



## Mathematics and Mother Tongue Academic Achievement: A Machine Learning Approach

Catarina Nunes <sup>1</sup>, Ana Beatriz-Afonso <sup>1</sup>, Frederico Cruz-Jesus <sup>1\*</sup>,  
Tiago Oliveira <sup>1</sup>, Mauro Castelli <sup>1</sup>

<sup>1</sup> NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal.

### Abstract

Academic achievement is of great interest to education researchers and practitioners. Several academic achievement determinants have been described in the literature, mostly identified by analyzing primary (sample) data with classic statistical methods. Despite their superiority, only recently have machine learning methods started to be applied systematically in this context. However, even when this is the case, the ability to draw conclusions is greatly hampered by the "black-box" effect these methods entail. We contribute to the literature by combining the efficiency of machine learning methods, trained with data from virtually every public upper-secondary student of a European country, with the ability to quantify exactly how much each driver impacts academic achievement on Mathematics and mother tongue, through the use of prototypes. Our results indicate that the most important general academic achievement inhibitor is the previous retention. Legal guardian's education is a critical driver, especially in Mathematics; whereas gender is especially important for mother tongue, as female students perform better. Implications for research and practice are presented.

### Keywords:

Academic Achievement;  
Education;  
Machine Learning;  
Networks.

### Article History:

<b>Received:</b>	12	June	2022
<b>Revised:</b>	18	August	2022
<b>Accepted:</b>	30	August	2022
<b>Published:</b>	10	September	2022

## 1- Introduction

Academic Achievement (AA), the extent to which a student is successful in school, is a topic that has gained growing interest among researchers in several areas of knowledge [1] as education has significant personal and social impacts. Among these impacts are the positive effects on fundamental social aspects such as developing countries and productivity indexes [2]. Societies constantly face economic and social challenges, and education helps overcome these challenges by acquiring knowledge and highly differentiated skills, thereby contributing to socioeconomic development and more equality. High education rates allow individuals to be well-prepared and to adapt to these changes and to the competition in the job market [3]. Therefore, it is of the utmost importance to understand AA drivers to provide policymakers with adequate strategies to promote academic success.

Traditionally, AA research has relied primarily on surveys and classic statistical analyses [4]. Recent papers, however, have explored macrolevel institutional data collected by schools [5], using machine learning (ML) techniques to extract insights from the data. Institutional data encompasses decades of student information that is essential for educational research, such as demographics, academic background, and academic records. Apart from this individual information, educational databases also collect information regarding schools, class composition, teacher characteristics, and parental features, providing a diverse and comprehensive set of variables that can be studied. Much of these data are typically not examined, however, as accessing and analyzing big data in a school context poses concerns regarding data availability

\* **CONTACT:** [fjesus@novaims.unl.pt](mailto:fjesus@novaims.unl.pt)

**DOI:** <http://dx.doi.org/10.28991/ESJ-2022-SIED-010>

© 2022 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

and privacy issues [6]. Nevertheless, studying such data can yield significant benefits, such as a higher predictability power [7], that can complement traditional methods and greatly contribute to research.

In this paper we use prototypes, i.e., fictional students that are scored with ML models, to measure how a standard deviation in each independent variable (i.e., AA driver) impacts a student's grade keeping all other parameters fixed with mean values. By applying this methodology, we can quantify exactly how much each AA driver impacts students' grades, so we understand what affects students' grades and in particular the extent to which it affects them, which is crucial to provide insights to improve AA. To the best of our knowledge, this is the first time a study can achieve this goal with this detail and sample size. More importantly, this study differs from earlier ones as we combine the best of two approaches, i.e., we combine the machine learning methods (superior) performance with the interpretability of some statistical methods. With this approach we draw theoretical and practical implications from the results. In particular, we answer the following research questions:

- What are the main AA drivers at upper-secondary level?
- How do AA drivers differ between Mathematics and mother tongue?
- How can education researchers and practitioners engender AA?

In answering these questions, this paper is organized as follows: Section two presents the theoretical background; Section three details the methodology; Section four presents the results; Section five has the discussion, implications, and future work; and Section six closes with the conclusions.

## 2- Theoretical Background

Academic Achievement (AA), the extent to which students are successful in school, is usually measured quantitatively in terms of performance: individual students' grades, school achievement exams, standardized test scores, and/or grade point average [8]. It is an expression of a cumulative learning experience across all school years. Understanding AA is crucial for educational decision-makers, since academic difficulties may lead to long-term dropout patterns, academic failure, and problems achieving a successful career in adulthood [9]. Struggling in school can also lead to externalizing problems [10], especially during adolescence and upper-secondary, which represent a critical period with considerable implications for AA. Even though upper-secondary completion is necessary to complete compulsory education in most developed countries, including in Portugal, many students quit school during this period. School dropout is linked to poverty and social exclusion [11] and difficulties entering a successful career in adulthood [12]. Portugal's upper-secondary dropout rate currently exceeds the European average (12.6% vs. 10.7%)., and a thorough understanding of the drivers of AA in upper-secondary can shed light on effective strategies to increase graduation rates and reduce dropout.

Most of the research done in AA has identified diverse factors that impact students' performance, namely characteristics of the students [13], parents [14], teachers [15], and the schools [16] – the primary agents related to the educational process. Other researchers described some influences that contribute to learning and achievement: the student, the home, the school, the classroom, the curricula, the teacher, and teaching and learning approaches [4].

### 2-1-AA Drivers

#### 2-1-1- Students' Characteristics

Students' characteristics are understandably a major factor in explaining AA and students' cognitive ability has been identified as a major determinant of AA [17]. However, this identification has been questioned [18] as a much wider number of factors exist that determine school success. Personal psychological attributes such as personality traits, motivation, academic engagement, self-efficacy beliefs, and grit also determine students' pursuit of learning goals and work efforts and are positive predictors of AA [13].

Motivational and personal predictors stem mostly from student progress across all school years and previous academic results achieved. In fact, a student's track record is a rich source of information and a strong predictor of current achievement. The previous achievement is a positive predictor of AA [19, 20], demonstrating the cumulative nature of learning and knowledge acquisition, and its effects on self-concept [21]. Likewise, the previous retention does not seem to provide great benefits to students with academic difficulties [22], but it even increases the probability that the student will fail a subsequent academic year [5]. Grade retention negatively impacts self-esteem, academic self-concept, and homework completion while contributing to higher maladaptive motivation and absenteeism [23]. Retention has been linked to peer stigmatization during early adolescence and may aggravate behavioral and socioemotional adjustment problems [24].

