Interpretable-Machine-Learning Evidence for Importance and Optimum of Learning Time

Ali Nadaf, Sebas Eliëns, and Xin Miao

***Abstract*—This study uses a machine learning technique, a boosted tree model, to relate the student cognitive achievement in the 2018 data from the Programme of International Student Assessment (PISA) to other features related to the student learning process, capturing the complex and nonlinear relationships in the data. The SHapley Additive exPlanations (SHAP) approach is subsequently used to explain the complexity of the model. It reveals the relative importance of each of the features in predicting cognitive achievement. We find that instruction time comes out as an important predictor, but with a nonlinear relationship between its value and the contribution to the prediction. We find that a large weekly learning time of more than 35 hours is associated with less positive or even negative effect on the predicted outcome. We discuss how this method can possibly be used to signal problems in the student population related to learning time or other features.**

***Index Terms*—Learning factor analysis, machine learning, SHAP values, PISA.**

I. Introduction

K12 education systems play a key role in empowering young citizens with the necessary knowledge, skills, mindsets and competencies ready for 21st century human capital workforce demand. Hence, understanding how well K12 education systems globally work to prepare young citizens for the 21st century workforce is very important. PISA, OECD's Programme for International Student Assessment, has been invented to drive this effort in the last decade.

In 2018, around 600,000 15-year old students from 79 countries and economies participate in PISA. The 2018 PISA assesses student cognitive learning outcomes in reading, math and science, and surveys factors relevant for academic learning from aspects of 1) non-cognitive and metacognitive constructs, 2) student background (i.e. socio-economic status, educational pathways in early childhood, etc.), 3) teaching and learning processes (i.e. teaching practices and classroom support, out-of-school experience, etc.), 4) school policies and governance (i.e. school climate, parental involvement, assessments, etc.) [1]. These surveys, by design, are to investigate the "why" behind what the cognitive achievement outcomes say, and are designed to allow educators and policy makers to make evidence-based decisions.

Given the comprehensive datasets collected by 2018 PISA and the promises of how such data could be utilized to drive data-driven education research and policies, this paper sets out to investigate the potential of using two recent advances in Machine Learning (ML) as methodological innovation for education research and policy implications. First, improved implementations of gradient boosted tree algorithms such as XGBoost [2], LightGBM [3] and CatBoost [4]; second, the SHapley Additive exPlanation framework (SHAP) [5]–[7].

The purpose of this paper is to investigate and measure important learning factors contributing to student cognitive achievement, using 2018 PISA data sets and ML approaches. Going beyond the global feature importance, we want to gain a further understanding of how a single feature contributes to the predicted cognitive achievement even on the single student level, or for subpopulations of interest, an exercise for which SHAP provides the necessary tools.

Other works have considered the applications of ML in the analysis of education data. The papers [8] and [9] advocate the use of simple decision trees and regression trees respectively. The attractiveness of these ML models in the context of education lies in its interpretability.

In [10] a regression tree, random forest model and a boosted regression tree model were used for estimation of the school added-value education production function (EPF) based on data from Hungary, while a similar work [11] focuses on Tunisia. A study applying boosted regression trees to estimate school added-value as well as the importance of school characteristics appeared in 2018 [12]. It uses PISA 2015 data for 9 selected countries. Another 2018 study [13] investigates the effect of mathematical dispositions on mathematical literacy, based on 2012 PISA data from Australia. Finally, a study [14] applied a different ML approach, based on Support Vector Machines, to study the importance of features impacting reading literacy in Singapore based PISA 2015 data.

The rest of this paper is organized as follows. In Sec. II we discuss SHAP in further detail. Sec. III contains the definitions of the data and the model. In Sec IV. we present the results. Finally, Sec. V provides concluding remarks.

II. SHapley Additive exPlanations (SHAP)

We use SHAP which does not suffer from issues with consistency in assigning feature importance faced by other measures of feature importance for tree models [7], [15].

SHAP [5]–[7] is a game-theory and a local explanation technique to quantify the contribution of each feature on the model outputs. SHAP values determine the significance of a feature by contrasting what a model predicts with and without the feature from every conceivable blend of features in the dataset.

In analyzing education data, having a black box that

Manuscript received January 1, 2021; revised June 4, 2021.

