De La Salle University
AKI
Angelo King Institute
for Economic and Business Studies

DLSU-AKI Working Paper Series
2022-09-085

Which Filipino Students are Being Left Behind in Mathematics? Testing Machine Learning Models to Differentiate Lowest-Performing Filipino Students in Public and Private Schools in the 2018 PISA Mathematics Test

By: Minie Rose C. Lapinid[1]
Macario O. Cordel II[2]
Jude Michael M. Teves[3]
Sashmir A. Yap[4]
Unisse C. Chua[5]
Allan B. I Bernardo[6]

[1] Department of Science Education, De La Salle University; minie.lapinid@dlsu.edu.ph
[2] Dr. Andrew L. Tan Data Science Institute, De La Salle University; macario.cordel@dlsu.edu.ph
[3] Dr. Andrew L. Tan Data Science Institute, De La Salle University; jude.teves@dlsu.edu.ph
[4] Dr. Andrew L. Tan Data Science Institute, De La Salle University; sashmir.yap@dlsu.edu.ph
[5] Dr. Andrew L. Tan Data Science Institute, De La Salle University; unisse.chua@dlsu.edu.ph
[6] Department of Psychology, De La Salle University; allan.bernardo@dlsu.edu.ph

DLSU - Angelo King Institute for Economic and Business Studies

Room 223, St. La Salle Hall
2401 Taft Avenue, Manila, 0922, Philippines

Visit Us

https://www.dlsu-aki.com/

Which Filipino Students are Being Left Behind in Mathematics? Testing Machine Learning Models to Differentiate Lowest-Performing Filipino Students in Public and Private Schools in the 2018 PISA Mathematics Test

Final Report

Minie Rose C. Lapinid[1]

Macario O. Cordel II[2]

Jude Michael M. Teves[3]

Sashmir A. Yap[4]

Unisse C. Chua[5]

Allan B. I Bernardo[6]

[1] Department of Science Education, De La Salle University; minie.lapinid@dlsu.edu.ph

[2] Dr. Andrew L. Tan Data Science Institute, De La Salle University; macario.cordel@dlsu.edu.ph

[3] Dr. Andrew L. Tan Data Science Institute, De La Salle University; jude.teves@dlsu.edu.ph

[4] Dr. Andrew L. Tan Data Science Institute, De La Salle University; sashmir.yap@dlsu.edu.ph

[5] Dr. Andrew L. Tan Data Science Institute, De La Salle University; unisse.chua@dlsu.edu.ph

[6] Department of Psychology, De La Salle University; allan.bernardo@dlsu.edu.ph

Abstract

Filipino students performed poorly in the PISA 2018 mathematics assessment, with more than 50% obtaining scores below the lowest proficiency level. Students from public schools also performed worse compared to their private school counterparts. We used machine learning approaches, specifically binary classification methods, to model the variables that best identified the poor performing students (below Level 1) vs. better performing students (Levels 1 to 6) using the PISA data from a nationally representative sample of 15-year-old Filipino students. We analyzed data from students in private and public schools separately. Several binary classification methods were applied, and the best classification model for both private and public school groups was the Random Forest classifier. The 10 variables with the highest impact on the model were identified for the private and public school groups. Five variables were similarly important in the private and public school models. But there were other distinct variables that relate to students' motivations, family, and school experiences that were important in identifying the poor performing students in each school type. The results are discussed in relation to the social and social cognitive experiences of students that relate to socioeconomic contexts that differ between public and private schools.

Keywords: mathematics achievement, machine learning, Philippines, public vs. private schools, school type, socioeconomic differences, PISA

Filipino students were among the lowest-performing groups of students among all the participating countries in the 2018 Programme for International Student Assessment (PISA). In mathematics, less than 20% of students demonstrated the minimum proficiency level (Level 2), whereas more than 50% showed very low proficiency (below Level 1). Scoring below the lowest level of proficiency in the PISA, these Filipino students have been clearly left behind in terms of mathematics education; more than half of this age group of Filipino students have inadequate mathematical skills compared to their peers in other parts of the world. The poor performance in mathematics also varied in degree between the students in public and private schools, where the means were 343 and 395, respectively (Department of Education, 2019).

The study aims to identify the factors (personal and contextual) that differentiate the lowest-performing students from the other Filipino students in mathematics in public and private schools in the Philippines. Previous studies have shown that public and private schools in the Philippines have very different environments for learning resources (Trinidad, 2020) and for supporting student motivation and engagement (Bernardo et al., 2015). We explore whether different factors identify low-performing students in each type of school. We use a range of machine learning approaches to analyze the Philippines 2018 PISA data from the student questionnaire and the school questionnaire and analyze the data of students from public and private schools separately. Although typical education research studies investigate predictors of achievement at one level of analysis, the machine learning approach allows researchers to consider factors at the student level, their family, the instructional experiences in school, and other school characteristics; thus, it reveals a more complex set of interrelated factors that identify the students that are left behind in mathematics education in the two types of school in the Philippines.

Filipino Students' Mathematics Proficiency in PISA 2018

Students' mathematics proficiency in the PISA assessment relates to the students' capacity to formulate, use, and interpret mathematics in different contexts, including familiar personal experiences and in broader and more abstract contexts of work, society, and science. Students who are assessed to have good mathematics proficiency are able "to reason mathematically and use mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena" (Organisation for Economic Co-operation and Development [OECD], 2019a, p. 104]. The test items were given in combinations of the different mathematical processes, mathematical content, and contexts. The mathematical processes included formulating situations mathematically; employing mathematical concepts, facts, procedures, and reasoning; and interpreting, applying, and evaluating mathematical outcomes. Underlying these mathematical processes were fundamental mathematical capabilities such as understanding a problem situation, its tasks, and questions; being able to present, explain, and justify a solution; translating and representing the problem and its quantities into a mathematical form; and utilizing mathematical content knowledge and tools to solve the problem and to communicate results (OECD, 2019a).

Six proficiency levels were described to represent the range of mathematics skills, knowledge, and understanding in the 2018 PISA mathematics assessment; the same six levels have been used because mathematics became a focal area of assessment in 2002 (OECD, 2019a). Level 2 is considered the minimum proficiency standard, and less than 20% of Filipino students attained Level 2 proficiency or better. Thus, an overwhelming majority of Filipino students score

below standard; more specifically, 27% scored at Level 1 proficiency and 54% scored below Level 1 (OECD, 2019a). According to the PISA mathematics proficiency guide, Level 1 means:

> students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are almost always obvious and follow immediately from the given stimuli. (OECD, 2019a, p. 105)

So less than three of every 10 15-year-old Filipino students can do math only at that level, and more than half of these students cannot even do those actions.

Although the results suggest that most Filipino high school students are not learning what they are supposed to in mathematics, the situation seems to be worse for the students in Philippine high schools. On average, private school students' scores were at Level 1 proficiency, whereas those from public schools were below Level 1. Although about three of every 10 private school students scored below Level 1 proficiency in mathematics, six out of every 10 public school students scored below Level 1.

In a sense, the results are not surprising as the Philippines had been consistently performing poorly in mathematics in the global assessments. It had not been able to improve from the bottom 5 ranks since it joined Trends in Mathematics and Science Study (TIMSS) in 1999 (Mullis et al., 2004). However, the Philippine government chose to participate in PISA 2018 with the aim of gaining knowledge from the international large-scale assessment to help improve the current educational system (National Economic Development Authority, 2020). Indeed, the PISA provides data on a wide range of variables that can be studied as possible predictors of successful (or unsuccessful) learning in the different domains. These variables might be

interacting in ways that predict either poor or good mathematics achievement. In the next section, we consider the possible types of variables known to be associated with students' mathematics learning.

Predictors of Mathematics Learning and Achievement

Research has revealed many important predictors of mathematics learning and achievement, and most of the predictors can be classified under one of five broad categories: student factors, family factors, teacher factors, classroom and school factors, and policy factors (Maamin et al., 2021). We will not attempt a comprehensive review of such factors but refer to some that were measured in the PISA 2018 questionnaire and that were included in the analysis for the current study; these factors fall under the first four broad categories, as no policy-related factors were included in the student questionnaire of PISA.

Beyond the typical student factors such as gender, cognitive abilities, and metacognitive (Desoete & De Craene, 2019; Lindberg et al., 2010), research has confirmed the importance of a range of non-cognitive social psychological factors as predictors of student academic success Kim & Choi, 2021; Lindberg et al., 2010). In mathematics achievement, these factors include motivation (Dela Rosa & Bernardo, 2013; Levpušček et al., 2013; Saw & Chang, 2018), attitudes, (Dela Rosa & Bernardo, 2013; Gjicali, & Lipnevich, 2021), self-beliefs (Damrongpanit, 2019; Szumski & Karwowski, 2019), and academic emotions (Villavicencio & Bernardo, 2013, 2016). There are more specific student factors that relate to these social psychological factors, such as the students' educational and career aspirations; students who have higher career aspirations that also require higher educational qualifications show stronger motivations related to achieving in mathematics (Watt et al., 2019; Webster & Fisher, 2000). On the other hand, poor motivation in learning is associated with students' absenteeism and

tardiness, which are also associated with lower mathematics achievement (Gottfried & Ansari, 2022; Vesić et al., 2021).