Research also shows significant differences in AA according to demographic characteristics of students, such as age [5], nationality [25], and gender [26]. Older students, meaning that they were likely previously retained, typically

perform more poorly [27]. The AA of non-national students or those whose parents are immigrants is contingent on how well they have adapted to school [25]. Concerning gender, girls typically attain higher scores than boys [28]. However, boys have better results in specific mathematics abilities [29], and girls outperform boys in reading [30], notably across OECD countries.

Computer use and internet access have increased in recent decades, especially among teenagers, demonstrating noticeable benefits of its applicability in many areas. Their impact on AA is still controversial as both benefits and losses have been identified in research. Although information and communication technologies (ICT) are an essential tool for education, its use for leisure has escalated, especially among children and teenagers. Excessive internet leisure use elicits lower academic results due to sleep disturbances and decreased study time [31]. However, other studies show that moderate computer use and video gaming have positively influenced upper-secondary AA [32]. Concerning internet access, studies have shown that students who more often use the internet tend to have higher scores [33]. Likewise, AA is higher for students who are not dependent solely on a cell phone (i.e., have computer access) for internet access and those with higher digital skills, mainly social media skills.

### ***2-1-2- Parents' Characteristics***

AA is influenced by family characteristics such as socioeconomic status and cultural capital. Family background is crucial for children's development and adjustment to school life. Students with socioeconomic disadvantages show lower communication, language, literacy, and mathematical development. Likewise, early motherhood, low maternal qualifications, low family income, and unemployment predict lower scores at school [28]. Children whose families receive financial support for educational purposes are more likely to have lower grades [34, 35]. AA is positively linked to higher parental income, high education levels, and the parents' occupation type [36, 37]. Young people with highly educated parents are likelier to attain higher AA levels than their peers with low-educated parents. Middle school students with university-educated parents achieve at much higher levels than their peers whose parents did not pursue post-secondary education [38], and parents from low educational backgrounds struggle to value education, give the necessary support, and provide the resources needed for improving AA [39, 40].

Parental involvement in school affairs has a noticeable relationship with AA [41, 42]. It enhances students' ability to cope with schooling activities and promotes appropriate behavioral attitudes that lead to success [43]. Parents who are more involved usually have higher expectations regarding their children's academic path, and these high expectations are linked to increased motivation for work and better results [42]. Participating in school activities seems to be more influential on AA for socioeconomically disadvantaged students [44, 45], and schools play an essential role in helping develop parenting skills [46], helping to diminish the gap derived from pre-existing conditions of students' background.

### ***2-1-3- Schools' Characteristics***

Research has also explored the role of schools' characteristics on student performance. Smaller schools seemingly benefit students with lower socioeconomic status and learning difficulties [47]. Likewise, schools with an average number of students, a higher percentage of teachers on a permanent contract, and a lower proportion of economically disadvantaged students tend to yield better results [4]. AA in Math and Reading falls as school size increases, and these adverse effects of large schools appear to be more salient in higher grades, which is also when schools tend to be the largest. Attending a school with a higher proportion of students from less-educated families harms AA [38].

Having smaller classes has often been suggested as a way of boosting academic success. In fact, smaller classes seem to improve AA, but they may also be more effective in diminishing the achievement gaps [48]. Nevertheless, it is not a consensual effect, as other studies demonstrate that reducing class size is not directly linked to better AA [5, 49, 50].

### ***2-1-4- Teachers' Characteristics***

Teachers' influence is also among the most significant determinants explaining the students' AA [51]. Adolescents' perceptions about teacher connection is the strongest predictor of growth in Math AA in middle and high school, particularly for adolescents from lower socioeconomic background families [52]. Teacher involvement is crucial for developing student perceptions of academic, social, and emotional support and helping students' career planning and decision-making [53]. Teachers with higher expectations about their students' AA help to improve their academic self-concept, which increases AA [44].

Previous research in Portuguese schools assessed the impact of teacher gender, teacher employment situation, education level, and experience and found that female teachers have a more substantial influence on students' AA than male teachers. Teachers working away from home have significant adverse effects on AA. Advanced degree teachers (with MSc or PhD) seemed not to affect AA compared to those with an undergraduate degree. Teachers with more experience seem to be more effective at increasing AA [54]. However, younger teachers typically work away from home in a more precarious work situation, which can account for this difference in teacher age and its influence on AA.

## 2-2-Application of Machine Learning on AA

Researchers and practitioners have resorted to novel data science and artificial intelligence techniques to shed light on AA drivers in recent years. It should be noted that most studies in this regard focus on classic, traditional methods. As more sophisticated data science methods tend to perform better than classic ones, our work uses the first to understand better what drives AA at the upper-secondary level.

One of the first studies using data science and artificial intelligence methods is Musso et al. [55]. The authors used data from 655 students from private universities in Buenos Aires to examine their AA using neural networks with error backpropagation. Despite its high performance, the ability of these authors to quantify each AA driver using neural networks is limited. Hence, to shed some light on AA drivers, some authors used more interpretable methods such as decision trees or random forests. Among these is Şen et al. [56] who examined data from 5,000 8th grade Turkish students to examine placement test scores. ML methods have been demonstrated to be highly effective, particularly decision trees. Abad and López [57] also used decision trees to study the AA of 18,935 upper-secondary students from 99 Mexican schools, which proved to be extremely efficient thanks to its simplicity and high interpretability in understanding the underlying conditions that drive AA. Asif et al. [58] provide another example of a study that used decision trees to examine AA drivers, this time from 210 Pakistani undergraduate students. Even in cases where decision trees or random forests are not the best methods for interpretation reasons, these end up being chosen for their advantages in terms of understanding the pattern between AA and its drivers. Delen [59], for example, compared which ML methods better predict school dropout —using data from 16,066 first-year students from a public university in the United States. As with other studies, although decision trees provided the second-best results, these methods were considered the most appropriate technique. Likewise, Miguéis et al. [60] used random forests, decision trees, and other variants to examine the AA of 2,459 students attending European public universities.

