NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Understanding Students' Academic Achievement in Public High School

Evidence for Portugal

Ana Filipa Rosa Louro

Dissertation as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

UNDERSTANDING STUDENTS' ACADEMIC ACHIEVEMENT IN PUBLIC HIGH SCHOOL: EVIDENCE FOR PORTUGAL

by

Ana Filipa Rosa Louro

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor: Frederico Miguel Campos Cruz Ribeiro de Jesus

Co Advisor*:* Jorge Nelson Gouveia de Sousa Neves

May 2018

ACKNOWLEDGEMENTS

This thesis would not be possible without the help of all that are dear to me and also those that worked beside me.

To my parents, my brother and my boyfriend, I'm especially grateful for all the patience and support during the accomplishment of this thesis.

To my friends that always encouraged me to do my best.

To my advisor, Professor Frederico Cruz Jesus, for all the help, support and wise words. Thank you for your advices, contributions and constructive feedback.

To Professor Jorge Neves for all the advices and availability.

To Direção-Geral de Estatísticas da Educação e Ciência, in particular, to Dra. Joana Duarte and Dra. Catarina Afflalo for all the availability and specially for creating the conditions to develop this work and provide me the access to the data, crucial to my work.

It was a pleasure to walk this path with all of you.

ABSTRACT

Several papers and studies have been conducted to better understand what are the main factors that influence students' academic achievement and what measures should be taken to improve it. Therefore, based on 383.560 students' observations, evaluated on secondary Portuguese public schools in 2014/2015 academic year, the purpose of this study is to provide a new approach to the collected data by using Data Mining predictive models. The results show differences on the academic achievement among females and male students, where females got better academic results. Access to computer and Internet found to be powerful tools in education that students can explore to their benefit and show to have a positive influence on academic results. Students benefiting from financial social support prove to have a lower performance in academic achievement. Results also point to the fact that the number of reproves still has a great negative impact on students' academic achievement. This is one of the first studies to the best of the authors knowledge to employ analytic techniques on such a large dataset on the context of academic achievement.

KEYWORDS

Academic Achievement; Predictive Models; Education

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

DGEEC	Direção-Geral de Estatísticas da Educação e Ciência
GLM	General Linear Model
GDP	Gross domestic product
GPA	Grade point average
IEA	International Association for the Evaluation of Educational Achievement
HLM	Hierarchical Linear Modeling
INE	National Institute of Statistics
IRT	Item Response Theory
MEC	Ministry of Education and Science
MLR	Maximum Likelihood
NCES	The National Center for Education Statistics
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
PISA	Programme for International Student Assessment
RDD	Regression Discontinuity Design
SASE	Serviços de Ação Social Escolar
SEM	Structural equation models
SES	Socioeconomic status
TIMSS	Trends in International Mathematics and Science Study

1. INTRODUCTION

Understanding the factors that lead to students' academic achievement is a timeless topic that is not only of universal concern to students, teachers, and families but also to society in general (Jayanthi, Balakrishnan, Ching, Latiff, & Nasirudeen, 2014; Maehr & Zusho, 2009). The role students' academic achievement has on the individuals and overall society is a matter that long concerns the researchers, mainly as a result of the positive effects that it has demonstrated and highlighted on key aspects of society such as development and productivity improvements (B. Spinath, 2012). In fact, it emphasizes the importance of human capital (Barro & Lee, 2001; Neamtu, 2015) increases the knowhow on new businesses and technologies and promote the spreading and transmission of information and knowledge (Hanushek & Wößmann, 2010). As so, students' academic achievement can be used to determine the variation of salaries, growth domestic product (GDP) rate and foster a higher rate of country development (Hanushek & Kimko, 2000; Hanushek & Woessmann, 2012), but also to fight social exclusion and discrimination of minority groups (Dronkers, Van Der Velden, & Dunne, 2012). In a more practice perspective the students' academic achievement at secondary level can act as a catalyst that will determine the progression to the next level of education and subsequently into the world of work (Abosede & Akintola, 2016).

There are several studies and theories that explain what factors influence students' academic achievement, one of the most well-known is promoted by the Organization for Economic Co-operation and Development (OECD) through the Programme for International Student Assessment (PISA), "*a triennial international survey which aims to evaluate education systems worldwide of 15-year-old students"*. The goal is to enhance the added value of this work by studying a global and borderless subject using data from Portuguese students attending public schools in the years 2014/2015 for 10th, 11th and 12th grades, evaluated and attending the 21 courses further carried out on the national examinations, provided by Directorate-General of Statistics of the Portuguese Ministry of Education (DGEEC). The data and the database collected have never been analyzed and studied in the same way as this study does and, as so, the findings will offer further insights that can help demonstrate, furthermore, the status of education and what possible factors may be influencing students' academic achievement in the context of an European country, facilitating, in the medium and long term our understanding of the measures that can be taken to help students achieve better results in the future. Additionally, the aim of this study is also to fill a gap that exists by analyzing academic achievement using data mining techniques, which is an innovative approach as these techniques permit handling large amount of data, finding rules, patterns, thus developing models that can make predictions and support new ideas and, ultimately, theories. In this particular case, the objective is to give evidence on which factors affect students' academic achievement by studying students, parents and school at the same time and provide evidence on why some students achieve better results than others leading to better understandings and conclusions (Hodis et al., 2015).

To pursue the objective expressed above, this work is organized as follows: the first section describes the importance of studying students' academic achievement and distinguish this from previous analysis. The second defines academic achievement and what has been studied on this topic. The third explains our conceptual model and displays the hypotheses tested. The fourth specifies the methodology and exposes our results. The fifth presents the discussion of findings. The sixth presents the main conclusions and finally the limitations and recommendations for future works.

2. THEORETICAL BACKGROUND

2.1. THE CONCEPT OF ACADEMIC ACHIEVEMENT

Academic achievement is an universal topic, and the study of tests related to school success has been initiated since the years of the Second World War mainly due to the correlation between the quality of an educational system and the socioeconomic development of the countries (Steinmayr, Meißner, Weidinger, & Wirthwein, 2016). Being a wide-opened topic, its definition depends and can be easily influence by the possible factors used to measure it. According to Steinmayr, Meißner and Weidinger (2016), "*academic achievement represents performance outcomes that indicate the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments, specifically in school, college, and university. Therefore, academic achievement should be considered as a multifaceted construct that comprises different domains of learning*". Hence, in a linear and more objective perspective is a cumulative function of current and prior family, community, and school experiences (Rivkin, Hanushek, & Kain, 2005). On the other hand, from a perspective of results obtained can be described as individual student marks in a given year, school achievement exams or standardized test scores in core subjects, grades or grade point average (GPA), or even teacher rating scales (Chowa, Masa, Ramos, & Ansong, 2015).

2.2. PRIOR RESEARCH ON ACADEMIC ACHIEVEMENT

Since the 1960s international agencies, such as the International Association for the Evaluation of Educational Achievement (IEA) and (OECD), have performed several international tests (e.g., TIMSS and PISA) regarding students' academic achievement in cognitive areas (Hanushek & Wößmann, 2010). According to what has been described, the literature review shows that one of the first social scientific studies that analysed which factors influenced students' academic achievement started with The Coleman Report that showed, on the one hand, little association between quantity and quality of school contributions and levels of educational attainment, and, on the other hand, the most important determinants of students' educational attainment were family background, but also to the backgrounds of other students in school (Coleman & Hopkins, 1966). Over the years many critiques and improvements were presented to the Coleman's Report particularly to have better insights on the impact of teachers' characteristics have on students' academic achievement (Bowles & Levin, 1968). Supporting this theory, Greenwald, Hedges and Laine (1996) argued that factors such as teacher education, experience, and smaller classes are positively related to students' academic achievements.

Since then, new approaches have emerged and started to create several new models and theories. Walberg (1984) designed a model defending that exists nine majors' factors directly influencing students' academic achievement: the ability or prior achievement, development, motivation, amount of time students engage in learning, quality of the instructional experience, home, classroom, peers and television, these last four factors influence psychologically the learning. Bourdieu's theory (Bourdieu, 1973, 1984) is the most well-known theory of cultural capital. He defends that children from high status backgrounds have an advantage since they share similar cultural understandings as those which inspired and guide the educational system (Marks, Cresswell, & Ainley, 2006). Coleman's theory (Coleman, 1988) is the most well-known theory of social capital struggles that defends that children have better results in schools with more inner network community around them, where

parents, teachers, and the local community interact and enhance educational success. Epstein's theory defends that the combination of psychological, educational, and sociological perspectives can lead the family, school, and local community to influence student achievement (Driessen, Smit, & Sleegers, 2005).