Family factors that relate to students' mathematics achievement include factors related to the family's socioeconomic background, which relates to parents' education and occupation, as well as the educational resources available in the home (Lam & Zhou, 2021; Lombardi & Dearing, 2021; Marks & Pokropek, 2019). The types of parental support for the students' learning are also important predictors of students' achievement (Bernardo et al., 2015; Soenens et al., 2007). Parental support also relates to the quality of the parent-child relationship (Christenson & Havsy, 2004), parental involvement in their children's learning in mathematics (Hyde, 2006; Jay et al., 2018), and expectations of their children's achievement (West, 1998).

Teachers' expectations of students also play an important role in students' achievement in mathematics (Szumski & Karwowski, 2019), as do other social and interpersonal teacher factors. Teacher characteristics relate to instructional quality, and both predict higher student achievement in mathematics (Toropova et al., 2019; Wayne & Youngs, 2003; Wedel, 2021). Teacher characteristics such as teacher preparation (Boyd et al., 2009; Fung et al., 2017), continuing professional development (Desimone et al., 2013; Harris & Sass, 2011), mathematical knowledge (Baumert et al., 2010), and teachers' self-efficacy (Fung et al., 2017; Zee et al., 2018) are some of the qualities that relate to their instructional performance.

Some teacher factors are also shaped by school-level factors, such as school policies on class sizes (Woessmann & West, 2002) and support for teachers' continuing professional development (Desimone et al., 2013). But other important aspects of the school environment also play an important role in predicting student achievement in mathematics.

The school environment can influence teachers' and students' behavior in the teaching and learning process and, eventually, students' achievement (OECD, 2019a). A school culture that promotes shared values and norms for learning, high academic standards (Jesse et al., 2004), and strong personal bonds between teachers and students showing genuine concern to students for academic success (Mateos et al., 2021) are said to be important predictors of student achievement. Other important predictors include orderly and highly structured schools, classes where rules and procedures are consistently and reasonably enforced (Ilg & Massucci, 2003; Pressley et al., 2004), and a school environment that encourages student participation in after-class activities (Wigfield et al., 2006).

Perhaps one of the most important school factors that predict student achievement relates to the schools' learning resources (Caponera & Losito, 2016; Levpušček et al., 2013). In the Philippines, for example, material constraints and lower teacher resources are associated with lower student attention, lower student respect, and more concerns with attendance, bullying, other problematic student behaviors, and student achievement (Trinidad, 2020). These resource constraints distinguish the school environments in public and private schools in the Philippines (Lockheed & Jimenez, 1994) and also other countries (OECD, 2019b), and more importantly, they are associated with achievement gaps (Braun et al., 2006; Carbonaro & Covay, 2010). Interestingly, one study showed that school-type differences were more pronounced in mathematics achievement compared to other subjects (Lubienski & Lubienski, 2006). Other studies found that the achievement gap between private and public schools in the Philippines is also associated with different levels of student motivations and perceived support from parents and teachers (Bernardo et al., 2015) and higher student selectivity in private schools (Yamauchi, 2005).

The Current Study

The various student, family, teacher, and school factors are also assumed to be interconnected in predicting students' achievement in mathematics. For example, individual students' career aspirations are related to their motivational beliefs about math, which are also related to how they perceive their classroom environment (Lazarides et al., 2020). Students' self-perceptions also interact with the school's social context in influencing students' engagement (Fall & Roberts, 2012; Wang & Eccles, 2013), and their self-beliefs also interact with their socioeconomic status in influencing their mathematics achievement (Bernardo, 2021). Thus, it is important to explore a range of predictors of students' mathematics achievement to see how they might work together.

In the current study, we wanted to study the factors that distinguish Filipino students who perform poorly in the PISA 2018 mathematics assessment from those who met the minimum performance standards. The PISA 2018 obtained self-report data on a wide range of factors—the students, their families, teachers, classes, and schools—that are possible predictors of students' proficiency in mathematics. Our objective was to identify the models that best identify the Filipino students who performed poorly in mathematics using machine learning approaches, and we wanted to identify the model for public school students and private school students. For this purpose, we trained and evaluated different machine learning models on the Philippine data to determine the best classifier for classifying poor and better-performing students. Eighty percent of the data were randomly sampled and used to iteratively adjust the model's parameters during the training phase. Training iterations were terminated based on any of the following conditions: (a) the training performance converges and is less than a preset value, (b) the validation

performance worsens, or (c) the validation performance does not improve. Each trained model was evaluated using region of convergence (RoC and ROC-AUC) scores to determine how well it separates the two categories, standard metrics, for example, average F1-score to measure its prediction performance, and cross-validation to demonstrate its performance on unseen data. By exploring models for identifying poor-performing students in mathematics in public and private schools, we hope to identify variables that will point to poor learners' vulnerabilities that could be the target of interventions.

Methods

The Dataset

The data used in the study were derived from the Philippine sample in the PISA 2018 database (publicly accessible at https://www.oecd.org/pisa/data/2018database/). The sample was obtained following a two-stage stratified random sampling system. First, 187 schools were randomly selected across the country's 17 regions, and the students were randomly sampled for each school. The sample comprised 7,233 15-year-old students, and of this sample, 18.5% meet the minimum standard defined in the PISA 2018 (i.e., Level 2 or higher) and 26.9% were assessed at Level 1 proficiency. The lowest proficiency group (below Level 1) comprised 54.6% of the sample.

From the dataset, 96 variables, including the estimate for the mathematics achievement (i.e., plausible values 1 or PV1MATH), the school type (SCHLTYPE), and other relevant student-, family-, teacher-, and school-related variables were considered for this exploratory study. We removed three variables with 100% missing values (these were not included in the Philippine version of the survey: ICTSCH, ICTHOME, and ST225Q03HA). We also excluded students with more than 50% missing values, decreasing the number of entries to 7,091 students. Of this total, 1,156 were from private schools (SCHLTYPE = 1 and 2) and 5,935 were from public schools (SCHLTYPE=3).

The remaining variables with missing values in the reduced data set were imputed using the k-nearest neighbor (kNN) algorithm, where k is empirically determined as being equal to 7. PVMATH1 variable was then transformed such that the lowest proficiency students (i.e., students with PVMATH1 < 357.7 or below Level 1) were set to 1, and all the remaining better-performing students (i.e., students with PVMATH1 ≥ 357.7 or Levels 1 to 6) is set to 0.

Normalization per variable, except for SCHLTYPE, was then performed such that each variable range is from 0 to 1. We further reduced the number of variables by removing variables with strong positive or negative correlation, that is, rho > |+/-0.75|, resulting in a more condensed dataset with 58 variables.

Machine Learning Modeling

Our objective was to discover the key variables that characterize the poor-performing students or, more specifically, that differentiate them from the better-performing Filipino students in mathematics. Machine learning (henceforth, ML) algorithms are typically used to discover the intrinsic and highly complex relationship between input and output data. An exhaustive search approach on the hyperparameters of different ML models, namely logistic regression, multilayer perceptron (MLP), support vector machine (SVM), decision tree, and random forest, were performed to zero in on the most optimal model for the classification task. Table 1 (next page) summarizes the hyperparameters considered in the exhaustive search.

The first ML model considered is the perceptron type which computes the activation outputs for each layer of the perceptron-type ML models given the previous activations. The second ML model considered is the kernel-based type ML model, known as SVM. SVM looks for the most optimal decision plane to optimally separate data into different categories. For non-linear data, the SVM is extended by transforming both the data and the SVM model to higher-order dimensions through the use of kernels. Finally, tree-based models are investigated for a more straightforward approach, more powerful for data that has no assumption on their normality. Tree-based models split from the top-down, grouping data into the most homogeneous

"sub-nodes" based on their attributes. (Please see the Annex for technical details on these machine learning modeling approaches.)

Table 1
List of the Considered ML Models and the Different Hyperparameters During the Grid Search. Hyperparameters Define the Complexity of the ML and Each Model's Learning Performance During the Training

ML models	Hyperparameters
Logistic Regression	solver: newton-cg, lbfgs, liblinear penalty: none, l1, l2, elasticnet c: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100
MLP	hidden layer sizes: (10, 30, 10), (10, 30), (32, 32), (10, 10, 10, 10) activation: tanh, relu, logistics solver: stochastic gradient descent, adam alpha: 1e-4, 5e-3, 5e-2 learning rate: constant, adaptive
SVM	kernel: radial basis function, polynomial gamma: 1, 1e-1, 1e-2, 1e-3, 1e-4 c: 1e-1, 1, 10, 100, 1000
Decision Tree	criterion: gini, entropy max depth: 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 30, 40, 50, 70, 90, 120, 150
Random Forest	criterion: gini, entropy number of estimators: 200, 500 max features: auto, sqrt, log2 max depth: 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 30, 40, 50, 70, 90, 120, 150

The classification task was performed for 7,091 participants from private schools and public schools. For each group of participants, training and test samples were randomly selected. For private schools, there were 1,238 training samples after data balancing using oversampling and undersampling (i.e., 619 for each class, 0 and 1) and 232 test samples (72 for class 1 and 160 for class 0). For public schools, 5,316 training samples were used after data balancing using

oversampling and undersampling (i.e., 2658 each for class 0 and 1), and 1,419 testing samples were set aside. Finally, the exhaustive search for the best ML model and the corresponding set of hyperparameters was performed. Each training in the exhaustive search carried out five-fold cross-validation, used 600 iterations, and reported the average precision, recall, F1-score, and accuracy.