In addition to the methodological issue regarding the trade-off between prediction power and interpretability discussed above, it should also be noted that most studies using sophisticated data science methods use sample data. In fact, this is the case in every example mentioned above. Although using more sophisticated methods has proven to be worthwhile, it is still unquestionable that using sample data is an explicit limitation. There are a few studies using data that do not come from samples. One example is provided by Costa-Mendes et al. [61], who used neural networks, support vector machines, regression, random forests, and an extreme gradient boosting machine, discovering that the last yields the best performance. Another example is given by Cruz-Jesus et al. [62], who compared the performance of a classic method with ML techniques to understand the factors leading to the failure or success of a school year. Random forests proved to be the most accurate procedure.

By examining previous research on AA drivers using machine learning methods, two facts are immediately noticeable: first, despite the most sophisticated approach, several studies using machine learning methods in the context of AA still use samples, which brings some representability and generalization constraints; second, because of the machine learning's "black-box" effect, none of these studies in the author's best knowledge is able to quantify precisely how much each potential driver impacts AA. Our study bridges this gap by combining these two limitations, i.e., we use a machine learning method with data from virtually every Portuguese public upper-secondary student and provide an estimate of how much each AA driver impacts Mathematics and mother tongue grade, providing also a comparison between the two.

## 3- Methodology

This section describes the methodology used in this work and is divided into three subsections. The first one presents a brief overview of the well-known machine learning method used in this paper - neural networks. The second focuses on a typical data science problem related to feature selection, i.e., choosing a subset of the independent variables. Finally, we present the data used in this work. The research methodology is displayed in Figure 1.



**Figure 1.** Predicted impacts of changing a unit in each variable, *ceteris paribus*, using prototypes

### 3-1- Machine Learning Methods

Artificial neural networks, or just neural networks (NN), are biologically inspired methods mimicking the structure of a human brain [63].

Multilayer perceptron (MLP) is an architecture with an input layer, multiple hidden layers, and an output layer. All these layers are formed by neurons, and each neuron in the input layer connects to every neuron in the first hidden layer, which in turn connects to every neuron of the next hidden or output layer. The number of layers and their respective number of neurons define the topology of the neural network. Each connection is associated with a weight, and the set of weights represents the parameters of the model. Thus, the learning process aims at determining the optimal values for the set of weights.

The computation is performed as follows: the neurons in the input layer collect the predictor variables' values and send a signal to every neuron in the first hidden layer. At this point, the inputs of the first hidden layer neurons are linear combinations of the signals received plus a bias. The neurons in the first hidden layer produce, by using an activation function, an output vector that represents the input for the second hidden layer, and the feed-forward process goes on up to the output layer. The output layer, for the considered problem, contains a single layer that is responsible for outputting a prediction for the considered sample in input.

In the second step of the learning procedure (i.e., the backpropagation), the network's parameters (i.e., the weights) are updated based on the feedback signal obtained by comparing the predictions of the neural network against the target values. In particular, backpropagation aims at minimizing the following error function with respect to the neural network's weights:

$$E(X, \theta) = \frac{1}{2N} \sum_{r=1}^N (\hat{y}_r - y_r)^2 \quad (1)$$

where  $X$  is the set of samples,  $N$  denotes the number of observations,  $\hat{y}_i$  the prediction of the model for the  $i^{\text{th}}$  observation, and  $y_i$  is the target value. This is achieved by calculating, for each weight  $w_{ij}^k$ , the derivative of the error function (with respect to the specific weight)  $\frac{\partial E}{\partial w_{ij}^k}$ . where  $w_{ij}^k$  denotes the weight for the neuron  $j$  in the layer  $k$  for the incoming neuron  $i$ . when the calculation of the gradient of the error function is completed, the weights are modified by a quantity equal to  $\Delta w_{ij}^k = -\alpha \frac{\partial E(X, \theta)}{\partial w_{ij}^k}$  where  $\alpha$  is a parameter called learning rate. This process can be repeated until the error is below a predefined threshold or the maximum number of iterations is reached [63].

### 3-2- Feature Selection

Feature selection is a process of reducing the independent variables based on relevance. This process allows for the creation of simpler models, shorter the training times, increases precision and avoids the curse of dimensionality.

Recursive Feature Elimination (RFE) [64] is used to select the optimal number of independent variables (N). The second step was to select the top N features, accordingly to their importance in predicting the target, using RFE, Lasso and Ridge Regressions. By combining the selected variables of these three methods, it was possible to reduce the number of variables on the datasets, from over 50 variables to 12-15 independent variables, depending on the dataset.

#### 3-2-1- Recursive Feature Elimination

RFE is a feature selection method that fits a model and removes the less relevant features until the specified number of features is reached. In particular, RFE selects features by recursively considering smaller and smaller sets of features. First, the model is trained on the initial set of features and the importance (e.g., the absolute value of the coefficients for a linear model) of each feature is obtained. Subsequently, the least important features are removed from the current set of features and the procedure is recursively repeated on the pruned set until the desired number of features to select is reached.

#### 3-2-2- Ridge Regression

Ridge regression penalizes the coefficients of the features allowing for minimizing the error. The coefficients are shrivelled toward zero (and each other), which grants the reduction of complexity and multicollinearity. The ridge estimate is specified as:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j) \text{ subject to } \sum_{j=1}^p \beta_j^2 \leq t \quad (2)$$

where  $N$  denotes the number of samples,  $p$  represents the number of variables, and  $t \geq 0$  is a complexity parameter that controls the amount of shrinkage: the larger the value of  $t$ , the greater the amount of shrinkage [63]. The features selected to be part of the model have higher coefficients.

### 3-2-3- Lasso Regression

Lasso is a form of linear regression that uses a shrinkage and minimizes the prediction error by also considering a constraint on the value of the coefficients. In particular, the model shrinking process penalizes the regression coefficients, and the higher the penalty, the further the estimates are shrunk towards zero. After the shrinking process, the variables shown to be the weakest, the zeros, are eliminated. The solution to the defined Lasso estimate  $L_1$  optimization problem is:

$$\hat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}} \left( \frac{\|Y - X\beta\|_2^2}{n} + \lambda \|\beta\|_1 \right) \quad (3)$$

where  $\|Y - X\beta\|_2^2 = \sum_{i=0}^n (Y_i - (X\beta)_i)^2$ ,  $\|\beta\|_1 = \sum_{j=1}^k |\beta_j|$ , and  $\lambda \geq 0$  is the parameter that determines the strength of the penalty; the larger the value of  $\lambda$ , the greater the amount of shrinkage [64]. In more detail, when  $\lambda = 0$ , no parameters are eliminated and, in this case, the estimate is equal to the one found with linear regression. On the other hand, as  $\lambda$  increases, more and more coefficients are set to zero and eliminated.

