origin. This constituted the original sample of 612,004 students from 74 countries and regions.

3.2 Tree-based Algorithms

We employed two tree-based algorithms to predict students' self-efficacy. Tree-based algorithms use a series of if-then rules to generate predictions. In each step of the series, the if-then rules separate data points into subsets according to a node where the prediction has the lowest error rate. By repeating the step, the split will eventually terminate when reaching the stopping criterion. Although singular tree models can be interpreted straightforwardly and work well with nonlinear relationships between predictors and the target variable, they usually have weaker predictive performance given that they are prone to overfitting, a situation where the supervised learning model fits too close to the training data to be able to generalize well and predict future data.

Tree-based ensemble learning methods are alternatives to singular decision trees by combining decision trees. The algorithms used in this study are RF and XGboost. RF is an ensemble learning algorithm developed based on two algorithms: decision tree and bootstrapping. Bootstrapping resamples data with replacement and it is used to repeatedly split the same dataset into bootstrapped samples based on which multiple decision trees can be built. Each of the trees built can generate a result. Then the algorithm makes the final decision by aggregating the results of all singular trees. In predicting the numerical values, the final result is calculated by averaging the results of all individual trees. The advantage of RF is that it is less prone to overfitting, which lifts the accuracy and stability of prediction to a much higher level.

XGBoost is another ensemble learning algorithm based on decision trees [7]. In contrast to RF, XGBoost employs boosting, a technique of correcting the errors of existing models by adding new models sequentially to predict the residuals of the existing model and, then, along with the existing model make another prediction. Through sequential iterations, each execution is completed on the same dataset and later models are improvements of prior models. Eventually, the errors will be gradually minimized; the algorithm stops when the model performance converges to a stable state.

3.3 Focal Variables

In this study, self-efficacy is the response or target variable (i.e., the variable to be predicted in the current supervised learning task). In the survey of PISA 2018, students' self-efficacy was measured by five items, namely, "I usually manage one way or another", "I feel proud that I have accomplished things", "I feel that I can handle many things at a time", "My belief in myself gets me through hard times", and "When I'm in a difficult situation, I can usually find my way out of it". Available responses were "Strongly disagree", "Disagree", "Agree", and "Strongly agree".

The predictors in the current study are the variables collected in the mandatory parts of students' self-reported questionnaire and their test results in PISA 2018, classified as home factors, students' well-being, motivational factors, other non-cognitive constructs, school climate, teacher-related variables, personal experiences, as well as the PISA 2018 test performance [19]. Table 1 lists the predictors included and their dimensions. The reliability of students' self-reported scale scores was available in PISA technical reports [18]. All scales achieved at least an acceptable reliability.

Table 1: Predictors used in the tree-based models

Dimension	Variable name
Home factors	Home possessions
	Parents' professions and qualifications
	Parents' education backgrounds
	Parental support
School climate	Cooperation climate
	Disciplinary climate
	Competition climate
Teachers	Teacher support
	Teacher understanding
	Adaptive instruction
	Teacher feedback
	Teacher enthusiasm
	Teacher directed instruction
Well-being	Meaning in life
	Life satisfaction
	Positive affective states
	Lively
	Miserable
	Proud
	Afraid
	Sad
	Fear
	Sacred
Motivational factors	Learning interests
	Learning aspiration
	Value of school
	Motivation for mastering tasks
	Motivation for competition
Other non-cognitive constructs	Reading self-concept
	Fixed mindset
	Empathy
	Attitude toward bullying
	Sense of belonging
Personal experiences	Exposure to bullying
	Skipped class or being late
	The age of early childhood education
	The age of pre-primary education
	Grade repetition
PISA 2018 test performance	Reading performance
	Math performance
	Science performance
Other	Gender

3.4 Data Preprocessing

A two-stage method was adopted to deal with missing values. In the first step, the entire row of the data entry was excluded if there were missing data on any of the five items measuring self-efficacy. Listwise deletion was used because it does not introduce new errors to the outcome variable as replacing the missing data with other values. This step excluded 84,179 instances, so 527,825 instances remained. In the second step, the missing values of the predictors were replaced with their column medians.