Results

Machine Learning Modeling Results

The results suggest that the best classifier for the task for both private and public schools is the random forest classifier. Table 2 and Figure 1 summarize these results.

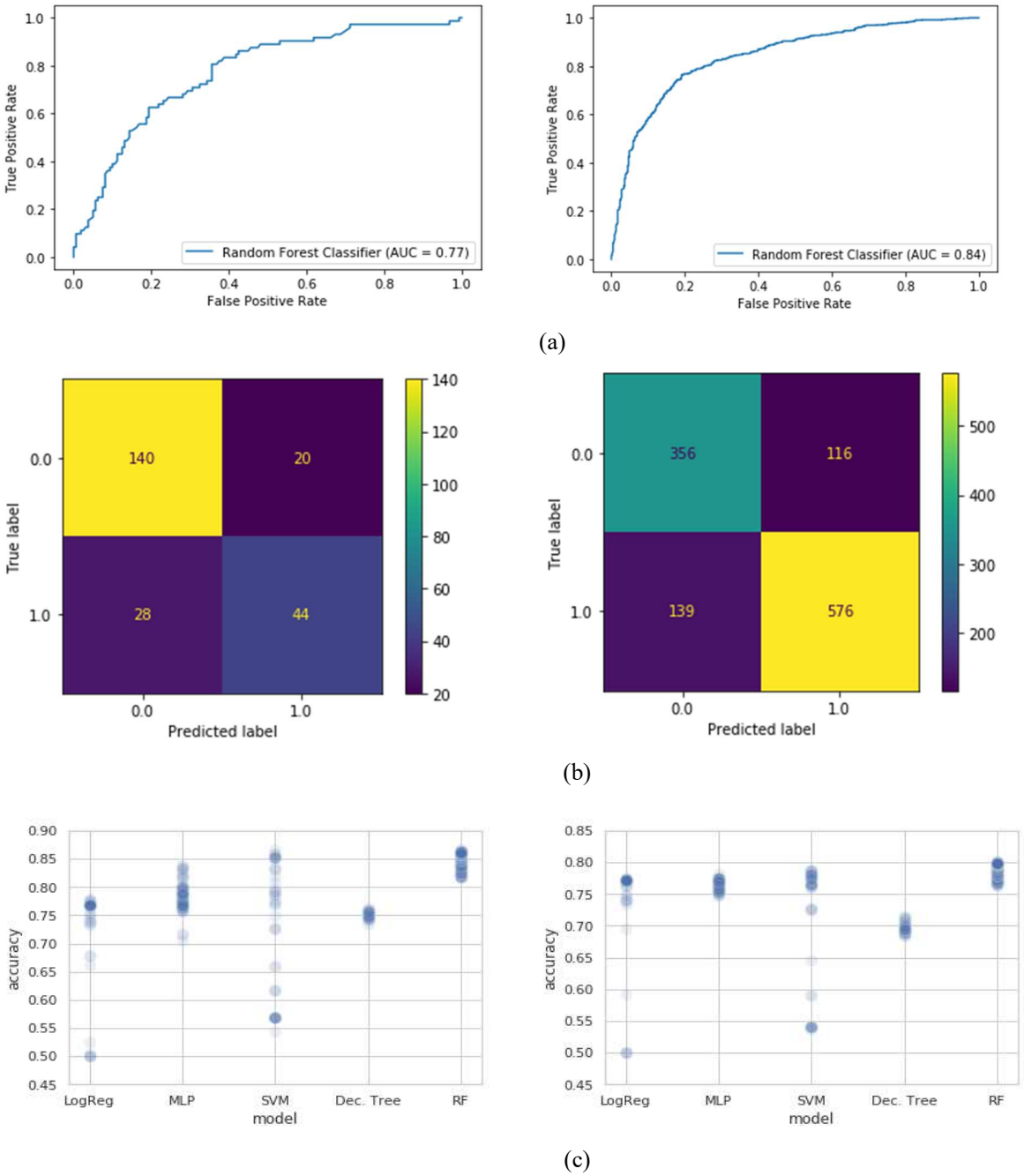
Table 2
Summary of Best Validation Performance Per ML Model After Grid Search

School type	ML model	Validation performance				Hyperparameter
		Precision	Recall	F1-score	Acc	
Private	Logistic regression	0.63	0.75	0.68	0.74	C: 1; penalty: l2; solver: newton-cg
	MLP	0.67	0.56	0.61	0.73	activation: 'relu' ; alpha: 0.005, hidden_layer_sizes: (32, 32) learning_rate: 'constant', solver: 'adam'
	SVM	0.67	0.02	0.04	0.63	C: 10; gamma: 1; kernel: rbf
	Decision tree	0.54	0.54	0.54	0.72	criterion: gini; max_depth: 12
	Random forest	0.69	0.61	0.65	0.79	criterion: 'gini'; max_depth: 20 max_features: log2 n_estimators: 500
Public	Logistic regression	0.81	0.75	0.78	0.75	C: 1; penalty: l1; solver: liblinear
	MLP	0.80	0.75	0.78	0.74	activation: 'relu' ; alpha: 0.05; hidden_layer_sizes: (32, 32) learning_rate: 'constant', solver: 'sgd'
	SVM	0.75	0.76	0.75	0.70	C: 100; gamma: 0.1; kernel: rbf
	Decision tree	0.76	0.76	0.76	0.71	criterion: gini; max_depth: 6
	Random forest	0.81	0.78	0.79	0.79	criterion: 'gini' ; max_depth: 15 max_features: auto n_estimators: 200

Note: Text in bold indicates the best ML model performance for a specific metric and school type. For both participants from private and public schools, the best classifier is the random forest in terms of accuracy.

Figure 1

Test for Best Clarifier Using Different ML Models



Note: (a) Area under the ROC curve (AUC) indicators for the private (left) and public (right) school participants. AUC score indicates how well separated are the classes 0 and 1 in the random forest classifier. (b) Confusion matrix for the random classifier model for the private (left) and public (right) school participants. (c) a cursory look at the accuracy of the different ML models in the exhaustive search for the best. hyperparameters for the private (left) and public (right) school participants. Note that RF performs better than other ML models in terms of performance consistency regardless of the hyperparameters.

Most Important Variables

To identify the level of importance with which poor and better performers in mathematics were classified, we used Shapley values, which tell how to fairly distribute the prediction outcome among the features (Lundberg & Lee, 2017). The Shapley value is the mean marginal contribution of a feature value across all possible feature groups. It produces a ranked list of several features in descending order, indicating the degree of significance of the features.

Initial works used top 10 (Chen et al., 2021), top 15 (Dong & Hu, 2019; Bernardo et al., 2021), and top 20 (Chen et al., 2021; Dong & Hu, 2019) variables in their feature importance analysis. In this work, to manage complexity in comparing the key variables for private and public student performance classification, the 10 most significant features for the public and private school groups are analyzed and illustrated in Figure 2.

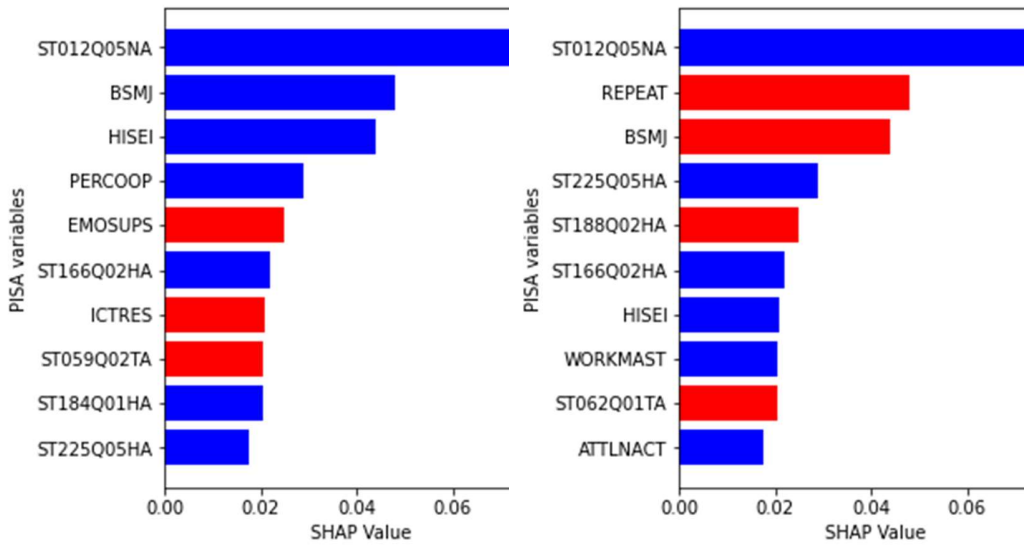
Four variables were consistent significant features for both models for the private and public school students: ST012Q05NA, ST225Q05HA, ST166Q02HA, and HISEI. All four had inverse relationships with identifying poor-performing students, which means that lower scores in the variables were associated with better identification of poor-performing students in mathematics.