### 3-3-Data

An anonymized dataset from the 2018-2019 academic year provided by the Directorate-General of Statistics for Education and Science of the Portuguese Ministry of Education was used. The data include virtually every student attending a public upper-secondary year in Portugal. We include data from the mathematics and mother tongue (Portuguese) national exams used in Portugal for admission to tertiary education. The dataset includes the grades of 19,445 unique students who performed 35,780 exams (14,207 for mathematics exams and 21,573 for Portuguese). We created three target variables measuring AA: the mathematics and the Portuguese exams; and an aggregate grade that combines the two.

## 4- Results

A parameter tuning phase was applied in all the data science methods for the learning phase to yield better results. The parameters that were tested to tune the NNs were the hidden layers, the activation function, the solver, and the learning rate, using a grid search and cross-validation provided by scikit-learn to choose the optimal option. Then the models were trained using the chosen parameters. Cross-validation with repeated K(ten)-fold was implemented to assess each technique's performance and to prevent overfitting. The performance of each method was assessed using the mean R-Squared of all the iterations for each model, i.e., from each K in the cross-validation.

### 4-1-Feature Importance

To examine the features (i.e., AA drivers) included, we started by looking at the significance of the top 10 features of the best models. Feature importance was also assessed using scikit-learn. This method randomly shuffles each feature and computes the models' performance changes. Features significantly impacting the model's performance are considered the most important (see Table 1).

As mentioned above, one of the main contributions of this work is that it provides an exact measure of how much each driver impacts AA. In this sense, while feature importance provides a measure regarding the importance of each feature, it does not show the extent to which each variable impacts the grade. In other words, although it allows us to have an ordinal measure (rank) of importance, it is not particularly useful for understanding how much AA is affected and even less to draw policies to engender AA. This limitation of data science methods is commonly known as the "black-box effect", which is why sometimes researchers prefer to use less powerful, but more interpretable methods, as discussed in Subsection 2.2.

To mitigate this limitation, we use a new approach, creating a dataset made with "prototypes" to understand the importance of the variables in the neural networks' scoring process. A prototype dataset was built for each target (Portuguese, mathematics, and aggregated grades). The baseline is given by a prototype (fictional student) for whom all variables have the average dataset value. Then, for each variable in the dataset, we add two rows (prototypes): one with one standard deviation above the mean; the other with one standard deviation below. For both cases, all other variables are fixed with average dataset values. For binary variables, the rows are added with one or zero. As an example, to quantify the effect of age on AA, we add two prototypes, one with average values for all the variables, except age with one standard deviation below and one above the mean, respectively. For gender, as another example, we also created two prototypes (i.e., rows), one as gender=0 and the other as gender=1. In both cases, the baseline is given by a row with average values for all variables. The score, i.e., the predicted grade that the neural network assigns to each prototype, is then compared with that of the baseline. This difference gives us the weight that each AA driver has on the students' grades (see Table 1).

**Table 1. Predicted impacts of changing a unit in each variable, ceteris paribus, using prototypes**

RANK	Aggregate Grade			Portuguese Exam			Mathematics Exam		
	Variable	Unit Change	$\beta$	Variable	Unit Change	$\beta$	Variable	Unit Change	$\beta$
#1	Stu_Age	0.58	-1.2	Stu_Age	0.61	-1.1	Stu_Age	0.54	-1.5
#2	LG_Educ	3.30	0.9	LG_Educ	3.29	0.9	LG_Educ	3.27	1.5
#3	Sch_FailR	0.08	-0.7	Stu_Fem	1 (Yes)	0.6	Sch_Elem&High	1 (Yes)	-0.9
#4	Stu_Fem	1 (Yes)	0.5	Sch_FailR	0.08	-0.5	Sch_FailR	0.08	-0.8
#5	Stu_EnrolNum	0.19	-0.3	Sch_SocSup	0.06	-0.2	Prof_MscPhD	0 (No)	-0.5
#6	Sch_SocSup	0.06	-0.1	Stu_SocSup	1 (Yes)	-0.2	LG_mother	1 (Yes)	0.3
#7	Sch_ProfDeg	0.20	-0.1	LG_mother	1 (Yes)	0.2	Stu_Fem	1 (Yes)	0.2
#8	Sch_Elem&High	1 (Yes)	-0.1	Sch_Elem&High	1 (Yes)	-0.2	Sch_SocSup	0.06	-0.1
#9	LG_mother	1 (Yes)	0.03	Sch_ProfDeg	0.2	0.1	LG_own	1 (Yes)	0.1
#10	LG_father	1 (Yes)	0.03	Stu_Net	1 (Yes)	0.1	Stu_Net	1 (Yes)	0.1

**Note:** Rate of students who failed the year in that school (Sch\_FailR), student's age (Stu\_Age), education years the legal guardian completed (LG\_Educ), rate of students with school social support (Sch\_SocSup), rate of teachers with MSc or PhD at the school (Sch\_ProfDeg), the student being female (Stu\_Fem), number of enrolments that the student attempted (Stu\_EnrolNum), students having the highest level of social support (Stu\_SocSup), the legal guardian being the mother (LG\_mother), the father (LG\_father), or being the student himself (LG\_own), teachers holding an MSc or a PhD (Prof\_MscPhD), schools offering elementary and upper-secondary (Sch\_Elem&High), and student's access to the Internet (Stu\_Net).

## 5- Discussion, Implications, and Limitations

### 5-1-Discussion of Findings

We assessed the impact of several potential AA drivers in Portuguese public upper-secondary schools. To the best of our knowledge, this work is the first, at least in this research stream, to use prototypes to quantify the impact of each feature on a specific target, i.e., AA drivers on AA. Hence, we bring to sophisticated data science methods some advantages of traditional techniques, i.e., interpretability. Besides the methodological aspect, in this work we used data from the whole national reality and not data based on a sample, which is always a partial representation of a larger reality. We found that the three most crucial AA drivers are student age, the rate of failing students at the school, and the legal guardians' education. More importantly, we can indicate exactly by how much each driver affects mother tongue (PT) and mathematics (MAT) national exams.