First, the responses of reverse worded items were reverse coded. Second, if the predictor is a categorical variable and is not grouped with other predictors (single-item scale), k-1 dummy variables (k is the number of categories) were created to replace the original categorical variable. Third, for multi-item scale response data (e.g., self-efficacy, measured by five items), a polytomous item response theory (IRT) model, the generalized partial credit model (GPCM) was used to transform the data to IRT scores ranging from negative to positive. In this way, self-efficacy scores became truly continuous data. This is a similar method that was adopted in the PISA 2018 Technical Report [18]. In order to facilitate meaningful interpretations, the IRT scores were linearly transformed using a formula: $X' = X \times 15 + 100$. The choice of multipliers for mean and standard deviation was arbitrary, just for the ease of interpretation (i.e., no negative self-efficacy scores). Such transformation does not alter the true comparative values of the measured constructs.

Upon the completion of data preprocessing, the dataset contained 64 individual predictors (including coded categorical variables) and one target variable (self-efficacy).

3.5 Model Training, Validating, and Testing

The *xgboost* and *scikit-learn* libraries in Python 3.7 were used to build the XGBoost and RF regressors. Model training, validating, and testing were conducted using the *scikit-learn* library. A nested cross-validation with grid search algorithm was used to obtain a robust and trustworthy estimation of the model tuning and performance [16]. As shown in Figure 1, the nested cross-validation algorithm has two layers: an outer three-fold cross-validation and an inner three-fold cross-validation. There was a total of nine distinct folds of inner cross-validation and three folds of outer cross-validation. The goal of the inner cross-validation was to find the hyperparameters yielding the best model performance, while the outer cross-validation was to test the generalizability of the tuned model performance to a new dataset.

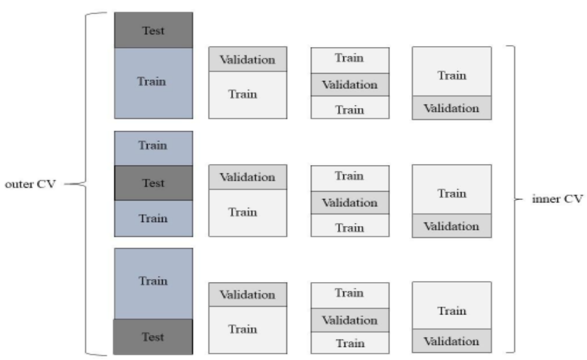

Figure 1: Nested cross-validation

Hyperparameters tuned for both RF and XGBoost regressors included the number of trees (n_estimators) and the maximum depth of the trees (max_depth). The maximum depth was selected in a range of 5 to 25, with a step of 5, whereas the number of trees could be 100, 150, or 200. Mean absolute error (MAE), Root mean square error (RMSE), and R^2 were used as evaluation metrics for both model validation and testing.

4. RESULTS

4.1 Descriptive Statistics of Self-Efficacy

The mean self-efficacy score of this sample was 99.99, ranging from 57.60 to 128.32, with a standard deviation of 13.28. Figure 2 shows the histogram of self-efficacy scores that indicates a leptokurtic distribution. Thus, more students scored extremely high or low compared to a normal distribution.

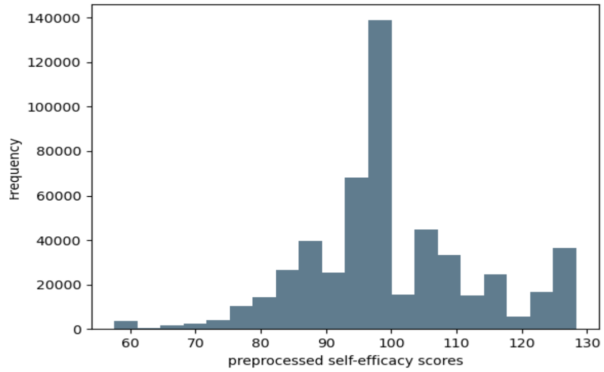

Figure 2: The distribution of self-efficacy IRT scores

4.2 Evaluation Results

Table 2 shows the training, validation, and test accuracy, for each set of the RF and XGBoost models, respectively. On the test set, R^2 of both prediction models was at least 0.447, suggesting that the two tree-based learning models could explain nearly half of the variability in students' self-efficacy. With reference to the range and standard deviation of self-efficacy scores, the MAEs and RMSEs indicated that both trained models achieved reliable prediction results.