The authors are with Alef Education, UAE (e-mail: ali.nadaf@alefeducation.com).

predicts a student performance $f(\boldsymbol{x})$ score based on some inputs $\boldsymbol{x}$ is of not much use. The rough idea of *model explanation* behind SHAP is to assign values to each of the features such that the prediction corresponds to the mean φ_0 plus the contributions from each of the features

$$f(x) = \varphi_0 + \sum_j \varphi_j$$

such that certain natural conditions are satisfied [5]–[7], what the authors call *additive feature attribution methods*.

To make the discussion a little less abstract, let us connect it back to 'the game' of predicting student performance score. The best prediction without any further information about the student would be the mean of the data. Suppose one has information about the socio-economic background. This will change the prediction. If one then includes some information about the students' behavior, the prediction would change again. At any stage, forgetting or including information about a certain feature will lead to a difference in the prediction. These differences quantify the effect of the particular feature on the final prediction. To arrive at SHAP values, one needs to take a certain average over these values and use conditional expectation values to define them.

As mentioned, SHAP values satisfy uniquely certain desirable requirements under the names, *local accuracy*, *consistency* and *missingness* [6]. Local accuracy states that for each input $\boldsymbol{x}$, the prediction $f(\boldsymbol{x})$ is equal to the mean plus the sum of SHAP values. Consistency is the requirement that if a feature is made to be less important in a model its SHAP value will also decrease. The latter is clearly important for a viable measure of feature importance. The locality of explanation indicates that the predicted performance of an individual student can be attributed to the different features. The missingness property is not so relevant for our purposes. It essentially means that when features play no role their SHAP value should be zero.

We do feel that drawing policy conclusions from SHAP values warrants further investigation, deferred to future work, in which issues relating to data quality as well as the intimate but subtle connection with questions concerning causality [16], [17] should be more rigorously addressed. But we note that using simpler models, which on first sight may be thought to give better interpretable results, can in fact be deceivable due to artefacts introduced by the simplicity of relations in the model not representative of the real data. In [15], it is shown that a linear model can be less interpretable than a boosted regression tree if nonlinearity in the data becomes important.

III. Data and Model

Here we provide details on our methods for selecting the data and boosted regression tree model we use in this study.

A. Model Inputs — Features

For model inputs, we use student survey responses regarding factors related to learning within the whole education systems (e.g. students, parents, teachers and principals) from the PISA 2018 database [18]. Teacher response data are not included because teacher-level data cannot be identified with individual students. A few indices were excluded for reasons of being highly correlated or containing overlapping information, namely we excluded indices indicating learning time for the main subjects LMINS, SMINS and MMINS in favor of the total learning time TMINS. Similarly, we excluded indices HOMEPOS, PARED, HISEI that factor into the computation of the index for socio-economic status (ESCS), as well as other indices related to occupation or education of the parents. We constructed some simple indices straight from the questionnaire data, following definitions indicated in the OECD results reports [19]-[20]. Other features were excluded, for instance those with more than 50% missing values over the whole dataset. The complete list of 80 features used in the analysis with the definitions of the simple indices is given in the appendix.

In this study, we focus on the total learning time given by the variable TMINS in the PISA data. This is computed using the student-reported information about the average minutes in a class period (question ST061) in relation to information about the number of class periods per week attended in total (question ST060) [15]. For ease of interpretation, we report total learning time in hours per week, while TMINS in the data is given in minutes.

We included all countries and economies (education systems) that participated in PISA 2018. In the final analysis we excluded Lebanon and North Macedonia which did not report any TMINS data.

B. Model Output — Cognitive Achievement Score

For the model output, we use a per-student estimate of cognitive achievement score for the three main subjects reading, mathematics and science.

The PISA framework does not estimate a single score for the cognitive tests for each student, but rather estimates a probability distribution for the student achievement and draws a number of plausible values (PVs) for the student's performance from this probability distribution. In order to compute unbiased statistics for the performance, one has to take a weighted average for the subpopulation for each single PV before taking the average over the PVs.

The cognitive achievement score per student that we define as model output corresponds to taking the mean of all PVs of the cognitive tests for the three main subjects. All averaged scores are computed by taking the weighted average first for each PV and then take the mean over PVs.