ST012Q05NA is the questionnaire item that inquired about how many mobile phones that have internet access there are in the student's home. So having lower values on this item strongly identified poor-performing students in both private and public schools; presumably, these are students with no internet access or no mobile phones at home. ST225Q05HA is the item that asked the students if they expect to complete a vocational degree after high school. So students in both private and public schools who do not expect to complete this post-secondary credential

are more likely to be identified as poor performing in mathematics. ST166Q02HA is a specific item in a set of variables that assess students' view of appropriate responses to receiving a possible SPAM email message. This item refers to checking the email address of the email's sender. Students who say that checking the email address is not an appropriate response are more likely to be identified as poor performing in mathematics, both for private and public school groups. The final important variable was HISEI or the parents' occupational status, which was scored using the international socioeconomic index of occupations (Ganzeboom, 2010). Students whose parents had lower status occupations were more likely identified as poor performing in mathematics in both public and private schools.

Figure 2

Top 10 Most Significant Variables (in Descending Order) in the Random Forest Model Classifier for (a) Private School Participants and (b) Public School Participants



(a) (b)

Note: Red bars represent direct relationships with identifying the poor performing students, whereas blue bars represent inverse relationships with identifying poor performing students. SHAP values represent the level of variable importance relative to other variables.

One variable—BSMJ—was important in both private and public school models but in different directions. BSMJ is the variable that asked the students to indicate their expected job when they are 30 years old, and this was also scored using the international socioeconomic index of occupations (Ganzeboom, 2010). For the private school group, higher expected occupational status negatively indicated the poor performing students; that is, students who indicated lower expected occupations were more likely to be poor performing studies. However, for the public school, the result was reversed. Higher expected occupational status directly indicated the poor performing students; students who expected higher occupational status were more likely to be poor-performing students in mathematics. This is an unexpected but interesting finding that might reflect on how high school students from the two school types think about how their education (and possibly how their mathematics education) relates to the jobs they are likely to have in the future. We discuss this result in more detail in the Discussion section but use this divergent result to begin presenting the other different important variables for private and public school students.

For public school students, three non-cognitive motivation-related variables were important in identifying the poor-performing students in mathematics: WORKMAST, ATTLNACT, and ST188Q02HA. WORKMAST and ATTLNACT are both indexes computed based on responses to a set of items, and both are inversely related to identifying poor-performing mathematics students. WORKMAST represents the motivation and persistence to master given learning tasks, whereas ATTLNACT represents the value of schooling, specifically, the importance of trying hard at school to get a good job or into a good college in the future. So students who had low scores in these two motivational variables are more likely to be identified as poor performers. ST188Q02HA is a specific self-efficacy item that states, “I feel proud that I have accomplished

things" and students who had a high score on this item were more likely to be identified as poor performing. The result seems odd as feeling proud about one's accomplishment is not an emotion that one would associate with poor performance, but that is what the results indicate. Two other important variables are related to the students' school records. REPEAT was a categorical variable that indicated whether the student had previously repeated a grade level, whereas ST062Q01TA referred to how often the student skipped a whole day in school during the past two weeks. Both variables positively identified poor-performing students in public schools.

For private school students, a different set of non-cognitive engagement-related variables were important in identifying the poor-performing mathematics students: PERCOOP, EMOSUPS, and ST184Q01HA. PERCOOP represents the students' perception that cooperation is encouraged in their school and was inversely related to identifying poor-performing students; students' who reported that cooperation is not encouraged in their private school were identified as poor performing in mathematics. EMOSUPS was the index of emotional support from parents, which was directly related to identifying poor-performing students; students who reported having parents who were emotionally supportive were likely to be identified as low performing. ST184Q01HA is the single-item measure of fixed mindset for intelligence (i.e., the belief that one's intelligence cannot be changed) and was inversely related to identifying poor-performing students. Therefore, students who do not believe that intelligence is fixed are more likely to be identified as poor performing in mathematics. The other two important variables were both directly related to identifying poor-performing students. ICTRES was an index of available ICT resources in the students' homes, and this is a broader set of resources compared to ST012Q05NA, which were mobile phones with internet. Interestingly, students with more ICT

resources were identified as poor performing, which may indicate that the ICT devices may not necessarily be used to support learning in mathematics. Finally, ST059Q02TA referred to the number of required class periods for mathematics per week; fewer required class periods identified poor-performing students.

Discussion

The aim of the study was to use machine learning approaches to pinpoint important variables that can be used to identify the poor-performing Filipino students in public and private schools, with the goal of possibly identifying those factors that make students more vulnerable to poor achievement in mathematics. We analyzed data from students in public and private schools separately, assuming that there might be different identifying factors given the different environments and contexts of the two types of schools in the Philippines. Random forest classifiers generated the best performing models for both private and public school groups, and SHAP analysis pointed to notable similarities and differences in the top 10 variables that identified poor-performing students in each school type.

For students in both private and public schools, variables that indicate resource constraints identify poor-performing students, but the constraint goes beyond material disadvantage and relates to aspirational constraints as well. However, the poor-achieving Filipino students are also identified as having lower expectations of completing a post-secondary vocational degree and lower expected occupations when they become adults and also have parents who have low-status occupations. Previous studies have noted how the occupational status of parents also tends to be associated with students' own educational and occupational aspirations (Al-Bahrani et al., 2020; Gutman & Schoon, 2018), with parents' occupational status typically associated with socioeconomic status, as well (Lee & Byun, 2019; Ng & Choo, 2021). The relationship of lower student educational and occupational aspirations with lower achievement is typically associated with less positive motivations and less engagement (Al-Bahrani et al., 2020; Watt et al., 2019).

In the case of the public school students, these less positive motivations were among the important variables in the model; poor-performing public school students were identified by the

low importance they ascribe to trying hard at school to get a good future (ATTLNACT) and lower persistence to master given tasks (WORKMAST). In the contexts of the poor-performing students' low educational and occupational aspirations, it seems to make sense that feeling proud of their school achievement (ST188Q02HA) even if their performance is poor also identifies the poor-performing students in public schools. That is, given their limited expectations, they may also be quite satisfied by their limited achievement.

The variables associated with low educational/occupational aspirations and poor performance in mathematics have a different dynamic among private school students. As mentioned earlier, higher emotional support from parents identified the poor-performing private school students. This might reflect a parenting style that provides unconditional emotional support to the children, which has been shown to be an important factor in academic success among disadvantaged students (Osman et al., 2021). A low fixed mindset (or higher growth mindset) is also said to be associated with higher achievement in mathematics (Hwang et al., 2019), but it also identified poor-performing mathematics students in private schools. It is possible that unconditional emotional support and the growth mindset are constructed differently by these private school students in ways that do not relate to being more motivated and engaged to achieve; these possibilities could be investigated in future research.

The SHAP analysis for both private and public schools shows that fewer (or no) mobile phones with access to the Internet had a very strong impact on identifying the poor-performing student in both private and public schools. This result is consistent with studies showing how mobile phones can promote more positive motivations and higher achievement in mathematics (Güler et al., 2022; Yoon & Yun, 2021). Among private school students, having more ICT resources at home also identifies the poor-performing students, which may be explained by how

such higher access to ICT is used for non-educational purposes. There are some studies that do show how the use of ICT for leisure is negatively associated with mathematics achievement (Petko et al., 2017; Skryabin et al., 2015). Thus, although lack of access to a specific form of ICT resources seems to adversely affect students, in the case of private students, having more ICT resources seems to also have adverse effects on achievement, or at least seems to identify some of the poor performing students in mathematics.

Among public school students, skipping class days was also associated with identifying poor performance in mathematics. A qualitative study of absenteeism in the Philippines found varied reasons why students skip their classes, including feeling helpless in their classes, having mixed priorities, and unappealing learning environments (Clores, 2009), but among students from lower-income families, one reason for skipping classes is not having money for transportation to go to school, or they may be too hungry to go to school on an empty stomach (Jabar, 2021). In the case of public school students, these different reasons might be intersecting and have an adverse impact on the students' learning of mathematics.

We also note that there is one variable related to the students' metacognitive abilities related to potential misinformation on emails. This result that was found in both private and public school students suggests that there might be some specific metacognitive skills that are lacking or not well developed among the Filipino students who are poor in mathematics (Desoete & De Craene, 2019; Lindberg et al., 2010).

Note that other than the last mentioned variable, most of the important variables that identify the poor-performing students in mathematics are variables that relate to resource limitations (i.e., associated with more disadvantaged socioeconomic conditions) and associated motivations and aspirations. Indeed, the poor-performing students tend to have parents with low-

status occupations, who do not have mobile phones that have internet access and may not have as sharp metacognitive skills in dealing with possible false information online, and who have lower educational and occupational aspirations for themselves. In the case of the students from public schools, the relative deprivations seem to be also associated with skipping classes, weaker motivations to persist in task master, lower appreciation of the value of education to succeed in the future, and engagement.

In the case of private school students, there are identifying variables that do not relate to resource limitations and are instead known to be associated with higher achievement in the research literature (e.g., ICT resources at home, emotional support from parents, growth mindset, and classrooms that encourage cooperation). Thus, among private school students, there might be students who perform poorly in mathematics for other reasons. That is, these students are doing badly in mathematics even as there are aspects of their learning experiences and environments that, in theory, should be helping them do well. Earlier, we suggested that these students might be giving different meanings to these positive aspects. For example, the emotional support from their parents might be constructed as being unconditional regardless of how well they do in school and a signal not to try to work harder to achieve. Their low fixed mindset about intelligence might suggest that they do not see their poor performance in mathematics as defining their intelligence and sense of worth. And their perception that the school encourages cooperation among learners might be recognized as an opportunity to rely on others to get by in their mathematics courses. These interpretations are speculative and will need to be further probed in future research studies.