As a result of numerous empirical studies over a very long period of time and join all of the perspectives, the best way to perceive the main causes of students' academic achievement is to understand and study the characteristics of students, parents and schools (Chowa et al., 2015) as conclude when applying hierarchical linear modeling (HLM) due to the multileveled nature of the data, but also teachers' characteristics, although we were not able to retrieve data from teachers resulting in a limitation of this study. In order to clarify the importance of the four characteristics presented before, each of the perspectives and the respective methodology used will be described in detailed below.

Students' Characteristics

Students' characteristics have been studied as a major factor to explain students' academic achievement (C. L. Lee & Mallik, 2015; Patterson & Pahlke, 2011). To obtain these results, the researchers used OLS regressions collecting multi-year data set over 2007-2012 and regression model using data from 2007 from two periods of time, respectively. One of the most studied determinants is gender, where according to research, exist significant gender differences in students' academic achievement, since females in almost all cases outperform males in school results (Mensah & Kiernan, 2010; Steinmayr & Spinath, 2008; Wally-Dima & Mbekomize, 2013) with higher highlight among different scientific areas of studies (Brunner et al., 2013; Ghazvini & Khajehpour, 2011; Steinmayr & Spinath, 2008). While Mensah & Kiernan (2010) used Tobit regression for both univariate and multivariate analysis, Steinmayr & Spinath (2008) applied multiple regression analyses and structural equation models (SEMs) and Wally-Dima & Mbekomize (2013) descriptive statistics. Brunner chose to apply multiple-group factor-analytic models to examine the standard model, the nested-factor model and the non-independence of the student data was estimated by means of the full maximum likelihood method (MLR). However, male students outperformed female students in specific mathematics ability (Brunner, Krauss, & Kunter, 2008), and in the opposite way girls outperform boys in reading (Brunner et al., 2013), specifically across OECD countries (2014). There are still those who conclude that female students also graduate from high school with higher grade point averages (GPAs) than male colleagues (Perkins, Kleiner, Roey, & Brown, 2004). This conclusion came from The National Center for Education Statistics (NCES) of United States using as methodology linear regression parameters, and logistic regression parameters.

Another sociodemographic characteristic that has been pointed as key for this topic is ethnicity, particularly in which way students' ethnicity affects their school performance. In this case Lee (2007) applied multi and individual-level analysis using the classical linear regression model. It's also important to understand whether immigrant children or those whose parents are emigrants have adapted well in schools and how this has influenced their school performance as well as their adaptation in adulthood (Portes & Rumbaut, 2005). In Netherland, for example, *"(non-Western) ethnic minority pupils start and finish primary education with considerable arrears in mathematics and (Dutch) language compared to the native Dutch population*" (Stevens, Clycq, Timmerman, & Van Houtte, 2011). According to OECD results (2012) this behavior is recorded in most OECD countries

where the results shows clearly that students with an immigrant background tend to have lower education performance than native students.

Students are probably the most frenetic and high consumable users of contemporary digital communication technologies (Wentworth & Middleton, 2014). To have a clear understanding on how it affects their actions and consequently their performance at school, it seems of crucial concern and at the same time vital to comprehend if the new teaching approach brings advantages to student performance. With this in mind, technology is being taken in consideration on the traditional indicators of students' academic achievement, including not only the access to internet but also the way students are using it (Torres-Díaz, Duart, Gómez-Alvarado, Marín-Gutiérrez, & Segarra-Faggioni, 2016). For this study, the researchers used a random sample and categorized it in two groups using factor and cluster analysis, for the results they applied a multinomial logistic regression model. As a result, researchers also felt the need, to create a new concept and approach that is characterized by the fusion of the terms education and entertainment – Edutainment, being describe as "*a very interesting combination of traditional content and teaching methods in the context of new technologies*" (Oksana & Elena, 2015). This still however, an underdeveloped topic and therefore it is not permissible to draw enough conclusions from it (Okan, 2003).

Some authors have concluded, through interviews methods, that internet access has shown to be highly correlated with impaired and poor students' academic achievement (Kubey, Lavin, & Barrows, 2001; Liebert & Chou, 2001), although more recent studies have proved that these conclusions are not so clear since children who use Internet with higher frequency have higher scores (Jackson et al., 2006; Torres-Díaz et al., 2016).

Owning and using a home computer is one of those factors that, after literature review, leads to inconclusive results as there are studies that indicate that there is a clear positive relation between having access to the computer at home and better students' academic achievement (Borzekowski & Robinson, 2012; Gil-Flores, 2009). At the same time, there are researchers arguing that using computers in a more frequent way do not necessarily lead to better or higher students' academic achievement (Lei & Zhao, 2007), it can actually lead to a broaden, rather than narrow, math and reading achievement gap (Vigdor, Ladd, & Martinez, 2014), and there are even those who criticize and defend that those students who spent more time on their computer, compared to those who spent less time, have lower GPAs and spend less time studying (Wentworth & Middleton, 2014).

Parents' Characteristics

Parents' characteristics and their involvement has been identified as a possible important key factor for students' academic achievement as demonstrated by Fan & Chen (2001) when applying regressions using GLM or by Hill & Taylor (2004) when referring that parents' involvement "*has a positive influence on school-related outcomes for children*". In fact, this influence occurs at both quantitative and qualitative level. One example is the positive relation between parent involvement at school and parents' higher expectations being associated with higher students' academic achievement, demonstrated after recurring to t tests, chi-square statistics, and hierarchical regressions (J.-S. Lee & Bowen, 2006). Concerning the effect of parental participation and its effect on achieving better results, the literature also reveals a positive correlation between home climate and environment influence on students' academic achievement, being observed by means of interview techniques carried out on the students, through regression analysis and qualitative

research respectively (Codjoe, 2007; Jeynes, 2007; Wilder, 2014). When applying a longitudinal study, by using hierarchical regression analysis and logistic regression analysis, Miedel and Reynolds (1999) demonstrated that parent involvement independently led to greater achievement for children and adolescents. In this sense Barnard (2004) also explained that parents have a positive influence on students as "*parent involvement in school was significantly associated with lower rates of high school dropout, increased on-time high school completion, and highest grade completed*", when a logistic regression and hierarchical linear regression were adopted as well as univariate analysis.

Another well-known factor that has been demonstrated, when recurring to regression techniques, to be relevant and having a positive relationship (Caro, Mcdonald, & Willms, 2009) to the explanation of students' academic achievement is parents' socioeconomic status (SES). Indeed, there is support that this factor is influenced by parents' income, occupation and education level (J.-S. Lee & Bowen, 2006; Sirin, 2005; Steinmayr, Dinger, & Spinath, 2010). However, the research is more expansive with respect to the influence of mother's education on students' academic achievement, presenting that lower maternal education remained significantly linked with lower children' school outcome (Hartas, 2011; Mensah & Kiernan, 2010) as both researches concluded after using mainly univariate analysis. With the current recession, more students are now living in households where their parents are unemployed. As result, does this cause any effect on their educational outcomes? Even though there weren't found many studies on this topic there are those who argue that "*long-term parental unemployment predicts lower educational attainment for children*" (Sandstrom & Huerta, 2013).

Schools' Characteristics

Regarding schools' characteristics, the first focus is mainly in the school and class size due to the fact that it is not difficult to find claims for both sides of the argument about whether their influence can lead to enhancements in learning outcomes (Leithwood & Jantzi, 2009). Class size is one of the topics that has elicited more debates as a result of dubious and different conclusions but also its associated costs (Schanzenbach, 2014). Smaller classes size has been frequently suggested as a way to enhance students' academic achievement when applying OLS regression analyses (Krassel & Heinesen, 2014). In fact, according to Bosworth (2014) results show that not only it improves students' achievement but may also be relatively more effective closing the achievement gaps after using regressions and chi-square tests. A similar perspective provided by Rivkin, Hanushek and Kain (2005) reinforcing, through regression tests, prove that class size reduction, allow students to have better progression at school. A contradictory view is given by other studies proving that using linear regression analysis, the reduction of class size is not directly linked to the better students' academic achievement or school performance (Hoxby, 2000; Wößmann & West, 2006).