C. Boosted Regression Tree Model

The feature values for all the data points are encoded in a matrix $\boldsymbol{X}$ of 600821 rows corresponding to the total number of students that participated in PISA 2018 with the exception of students from Lebanon and North Macedonia and 80 columns representing number of learning features. Together with the vector $\boldsymbol{y}$ of length 600821 of student cognitive achievement scores, this forms the input data for our model.

We have trained a boosted-regression-tree model with data $\boldsymbol{X}$ **and** labels $\boldsymbol{y}$ using the CatBoost package in python [21]. Before training, we split the data into training and validation sets with a ratio 70/30. Next, we trained a CatBoost model

with 2000 estimators (regression trees) with a tree depth of 10 and learning rate η=0.05. We chose root-mean-squared-error (RMSE) as the loss function. SHAP values were computed using the TreeExplainer from the SHAP package for python [22].

IV. Results

In this section, we present the results obtained from the model and its analysis.

A. Model Performance

We start by examining the model's capabilities in correctly predicting student score values. In Fig. 1, several aspects of the model performance are presented. After training, applying the model on the validation set gives RMSE=47.74 and r^2=0.75 (coefficient of determination, measuring the correlation of predicted and actual target variable). Hair et al. (2011) suggested to consider r^2 of 0.75 to be a substantial level of predictive accuracy [23]. For interpretation, the mean-absolute error (MAE) is more intuitive than the RMSE. We found MAE=37.2 on the validation set. For comparison, [13] reports a MAE of 51 mathematical literacy points and $r^2=0.51$ in their study.

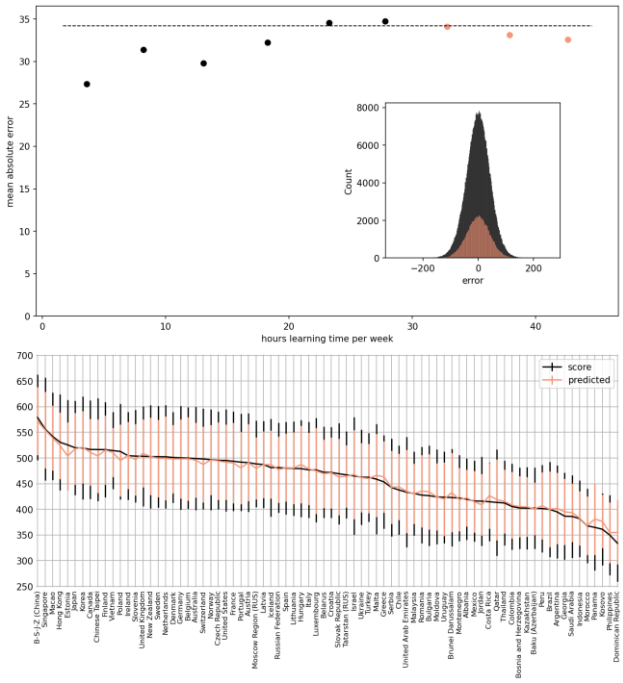

Fig. 1. Model evaluation. Top: mean-absolute error for binned data. We use bins of 5 hours. Mean-absolute error of full data is indicated (dashed line). High learning time, larger than 30 hours, is indicated in orange. Inset shows error for full data and high-learning time subset. Bottom: Score per education system (error bars indicate ∓ one standard deviation).

After performing the analysis of feature importance (see Sec. III A), we found learning time (TMINS) to be the most important feature contributing to student cognitive achievement. To check that there is no structural error in the prediction related to this variable, for instance a biased prediction for large learning time, we also plotted the MAE for data binned in terms of learning time in five-hour bins (Fig. 1 top) as well as the error for the full data and data with learning time value larger than 30 hours per week). There is no sign of worse performance on any specific bin compared to the full data.

Finally, we have compared the country-averaged prediction with the actual performance score. Here we did find some biases. We checked whether this may be explained by the bias introduced when one averages over PVs first and then takes the country average, which the prediction emulates, but it is not the case as these biases were of the order of 10^{-13}.

B. SHAP Summary Plot of Feature Importance

Having gained confidence in the model, we examined the SHAP values for all feature inputs. Fig. 2 shows the summary plot of top twenty features in terms of importance. Here, importance is measured as the mean absolute SHAP values [7], [15]. The feature importance is given in terms of PISA points. An importance of, say, 10 points tells us that the SHAP algorithm on average estimates a 10 point contribution, negatively or positively.