But what the above discussion points to is that although there are common variables that identify poor performing students across schools, there are specificities in the experience and

context of private school students that suggest different identifying variables and vulnerabilities. Indeed, it is possible that if we look at a longer list of important variables, we might also find such specificities in the vulnerabilities in different public school contexts, as well.

Some Recommendations

In this section, we discuss some recommendations that are derived from the findings of the study. We should be clear that the recommendations are inferences that derive from specific findings and not recommended actions that have been directly tested in the current study. We do refer to suggested evidence-based actions that are based on other previous research studies but not based on the current research using the PISA data.

Identification for Prevention and Intervention

The main findings of the study refer to factors that best identify the group of students who are attaining very low levels of achievement in mathematics. Perhaps the most basic recommendation is to pay attention to these students by first identifying them and recognizing their vulnerability in the area of mathematics. Indeed, it is likely that these are the students who typically do not get the mathematics teachers' attention until it is examination time. But knowledge about the profile of these students who are not learning mathematics should be used to help teachers and schools identify who needs attention and possibly different forms of intervention to mitigate the likelihood of failure in mathematics.

Some of the variables that strongly identify poor-performing students are likely to be long-standing conditions, such as having no mobile phone with internet access at home and having parents who have low occupation status. Teachers and schools must recognize the disadvantages these variables bring to the learning experience of particular students and attempt to mitigate the

risks. The mitigation could focus on other important variables that are associated with the resource constraints that the students experience, such as lower educational and occupational aspirations, weaker positive motivations and self-beliefs in schools, among others, which we will discuss further later. But the first point we suggest is that identification of long-standing risk factors associated with poorer socioeconomic and resource constraints should be a cue to action.

As the results show, some of the best identifiers of the students are motivational and behaviors that likely emanate from the constrained economic conditions. Students who indicate no intentions to complete post-secondary degrees are likely to be poor-performing. Students who do not see the value of school for their future success and who do not persist in trying to master tasks can also be identified as at risk of performing poorly in mathematics. Such variables may not be known to the mathematics teachers, but there are other educators in the school, such as the homeroom teacher, guidance counselor, school psychologists, and other teachers who could recognize these other identifying variables in students.

We note that some of the identifiers are different for public and private schools. For example, having a lot of ICT resources at home is also an identifying factor for those in private schools. In this case, the identifier is not associated with resource constraint, but perhaps too much of a resource that it is used for purposes that are not supporting learning in mathematics education. Students who report aspiring for lower status professions are identifiers of poor-performing students in private schools, but the reverse is true among those in public schools, so the use of these identifiers should be done with careful attention.

We note that it is also possible to identify schools or classes within schools that have a higher proportion of vulnerable students. Thus, decision-makers could also rationalize their programs or allocation of resources towards those classes or schools that are more likely to be

hosting more at-risk students based on the identified important factors. The point is that knowledge of the factors that identify poor-performing students should be used to guide actions and decisions at many levels as possible.

Access to Mobile Phones With Internet Access

The issue of lack of internet access and available digital devices was highlighted in the two years of the COVID-19 pandemic. But the results show that two years before the schools were closed for the COVID-19 pandemic, students without mobile phones with internet access were already at risk for poor performance in mathematics. During the pandemic, some local government units and private sector groups addressed this problem by providing devices, while telecom companies tried to expand access to the internet in remote areas. These pandemic-related interventions should be sustained and expanded, as they are likely to benefit more students even post-pandemic. The expansion is important, as it is very likely that this identifying variable is "worse" in remote areas where the access of the telecom companies is poor (perhaps because the "market" is small and unprofitable in these areas. Finally, simple access to devices and the internet is not likely to be sufficient, as students will need to be taught how to use the internet in ways that help them in their mathematics learning.

Metacognitive Ability to Verify Fraudulent Emails

Related to the preceding statement, the results also show that one important identifying variables for both public and private schools is the metacognitive skills ability for dealing with possible false information online. Perhaps because of lack of experience, students who do not think that checking the email address of the sender of an email that contains fraudulent or incorrect information is a good strategy are also identifies as poor in mathematics. Beyond lack of experience, this result might also reflect a deficit in how students are taught to verify

information online. Better students might have the experience and wherewithal to figure things out, but the weaker students may not have the metacognitive skills to engage in this tasks. Perhaps, one possible line of interventions relates to developing the range of metacognitive skills that are useful for sorting out useful information online (and for verifying potentially fake information) might help in students' mathematics learning.

Access to Guidance and Counseling, Coaching, and Psychological Services

Many of the identifying factors related to students' beliefs and goals that are shaped by the students' life experiences, but they can also be shaped and strengthened with the help of professionals trained to help more positive motivations, mindsets, goals, and life strategies. For example, professionals like counselors and psychologists in schools can help students set higher achievement goals and educational aspirations and also guide them to develop more effective learning strategies to engage their mathematics tasks more effectively. Counselors and school psychologists can help students rethink and plan their school and life goals and also create activities in school that will help more students consider and strive toward higher educational targets. Such professionals can also create interventions that can help students plan alternative professional paths that target higher status occupations, which typically might require stronger mathematical proficiency, and also help students develop concrete plans and strategies to attain such targets.

Part of the important work of school counselors and psychologists is to do some basic research to understand where the lower aspirations and less positive motivations in school come from. Are these lower self-expectations coming from models in the family? Or are they messages coming from their teachers or peers in school? Are there school practices that implicitly tell students to set lower aspirations for themselves? Or are these beliefs held by students reflect

norms in their own communities or perhaps based on a lack of information? Before appropriate guidance and interventions can be developed, it is important to know the roots of students' lower personal aspirations and less positive motivations.

Ideally, access to these professional services means access within every school. Indeed, there have been moves to increase the number of counselor and psychologist positions in schools. However, if such is not possible, there can be plans to have such professional services available shared across schools within a district or a division of schools.

Teachers as Facilitators of Positive Motivations and Self-Beliefs

Teachers' instructional and assessment practices are known to shape students' motivation and engagement and also the way students see themselves as learners. More than school counselors and psychologists, mathematics teachers probably have more direct interactions with students that shape their motivations and self-beliefs related to mathematics. The manner in which teachers design the learning activities in mathematics classes can influence students' engagement and persistence, and the way they present mathematics lessons may or may not communicate the value of mathematics in the future lives of students. The way mathematics teachers give feedback to students and communicate the results of mathematics assessments can influence students' beliefs about themselves as mathematics learners and about mathematics as a subject matter. There could be continuing professional education for mathematics teachers, particularly those who are working with students who are identified as at risk of performing poorly. Indeed, it is much easier for mathematics teachers to focus on the students who are doing well, but they need to be capacitated to work more effectively with those at risk of failing, especially in keeping them engaged and motivated to learn in the subject matter.

Parents as Partners in Strengthening Student School Engagement

Although the results only refer directly to parents' occupational status, it is not unreasonable to infer that many of the students' educational and occupational aspirations are also influenced by models and dynamics within their families. Schools can engage the parents as partners in keeping the students engaged in school and in having the students understand the value of persistence in difficult subjects like mathematics. Schools can also help parents appreciate higher educational and occupational goals for their children, which may be challenging when there are limited positive models of such in the children's immediate family and community when the models of success in the community are persons who succeeded without the benefit of educational achievement. Thus, it is important for schools to engage parents to more concretely envisage positive futures for their children that can be attained through educational achievement.

Parents can also help in monitoring students' learning behaviors at home. As the results among the private students suggest, students with more ITC resources at home are more likely to be identified as poor learners in mathematics, and we speculate that this might be due to students using such resources for non-educational purposes (i.e., more leisure). Thus, it is important that parents and other guardians at home be engaged as partners to help address the variables that make particular students more likely to achieve poorly in mathematics.

Review of Pertinent Policies

Some of the above recommendations are likely to be shaped by present policies and practices in the basic education system. The structure and density of the curriculum and the allocation of time for the subject can determine how deeply (or superficially) students are engaged in mathematical concepts. The features of the assessment system, the retention and promotion systems, and other rewards and incentives in schools are likely to shape teachers'

behaviors towards students and students' personal constructions of their own learning experiences and abilities for learning in mathematics. Teachers' workload and class sizes can either constrain or facilitate the level of attention and engagement that teachers have with poor-performing students. The performance appraisal and incentive system for teachers (and school heads and other administrators) could influence how they engage the poor-performing student relative to those who are doing well. We can list so many other practices and procedures, but the point we wish to underscore is that there is a whole system that bears on what goes on in mathematics classrooms and schools that shape how students experience their learning of mathematics.

Conclusions

We applied machine learning techniques to try to identify the variables that identify the poor-performing Filipino students in mathematics, with the assumption that such variables will point to possible vulnerabilities or risk factors associated with poor learning in mathematics. Our study points to a cluster of resource constraint-related variables that include motivational and social cognitive elements and also possible distinct convergences of factors for those in public and in private schools. We note that in the Philippines' Department of Education's report on the 2018 PISA results (Department of Education, 2019), the country's educational policy decision-makers reiterated their four focal thrusts in the efforts to improve student learning: curriculum review and update, improving the learning environment, teacher "upskilling and reskilling," and engaging stakeholders for support. These thrusts have been the focal points of improving mathematics education in the Philippines for many years now (Ogena et al., 2018). Our results call attention to the need to go beyond curricular and instructional factors, as there are elements of the students' social and psychological experiences in school that are important identifiers of

poor-performing students. Although improving the learning environment might be a good entry point to begin addressing these vulnerabilities, the specific ways of enhancing these learning environments might require a deeper understanding of the particular social and psychological factors that make school environments less effective.