According to the existing link between school size and students' academic achievement the truth is the remaining studies reported a negative relationship between increasing school size and achievement supported by regression and cross-sectional regression analysis (Archibald, 2006; Welsch & Zimmer, 2016). Aligned with this theory Egalite and Kisida (2016) expose a clear evidence that students' academic achievement in math and reading declines as school size increases. Opposed to the previous theory there are also those who have found evidence of a positive relationship between larger school size and academic achievement after using linear programming techniques (Barnett, Glass, Snowdon, & Stringer, 2002).

Teachers' Characteristics

The literature also reports to be fairly reasonable, through regression analysis, to assume that teachers' influence is among the most significant determinant to explain the students' academic achievement (Rockoff, 2004) leading to an emergent interest and therefore to a growth on the number of studies on how teachers' characteristics affect the students' academic achievement (Buddin & Zamarro, 2009; Clotfelter, Ladd, & Vigdor, 2006; Goldhaber & Hansen, 2013; Guarino, Reckase, Stacy, & Wooldridge, 2015; Rivkin et al., 2005). It is without surprise that Hanushek (2011) reports teachers as one the most crucial factors to students' academic achievement. However, it is important to know which characteristics most likely explain the teacher' impact on students' academic achievement. A recent research conducted in Portugal (Sousa, Portela, & Sá, 2003), using data from the period between 2010 and 2012, studied the impact of gender, teacher situation, education level and experience by using as methodology OLS regression analysis, concluded that female teachers have higher influence on students' academic achievement than males' teachers and that teachers working away from home have significant negative effects on students' academic achievement. Advanced degree teachers (Masters or PhDs) seemed to have no effect on the lower or higher students' performance when compared to those with a graduation degree. Finally, it's also pointed that teachers with more experience are more effective increasing student achievement gains than those with less experience. The teacher characteristics presented previously suggested that there's a positive correlation between teacher experience on math and reading results when applied mostly OLS regression analysis (Clotfelter et al., 2006; Croninger, Rice, Rathbun, & Nishio, 2007). In addition, students taught by female teachers scored significantly higher than those taught by male teachers in both mathematics and science as conclude Wößman (2003) after applying WLS regressions. The position of teacher education level is an aspect that does not present a clear consensus, since there are different conclusions. For instance, Croninger (2007) reports positive effects between teachers' education level and students' achievement, however, Rivkin (2005) raise serious doubts on this topic.

As displayed in the table below, there are few studies applying students, parents, schools and teacher's characteristics to explain students' academic achievement. As a way to surpass this lack of information, the purpose of this study is to analyze each of these four dimensions at the same time, although there is no information about teachers' characteristics as mentioned before.

Ref	Data	Methods	Students	Parents	Schools	Teachers	Findings
(Hanushek & Kimko, 2000)	Cognitive skills for 39 countries, only 31 countries have the measurement of economic performance	Regression models	x		x		• International mathematics and science test scores are strongly related to growth of nations. • Direct spending on schools has no relationship to student performance differences. • Home-country quality differences of immigrants are directly related to U.S. earnings. • Mathematics and science skills are relevant for the labor force.
(Hoxby, 2000)	Connecticut, USA: 649 elementary schools with data from 1992-1993 to 1997-1998 and 146 elementary districts with data from 1986-1987 to 1997-1998	Regression models	x		x		• Class size does not have a statistically significant effect on student achievement. • Class size reduction has greater effect in schools with more low income or African-American students. • Policy experiments containing incentives produce better results than class reduction.
(Fan & Chen, 2001)	Meta-analysis from 25 different studies	General linear model (GLM)	x	x			• Positive relationship between parental involvement and students' academic achievement, when applying GPA. • Parental home supervision has very low relationship with students' academic achievement. • Strong relationship between parents' aspiration/expectation and students' academic achievement • Low relationship between parental home supervision and students' academic achievement.
(Barnett et al., 2002)	152 secondary schools from Northern Ireland, between 1994-1995 and 1995-1996 academic years	Linear Programming techniques			x		• Positive relationship between effectiveness-efficiency performance scores and secondary school size. • Larger secondary schools perform better than smaller ones.
(Rockoff, 2004)	10,000 elementary-school students and 300 teachers from two districts in New Jersey. In district A between 1989-1990 to 2000-2001 and district B between 1989-1990 to 1999-2000 academic years	Regression models				x	• Large differences in quality among teachers within schools. • Teachers' experience increases student test scores, particularly in reading subject areas.
(Driessen et al., 2005)	Primary school from the Netherlands, with more than 500 schools and 12,000 students, in 1994-1995 academic year	Frequency, Variance and Structural models	x	x	x		• No direct effect of parental involvement on students' academic achievement. • No direct effect on schools with numerous minority pupils where they appear to provide a considerable amount of extra effort with respect to parental involvement.
(Rivkin et al., 2005)	Public school students from Texas. Data for three cohorts between 1993-1995 academic year	Regressions models			x	x	• Class size reduction is not a good predictor to explain students' academic achievement. • Teacher is an important factor to explain school quality.
(Archibald, 2006)	Elementary schools from Nevada, USA, with more than 60,000 students, between 2002-2003 academic year	Hierarchical linear models (HLM)	x		x	x	• Teacher performance is positively related to students' academic achievement. • Per-pupil expenditure at the school level is positively related to students' academic achievement in reading as it indicates what resources matter for education. • Student background characteristics matter, at the student level and school level. • School size and school level poverty have negative impacts on both math and reading results.
(Jackson et al., 2006)	140 children from USA, between December 2000 and June 2002 with an average age of 13.8 years	Internet recorded	x				• Children using more internet have better results in reading achievement than children who used it less. • Despite the age, the use of internet has no effect on students' academic performance.
(J.-S. Lee & Bowen, 2006)	415 children of 3rd until 5th grade from the southeastern United States in 2004 academic year	Hierarchical linear models (HLM)	x	x			• Parents with different demographic characteristics and different types of involvement from dominant groups had the strongest association with achievement. • Parental homework help was negatively associated with European American students' academic achievement. • Parent involvement at school and high educational expectations, displayed the strongest relationship with achievement.
(Marks et al., 2006)	PISA 2000, over 6,000 schools across 32 countries	Item Response Theory (IRT) Regression models	x	x	x		• Cultural factors show to be important to explain socioeconomic inequalities in education. • Cultural resources play a more important role in socioeconomic inequalities in students' academic achievement than material resources at home. • Material and learning infrastructure play a more important role for student performance in mathematics and science than for reading.
(Jeynes, 2007)	Meta-analysis, from 52 studies, between 1972-2000	Regression models		x			• Parental involvement has a positive impact on secondary students' academic achievement.
(Codjoe, 2007)	Sample from black students in Edmonton, Canada	Interviews	x				• Home environment and parental support contribute to students' academic achievement.
(Croninger et al., 2007)	From Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999	Hierarchical linear models (HLM)	x			x	• Teachers' degree type and experience positively affect students' reading achievement. • Teachers' qualifications influence students' academic achievements on reading and mathematics.
(H. Lee, 2007)	80 high schools and 52 middle schools, from USA, with students grades 7 to 12, in 1994	Hierarchical linear models (HLM) Classic lineal regression model	x	x	x		• Peer racial/ethnic composition do not mediate the relationship between school racial/ethnic composition and achievement. • Racial/ethnic composition of schools matters for educational achievement in the USA.
(Lei & Zhao, 2007)	Middle school from Ohio, USA with 237 students, between 2003–2004 academic year	Hierarchical linear models (HLM) ANOVA tests	x				• The quantity of technology use, per itself, is not critical to student learning. • When the quality of technology use is not ensured, more time on computers may cause more harm than benefit. • When GPA changes, technology with higher impact on students were those related to specific subject areas and student development.