Students' age is a proxy for a driver of utmost importance in AA research, i.e., previous retention, as older than average students in a class are synonymous with previous retention. We demonstrate that older students perform more poorly in terms of AA, especially in Mathematics ( $MAT: \beta = -1.5$ ;  $PT: \beta = -1.1$ ). This point is absolutely critical, as it seems that preventing a student from progressing in his/her school path does not prevent future retentions – perhaps the contrary. Moreover, it is known that students who experience retention also suffer from a psychosocial point of view. Although the previous retention is a fairly known AA driver, in this case inhibitor, our study sheds further light on this finding as it indicated that it is even more important for Mathematics. Besides the impact on the individual students, it should be noted that the failing rate at the school level is also among the most important AA (negative) drivers we found ( $MAT: \beta = -0.8$ ;  $PT: \beta = -0.5$ ). Hence, one may question the need to revise the process of failing students across all school years from a pedagogical and policy point of view. It is interesting to notice that the effect of retention is coherent at individual (student) and at school level, as in both cases, it is particularly harmful for mathematics. We should note that we found evidence that retention has deleterious effects on AA for the failing student's progress and the entire school community.

Our results also indicate that the legal guardians' education level is among the most critical AA drivers. The legal guardian's education has a strong positive effect on exam results, as students whose legal guardians have more than 12 years of education (tertiary education) outperform their peers by as much as 1.5 points out of 20. This driver is especially significant in Mathematics ( $MAT: \beta = 1.5$ ;  $PT: \beta = 0.9$ ), an important result to consider as there is a high and rising demand for STEM (science, technology, engineering, and mathematics) skills in the job market [65]. It seems that difficulties in Mathematics are easier to overcome when the legal guardians have tertiary education. Likewise, when this is not the case, a student who lags behind on this subject is significantly penalized, especially considering mathematics' importance to the job market and society[66]. Hence, we posit that enabling teachers to know the degree of education of the legal guardians may be an effective tool to prevent poor AA. We believe this finding is, perhaps, one of the most important of our study. The fact that legal guardian's (often parents) education is a critical AA driver means that education has can promote social mobility but can also inhibit it. A student whose legal guardian has a low level of education is also more likely to underperform in terms of AA as well, thus creating a snowball effect for further generations. Like with previous retention, our study indicates this is special true for mathematics as well.

School-related characteristics are also relevant in the AA context. Starting with the schools' sizes, our results show that in bigger schools, where middle and upper-secondary students are together, upper-secondary exam scores tend to be lower ( $MAT: \beta = -0.1$ ;  $PT: \beta = -0.2$ ), in line with previous research [67]. Schools with higher rates of underprivileged students (e.g., receiving government support) are more likely to present lower levels of AA ( $MAT: \beta = -0.1$ ;  $PT: \beta = -0.2$ ). Likewise, schools with a higher rate of teachers holding an MSc or PhD tend to present slightly higher AA levels, although only in Mathematics ( $MAT: \beta = 0.5$ ). In this context, it is worth mentioning that in Portugal, (public) school enrolment is determined by students' home addresses because students should be as close to home as possible. This criterion is a disadvantage for those living in underprivileged neighborhoods, creating a snowball effect that is challenging to escape. One may question if it would make sense to develop specific learning programs in these locations to prevent the formation of underprivileged clusters.

Additional findings of this study show other relevant AA drivers. We find that females tend to outperform boys, especially in Portuguese ( $MAT: \beta = 0.2$ ;  $PT: \beta = 0.6$ ). Previous studies also report this link [29]. The use of the internet seems to be a positive driver of AA, albeit with a small impact. Support for this link between the internet and AA is far from unanimous in the literature. Some authors find that the internet positively affects AA [32], while others report an adverse effect [68].

It is interesting to note that education itself plays a decisive role in AA. Note that the legal guardians' and teachers' education are among the leading AA drivers. Hence, students with more educated legal guardians and teachers are also more likely to perform better at school. Hence, it is absolutely critical that education stakeholders (researchers, teachers, and policymakers) try to avoid a contagion effect in which underprivileged students will also be legal guardians of underprivileged students in the future. It should be noted that higher education levels lead to more personal development and improved life conditions (i.e., a better job, higher S.E.S., and cultural capital) and that legal guardians in these conditions also help their children succeed at school, it is understandably the path to follow by decision-makers in the education field.

### ***5-2-Theoretical Implications***

Our results contribute to the literature by using sophisticated methods to understand AA at the upper-secondary level. We develop an innovative approach based on prototypes to quantify how much each driver influences AA, comparing the results between what are arguably the two most essential subjects in upper-secondary: mathematics and mother tongue.

Our results are mainly in line with previous research on this topic, as they indicate that the effects of school retention are mainly adverse on AA itself [69]. Our results demonstrate that older (previously retained) students perform more poorly, with greater prejudice toward mathematics. Anderson et al. [24] also posit that retention is a strong predictor of upper-secondary school dropout, mainly because there is an absence of effective strategies to increase competencies and peer stigmatization that may exacerbate behavioral and socioemotional adjustment problems. Any short-term positive academic results from retention tend to fade, and failing students fall further behind their promoted counterparts within a few years [27]. Retention is also related to lower academic self-concept and higher maladaptive motivation and absenteeism [23].

We also show that upper-secondary AA is affected by legal guardians' education, as students with legal guardians with tertiary education have better results in final exams, particularly in mathematics. Despite the fact that the literature shows evidence of the positive impact of legal guardians' education level [36], identifying a university or post-secondary degree as one of the most significant aspects to enhance student AA is a very important and novel finding. According to previous research, legal guardians' education seems to be a driver of AA across all levels, evidencing positive effects from as early as kindergarten [70]. Highly educated legal guardians are typically more involved and supportive in school matters, and they can help students in a way that parents with less cultural capital cannot, by engaging in academic discussions with their children, for instance (Tan et al., 2019). Hence, it is shown that students' home contexts play a decisive role in school experiences and AA. This dynamic has been studied at both the classroom level, where the higher education level of peers' parents improves AA [71], and at the school level, as in middle school, where attending a school with a higher proportion of students from educationally disadvantaged families harms AA [38].