Acknowledgment

This research was funded by a research grant from the De La Salle University-Angelo King Institute for Economic and Business Studies (AKI Research Grants 2020-2021 Project No. 500-138), and the research team wholeheartedly acknowledges the support (and patience) of the leadership and staff of the Institute throughout the project.

Declarations

Institutional Review Board Statement and Informed Consent Statement. The study involved secondary analyses of the officially published PISA 2018 dataset. This dataset was downloaded as a public use file from the the website of the Organisation for Economic Co-operation and Development at https://www.oecd.org/pisa/data/2018database/. Therefore, neither consent to participate nor ethics approval were required for the reported analyses.

Data Availability Statement. The data analyzed in this study are available in the PISA 2018 Database page on the website of the Organisation for Economic Co-operation and Development at https://www.oecd.org/pisa/data/2018database/

Conflicts of Interest. The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; and in the writing of the manuscript, or in the decision to publish the results.

References

Al-Bahrani, M. A., Allawati, S. M., Abu Shindi, Y. A., & Bakkar, B. S. (2020). Career aspiration and related contextual variables. *International Journal of Adolescence and Youth, 25*(1), 703–711. https://doi.org/10.1080/02673843.2020.1730201

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Bernardo, A. B. I., Ganotice, F. A., & King, R. B. (2015). Motivation gap and achievement gap between public and private high schools in the Philippines. *The Asia-Pacific Education Researcher, 24*(4), 657-667.

Bernardo, A. B. I. (2021). Socioeconomic status moderates the relationship between growth mindset and learning in mathematics and science: Evidence from PISA 2018 Philippine data. *International Journal of School Educational Psychology, 9*(2), 208–222. https://doi.org/10.1080/21683603.2020.1832635

Bernardo, A. B. I., Cordel, M. O., II, Lucas, R. I. G., Teves, J. M. M., Yap, S. A., & Chua, U. C. (2021). Using machine learning approaches to explore non-cognitive variables influencing reading proficiency in English among Filipino learners. *Education Sciences, 11*(10), 628. https://www.mdpi.com/2227-7102/11/10/628

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440.

Braun, H., Jenkins, F., & Grigg, W. (2006). *Comparing private schools and public schools using hierarchical linear modeling*. U.S. Department of Education, National Center of Statistics, Institute of Education Sciences, U.S. Government Printing Office. https://nces.ed.gov/nationsreportcard/pdf/studies/2006461.pdf

Carbonaro, W., & Covay, E. (2010). School sector and student achievement in the era of standards based reforms. *Sociology of Education, 83*(2), 160–182. https://doi.org/10.1177/0038040710367934

Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. *Large-scale Assessments in Education, 4,* Article No. 12. https://doi.org/10.1186/s40536-016-0030-6

Chen, J., Zhang, Y., Wei, Y., & Hu, J. (2021). Discrimination of the contextual features of top performers in scientific literacy using a machine learning approach. *Research in Science Education, 51*, 129–158. https://link.springer.com/article/10.1007/s11165-019-9835-y

Christenson, S. L., & Havsy, L. H. (2004). Family–school–peer relationships: Significance for social, emotional, and academic learning. In J. Zins, R. Weissberg, M. Wang, & H. Walberg (Eds.), Building academic success on social and emotional learning (pp. 59–75). Teachers College Press.

Clores, M. A. (2009). A qualitative research study on school absenteeism among college students. *The Asia-Pacific Education Researcher, 18*(2), 151–165.

Damrongpanit, S. (2019). From modern teaching to mathematics achievement: The mediating role of mathematics attitude, achievement motivation, and self-efficacy. *European Journal of Educational Research, 8*(3), 713–727. https://pdf.eu-jer.com/EU-JER_8_3_713.pdf

Dela Rosa, E. D., & Bernardo, A. B. I. (2013). Are two achievement goals better than one? Filipino students' achievement goals, deep learning strategies and affect. *Learning and Individual Differences, 27*, 97–101. https://doi.org/10.1016/j.lindif.2013.07.005

Department of Education. (2019). *PISA 2018 Philippine national report*. Department of Education. https://www.deped.gov.ph/wp-content/uploads/2019/12/PISA-2018-Philippine-National-Report.pdf

Desimone, L., Smith, T., & Phillips, K. (2013). Linking Student Achievement Growth to Professional Development Participation and Changes in Instruction: A Longitudinal Study of Elementary Students and Teachers in Title I Schools. *Teachers College Record, 115*(8), 1–53. https://doi.org/10.1177/016146811311500805

Desoete, A., & De Craene, B. (2019). Metacognition and mathematics education: An overview. *ZDM Mathematics Education, 51*(4), 565–575. https://doi.org/10.1007/s11858-019-01060-w

Dong, X., & Hu, J. (2019). An exploration of impact factors influencing students' reading literacy in Singapore with machine learning approaches. *International Journal of English Linguistics, 9*(5), 52–65. https://www.ccsenet.org/journal/index.php/ijel/article/view/0/40486

Fall, A. M., & Roberts, G. (2012). High school dropouts: Interactions between social context, self-perceptions, school engagement, and student dropout. *Journal of Adolescence, 35*(4), 787–798. https://doi.org/10.1016/j.adolescence.2011.11.004

Fung, D., Kutnick, P., Mok, I., Leung, F., Lee, B., Mai, Y., & Tyler, M. T. (2017). Relationships between teachers' background, their subject knowledge and pedagogic efficacy, and pupil achievement in primary school mathematics in Hong Kong: An indicative study. *International Journal of Educational Research, 81*, 119–130. https://doi.org/10.1016/j.ijer.2016.11.003

Ganzeboom, H. B. G. (2010). *A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007*. Paper presented at the Annual Conference of International Social Survey Programme, held in Lisbon, Portugal, on May 1, 2010.

Gjicali, K., & Lipnevich, A. A. (2021). Got math attitude? (In)direct effects of student mathematics attitudes on intentions, behavioral engagement, and mathematics performance in the US PISA. *Contemporary Educational Psychology, 67*, 1–14. https://doi.org/10.1016/j.cedpsych.2021.102019

Gottfried, M., & Ansari, A. (2022). Classrooms with high rates of absenteeism and individual success: Exploring students' achievement, executive function, and socio-behavioral outcomes. *Early Childhood Research Quarterly*, *59*, 215–227. https://doi.org/10.1016/j.ecresq.2021.11.008

Güler, M., Bütüner, S. Ö., Danişman, Ş., & Gürsoy, K. (2022). A meta-analysis of the impact of mobile learning on mathematics achievement. *Education and Information Technologies, 27*(2), 1725–1745. https://doi.org/10.1007/s10639-021-10640-x

Gutman, L. M., & Schoon, I. (2018). Emotional engagement, educational aspirations, and their association during secondary school. *Journal of Adolescence, 67*, 109–119. https://doi.org/10.1016/j.adolescence.2018.05.014

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality, and student achievement. *Journal of Public Economics, 95*(7–8), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Hwang, N., Reyes, M., & Eccles, J. S. (2019). Who holds a fixed mindset and whom does it harm in mathematics? *Youth & Society, 51*(2), 247–267. https://doi.org/10.1177/0044118X16670058

Hyde, J. S., Else-Quest, N. M., Alibali, M. W., Knuth, E., & Romberg, T. (2006). Mathematics in the home: Homework practices and mother-child interactions doing mathematics. *Journal of Mathematical Behavior, 25*(2), 136–152. https://doi.org/10.1016/j.jmathb.2006.02.003

Ilg, T. J., & Massucci, J. D. (2003). Comprehensive urban high school: Are there better options for poor and minority children. *Education and Urban Society, 36*(1), 63–78. https://doi.org/10.1177/0013124503256941

Jabar, M. A. (2021). Qualitative inquiry on parental involvement in children's education: Perspectives of parents, children, and teachers in select elementary schools in the Philippines. *Asia Pacific Journal of Education, 41*(3), 488–502. https://doi.org/10.1080/02188791.2020.1806035

Jay, T., Rose, J., & Simmons, B. (2018). Why is parental involvement in children's mathematics learning hard? Parental perspectives on their role supporting children's learning. *Sage Open, 8*(2), 1–13. https://doi.org/10.1177/2158244018775466

Jesse, D., Davis, A., & Pokorny, N. (2004). High achieving middle schools for Latino students in poverty. *Journal of Education for Students Placed at Risk, 9*(1), 23–45. https://doi.org/10.1207/S15327671ESPR0901_2

Kim, S., & Choi, J. (2021). A cross-cultural examination of socio-psychological resources in mathematics achievement between Korea and the US. *Social Psychology of Education, 24*(4), 1043-1064. https://doi.org/10.1177/0013124519842654

Lam, S. M., & Zhou, Y. (2021). SES-achievement gaps in East Asia: Evidence from PISA 2003–2018. *Asia-Pacific Education Researcher*, 1–20. https://doi.org/10.1007/s40299-021-00620-7