(Steinmayr & Spinath, 2008)	342 students from a German school in 11th and 12th graders	Regressions models	x				• Gender differences are presented in most of the variables studied. • Girls' grades were significantly better than boys'. • Personality and motivation play important roles in gender differences in school attainment. • School attainment is a better predictor for girls than for boys to explain gender differences in academic achievement.
(Caro et al., 2009)	Canada's National Longitudinal Study with a sample of 6290 students between 1994-2001 academic years	Hierarchical linear models (HLM) Panel data models	x				• Higher discrepancy in mathematics achievement among students with higher and lower SES families.
(Mensah & Kiernan, 2010)	Millennium Cohort Study, with children in the primary year of school, England, between 2005-2006 academic year	Tobit regression models Univariate and Multivariate analyses	x	x			• Students socioeconomic disadvantages show lower attainment in communication, language and literacy, and mathematical development. • Early motherhood, low maternal qualifications, low family income and unemployment predict lower scores at school. • Gender differences are identified for students in families where: mothers are young, lack of maternal qualifications, or they are living in poor quality areas.
(Hanushek, 2011)	Hanushek and Rivkin (2010)	Regression models				x	• Positive correlation between teachers' effectiveness and marginal gains in students' future earnings.
(Hartas, 2011)	Longitudinal sample from Millennium Cohort Study, from England, for child with 3 and 5 years	Univariate analyses of variance Chi-square tests		x			• Social-economic status does not affect parents' participation in learning activities. • Families' income and parents' education have a strong effect on children's language/literacy (maternal education has a stronger effect). • Socioeconomic disadvantage and lack of maternal educational qualification strongly influence children competencies.
(Patterson & Pahlke, 2011)	Public middle school, in the southwestern United States, with 211 students, between 2007-2011 academic years	Regression models	x	x			• Student characteristics are associated with students' academic achievement. • African American and Latina students tend to have lower grades than other students. • Prior achievement show to be a significant predictor of students' academic achievement. • Gender stereotyping is a significant predictor of students' academic achievement.
(Hanushek & Woessmann, 2012)	64 different countries between 1964 and 2003 years	Regression models	x		x		• School policy can be a key instrument to spur growth. • Differences in cognitive skills lead to differences in economic growth.
(Brunner et al., 2013)	PISA 2003, with 275,369 15th years old students from 41 nations	Multiple group factor analytic models Full maximum likelihood method "MLR"	x				• Girls outperformed boys in reading achievement in all countries studied. • Boys outperformed girls in mathematics achievement in almost all countries studied. • A fully hierarchical conceptualization of achievement, contributes to a better understanding of gender differences.
(Wally-Dima & Mbekomize, 2013)	660 Students from Bachelor of Accountancy degree program at the University of Botswana in 2011-2012 academic year	Descriptive statistics T tests	x				• Individual's commitment and right attitude toward accounting studies are the key factors to explain academic performance. • Female students perform better than male students.
(Bosworth, 2014)	Public school from North Carolina, USA, with 4th and 5th grade students, for 2000-2001 academic year	Regression models	x		x		• Students are assigned to classrooms based on students' characteristics (Gender, Ethnicity, Parents education, others). • Students who struggle in school benefit more from class size reductions when compared with those on the top of the achievement distribution. • Smaller classes have smaller achievement gaps. • Class size reduction is more effective at closing achievement gaps than raising achievement. • Class size effects on both average achievement and achievement gaps are small.
(Krassel & Heinesen, 2014)	Secondary school from Denmark with students of 9th and 10th grade, between 2003-2006 academic years	Regression discontinuity design (RDD) Control for school fixed effects (SFE) Ordinary Least Squares (OLS)	x	x	x		• Negative effects of class size on students' academic achievement.
(Vigdor et al., 2014)	Public school students from 5th to 8th grade, in North Carolina, between 2002-2005 academic years	Probit regression Regression models	x				• Home computer technology is associated with negative impacts on student math and reading scores. • Providing universal access to home computers and high-speed internet access would broaden, rather than narrow, math and reading achievement gaps.
(Hodis et al., 2015)	Secondary schools, from New Zealand, with a sample of 782 students	Hierarchical linear models (HLM)	x				• Maximal levels of aspiration and minimal boundary goals predict students' academic achievement. • Maximal levels of aspiration, minimal boundary goals and students' academic achievement are moderated by the type of assessment tasks.
(C. L. Lee & Mallik, 2015)	Students from the University of Western Sydney, between 2007–2012 academic years	Ordinary Least Squares (OLS)	x				• Positive association between university entry scores and students' academic achievement. • Student performance is related to age and students' grades.

Table 2.1 – Review of prior research on academic achievement

3. CONCEPTUAL MODEL FOR UNDERSTANDING ACADEMIC ACHIEVEMENT

The literature review conducted on the previous section allowed us to be aware of the main antecedents of academic achievement. By combining the results of multiple past studies and theories that supported them, we have built a comprehensive research model to shed some light on what drives academic achievement. Based on the literature, we have identified four contexts that may affect academic achievement, namely the characteristics of students, parents, schools and teachers. However, as there are contradictory findings on the literature, and due to data availability, mentioned below, the last one (teachers) was excluded from the context of this study. Hence, within each of the three constructs that are likely to influence academic achievement, some relationships are hypothesized.

Gender differences is one of most studied characteristics over the years. In fact, the main conclusions refer that female students obtain better academic results when compared with male students (Mensah & Kiernan, 2010; Steinmayr & Spinath, 2008; Wally-Dima & Mbekomize, 2013) despite these conclusions being more pronounced in some academic areas (Brunner et al., 2013; Ghazvini & Khajehpour, 2011; Steinmayr & Spinath, 2008). Usually females tend to have better academic performance. Therefore, we hypothesize:

H1: Gender will have an impact on students' academic achievement as females will perform better.

Students' sociodemographic characteristics, more specifically if the student is native (has the same nationality as the country under study) or immigrant (has other nationality), have been studied from the moment native students have presented better results than immigrant ones (Strand, 2011). In this study, will only be presented if the student has a Portuguese nationality or other. Therefore, we hypothesize:

H2: Native students will perform better on academic achievement.

Computer access is another characteristic that triggers different conclusions. In this context some researchers are less optimistic about the relationship between academic achievement and access to computers (Lei & Zhao, 2007; Vigdor et al., 2014; Wentworth & Middleton, 2014), while the more optimistic consider that computers as a working tool are a benefit for students since we currently live in a digital age and information systems (Borzekowski & Robinson, 2012; Gil-Flores, 2009; Lei & Zhao, 2007; Vigdor et al., 2014). Therefore, we hypothesize:

H3: Students with computer access will perform better on academic achievement.

Internet access is one of the characteristics that has generated more contradictions or less clear conclusions since it can be seen as a distraction, when used excessively (Kubey et al., 2001; Liebert & Chou, 2001), and not for academic purposes, but on the other hand, it can be seen as an added value for the students, providing a wider learning network (Jackson et al., 2006; Torres-Díaz et al., 2016). Therefore, we hypothesize:

H4: Students with internet access will perform better on academic achievement.

Although no reference was found between the number of previous students reprove years and the academic achievement in this study, what is being studied is whether the fact that the student having reproved in previous years or not could in fact be a factor that influences students' academic achievement. Therefore, we hypothesize:

H5: Students that have reproved in the pass will present lower levels on academic achievement in the future.

Family support from Social Services (SASE) is a social benefit which the main purpose is supporting underprivileged families who have children of school age, guaranteeing equal access opportunities and school success for all students in primary and secondary education levels. It also tries to promote socio-educational support measures for the students of households whose economic situation determines the need for financial contributions for school expenses such as the purchase of books and school supplies, meals and transport (DGE, 2018). What we want to find out is whether the students that receive this kind of support, will be impacted on their academic achievement. Therefore, we hypothesize:

H6: Students that receive support from social services (SASE) will have lower levels on academic achievement.

Family financial support is a social benefit attributed monthly to families. The objective is to compensate households' expenses related to the sustenance and education of children and young people. What we want to find out is whether the students that receive this kind of support, will be impacted on their academic achievement (Segurança Social, 2018). Therefore, we hypothesize:

H7: Students that receive family financial support will have lower levels on academic achievement.

Mother education level refers to the level of the academic degree of the mother and is one of the characteristics that present a clear impact on academic achievement (Hartas, 2011; Mensah & Kiernan, 2010), being also one of the variables that is directly linked to the parental SES (Caro et al., 2009; Sirin, 2005; Steinmayr et al., 2010). Father education level refers to the level of the academic degree of the father and is one of the characteristics that seems clear to be interesting to know how can impact the academic achievement, being also one of the variables that is directly linked to the parental SES (Caro et al., 2009; Sirin, 2005; Steinmayr et al., 2010). Therefore, we hypothesize:

H8: Parents education level will have a positive impact on academic achievement.

Class size is the number of students per class. There are several (and to some extent contradictory) conclusions on this topic diverging when it comes to an overall agreement. Although there are researchers who defend that there is no direct relationship between the reduction of number of students per class and the increase of students' academic achievement (Hoxby, 2000; Wößmann & West, 2006), there are those who argue that students benefit a lot when implementing this measure (Bosworth, 2014; Krassel & Heinesen, 2014; Rivkin et al., 2005). Therefore, we hypothesize:

H9: Class size will have a negative impact on academic achievement.