Table 2: Results of model performance in training

Model	Data	RMSE	MAE	R^2
RF	Training set	4.240	3.279	0.898
	Validation set	9.898	7.373	0.444
	Test set	9.878	7.354	0.447
XGBoost	Training set	<.001	<.001	1
	Validation set	10.760	8.030	0.344
	Test set	9.776	7.271	0.458

4.3 Relative Importance of Predictors

The importance of the 64 predictors of students' self-efficacy was ranked. Figure 3 and Figure 4 present the top ten predictors and their weight contribution to the predictive power for the two models. In the RF model, the motivation for mastering tasks appeared to be the most powerful predictor with a relative importance of 19.5%, followed by meaning of life (10.8%), reading self-concept (5.1%), learning aspiration (4.4%), motivation for competition (3.2%), positive emotions (3%), empathy (2.9%), always feel proud (2.5%), and fear (2.5%).

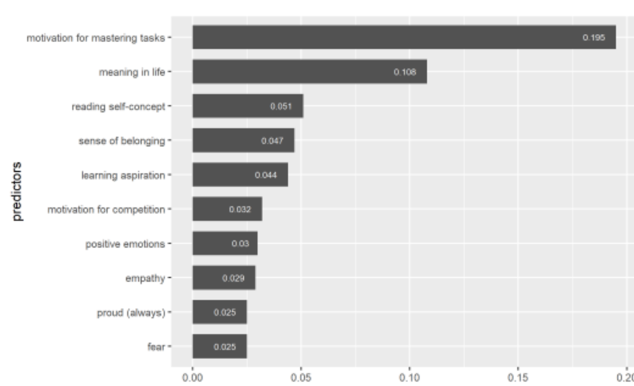

Figure 3: Relative predictor importance of the RF model

The ten most important predictors of the XGBoost model were the motivation for mastering tasks (14.5%), meaning in life (11.1%), proud (sometimes, always, rarely, and never; a total of 26.5%), motivation for competition (4%), learning aspiration (3.9%), positive emotions (3.2%), and reading self-concept (3.1%). These predictors contributed a total of 66.3% of the model prediction power. Notably, numerous variables were ranked in the top ten in both RF and XGBoost models, with motivation for task mastery and purpose in life maintaining the top two positions in both models. However, all the highly ranked predictors in both models appeared to be students' non-cognitive constructs including well-being and motivational factors. There were no variables of home factors, school climate, teachers, experience, and PISA 2018 test performance in the top 10 list.

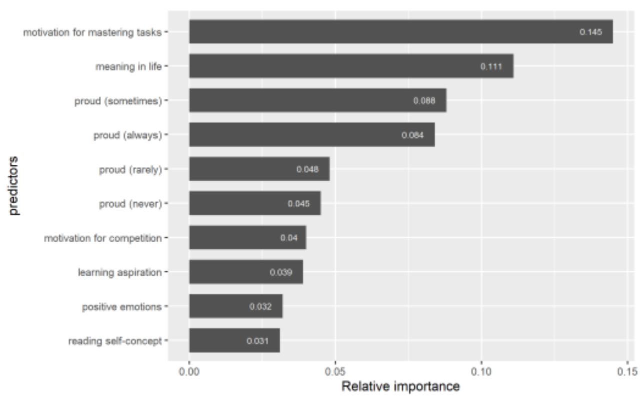

Figure 4: Relative predictor importance of the XGBoost model

5. DISCUSSION

The present study employed the RF and XGBoost algorithms to predict students' self-efficacy. The results suggest that the two tree-based algorithms could predict students' self-efficacy with small error sizes based on their self-reported survey data and test data. The XGBoost model seems to slightly outperform the RF model with respect to all chosen evaluation metrics on the test data.