Lazarides, R., Dicke, A. L., Rubach, C., & Eccles, J. S. (2020). Profiles of motivational beliefs in math: Exploring their development, relations to student-perceived classroom characteristics, and impact on future career aspirations and choices. *Journal of Educational Psychology, 112*(1), 70–92. https://doi.org/10.1037/edu0000368

Lee, B., & Byun, S. Y. (2019). Socioeconomic status, vocational aspirations, school tracks, and occupational attainment in South Korea. *Journal of Youth and Adolescence, 48*, 1494–1505. https://doi.org/10.1007/s10964-019-01056-5

Levpušček, M. P., Zupančič, M., & Sočan, G. (2013). Predicting achievement in mathematics in adolescent students: The role of individual and social factors. *Journal of Early Adolescence, 33*(4), 523–551. https://doi.org/10.1177/0272431612450949

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Lockheed, M., & Jimenez, E. (1994). *Public and private secondary schools in developing countries: What are the differences and why do they persist?* (ESP Discussion Paper No. 33). Education and Social Policy Department, The World Bank. https://documents1.worldbank.org/curated/en/712691468743991134/pdf/multi-page.pdf

Lombardi, C. M., & Dearing, E. (2021). Maternal support of children's math learning in associations between family income and math school readiness. *Child Development, 92*(1), e39–e55. https://doi.org/10.1111/cdev.13436

Lubienski, C., & Lubienski, S. T. (2006). *Charter, private, public schools and academic achievement: New evidence from NAEP mathematics data* (Vol. 16). National Center for the Study of Privatization in Education, Teachers College, Columbia University. https://nepc.colorado.edu/publication/charter-private-public-schools-and-academic-achievement

Lundberg, S., & Lee, S. (2017, December 4–9). A unified approach to interpreting model predictions. In *Proceedings of the 31st Conference on Advanced Neural Information Processing Systems* (pp. 4766–4775). https://dl.acm.org/doi/10.5555/3295222.3295230

Maamin, M., Maat, S. M., & Iksan, Z. H. (2021). Analysis of the factors that influence mathematics achievement in the ASEAN countries. *Cypriot Journal of Educational Science, 16*(1), 371–388. https://doi.org/10.18844/cjes.v16i1.5535

Marks, G. N., & Pokropek, A. (2019). Family income effects on mathematics achievement: Their relative magnitude and causal pathways. *Oxford Review of Education, 45*(6), 769–785. https://doi.org/10.1080/03054985.2019.1620717

Mateos, N. E., Zabala, A. F., Palacios, E. G., & Dias-de-Cerio, I. D. (2021). School climate and perceived academic performance: Direct or resilience-mediated relationship? *Sustainability, 13*(1), 1–14. https://doi.org/10.3390/su13010068

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report*. Boston College. https://timss.bc.edu/pdf/t03_download/t03intlmatrpt.pdf

National Economic Development Authority. (2020). *Updated Philippine development plan 2017-2020*. NEDA.

Ng, I. Y. H., & Choo, H. (2021). Parental education and youth educational aspiration in Singapore: A path analysis in institutional and psychological context. *Asia Pacific Journal of Education, 41*(1), 55–69. https://doi.org/10.1080/02188791.2020.1770691

Ogena, E. B., Ubiña-Balagtas, M., & Diaz, R. V. (2018). Philippines: Mathematics and its teaching in the Philippines. In J. Mack & B. Vogeli (Eds.), Mathematics and its teaching in the Asia-Pacific region, (pp. 157–190). World Scientific Publishing Company Pte. Limited. https://www.worldscientific.com/doi/abs/10.1142/9789813272132_0005

Organisation for Economic Co-operation and Development. (2019a). *PISA 2018 results (Vol I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Organisation for Economic Co-operation and Development. (2019b). *PISA 2018 results (Volume V): Effective policies, successful schools*. OECD Publishing. https://www.oecd-ilibrary.org/sites/14bbef20-en/index.html?itemId=/content/component/14bbef20-en

Osman, A., Ydhag, C. C., & Månsson, N. (2021). Recipe for educational success: A study of successful school performance of students from low social cultural background. *International Studies in Sociology of Education, 30*(4), 422–439. https://doi.org/10.1080/09620214.2020.1764379

Petko, D., Cantieni, A., & Prasse, D. (2017). Perceived quality of educational technology matters: A secondary analysis of students' ICT use, ICT-related attitudes, and PISA 2012 test scores. *Journal of Educational Computing Research, 54*(8), 1070–1091. https://doi.org/10.1177/0735633116649373

Pressley, M., Raphael, L., & Gallagher, J. G. (2004). Providence-St. Mel School: How a school that works for African American students works. *Journal of Educational Psychology, 96*(2), 216–235. https://doi.org/10.1037/0022-0663.96.2.216

Saw, G., & Chang, C. N. (2018). Cross-lagged models of mathematics achievement and motivational factors among Hispanic and non-Hispanic high school students. *Hispanic Journal of Behavioral Science, 40*(2), 240–256. https://doi.org/10.1177/0739986318766511

Skryabin, M., Zhang, J., Liu, L., & Zhang, D. (2015). How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education, 85*, 49–58. https://doi.org/10.1016/j.compedu.2015.02.004

Soenens, B., Vansteenkiste, M., Lens, W., Luyckx, K., Goossens, L., Beyers, W., & Ryan, R. (2007). Conceptualizing parental autonomy support: Adolescent perceptions of promotion of independence versus promotion of volitional functioning. *Developmental Psychology, 43*(1), 633–646. https://doi.org/10.1037/0012-1649.43.3.633

Szumski, G., & Karwowski, M. (2019). Exploring the Pygmalion effect: The role of teacher expectations, academic self-concept, and class context in students' math achievement. *Contemporary Educational Psychology, 59*, 1-10. https://doi.org/10.1016/j.cedpsych.2019.101787

Toropova, A., Johansson, S., & Myrberg, E. (2019). The role of teacher characteristics for student achievement in mathematics and student perceptions of instructional quality. *Education Inquiry, 10*(4), 275–299. https://doi.org/10.1080/20004508.2019.1591844

Trinidad, J. E. (2020). Material resources, school climate, and achievement variations in the Philippines: Insights from PISA 2018. *International Journal of Educational Development, 75,* 102174. https://doi.org/10.1016/j.ijedudev.2020.102174

Vesić, D., Džinović, V., & Mirkov, S. (2021). The role of absenteeism in the prediction of math achievement on the basis of self-concept and motivation: TIMMS 2015 in Serbia. *Psihologija, 54*(1), 15–31. https://doi.org/10.2298/PSI190425010V

Villavicencio, F. T., & Bernardo, A. B. I. (2013). Negative emotions moderate the relationship between self-efficacy and achievement of Filipino students. *Psychological Studies, 58*, 225–232. https://doi.org/10.1007/s12646-013-0193-y

Villavicencio, F. T., & Bernardo, A. B. I. (2016). Beyond math anxiety: Positive emotions predict mathematics achievement, self-regulation and self-efficacy. *Asia-Pacific Education Researcher, 25*, 415–422. https://doi.org/10.1007/s40299-015-0251-4

Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning Instruction, 28*, 12–23. https://doi.org/10.1016/j.learninstruc.2013.04.002

Watt, H. M., Bucich, M., & Dacosta, L. (2019). Adolescents' motivational profiles in mathematics and science: Associations with achievement striving, career aspirations and psychological wellbeing. *Frontiers in Psychology, 10*, 1–23. https://doi.org/10.3389/fpsyg.2019.00990

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122. https://doi.org/10.3102/00346543073001089

Webster, B. J., & Fisher, D. L. (2000). Accounting for variation in science and mathematics achievement: A multilevel analysis of Australian data Third International Mathematics and Science Study (TIMSS). *School Effectiveness and School Improvement, 11*(3), 339–360. https://doi.org/10.1076/0924-3453(200009)11:3;1-G;FT339

Wedel, K. (2021). Instruction time and student achievement: The moderating role of teacher qualifications. *Economics of Education Review, 85*, 1-16. https://doi.org/10.1016/j.econedurev.2021.102183

West, A., Noden, P., Edge, A., & David, M. (1998). Parental involvement in education in and out of school. *British Educational Research Journal, 24*(4), 461–484. https://www.jstor.org/stable/1502014

Wigfield, A., Byrnes, J. P., & Eccles, J. S. (2006). Development during early and middle adolescence. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology*, (2nd ed.; pp. 87–114). Lawrence Erlbaum Associates Publishers.

Woessmann, L., & West, M. R. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review, 50*(3), 695–736. https://doi.org/10.1016/j/euroecorev.2004.11.005.