### ***5-3-Implications for Practice***

Besides its theoretical implications, this study also helps shape education practices. By quantifying the impact of each driver on AA precisely, we provide meaningful insights for education policymakers and practitioners to implement substantial adjustments and improve school success. The first practical implication is related to the retention criteria and process. We believe this could be revisited, as school retention has detrimental effects on AA. Interestingly, retention policies have been under recent debate in Portugal. Retention rates have been falling, but further work at the student level (specialized tutoring at school for students at risk of failing, for example) and at the institutional level, namely changing retention criteria, is recommended, according to this study. Eurydice [72] found a prevailing culture of retention in some European countries. In countries where this culture exists, the dominant belief is that repetition benefits learning,

even though recent research and results show the contrary. Some teachers still share this belief, along with school communities and legal guardians, even though it has been repeatedly called into question. Therefore, the challenge may reside more in questioning such assumptions before implementing legislative changes.

Concerning legal guardians' education (a main positive driver of AA found in this study), some measures can be suggested, such as considering the education of the legal guardian when assigning students to classes. This measure will create heterogeneous groups that integrate students with low and high-educated parents or provide teachers with information about parents' education level at the beginning of each academic year, helping to flag students who may need extra support in class, considering the strong impact of parental education on exam grades.

The teacher education level (undergraduate degree vs. MSc or PhD) is also a promising positive driver of AA, so policymakers could facilitate career opportunities for teachers who pursue post-graduate education and motivate schools to hire post-graduate teachers. Policies at the school level are also contemplated: our results show that Internet use can help boost AA, and therefore schools could implement an effective digital reinforcement, allowing all students to have internet access, granting equality of access, and minimizing any prior disadvantages (e.g., students who do not have a computer or internet access at home). Likewise, because our results also show lower exam scores in schools with higher rates of students who receive an allowance, policymakers may want to implement measures to minimize the effects of living in socially and economically disadvantaged territories.

## 6- Conclusion

This study allowed us to understand the primary drivers of AA in upper-secondary public schools in Portugal. Analyzing data from virtually all Portuguese public upper-secondary schools, we verified the negative effect of previous retention on AA and, more importantly, identified new AA drivers, such as legal guardians' tertiary degrees and schools with highly educated teachers. AA has a contagion effect, as parents with university education and teachers with higher degrees (MSc or PhD) are positive AA drivers. Higher education levels lead to more personal development and improved life conditions (i.e., a better job, higher socioeconomic status, and more cultural capital) and seemingly equip parents to better help their children succeed at school. Our results highlight the importance of promoting AA, and provide valuable insights for theory by identifying new determinants and pinpointing their effect, and also for practice, helping to define the most effective strategies to promote success – raising compulsory graduation rates and increasing university enrolments and graduates.

### 6-1-Limitations and Future Work

As with any other work of this nature, ours has some limitations that need to be acknowledged. First, although we use data from virtually every upper-secondary student attending a public school in Portugal, our data still pertain to only one moment (academic year) in time. Hence, changes that may occur in AA drivers may be partially overlooked as time goes by. The second limitation is also related to the data. It should be noted that some potentially important AA drivers have not been included in our study. There is a trade-off between the depth and width of the data, i.e., the more observations we include, the fewer the available variables. Because we use data from the Portuguese Ministry of Education, we were not able to collect specific variables of interest, which could be collected only using surveys and, therefore, through samples. Future research should aim to include multiple years and seek to expand the types of variables used (e.g., sleeping habits or personality traits).

## 7- Declarations

### 7-1-Author Contributions

Conceptualization, C.N., A.B.A., and F.C.J.; methodology, A.B.A., F.C.J., T.O., and M.C.; validation, F.C.J., T.O., and M.C.; writing—original draft preparation, C.N., A.B.A., and F.C.J.; writing—review and editing, T.O. All authors have read and agreed to the published version of the manuscript.

### 7-2-Data Availability Statement

Data were obtained from DGEEC- Direção Geral de Estatísticas da Educação e da Ciência and are available from the authors upon reasonable request with the permission of DGEEC- Direção Geral de Estatísticas da Educação e da Ciência.

### 7-3-Funding

This study was funded by FCT – Fundação para a Ciência e Tecnologia (DSAIPA/DS/0032/2018).

### 7-4-Acknowledgements

We also gratefully acknowledge financial support from FCT Fundação para a Ciência e a Tecnologia (Portugal), national funding through research grant Information Management Research Center – MagIC/NOVA IMS (UIDB/04152/2020).

### 7-5- Informed Consent Statement

Not applicable.