The results also revealed the most salient predictors of both ML models. According to the rank of the relative importance, the best predictors for both models appeared to be students' non-cognitive factors including well-being and motivation. This is consistent with theories and empirical evidence [2, 13, 22, 27] supporting the close relationships between one's self-efficacy and their other non-cognitive constructs. On the other hand, gender was not a very important predictor. This is in line with a previous meta-analysis which reveals only a slight difference in self-efficacy between genders [13]. A surprising finding, however, is that students' test performances in PISA 2018 did not strongly predict self-efficacy. In a number of previous studies, researchers often consider self-efficacy as one of the strongest predictors for academic achievement [10]. However, this study revealed that predicting self-efficacy based on academic achievement seemed to be less unsuccessful. Another finding that is beyond our expectation is that home factors including parents' education and qualifications and home possessions contribute poorly to the predictive power in both models. This contradicts other studies which suggest the strong associations between self-efficacy and socioeconomic status [12, 15]. In addition, teachers-related variables and factors of school climate were not highly ranked predictors. We attribute the relatively weak predictive power of these predictors to their indirect relationship with students' self-efficacy. Home factors, academic achievement, as well as school and teacher factors, usually shape students' self-efficacy through other non-cognitive constructs. This also explains why the non-cognitive constructs are better predictors in the current models.

The present study has three major implications. First, it provides a successful example of predicting students' self-efficacy, expanding the body of literature on self-efficacy modeling. Second, it ranks the relative importance of predictors for students' self-efficacy, paving the way for future studies to further examine the relationships between self-efficacy and its best predictors. This may advance theories of self-efficacy as such expanding the model of how one's self-efficacy is formed. Third, it predicts and models students' self-efficacy at scale, using data from an international assessment. Because high self-efficacy is beneficial to students' motivation and learning experience [8], while low self-efficacy is associated with many mental and behavioral problems [24], early identification of low efficacious students is critical to students' educational careers and global well-being. This study suggests that it is feasible for education systems to use ML approaches to identify low self-efficacious students at scale.

A limitation of this study is that the family-level, school-level, and country-level factors used to predict students' self-efficacy are not exhaustive in the current ML models. Although our findings indicate that factors such as students' home factors and learning environment have a negligible effect on the model's performance, the current study was not able to examine a number of other potentially significant features. For example, parenting styles may be predictive of students' self-efficacy at the family level; at the school level, the predictive effect of geography, socioeconomic position, and school resources remains unknown. Another limitation is that the dataset mainly relied on students' self-reported questionnaires. Due to the subjective nature of self-reported data, the quality of students' responses may be subjectively biased. Finally, more tuning is needed for these models to address the overfitting issue inherent with tree-based models.

6. CONCLUSION AND FUTURE WORK

Noting that very limited ML research has been conducted to model students' self-efficacy, this study is the first to establish tree-based models that successfully predict students' self-efficacy at a large scale. This study also identified important predictors of students' self-efficacy, which helps to identify students with low self-efficacy and develop targeted programs to potentially improve self-efficacy. In responding to the limitations of the current studies, future studies can seek to use a more comprehensive feature set that includes more family level, school level, and country-level variables. In addition, for objective and real-time monitoring of self-efficacy, future studies may use other methods to gather objective, real-time indicators for predicting students' self-efficacy. In the future, other ML algorithms (e.g., lasso, deep learning) will be employed to tackle this task. Thus, applying ML approaches to predicting students' self-efficacy is feasible and constitutes an important undertaking.

7. ACKNOWLEDGMENTS

We would like to thank the reviewers for their feedback. We are grateful to the following granting agencies for supporting this research: the Social Sciences and Humanities Research Council of Canada - Insight Grant (SSHRC IG) RES0048110 and the Natural Sciences and Engineering Research Council Discovery Grant (NSERC DG) RES0043209.

8. REFERENCES

[1] Farooque, A., & Rizk, A. (2020). *Using k-nearest neighbors to classify undergraduate female self-efficacy in computer science.* [Poster Presentation]. Summer Undergraduate Research Fellowship, University of Houston.

[2] Alexopoulos, G. S., Raue, P. J., Banerjee, S., Mauer, E., Marino, P., Soliman, M., Kanellopoulos, D., Solomonov, N., Adeagbo, A., & Sirey, J. A. (2021). Modifiable predictors of suicidal ideation during psychotherapy for late-life major depression. A machine learning approach. *Translational Psychiatry, 11*(1), 1–8. https://doi.org/10.1038/s41398-021-01656-5

[3] Areepattamannil, S., Freeman, J. G., & Klinger, D. A. (2011). Intrinsic motivation, extrinsic motivation, and academic achievement among Indian adolescents in Canada and India. Social Psychology of Education, 14(3), 427–439. https://doi.org/10.1007/s11218-011-9155-1

[4] Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

[5] Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development, 72*(1), 187–206. https://doi.org/10.1111/1467-8624.00273

[6] Bassi, M., Steca, P., Fave, A. D., & Caprara, G. V. (2006). Academic self-efficacy beliefs and quality of experience in learning. *Journal of Youth and Adolescence, 36*(3), 301–312. https://doi.org/10.1007/s10964-006-9069-y

[7] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* (pp. 785–794). New York, NY, USA.