School size is the characteristic that measures the number of students per school. Even though there might be no direct connection between the students' academic achievement and school size, there are those who defend that larger schools have better student´s results (Barnett et al., 2002). The truth is that a link between the dimension of the schools and students' results can be found: bigger schools will have bigger classes. With this in mind, we found evidences that prove the opposite: a negative relationship between school size and students' academic achievement (Archibald, 2006; Egalite & Kisida, 2016; Welsch & Zimmer, 2016). Therefore, we hypothesize:

H10: School size will have a positive impact on academic achievement.

4. METHODOLOGY AND RESULTS

4.1. DATA

To reach the proposed objectives in this study, with the most reliable and complete data, we used data from MISI data base. The MISI database is the information system where educational data concerning pre-scholar, basic, secondary from public schools under MEC (Ministry of Education and Science) and some types of private schools is collected. Its purpose is to centralize all educational data collection from pre-schools, basic and secondary, as well as provide to the respective institutes the necessary information that will serve as basis to the production of educational statistics to the decisions-making processes. The public education context comprises four programs: employees, accounting, students and school social actions. In the context of this paper, the MISI data base was used to collect all the data from students, parents and schools, however we were not able to retrieve data from teachers resulting in a limitation of this study. To better contextualize the data collected, we also used data from Portuguese National Institute of Statistics (INE) to gather information on the students' residence area, specifically on population density, monthly average income, percentage average on culture expenses, aging index, residence population and unemployment rate.

All data from the MISI database regarding students, parents and schools was collected at the DGEEC facilities in Lisbon between November 2016 and January 2017. Programming techniques were used to collect them, in this case SQL, namely SQL Server Management Studio tool. For the treatment of the data collected, data analysis techniques were used, recurring to SAS software, more specifically SAS Guide and SAS Miner tools.

After the appropriate data processing it, was also added data from INE source to better contextualize the data. The final database contains a total of 383560 observations, from Portuguese students attending public schools in the years 2014/2015 for 10^{th}, 11^{th} and 12^{th} grades evaluated and attending the 21 courses where national exams take place, since was excluded Portuguese course reserved to students with severed to profound deafness, so that the analysis would be the most accurate as possible as these comprise the majority of registered students (Gabinete do Secretário de Estado da Educação, 2017), this means: Biology and Geology, Drawing A, Economics A, Philosophy, Physics and Chemistry A, Geography A, Descriptive Geometry A, History A, History B, History of Culture and Arts, Latin A, Portuguese Literature, Foreign Language - German, Foreign Language - Spanish, Foreign Language - French, Foreign Language - English, Mathematics A, Mathematics Applied to Social Sciences, Mathematics B, Portuguese and Portuguese Non Maternal Language.

4.2. DESCRIPTIVE STATISTICS AND NON-PARAMETRIC TESTS

To better understand the conclusions found on the literature review it is vital to perform tests so that our analysis can be as much accurate and complete as possible. However, it is important to understand the correct distribution followed by the data collected, as it is incorrect to assume that all data follows at first sight a normal distribution. As there were suspicions that our dependent variable, final grade/score, did not follow a normal distribution, a Kolmogorov-Smirnov test was used to verify this hypothesis and the results showed that, in fact, there is a high statistical proof that our dependent variable does not follow a normal distribution as the null hypothesis was rejected with a

significance of 1%. Furthermore, it was observed that the variable distribution histogram is asymmetrical, and as so, it supports the previous statement that the variable doesn't follow a normal distribution.

Test for Normality		
Test	Statistic	P-Value
Kolmogorov-Smirnov	D = D = 0,099641	Pr<D <0,0100

Table 4.1 – Normality Test

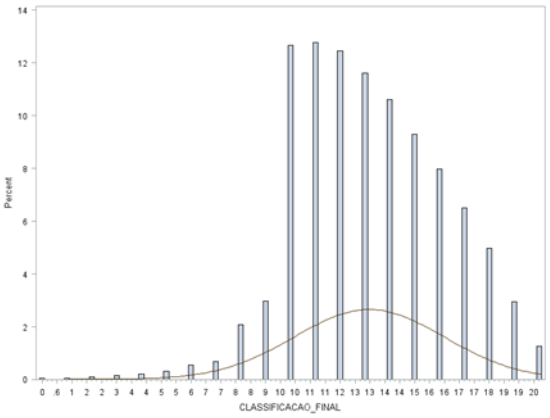

Figure 4.1 – Final Classification do not follow a normal distribution

As we rejected the hypothesis that students' grade follows a normal distribution, to assure there is no violation of statistical tests' assumptions, the choice was to analyse the data through non-parametric tests. In this situation, the Mann-Whitney test was used to compare two independent samples, the Kruskal-Wallis to compare more than two independent samples, but also the variances test, called the Conover test that measures if two or more samples have the same variance, i.e., the same asymmetry of final classifications, W.J. Conover said that "*nonparametric methods use approximate solutions to exact problems, while parametric methods use exact solutions to approximate problems*" (K. M. Ramachandran & Tsokos, 2015).

To begin the explanatory data analysis, we can start by analyzing students' characteristics on this study: students from Portuguese public high-school, in the years 2014/2015 for 10th, 11th and 12th grades, evaluated and attending the 21 courses where the national exams take place, as showed on table 4.2.

As it can be observed, more than half the students is female (55.9%) and those whose age is comprised between 16 and 18 years old are the majority (81.5%). Focusing on the number students reprove years, it is possible to observe that from the full student's sample, the majority has never reproved before 10th (9.4%), 11th (32.5%) and 12th grades (28.5%). The next highest score belongs to students from 10th, 11th and 12th grades that reproved once (2.6%, 11,8% and 12.6%, respectively). It is also possible to conclude that the students who present the lowest rate in this variable are registered in the 10th grade students who reproved twice or more (0.2%). Concerning to nationality, 97.2% of the students are Portuguese and only 2.8% come from other country. Regarding computer

access, 71.2% of the students have access to computers, while 28.8% confirm not have computer access. Internet access has very similar results, with 68.6% of students claiming to have access to the Internet, compared to 31.4% who said they did not.

Regarding the results obtained in the Mann-Whitney test, it is verified that there is statistical evidence of the difference on academic achievement between gender, in other words, female students tend to obtain better results on mean values of academic results (13.38) than their male colleagues (12.90). The results obtained in the Conover test indicate that there are greater asymmetries in the final average score of the male students, this means that the scores obtained by male students are more irregular (females with 2.98 and males with 3.04), there are a great number of students who obtained good results and at the same time others who got poorer results.

Concerning the age variable, it is observable, through Kruskal-Wallis test that at least one of the classes tends to achieve statistical different values than at least one of the other classes. Nevertheless, through the Conover, i.e., Variance test, it is possible to conclude that the lowest asymmetries in students' academic achievement are recorded in the class of students between 19 and 21 years old (2.67), followed by students between 16 and 18 years old (2.99).

Regarding the results obtained in variable N_Reprov, which determine the number of reprove years that the student has until the current year and academic period (2014/2015), Kruskal-Wallis test indicates that within the existing classes of each year there is at least one that registers a higher value, and it is clearly suspected that are the classes of the students that belong to 10^{th} (13.18), 11^{th} (13.77) and 12^{th} (13.64) that have never reproved a year before. On the other hand, the Conover test indicates that on the 10^{th} year there are greater discrepancies on the results for students who have never reproved (3.15), on the 11^{th} on the students who reproved more than twice (3.72) and on the 12^{th} again the category of students who reproved more than twice (3.51).

The statistics tests of the Mann-Whitney test, on variable nationality, show us that there are differences on student's mean values of academic results, with native Portuguese students (13.19) tending to obtain better results in comparison to students of other nationalities (12.37). According to Conover results, both Portuguese and immigrants have equal asymmetries in their school results (3.01).

Regarding students with computer and Internet access, it is possible to conclude that there are differences in their academic achievement when comparing to the ones that do not have access to both technologies. The Mann-Whitney results shows that students who have access to computer (13.17) and Internet (13.20) tend to have better mean values of academic results. However, on the Conover test we observe greater asymmetries in the results of the final average score, obtained among students who have access to computer or internet (3.08).