Yamauchi, F. (2005). Why do schooling returns differ? Screening private schools and labor markets in the Philippines and Thailand. *Economic Development and Cultural Change, 53*(4), 959–981. https://www.journals.uchicago.edu/doi/10.1086/429151

Yoon, M., & Yun, H. (2021). Relationships between adolescent smartphone usage patterns, achievement goals, and academic achievement. *Asia Pacific Education Review* https://link.springer.com/article/10.1007/s12564-021-09718-5

Zee, M., Koomen, H. M. Y., & de Jong, P. F. (2018). How different levels of conceptualization and measurement affect the relationship between teacher self-efficacy and students' academic achievement. *Contemporary Educational Psychology, 55*, 189–200. https://doi.org/10.1016/j.cedpsych.2018.09.006

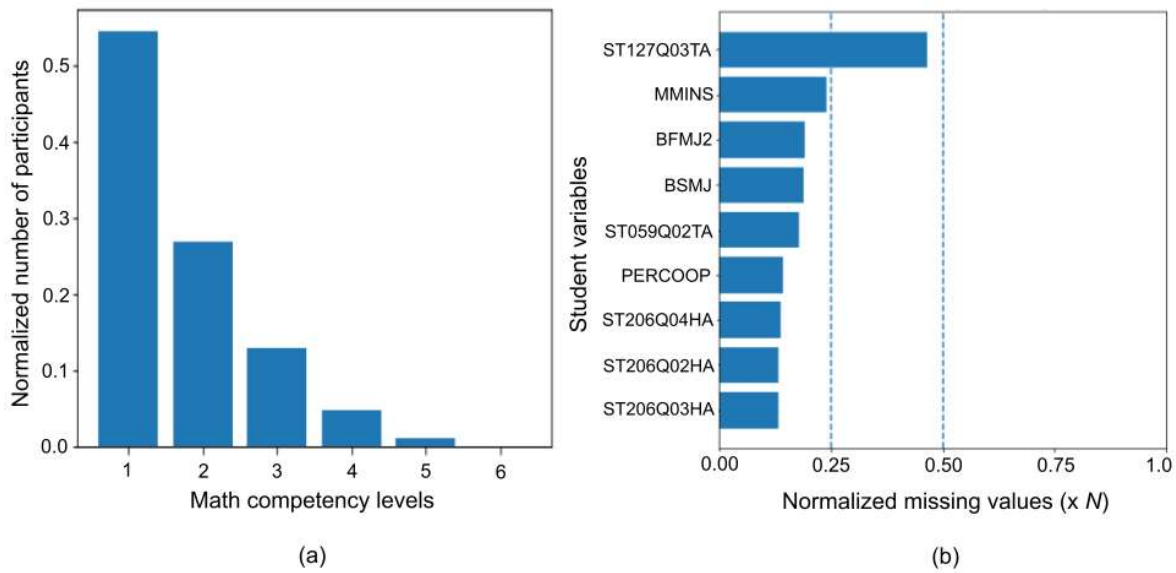
Annex

Additional Data Description and Machine Learning Training Information

The PISA dataset measuring the math proficiency of 7,233 Filipino students is used for this study. As summarized in Figure A1-(a), around 54.60% of the students belong to level 1 math competency, roughly 26.92% have level 2 math competency, while approximately only 18.50% of the students have higher than level 2 math competency. The top variables with the highest number of missing variables are shown in Figure A1-(b).

Figure A1

Normalized Distribution of Math Competency Levels of Filipino Students and Variables With Missing Values



Note: (a) Normalized distribution of math competency levels of FIlipino students. Around 54.60% of students have level 1 math competency while 45.50% have level 2 and higher math competency. (b) Variables with missing values. Note that these variables have missing values for less than 50% of the student participants.

We divided the Philippine data into two groups: data from private schools (SCHTYPE = 1 or 2) and data from public schools (SCHTYPE = 3). The number of students from each school type and the distribution of students with poor and good performance are summarized in Table A1. Each dataset was further split into training and test sets. The training data were used for training the machine learning models, whereas the test data were used for evaluation. To minimize the possibility of having high variance and high bias models, we performed data balancing by oversampling using SMOTE and undersampling using the Tomek links method. The final number of training data after balancing is detailed in Table A2.

Table A1

Data Distribution of Train and Test Sets With 80%-30% Split

School type	Data split	Good performance (Level ≥ 2)	Poor performance (Level = 1)	TOTAL
Private school	Training data	636	288	924
	Test data	147	85	232
	TOTAL	**783**	**373**	**1156**
Public school	Training data	1989	2759	4748
	Test data	489	698	1187
	TOTAL	**2478**	**3457**	**5935**

Note: The total number of processed data is 7,091. Note the imbalance in the number of training samples for the good and poor performing students.

Table A2

Training Data Distribution After Balancing Using the SMOTE-Tomek Links Algorithm. SMOTE and Tomek Links are Oversampling and Undersampling Methods, Respectively

	Poor performance (Level = 1)	TOTAL
Private school	619	1238
Public school	2658	5316

Machine Learning Approaches

In this work, our objective was to discover the key variables that characterize the low-performing and better-performing Filipino students in math through machine learning. Machine learning (ML) algorithms are typically used to discover the intrinsic and highly complex relationship between input data and output data. An exhaustive search approach on the hyperparameters of different machine learning models, namely logistic regression, multilayer perceptron (MLP), support vector machine (SVM), decision tree, and random forest, were performed to zero in the most optimal model for the classification task. Table A3 summarizes the hyperparameters considered in the exhaustive search.

The first two ML models considered (logistic regression and MLP) are the perceptron-type models, which compute the activation outputs $\hat{y}_l$ for each layer of the perceptron-type ML models given the previous activations $\hat{y}_{l-1}$, such that

$$\hat{\mathbf{y}}_l = f^l_{Hl}(\hat{\mathbf{y}}_{l-1}) \tag{1}$$

$$\hat{\mathbf{y}}_{l-1} = \mathbf{x} \quad \text{for } l-1=-1 \tag{2}$$

$$f^l_{Hl}(\hat{\mathbf{y}}_l) = a(\mathbf{w}^{\mathbf{T}}\hat{\mathbf{y}}_{l-1}) \tag{3}$$

where $l = 0$ to L, which denotes the number of hidden layers. Note that $L = 0$ for logistic regression. H is the number of hidden neurons which is essentially the size of the l th hidden layer. The f function denotes the inner product of the weight connection between (l-1)th layer and the l th layer, **w** and the activation function of the (l-1)th layer. Equations (1) and (3) are computed from input

to output. To adjust the weights during training, a backward pass (i.e., from output to input) is performed as guided by the learning hyperparameters, including the gradient descent algorithm, the learning rate, and regularization. See Table A3 for the list of the hyperparameters explored.

Table A3

List of Considered ML Models and the Different Hyperparameters During the Grid Search. Hyperparameters Define the Complexity of the ML Model and Each Model's Learning Performance During the Training

ML models	Hyperparameters
Logistic Regression	*solver*: newton-cg, lbfgs, liblinear *penalty*: none, l1, l2, elasticnet *c*: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100
MLP	*hidden layer sizes*: (10, 30, 10), (10, 30), (32, 32), (10, 10, 10, 10) *activation*: tanh, relu, logistics *solver*: stochastic gradient descent, adam *alpha*: 1e-4, 5e-3, 5e-2 *learning rate*: constant, adaptive
SVM	*kernel*: radial basis function, polynomial *gamma*: 1, 1e-1, 1e-2, 1e-3, 1e-4 *c*: 1e-1, 1, 10, 100, 1000
Decision Tree	*criterion*: gini, entropy *max depth*: 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 30, 40, 50, 70, 90, 120, 150
Random Forest	*criterion*: gini, entropy *number of estimators*: 200, 500 *max features*: auto, sqrt, log2 *max depth*: 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 30, 40, 50, 70, 90, 120, 150

Another ML model considered was the kernel-based type ML model, known as SVM. SVM looks for the most optimal decision plane to optimally separate data into different categories. The SVM decision plane is defined by $\mathbf{w}^T\mathbf{x} + b = 0$, where $\mathbf{x}$ is the input feature vector and $\mathbf{w}$ is the weight vector. SVM training objective looks for representative data or samples, called the support vectors,

that provide maximum margin (see Figure A2, left most plot) between the decision boundary and these support vectors.

For non-linear data, the SVM is extended by transforming both the data and the SVM model to a higher dimension (see Figure A2, right most plot) through the use of transformation kernels. The nonlinearity in kernels can be varied using the kernel parameters. See Table A3 for the evaluated hyperparameters.

Figure A2

SVM Illustration for Linearly Separable Data With Two Categories (Left Most Plot) and Non-Linearly Separable Data Whose Decision Boundary or Optimal Hyperplane (Middle Plot) was Determined by Transforming the Problem to Higher Dimension (Right Most Plot) Using a Kernel

Finally, tree-based models were investigated for the possibility of a more straightforward approach. Tree-based models are more powerful for data for which normality cannot be assumed. Tree-based models split from the top down to their decision nodes, grouping the data into the most homogeneous “sub-nodes” based on their attributes.

A decision tree, as illustrated in Figure A3 left, is a graphical representation of a decision and every potential outcome or result of making that decision. It is composed of the top most node called the *root node*, the decision nodes called the *leaf nodes,* and all other nodes as the *internal nodes*. Although decision trees are very intuitive, these models are prone to overfitting. The random forest model addresses this issue by utilizing several decision tree estimators (see Figure A3 right). The datasets are bootstrapped and features are randomly sampled per estimator to form its training data. The decisions of these trees are then combined using majority voting. The training is monitored using information gain (IG). IG measures the impurity reduction depending on the impurity criterion. See Table A3 for the values explored.

Figure A3

An Example of a Decision Tree (left) and a Random Forest With Four Estimators (Right). Shown in Red are Mock-Up Decisions Based on Input Data