[8] Chowdhury, M. S., & Shahabuddin, A. M. (2007). Self-Efficacy, Motivation and Their Relationship to Academic Performance of Bangladesh College Students. *College Quarterly, 10*(1), 1-9.

[9] Ezen-Can, A., & Boyer, K. E. (2014). Toward adaptive unsupervised dialogue act classification in tutoring by gender and self-efficacy. *Strategies, 8,* 24.

[10] Gabriel, F., Signolet, J., & Westwell, M. (2018). A machine learning approach to investigating the effects of mathematics dispositions on mathematical literacy. *International Journal of Research & Method in Education, 41*(3), 306–327. https://doi.org/10.1080/1743727X.2017.1301916

[11] Govorova, E., Benítez, I., & Muñiz, J. (2020). Predicting student well-being: Network analysis based on Pisa 2018. *International Journal of Environmental Research and Public Health, 17*(11), 4014. https://doi.org/10.3390/ijerph17114014

[12] Han, J., Chu, X., Song, H., & Li, Y. (2015). Social capital, socioeconomic status and self-efficacy. *Applied Economics and Finance, 2*(1), 1–10.

[13] Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*(1), 1–35. https://doi.org/10.1007/s10212-011-0097-y

[14] Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., & Hutchison, B. (2015). The relationship between self-efficacy and academic achievement: A 5-year panel analysis. *The Asia-Pacific Education Researcher, 25*(1), 89–98. https://doi.org/10.1007/s40299-015-0236-3

[15] Karaarslan, G., & Sungur, S. (2011). Elementary students' self-efficacy beliefs in science: Role of grade level, gender, and socio-economic status. *Science Education International, 22*(1), 72–79.

[16] Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics, 6*(1). https://doi.org/10.1186/1758-2946-6-10

[17] McQuiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction, 18*(1–2), 81–123. https://doi.org/10.1007/s11257-007-9040-y

[18] OCED (2019). *PISA 2018 technical report.* OECD publishing.

[19] OCED. (2019). *PISA 2018 assessment and analytical framework.* OECD publishing.

[20] OECD (2019). *PISA 2018 Database* [Data set]. https://www.oecd.org/pisa/data/2018database/

[21] Schunk, D. H., & Ertmer, P. A. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 631–649). Academic Press. https://doi.org/10.1016/B978-012109890-2/50048-2

[22] Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In *Development of achievement motivation* (pp. 15-31). Academic Press. https://doi.org/10.1016/b978-012750053-9/50003-6

[23] Schwarzer, R., & Luszczynska, A. (2006). Self-efficacy, adolescents' risk-taking behaviors, and health. Self-Efficacy *Beliefs of Adolescents, 5,* 139–159.

[24] Tahmassian, K., & Moghadam, N. J. (2011). Relationship between self-efficacy and symptoms of anxiety, depression, worry and social avoidance in a normal sample of students. *Iranian Journal of Psychiatry and Behavioral Sciences, 5*(2), 91.

[25] Valois, R. F., Zullig, K. J., & Hunter, A. A. (2013). Association between adolescent suicide ideation, suicide attempts and emotional self-efficacy. *Journal of Child and Family Studies, 24*(2), 237–248. https://doi.org/10.1007/s10826-013-9829-8

[26] Wirojcharoenwong, W., Luangnaruedom, N., Rattanasiriwongwut, M., & Tiantong, M. (2014). Decision Tree Classifier for Computer Self-Efficacy with Best First Feature Selection. *International Journal of the Computer, the Internet and Management, 22,* 62-67. https://doi.org/10.3390/su13169337

[27] Zimmerman, B.J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-Efficacy in Changing Societies* (pp.202-231). New York: Cambridge University Press.