Variables	n	%	Mean	SD	Mann-Whitney /Kruskal-Wallis (k)	Conover Variance Test
Gender						
Female	62.174	55.9%	13.38	2.98	-2548.4738***[1]	-2.5849***
Male	49.128	44.1%	12.90	3.04		
Age (k)						
[0-16[	153	0.2%	13.98	3.46	14072.9362***	3777.8620***
[16-18[	90.682	81.5%	13.36	2.99		
[19-21[	19.731	17.7%	11.45	2.67		
]>=21]	736	0.7%	11.74	3.66		
N_Reprov by year						
10th,0 rep	10.475	9.4%	13.18	3.15	4050.3054***	372.5845***
10th,1 rep	2.870	2.6%	11.40	2.93		
10th,2 reps	187	0.2%	10.82	2.91		
10th, +2reps	209	0.2%	12.08	2.95		
11th,0 rep	36.124	32.5%	13.77	2.89	16579.2848***	2644.0207***
11th,1rep	13.116	11.8%	11.95	2.65		
11th,2reps	976	0.9%	11.09	2.64		
11th, +2reps	444	0.4%	11.08	3.72		
12th,0 rep	31.725	28.5%	13.64	2.88	9025.3129***	1871.5462***
12th,1 rep	13.990	12.6%	11.61	2.67		
12th,2 reps	785	0.7%	10.93	2.86		
12th, +2 reps	401	0.4%	11.83	3.51		
Nationality						
Portugal	108.134	97.2%	13.19	3.01	730.1785***	-6.0427***
Other	3.168	2.8%	12.37	3.01		
Computer						
0	32.110	28.8%	13.14	2.84	44.3024***	-37.7989***
1	79.192	71.2%	13.17	3.08		
Internet						
0	34.922	31.4%	13.10	2.86	144.2750***	-37.2340***
1	76.380	68.6%	13.20	3.08		

Table 4.2 – Students' Characteristics

Regarding the descriptive analysis of the parents' socioeconomic characteristics, presented on table 4.3, it is verified that most students are not Beneficiary_SASE (73.5%), existing a similar distribution among the students with Beneficiary_SASE in levels 1 (13.7%) and level 2 (12.8%). The Kruskal-Wallis test indicates that, in fact, there are differences in students' academic achievement when comparing the results of mean values of academic results on the three levels mentioned above. According to the results it is suspected that students with no support from SASE are the ones who obtained better results on mean values of academic results (13.33). On the other hand, the Conover test shows that

[1] For a significance level of 1%, we reject the null hypothesis (p-value <0.0001)

the range where there are greater asymmetries in the results of mean values of academic results is recorded in the students that are not Beneficiary_SASE (3.05).

Observing students who received financial support (FFS), and similarly to the results obtained above, it is verified that 73.1% of the students do not receive any financial support. It is also possible to observe that 12% receive the supports that are established in level 1, with highest financial support, followed by the students from level two (13.7%) and after followed by the students that are included in level 3, the lowest support (12%). The Kruskal-Wallis test indicates that, in fact, there are differences in students' performance when the results of the mean values of academic results are acquired in the four classes presented. With the results presented above it is suspected that students with no financial support (FMS) are the ones who obtained better results on mean values of academic results. On the other hand, the Conover test indicates that the interval where there are greater asymmetries in the results of the final average classification is recorded in students who do not have financial support at all (3.05).

Variables	n	%	Mean	SD	Kruskal-Wallis	Conover Variance Test
Beneficiary_SASE						
No Support	81.787	73.5%	13.33	3.05	3119.6251***	1080.8923***
Level 1 (Highest support)	15.215	13.7%	12.89	2.88		
Level 2 (Highest support)	14.300	12.8%	12.59	2.87		
Family Financial support (FFS)						
No Support	81.406	73.1%	13.32	3.05	2920.3323***	1151.4959***
Level 1 (Highest support)	13.351	12.0%	12.59	2.85		
Level 2 (Medium support)	15.242	13.7%	12.89	2.88		
Level 3 (Lowest support)	1.303	12.0%	13.22	2.94		

Table 4.3 – Parents' Socioeconomic Characteristics

Although there are no results on variables parent's educational level and parents' professional characteristics, due to the high number of missing values, it's possible to conclude, with data provided by students who produced results, that students whose both parents have PhD degree are the ones that have the highest mean values of academic results, opposing those coming from the most disadvantaged socioeconomic families and with a lower educational level (middle school) as the ones that have the lowest mean values of academic results.

To what refers to schools' characteristics on table 4.4, more specifically to variable class size, it is verified that most of the students are included in classes between 26 and 30 students (46.8%). The second highest percentage is registered in groups that have between 21 and 25 students (23.3%), followed by classes that have between 31 and 40 students (15.4%). The Kruskal-Wallis test indicates that there are significant differences in the mean values of the students' classification where at least one of the classes obtains better results than the others. According to the results it is suspected that the groups composed between 26 and 30 students score the highest mean values (13.29 values) whereas the class with the lowest mean values is registered in classes with more than 40 students (12.15). The Conover test indicates that the class with the highest mean values asymmetries is the class composed by more than 40 students per class (3.44). On the opposite way, the one with the

smallest asymmetries is the smallest classes: 10 or less than 10 students and between 10 and 20 students per class (2.93), followed by classes composed between 21 and 25 students per class (2.97).

Focusing on variable school size, it is possible to conclude that the largest proportion of students is registered on schools with a composition between 301 and 500 (21.8%) followed by those registering more than 900 students (20.1%) and finally schools between 701 and 900 students (17.8%). The remaining students' population is equally spread on schools from other classes. It should be also noted that schools between 601 and 700 students contain 12.7% of students. As for the results obtained through the Kruskal-Wallis test, it is possible to conclude that at least one of the classes obtains higher mean values when compared to the others, and it is suspected that the classes where this occurs are in schools between 501 and 600 students, and with more than 900 students (13.34 for both). With a different outcome, the Conover test indicates that the highest asymmetries occur in classes corresponding to schools that have 100 or fewer students (3.14), followed by those between 701 and 900 students (3.08). In contrast, the range with the lowest asymmetries in the mean values is in schools between 301 and 500 students (2.92).

Variables	n	%	Mean	SD	Kruskal-Wallis	Conover Variance Test
Class Size						
[1-10]	864	0.8%	12.66	2.93	971.5192***	588.0421***
]10-20]	14.389	12.9%	13.01	2.93		
]20-25]	25.974	23.3%	13.16	2.97		
]25-30]	52.046	46.8%	13.29	3.05		
]30-40]	17.152	15.4%	12.93	3.02		
]>=40[	877	0.8%	12.15	3.44		
School Size						
[1-100]	2.469	2.2%	12.83	3.14	1092.5769***	634.9647***
]100-200]	8.954	8.0%	12.85	3.01		
]200-300]	9.293	8.3%	12.93	3.02		
]300-500]	24.223	21.8%	13.11	2.92		
]500-600]	10.073	9.1%	13.34	3.01		
]600-700]	14.119	12.7%	13.30	3.01		
]700-900]	11.762	17.8%	13.12	3.08		
]>=900[	22.352	20.1%	13.34	3.04		

Table 4.4 – Schools' Characteristics

From within the 21 courses where national exams take place on table 4.5, the analysis will be conducted only on those that constitute the ground basis for the four major type courses on the secondary degree: Science and Technologies, Social-Economic Sciences, Languages and Humanities and Visual Arts – Portuguese, Foreign Languages – English and Philosophy, as well as mandatory school disciplines on the 4 courses mentioned above – Mathematics A, History and Drawing A.

From here it is possible to conclude, through the descriptive analysis, as expected, that the subjects with the largest number of students enrolled, are the four subjects of the general formation of the students. The subject with the highest number of students, either female or male, is Portuguese

(11.3% and 8.6%, respectively), followed by Philosophy (8.5% and 6.7%, respectively) and Foreign Language – English (8,5% and 7,0%, respectively). The subject with the lowest number of students enrolled is Drawing A, with 1.1% female students and 0.5% male students. Regarding the analysis of the non-parametric Kruskal-Wallis test, it is possible to conclude that at least one of the classes obtains higher mean values, compared to the others, and it is suspected that this situation occurs in the subjects of Drawing A (14.75 for females and 14.14 for male students) and in Foreign Language – English (14.41 for females and 14.51 for male students). On the other hand, the subject that tends to obtain lower mean values is History A (12.50 for females and 12.09 for male students), followed by Mathematics A (13.01 for females and 12.43 for male students) and Portuguese (13.20 for females and 12.30 for male students). The Conover test indicates that, clearly, the subject that registers the greatest asymmetry of results is Mathematics A, for both genders (female with 3.49 and male with 3.56), followed by the Foreign Language – English (female with 3.16 and male with 3.00). Finally, the subject that presents the lowest asymmetry is Drawing A, for the female (2.32) and male (2.42).