### 7-6- Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

## 8- References

- [1] Koch, A., Nafziger, J., & Nielsen, H. S. (2015). Behavioral economics of education. *Journal of Economic Behavior and Organization*, 115, 3–17. doi:10.1016/j.jebo.2014.09.005.
- [2] Spinath, B. (2012). Academic Achievement. *Encyclopedia of Human Behavior*(2<sup>nd</sup> Ed.), 1-8. doi:10.1016/B978-0-12-375000-6.00001-X.
- [3] OECD. (2019). *OECD Employment Outlook 2019 - The future of work*. Organisation for Economic Co-operation and Development Publishing. doi:10.1787/9ee00155-en.
- [4] Hattie, J., & Clarke, S. (2018). *Visible learning: feedback*. Routledge, London, United Kingdom. doi:10.4324/9780429485480.
- [5] Al-Khafaji, M., & Eryilmaz, M. (2021, November). Using Artificial Intelligence Methods to Predict Student Academic Achievement. In *Proceedings of the Future Technologies Conference*, 403-414, Springer, Cham, Switzerland. doi: 10.1007/978-3-030-89880-9\_31.
- [6] Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, 44(1), 130–160. doi:10.3102/0091732X20903304.
- [7] Philippou, N., Ajoodha, R., & Jadhav, A. (2020). Using machine learning techniques and matric grades to predict the success of first year university students. *2nd international multidisciplinary information technology and engineering conference (IMITEC), IEEE*, 1-5. doi: 10.1109/IMITEC50163.2020.9334087.
- [8] Chowa, G. A. N., Masa, R. D., Ramos, Y., & Ansong, D. (2015). How do student and school characteristics influence youth academic achievement in Ghana? A hierarchical linear modeling of Ghana YouthSave baseline data. *International Journal of Educational Development*, 45, 129–140. doi:10.1016/j.ijedudev.2015.09.009.
- [9] Gottfried, M. A., & Plasman, J. S. (2018). Linking the Timing of Career and Technical Education Coursetaking With High School Dropout and College-Going Behavior. *American Educational Research Journal*, 55(2), 325–361. doi:10.3102/0002831217734805.
- [10] Ansary, N. S., & Luthar, S. S. (2009). Distress and academic achievement among adolescents of affluence: A study of externalizing and internalizing problem behaviors and school performance. *Development and Psychopathology*, 21(1), 319–341. doi:10.1017/S0954579409000182.
- [11] Jahnukainen, M. (2001). Social exclusion and dropping out of education. *International Perspectives on Inclusive Education*, 1, 1–12. doi:10.1016/s1479-3636(01)80003-9.
- [12] Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record*, 103(5), 760–822. doi:10.1111/0161-4681.00134.
- [13] Steinmayr, R., Weidinger, A. F., & Wigfield, A. (2018). Does students' grit predict their school achievement above and beyond their personality, motivation, and engagement? *Contemporary Educational Psychology*, 53, 106–122. doi:10.1016/j.cedpsych.2018.02.004.
- [14] Lee, J. S., & Bowen, N. K. (2006). Parent involvement, cultural capital, and the achievement gap among elementary school children. *American Educational Research Journal*, 43(2), 193–218. doi:10.3102/00028312043002193.
- [15] Ober, T. M., Coggins, M. R., Rebouças-Ju, D., Suzuki, H., & Cheng, Y. (2021). Effect of teacher support on students' math attitudes: Measurement invariance and moderation of students' background characteristics. *Contemporary Educational Psychology*, 66, 101988. doi:10.1016/j.cedpsych.2021.101988.
- [16] Kuziemko, I. (2006). Using shocks to school enrollment to estimate the effect of school size on student achievement. *Economics of Education Review*, 25(1), 63-75. doi: 10.1016/j.econedurev.2004.10.003.
- [17] Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21. doi:10.1016/j.intell.2006.02.001.
- [18] Galindo-Rueda, F., & Vignoles, A. (2005). The declining relative importance of ability in predicting educational attainment. *Journal of Human Resources*, 40(2), 335–353. doi:10.3368/jhr.xl.2.335.

- [19] Costa-Mendes, R., Oliveira, T., Castelli, M., & Cruz-Jesus, F. (2021). A machine learning approximation of the 2015 Portuguese high school student grades: A hybrid approach. *Education and Information Technologies*, 26(2), 1527–1547. doi:10.1007/s10639-020-10316-y.
- [20] Phan, H. P. (2012). Prior Academic Achievement, Effort, and Achievement Goal Orientations: A Longitudinal Examination. *Journal of Educational and Developmental Psychology*, 2(2), 57–71. doi:10.5539/jedp.v2n2p57.
- [21] Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, 98(4), 788–806. doi:10.1037/0022-0663.98.4.788.
- [22] Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420–437. doi:10.1080/02796015.2001.12086124.
- [23] Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal*, 37(5), 739–763. doi:10.1080/01411926.2010.490874.
- [24] Anderson, B. G. E., Whipple, A. D., & Jimerson, S. R. (2011). Mental health outcomes. *Learning Disability Practice*, 14(3), 11–11. doi:10.7748/ldp.14.3.11.s7.
- [25] Portes, A., & Rumbaut, R. G. (2005). Introduction: The second generation and the children of immigrants longitudinal study. *Ethnic and Racial Studies*, 28(6), 983–999. doi:10.1080/01419870500224109.
- [26] Patterson, M. M., & Pahlke, E. (2011). Student Characteristics Associated with Girls' Success in a Single-sex School. *Sex Roles*, 65(9–10), 737–750. doi:10.1007/s11199-010-9904-1.
- [27] Tolen, A., & Quinlin, L. (2017). The Efficacy of Student Retention: A Review of Research & Literature. Available online: <https://dokumen.tips/documents/the-efficacy-of-student-retention-a-review-of-research-remediation-for-students.html?page=1> (accessed on June 2022).
- [28] Mensah, F. K., & Kiernan, K. E. (2010). Gender differences in educational attainment: Influences of the family environment. *British Educational Research Journal*, 36(2), 239–260. doi:10.1080/01411920902802198.
- [29] Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, 36(5), 403–421. doi:10.1016/j.intell.2007.11.002.
- [30] Brunner, M., Gogol, K. M., Sonnleitner, P., Keller, U., Krauss, S., & Preckel, F. (2013). Gender differences in the mean level, variability, and profile shape of student achievement: Results from 41 countries. *Intelligence*, 41(5), 378–395. doi:10.1016/j.intell.2013.05.009.
- [31] Kubey, R. (2001). Internet use and collegiate academic performance decrements: early findings. *Journal of Communication*, 51(2), 366–382. doi:10.1093/joc/51.2.366.
- [32] Bowers, A. J., & Berland, M. (2013). Does recreational computer use affect high school achievement? *Educational Technology Research and Development*, 61(1), 51–69. doi:10.1007/s11423-012-9274-1.
- [33] Torres-Díaz, J. C., Duarte, J. M., Gómez-Alvarado, H. F., Marín-Gutiérrez, I., & Segarra-Faggioni, V. (2016). Internet use and academic success in university students. *Comunicar*, 24(48), 61–70. doi:10.3916/C48-2016-06.
- [34] Archibald, S. (2006). Narrowing in on educational resources that do affect student achievement. *Peabody Journal of Education*, 81(4), 23–42. doi:10.1207/s15327930pje8104\_2.
- [35] Arranz, C. F., Sena, V., & Kwong, C. (2022). Institutional pressures as drivers of circular economy in firms: A machine learning approach. *Journal of Cleaner Production*, 355, 131738. doi:10.1016/j.jclepro.2022.131738.
- [36] Steinmayr, R., Dinger, F. C., & Spinath, B. (2010). Parents' education and children's achievement: The role of personality. *European Journal of Personality*, 24(6), 535–550. doi: 10.1002/per.755.
- [37] Tomul, E., & Savasci, H. S. (2012). Socioeconomic determinants of academic achievement. *Educational Assessment, Evaluation and Accountability*, 24(3), 175–187. doi:10.1007/s11092-012-9149-3.
- [38] Chesters, J., & Daly, A. (2017). Do peer effects mediate the association between family socio-economic status and educational achievement? *Australian Journal of Social Issues*, 52(1), 65–77. doi:10.1002/ajs4.3.
- [39] Coleman, J. S. (1968). Equality of Educational Opportunity. *Equity and Excellence in Education*, 6(5), 19–28. doi:10.1080/0020486680060504.
- [40] Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. doi:10.3102/00346543075003417.
- [41] Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational psychology review*, 13(1), 1–22. doi:10.1023/A:1009048817385.