As shown in Fig. 2 (inset), the feature importance does not decay very rapidly. The top 20 features displayed in the bar plot account for 51% of the total feature importance. It is therefore not clear, based on this analysis, where to draw a line between important and not important features.

Among the top 20 important features, four categories are distinguished: 1) student-related; 2) teacher-related; 3) parents-related; 4) school & demographics related. student-related factors include student metacognition of reading (i.e. assessing credibility, summarizing, understanding and remembering), awareness and self-efficacy of explaining and discussing global issues, student perceived difficulty in reading, meaning of life and interest in ICT. Frequency of teachers applying teacher-directed instruction in the classroom is the only feature related to teachers. As for parental involvement, parents' emotional support and parents' enjoyment of reading are top important features. Lastly, total learning time spent per week, school size (i.e. total enrollment of boys and girls in the school), discriminating school climate, school type (i.e. public or private), ICT resources within school, student socioeconomic status and gender are top features related to school and demographics.

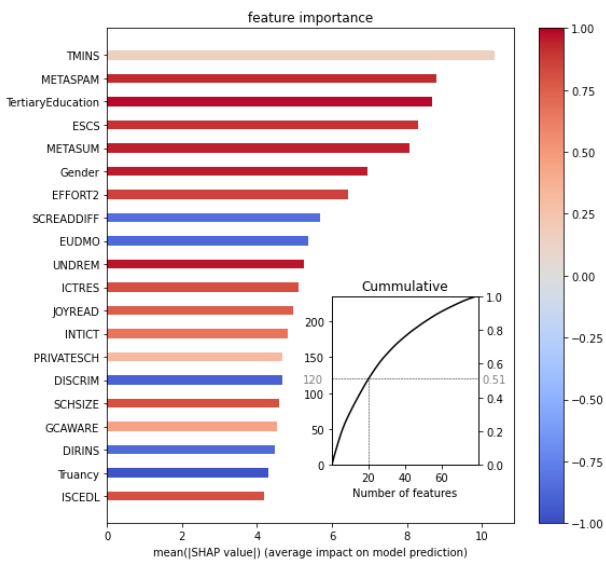

Fig. 2. SHAP summary plot of feature importance. Color indicates correlation coefficient with feature value. Inset shows cumulative importance.

We focus attention on the top 5 features. The most important feature, TMINS, indicating the total learning time, is assigned an importance of about 10 points in this analysis. Other features we find in the top 5 are TertiaryEducation (the student's self-reported estimate for finishing tertiary education[1]), ESCS (index for socio-economic status) and the indices for meta-cognitive skills METASPAM (ability to assess credibility) METASUM (summarizing).

From the importance alone, as measured by the average effect on the model prediction, we cannot see whether the effect is positive or negative. An aggregate understanding of the sign of the effect can be achieved by computing the linear correlation between the SHAP values and the feature value. Fig. 2 also shows the result of this computation in the color coding of the bars. Most of the features with large importance show a correlation either close to 1 (positively correlated) or -1 (negatively correlated). However, the most important feature, TMINS, shows a small correlation of 0.15. This indicates a nonlinear global relation between the SHAP values and the learning time values.

C. SHAP Values of TMINS with Four Education Systems Highlighted

In Fig. 3, we showed the aggregated averages of the cognitive achievement score plotted against the averaged learning time to represent the position of all PISA 2018 participating education systems. We divided figure 3 into four quadrants, using the OECD average value for both cognitive achievement score and learning time. In the high learning time, high score quadrant we picked B-S-J-Z (China) as a clear case of having very high values for both variables. Similarly, the Philippines appears as a clear pick for low cognitive score but high learning time. Vietnam and Kosovo are chosen to represent low learning time, with high and low performance respectively.

Fig. 3. Country scores and learning time (TMINS).

In Fig. 4, the dependence of the SHAP values on the feature values plotted for learning time. The four plots each show the full data with one of the selected education systems highlighted. We clearly see a nonlinear dependence of the SHAP value on the learning time value.

[1] Defined using International Standardised Classification of Education 1997 ISCED level 5A and/or ISCED level 6 (theoretically oriented tertiary and post-graduate).