Variables	n	%	Mean	SD	Kruskal-Wallis	Conover Variance Test
Courses by Gender						
Drawing A, Female	4.359	1.1%	14.75	2.32	89.1649***	1.9836***
Drawing A, Male	2.021	0.5%	14.14	2.42		
Philosophy, Female	32.715	8.5%	13.76	2.84	1038.7988***	-6.2612***
Philosophy, Male	25.705	6.7%	13.00	2.83		
History A, Female	13.506	3.5%	12.50	2.66	104.3030***	-8.9812***
History A, Male	6.465	1.7%	12.09	2.54		
Foreign Language – English, Female	32.701	8.5%	14.41	3.16	15.4312***	-14.2941***
Foreign Language – English, Male	26.645	7.0%	14.51	3.00		
Mathematics A, Female	23.757	6.2%	13.01	3.49	314.4330***	-1.5765***
Mathematics A, Male	23.302	6.1%	12.43	3.56		
Portuguese, Female	43.285	11.3%	13.20	2.47	2508.7112***	-5.6441***
Portuguese, Male	33.148	8.6%	12.30	2.46		

Table 4.5 – Courses by Gender

To what the non-parametric statistical tests is concerned, one limitation should be acknowledged, that is, naturally, also a limitation of the present study. Due to the high number of students, the sample size in the tests is very large, in the order of dozens of thousands. Hence, rejecting the null hypothesis in the non-parametric tests is something very likely to happen, even when the differences have no practical significance.

4.3. DECISION TREES

The data mining technique applied in this study are decision trees. Decision trees are a data mining technique that are very popular given the fact that there are very easy to interpret and to implement their findings in any human or automated decision-making process. Although they are not among the most powerful methods for prediction, decision trees entail several advantages, especially the one related with the fact that they are non-parametric models, i.e., they have virtually no assumptions whatsoever regarding the data type and its characteristics such as variables' distributions, outliers or missing values. Decision trees are hierarchical collections of rules that describe how to divide a large

collection of records into successful smaller groups of records consider the goal of maximizing the split in the dependent variable classes. First, the decision tree splits the data into smaller cells independently. To find a new split, the algorithm test splits based on all variables for every value possible. Secondly, it uses the target variable to determine how each input should be partitioned. In the end, each segment will form a decision tree (Berry & Linoff, 2011). As a non-parametric supervised learning method, the goal is to create a model that predicts the value of a target variable by learning simple decision rules provided by the data features (Cabot, De Virgilio, & Torlone, 2017). Finally, another important characteristic of decision tree is that the first variable is always the most important, or the most significant, to predict the target one. Hence a variable that is chosen to split the data first is always more relevant in the context of better defining individuals regarding the dependent variable under study, than the second. This is, to some extent, also a limitation as it tends to turn decision trees into "*greedy*" algorithms, in the sense that choosing the most important variable at the beginning does not assure that this would be the "optimal choice" once the tree is estimated. Nevertheless, this characteristic in the context of our study will be extremely useful as it will allow us to identify the most relevant hypothesized antecedent of academic achievement, when considered alone.

In this study decision trees were used in multiple ways. Several models were estimated in two different contexts: (i) one in which the target variable was if the student passed or not per course (Model 1); (ii) another in which passing, or reproving, was regarding the academic year (Model 2). Hence, in the first case each observation is the combination of student/course, whereas in the second each observation is a student. Moreover, different trees were estimated in each of these contexts with different aims and, therefore, different parameters. Whenever our goal was to predict the performance of each student, we let the decision tree to become bigger, i.e., with more levels and splits per parent node. However, for the sake of interpretation, we estimated smaller, and thus less powerful trees, to shed some light regarding which collected variables for predicting academic achievement are the most important, both per course and per year.

To avoid overfitting, i.e., the fact that the model "*memorizes*" the data instead of learning from it, the database was split in to train (70%) and validation (30%) sets. The train set comprises the data in which the model will be estimated, whereas the training set is used for constantly get a more realistic, although optimistic to some extent, sense of the error rate, as the records within this set are not used for training purposes. The algorithm stops the training, i.e., the tree's growth, when the error in the validation set stops to grow, as there is evidence that, behind this point, the tree will be "memorizing" the data, i.e., overfitting.

2

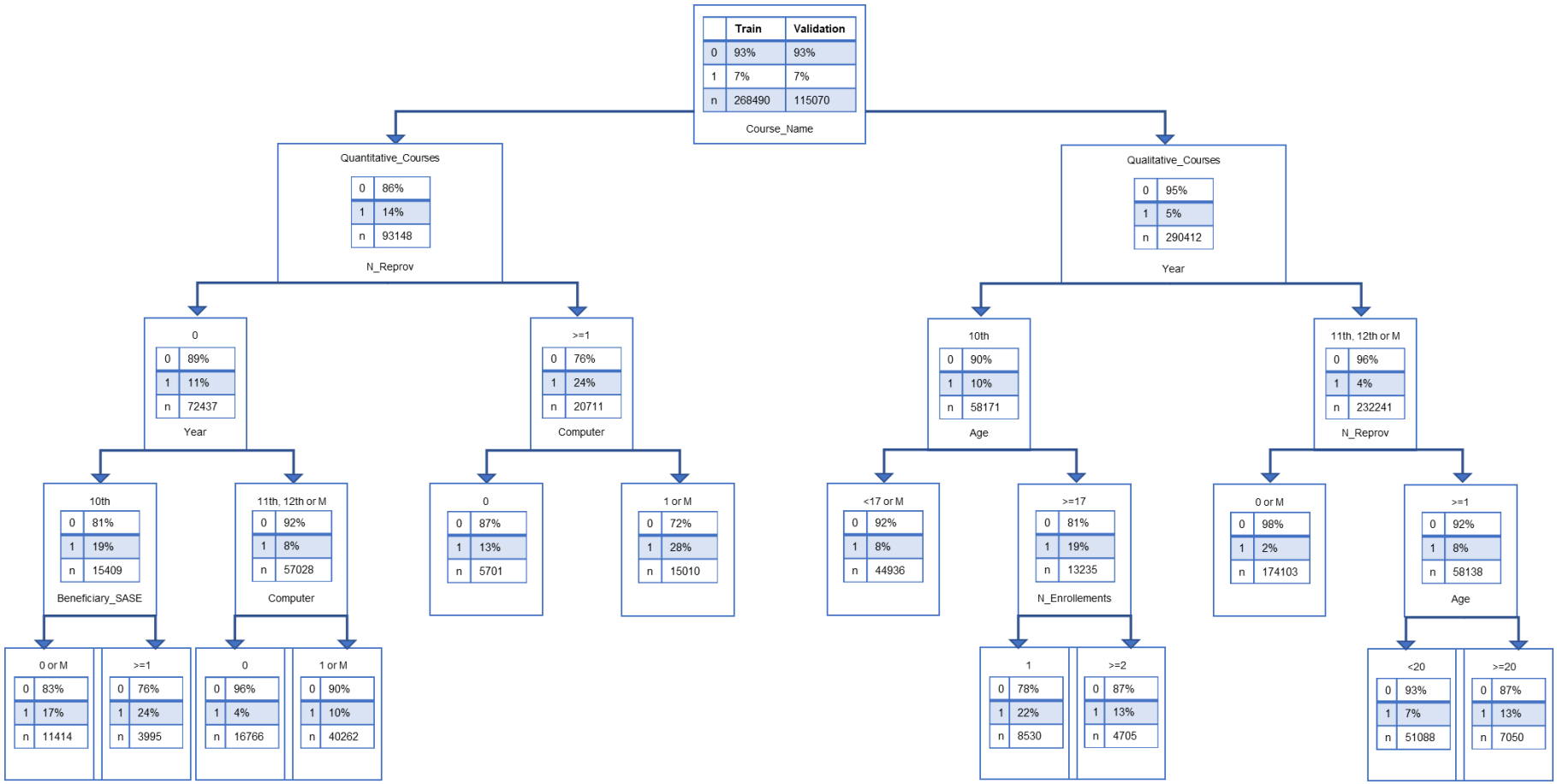

Figure 4.2 – Decision Tree for academic achievement at course level (Model 1)

Notes

- Quantitative Courses: Descriptive Geometry A, Mathematics A, Mathematics applied to Social Sciences, Mathematics B, Physics and Chemistry A, Portuguese as non Maternal Language
- Quantitative Courses: Biology and Geology, Drawing A, Economy A, Foreign Language – English, Foreign Language – French, Foreign Language – German, Foreign Language – Spanish, Geography A, History A, History B, History of Culture and Arts, Latin A, Philosophy, Portuguese, Portuguese Literature

[2] 0 – Students with positive approval rate; 1 – Students with reprove rate

To explain academic achievement at course level, which represents the reprove rate by course in the academic year 2014/2015, 383560 observations were analyzed, and our model indicates that 93% of the students have a positive approval rate while the remaining have a reprove rate of 7%. Here the variable carrying the biggest discriminating capacity for this result is Course_Name. The fact that this is the most important variable to understand if the student has a positive rate or not at course level, will immediately tell us that the first thing the model does is to slip the tree into two big groups, having the quantitative courses a higher reprove rate (14%), when compared to the group we called qualitative courses (5%).