- [42] Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., & Gaviria, J. L. (2015). Parental involvement on student academic achievement: A meta-analysis. *Educational research review*, 14, 33–46. doi:10.1016/j.edurev.2015.01.002.
- [43] Hill, N. E., & Taylor, L. C. (2004). Parental school involvement and children's academic achievement pragmatics and issues. *Current Directions in Psychological Science*, 13(4), 161–164. doi:10.1111/j.0963-7214.2004.00298.x.
- [44] Benner, A. D., Boyle, A. E., & Sadler, S. (2016). Parental Involvement and Adolescents' Educational Success: The Roles of Prior Achievement and Socioeconomic Status. *Journal of Youth and Adolescence*, 45(6), 1053–1064. doi:10.1007/s10964-016-0431-4.
- [45] Poon, K. (2020). The impact of socioeconomic status on parental factors in promoting academic achievement in Chinese children. *International Journal of Educational Development*, 75. doi:10.1016/j.ijedudev.2020.102175.
- [46] Ali, N., Ullah, A., Shah, M., Ali, A., Khan, S. A., Shakoor, A., Begum, A., & Ahmad, S. (2020). School role in improving parenting skills and academic performance of secondary schools students in Pakistan. *Heliyon*, 6(11). doi:10.1016/j.heliyon.2020.e05443.
- [47] Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research*, 79(1), 464–490. doi:10.3102/0034654308326158.
- [48] Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. doi:10.1111/j.1468-0262.2005.00584.x.
- [49] Wößmann, L., & West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3), 695–736. doi:10.1016/j.euroecorev.2004.11.005.
- [50] Damuluri, S., Islam, K., Ahmadi, P., & Qureshi, N. S. (2020). Analyzing Navigational Data and Predicting Student Grades Using Support Vector Machine. *Emerging Science Journal*, 4(4), 243–252. doi:10.28991/esj-2020-01227.
- [51] Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252. doi:10.1257/0002828041302244.
- [52] Gregory, A., & Weinstein, R. S. (2004). Connection and regulation at home and in school: Predicting growth in achievement for adolescents. *Journal of Adolescent Research*, 19(4), 405–427. doi:10.1177/0743558403258859.
- [53] Wong, L. P. W., Yuen, M., & Chen, G. (2021). Career-related teacher support: A review of roles that teachers play in supporting students' career planning. *Journal of Psychologists and Counsellors in Schools*, 31(1), 130–141. doi:10.1017/jgc.2020.30.
- [54] Sousa, S., Portela, M., & C. Sá, C. (2016). Teacher characteristics and student progress. *School of Economics and Management, University of Minho, Braga, Portugal*.
- [55] Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: a machine-learning approach. *Higher Education*, 80(5), 875–894. doi:10.1007/s10734-020-00520-7.
- [56] Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39(10), 9468–9476. doi:10.1016/j.eswa.2012.02.112.
- [57] Martínez Abad, F., & Chaparro Caso López, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, 28(1), 39–55. doi:10.1080/09243453.2016.1235591.
- [58] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. doi:10.1016/j.compedu.2017.05.007.
- [59] Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), 498–506. doi:10.1016/j.dss.2010.06.003.
- [60] Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51. doi:10.1016/j.dss.2018.09.001.
- [61] Costa-Mendes, R., Cruz-Jesus, F., Oliveira, T., & Castelli, M. (2021). Machine learning bias in predicting high school grades: A knowledge perspective. *Emerging Science Journal*, 5(5), 576–597. doi:10.28991/esj-2021-01298.
- [62] Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon*, 6(6), 4081. doi:10.1016/j.heliyon.2020.e04081.
- [63] Vanneschi, L., & Castelli, M. (2018). Multilayer perceptrons. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 612–620, Elsevier, Amsterdam, netherlands. doi:10.1016/B978-0-12-809633-8.20339-7.
- [64] Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *Research paper in business analytics*, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.

- [65] CEDEFOP. (2018). Insights into skill shortages and skill mismatch: Learning from Cedefop's European skills and jobs survey. European Centre for the Development of Vocational Training, Publications Office of the European Union, Luxembourg.
- [66] OECD (2022). Mathematics performance (PISA) (indicator). Organisation for Economic Co-operation and Development Publishing. doi: 10.1787/04711c74-en.
- [67] Egalite, A. J., & Kisida, B. (2016). School size and student achievement: a longitudinal analysis. *School Effectiveness and School Improvement*, 27(3), 406–417. doi:10.1080/09243453.2016.1190385.
- [68] Rozgonjuk, D., Täht, K., & Vassil, K. (2021). Internet use at and outside of school in relation to low- and high-stakes mathematics test scores across 3 years. *International Journal of STEM Education*, 8(1). doi:10.1186/s40594-021-00287-y.
- [69] Andrew, M. (2014). The Scarring Effects of Primary-Grade Retention? A Study of Cumulative Advantage in the Educational Career. *Social Forces*, 93(2), 653–685. doi:10.1093/sf/sou074.
- [70] Tan, C. Y., Peng, B., & Lyu, M. (2019). What types of cultural capital benefit students' academic achievement at different educational stages? Interrogating the meta-analytic evidence. *Educational Research Review*, 28, 100289. doi:10.1016/j.edurev.2019.100289.
- [71] Wang, T. (2021). Classroom Composition and Student Academic Achievement: The Impact of Peers' Parental Education. *The B.E. Journal of Economic Analysis and Policy*, 21(1), 273–305. doi:10.1515/bejeap-2020-0109.
- [72] European Commission. (2011). Grade retention during compulsory education in Europe : regulations and statistics., Publications Office of the European Union, Luxembourg.