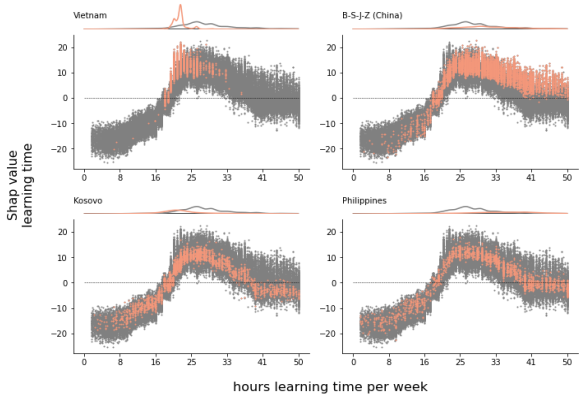

Fig. 4. SHAP values of TMINS with four education systems highlighted.

The range of learning times at which the SHAP value is switching between negative and positive SHAP values is defined as a cut-off range. As shown in Fig. 4, the cut-off range for the learning time is between 18 and 22 hours per week (presented as 20±2), below which the contribution of learning time is negative. When learning time is more than 33 hours per week, the contribution of learning time can be either positive or negative, depending on the values of other learning factors. When learning time is from 22 to 33 hours per week, contribution is positive with a peak around 25 hours. This is a strong evidence-based policy implication for education systems that 1) learning time has to reach the cut-off range of 20±2 hours per week; 2) more learning time can contribute negatively to student cognitive achievement.

Comparing the two education systems with high cognitive achievement scores, Vietnam and B-S-J-Z (China), it is obvious that learning time for Vietnam is tightly concentrated in the range of 18 to 33 hours per week and within this range the contribution to cognitive achievement is positive (Fig. 4 top). The distribution of learning time for B-S-J-Z (China) is much more spread out than that of Vietnam, and the majority of student learning time in B-S-J-Z (China) is reported to be 18 hours or more per week and considerable density for the high reported learning time. Yet, the SHAP value attributed to the students with learning time above 25 hours per week is still positive, with a decreasing trend.

As for the two education systems with low achievement scores, Kosovo and the Philippines, are characterized by a very broad spectrum of reported learning time values (see Fig. 3). The SHAP values of learning time for these two examples are negative in two segments of learning time values, 1) less than 20±2 hours per week, 2) more than 33 hours per week.

V. Conclusion

Using ML techniques and the PISA 2018 data, we studied the importance of 80 selected features in relation to predicting the score of students on the PISA cognitive test by measuring the feature contribution with SHAP values. We found that the average effect of the features on the model prediction decreases quite slowly from the most important to the least important feature. In other words, all the aspects of learning measured by PISA show some relation to the cognitive performance score, although some features are more important than others. Among the most important features we found measures of meta-cognitive constructs,

student expectations to finish higher education, socio-economic status and learning time.

In this analysis, we found that learning time was in fact attributed the largest importance as a predictor of cognitive achievement, and the relation is nonlinear. This corroborates evidence for diminishing returns of learning time and its complex nonlinear relation to academic performance [24]. In our analysis, there are even parts of the student population with large learning time values for which this feature value is associated with a negative contribution to predicted achievement. This is an illustration of how this type of analysis, based on SHAP values, can be valuable in clarifying the complex relations between different aspects of the learning process and learning outcome, in particular when such relations are nonlinear.

APPENDIX

The full list of features from the PISA 2018 student and school datasets used in the analysis is: ADAPTIVITY, ATTIMM, ATTIMMP, ATTLNACT, AUTICT, AWACOM, BEINGBULLIED, BELONG, BODYIMA, BSMJ, CHANGE, CLSIZE, COGFLEX, COMPETE, COMPICT, CREACTIV, CULTPOSS, CURSUPP, DIRINS, DISCLIMA, DISCRIM, DURECEC, EDUSHORT, EFFORT1, EFFORT2, EMOSUPP, EMOSUPS, ENTUSE, ESCS, EUDMO, FCFMLRTY, FLCONFIN, FLCONICT, FLFAMILY, FLSCHOOL, GCAWARE, GCAWAREP, GCSELFEFF, GFOFAIL, GLOBMIND, HEDRES, HOMESCH, ICTCLASS, ICTHOME, ICTOUTSIDE, ICTRES, ICTSCH, IMMIG, INFOCAR, INFOJOB1, INFOJOB2, INTCULT, INTCULTP, INTICT, ISCEDD, ISCEDL, ISCEDO, JOYREAD, JOYREADP, MASTGOAL, METASPAM, METASUM, PASCHPOL, PERCOMP, PERCOOP, PERFEED, PERSPECT, PQSCHOOL, PRESUPP, PRIVATESCH, PROATCE, RATCMP1, RATCMP2, RESILIENCE, RESPECT, SCHSIZE, SCMCEG, SCREADCOMP, SCREADDIFF, SOCONPA, SOIAICT, STAFFSHORT, STIMREAD, STRATIO, STUBEHA, STUBMI, SWBP, TEACHBEHA, TEACHINT, TEACHSUP, TMINS, UNDREM, USESCH, WEALTH, WORKMAST, Gender, GCLearning, Lateness, TakeAction, TertiaryEducation, TimeOnline, Truancy.

ALL CAPS features are indices in the data. We refer to the codebook and results volumes published by the OECD for definitions. The CamelCase indices are simple indices defined in the results volumes obtained from questionnaire answers. For clarity we give their definitions.

Gender: Student standardized gender, response to question ST004D01T (1: Female, 2: Male).

GCLearning: Number of learning activities attended by students concerning global issues, sum of responses to ten questions (ST221).

Lateness: Arriving late in the last two weeks prior to PISA test, derived from question ST062Q03TA. Value 0 if student reported not to have arrived late, value 1if student reported to have arrived late at least once.

TakeAction: Student willingness to take action, using a series of eight yes-or-no statements (ST222). Actions concern environmental protection, gender equality and interest in international and social issues, such as poverty and human rights.

TertiaryEducation: Indicates whether the student expects to complete tertiary education (from response to question ST225Q06HA; 0: no, 1: yes).

TimeOnline: Time spent online outside of school based on questions IC006 and IC007 asked in 51 countries. Five categories of internet users: “low Internet user” (0-9 hours per week); “moderate Internet user” (10-19 hours per week); “average Internet user” (20-29 hours per week); “high Internet user” (30-39 hours per week); and “heavy Internet user” (more than 40 hours per week).

Truancy: Students reported whether they had skipped days (ST062Q01TA) or classes (ST062Q02TA) of school in the two weeks prior to test? Value 0 if not, 1 if yes.

TMINS: Total learning time (minutes per week). Computed using information about the average minutes in a <class period> (ST061) in relation to information about the number of class periods per week attended in total (ST060).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ali Nadaf: Conceptualization, Methodology, Validation, Review & Editing, Interpretation.

Sebas Eliëns: Software and development, Formal Machine Learning analysis, Data Curation, Writing - Original Draft, Visualization.

Xin Miao: Review and editing of K12 education, PISA analytical framework, results interpretation, figure visualization and story flow for results.

REFERENCES

[1] OECD. (2020). *PISA 2018 Assessment and Analytical Framework*. [Online]. Available: https://www.oecd-ilibrary.org/education/pisa-2018-assessment-and-analytical-framework_b25efab8-en

[2] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[3] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 3146–3154, 2017.

[4] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: Unbiased boosting with categorical features,” Jun. 2017.

[5] S. M. Lundberg *et al.*, “From local explanations to global understanding with explainable AI for trees,” *Nat. Mach. Intell.*, vol. 2, no. 1, Art. no. 1, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[6] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 4765–4774, 2017.

[7] S. M. Lundberg, G. G. Erion, and S.-I. Lee, *Consistent Individualized Feature Attribution for Tree Ensembles, ArXiv180203888 Cs Stat*, Mar. 2019.

[8] L. S. Almeida and C. M. A. Gomes, “Advocating the broad use of the decision tree method in education,” Nov. 2017.

[9] C. M. A. Gomes and E. Jelihovschi, “Presenting the regression tree method and its application in a large-scale educational dataset,” *Int. J. Res. Method Educ.*, vol. 43, no. 2, pp. 201–221, Mar. 2020, doi: 10.1080/1743727X.2019.1654992.

[10] F. Schiltz, C. Masci, T. Agasisti, and D. Horn, “Using machine learning to model interaction effects in education: A graphical approach,” *Budapest Working Papers on the Labour Market, Working Paper BWP-2017/4*, 2017.

[11] S. Rebai, F. Ben Yahia, and H. Essid, “A graphically based machine learning approach to predict secondary schools performance in

Tunisia," *Socioecon. Plann. Sci.*, vol. 70, p. 100724, Jun. 2020, doi: 10.1016/j.seps.2019.06.009.

[12] C. Masci, G. Johnes, and T. Agasisti, "Student and school performance across countries: A machine learning approach," *Eur. J. Oper. Res.*, vol. 269, no. 3, pp. 1072–1085, Sep. 2018, doi: 10.1016/j.ejor.2018.02.031.

[13] F. Gabriel, J. Signolet, and M. Westwell, "A machine learning approach to investigating the effects of mathematics dispositions on mathematical literacy," *Int. J. Res. Method Educ.*, vol. 41, no. 3, pp. 306–327, May 2018, doi: 10.1080/1743727X.2017.1301916.

[14] X. Dong and J. Hu, "An exploration of impact factors influencing students' reading literacy in singapore with machine learning approaches," *Int. J. Engl. Linguist.*, vol. 9, no. 5, Art. no. 5, Aug. 2019, doi: 10.5539/ijel.v9n5p52.

[15] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, Art. no. 1, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[16] D. Janzing, L. Minorics, and P. Blöbaum, "Feature relevance quantification in explainable AI: A causal problem," *ArXiv191013413 Cs Stat*, Nov. 2019.

[17] H. Chen, J. D. Janizek, S. Lundberg, and S.-I. Lee, "True to the model or true to the data?" *ArXiv200616234 Cs Stat*, Jun. 2020.

[18] OECD PISA 2018. *2018 Database - PISA*. [Online]. Available: http://www.oecd.org/pisa/data/2018database/

[19] OECD. *PISA 2018 Results Vol I-VI*. [Online]. Available: https://www.oecd.org/pisa/publications/

[20] OECD. (2020). *PISA 2018 Results V. Effective Policies, Successful Schools*. [Online]. Available: https://ictlogy.net/bibliography/reports/projects.php?idp=4269

[21] CatBoost. (Dec. 12, 2020). github.com/catboost/catboost. [Online]. Available: https://github.com/catboost/catboost

[22] S. Lundberg, *slundberg/shap*,2020.

[23] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet," *J. Mark. Theory Pract.*, vol. 19, no. 2, pp. 139–152, Apr. 2011, doi: 10.2753/MTP1069-6679190202.

[24] A. Gromada and C. Shewbridge, "Student learning time: A literature review," Feb. 2016, doi: https://doi.org/10.1787/5jm409kqqkjh-en.

Copyright © 2021 by the authors. 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 (CC BY 4.0).

Ali Nadaf is the head of data science at Alef Education. He holds a Ph.D. in mathematics from Simon Fraser University, Canada and held postdoctoral fellowships in machine learning. His research spans a wide range of fields within the mathematical sciences and machine learning, and has received numerous academic awards and demonstrated record of scholarly accomplishments with published papers. He has more than 15 years of working experience in the United Nations and PayPal. His current interest is to develop an AI solution to identify and measure students' motivational and meta-cognitive behavior using students' interaction data within an educational platform.

Sebas Eliëns is a data-science freelancer and researcher at Alef Education. He holds a PhD in theoretical physics from the University of Amsterdam and held a postdoc at the International Institute of Physics, Natal Brazil. His research focused on quantum many-body physics and quantum information theory. Now he is interested in the application of machine learning methods in other areas, like education, while maintaining an interest in the deep connection between causal inference and the foundations of quantum physics.

Xin Miao is the lead researcher at Alef Education. She holds a master's degree in International Education Policy from Harvard University and a graduate certificate in International Development from Johns Hopkins University SAIS Nanjing Center. Xin Miao has both bottom–up K12 practice experience and top–down education research and policy training. She has worked in China's highly professionalized K12 education system, explored North America and international organizations (such as UNESCO, World Bank Group), and has been working in R&D in the EdTech industry in the UAE, focusing on education research using AI methodologies, impact evaluation, education policy analysis, agile R&D in interaction design for children, public–private collaboration in K12 education reforms.