In the quantitative courses we can observe that there is an average approval rate of 86% on this set of courses when compared with the 14% average reprove rate. The variable with higher impact on the quantitative courses is N_Reprov, which determine the number of reprove years that the student has until the current year and academic period (2014/2015). Here we can observe a clear distinction between students that have never reproved a year before and students that reproved at least once, having the students that reproved at least once a higher probability of reproving at course level (24%) when compared to students that have never reproved a year before (11%). Concerning the group of students that never reproved before the variable with greater influence is variable Year, the academic year where the student stands. Here the two sets formed are the students on 10^{th} and students on 11^{th} or 12^{th} grade. Students enrolled on 10^{th} grade two times more probability of reproving at level course (19%) when compared to those on the 11^{th} or 12^{th} grades (8%). For those students on the 10^{th} grade, the most explainable variable is Beneficiary_SASE, meaning, if the student is granted any type of social support due to low economical resources at the household. Those entitled to this social benefit a higher probability of reproving at course level (24%), when comparing to those with no social support have (17%). According to the students on 11^{th} or 12^{th} grade the variable with the highest influence on this score is Computer, i.e., if the student has or not access to a computer. Computer access will result in a higher probability of reproving at course level (10%), whereas no access to this tool only a 4% average rate. To what refers to students that enrolled the quantitative courses but that reproved at least one year before, it is fair to claim that Computer is once again the most significant variable to the results, having the students with computer access twice more probability of reproving at course level (28%), when comparing to those without access to this tool only a 13% average rate.

Secondly, we have qualitative courses. Here it seems possible to say that there is an average approval rate of 95% on this set of courses when compared with the 5% average reprove rate. The variable Year stands out, and the two sets formed are students on 10^{th} and students on 11^{th} or 12^{th} grade. In this model this is the normal separation that exist. Looking for the two groups formed, students on 10^{th} have a higher probability of reproving at course level (10%), when comparing to those that enrolled on 11^{th} or 12^{th} grade (4%). Focusing on students on 10^{th}, the most significant variable is Age, meaning age at the beginning of the school year, when those students with or more than 17 years old have a higher probability of reproving at course level (19%) in comparison with those under 17 (8%). Note that students that are under 17 years old at the beginning of the 10^{th} grade are those that, very likely, have never reproved a year before, as well as students who have 17 years or more, most probably already have reprove at least one year before, even if not in the high school. For the set of students on the 10^{th} grade under 17 years it is possible to see that 92% of the students have a positive approval rate while the remaining have a reprove rate of 8%. For the students enrolled on the 10^{th} grade with 17 years old or more the variable with more explanatory

power is N_Enrollments, the number of previous enrollments each student did on a certain academic year (10^{th} ,11^{th} and 12^{th} grades). contrary to what could be expected, the students with two or more enrollments have a lower probability of reproving at course level (13%) when comparing to those with just one enrollment (22%). Regarding to the students that enrolled the qualitative courses and are from the 11^{th} or 12^{th} grade, the most discriminating variable is N_Reprov, where students that reprove at least one year before have a higher probability of reproving at course level (8%) in comparison with those that have never reprove a year before (2%). To what concerns students that had at least reproved a year before, the age of the students it is once again the most explainable variable, having the students with more than 20 years old, a higher probability of reproving at course level (13%) when comparing to students under 20 years old (7%).

3

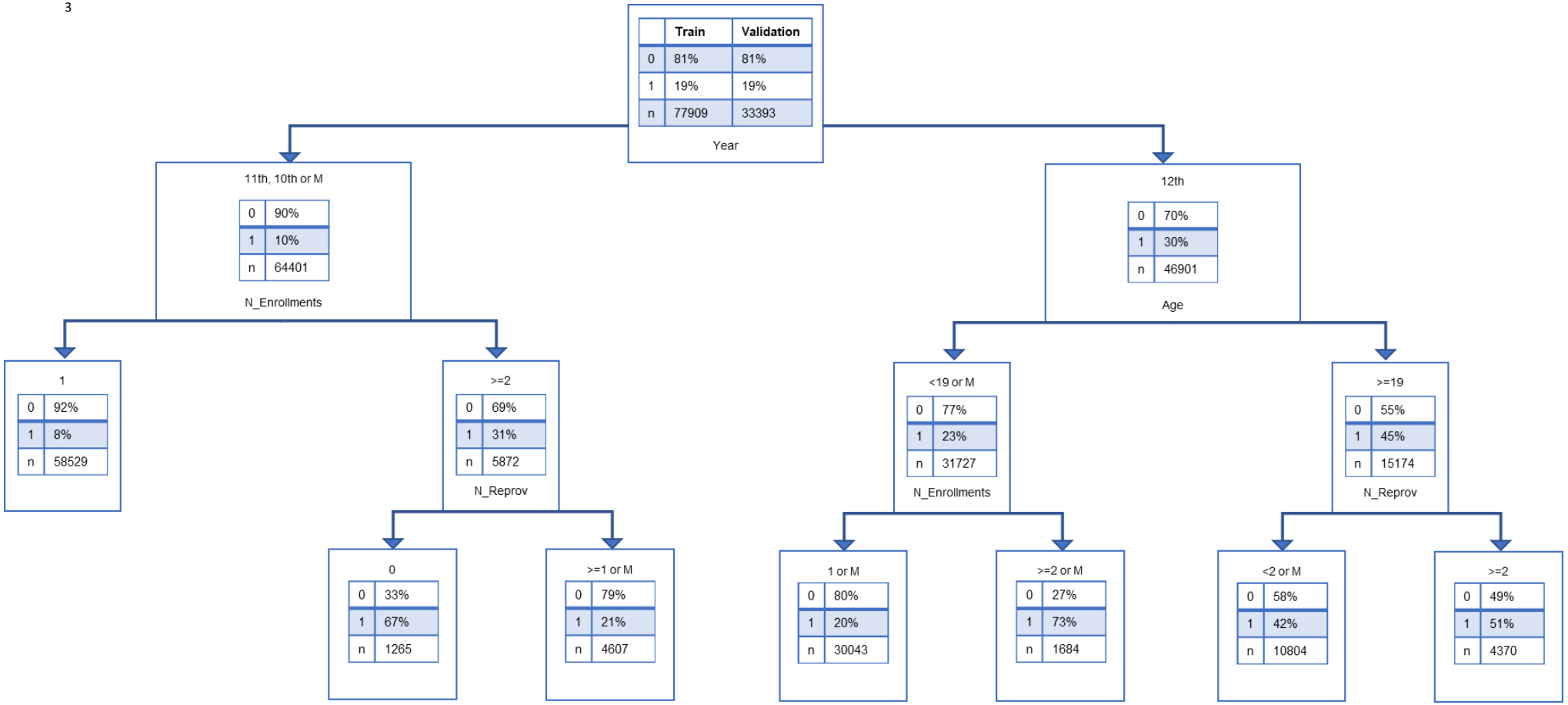

Figure 4.3 – Decision Tree for academic achievement at year level (Model 2)

[3] 0 – Students with positive approval rate; 1 – Students with reprove rate

To explain academic achievement at year level, which represents the reprove rate in the academic year 2014/2015, 111302 observations were analyzed, and our model indicates that 81% of the students show a positive approval rate and 18% while the other students present a reprove rate of 19%. Here the variable bringing the highest relevance to this model is Year. The fact that this is the most important variable to understand if the student has a positive rate or not at year level, will immediately tell us that the first thing the model does is divide the tree into two sets, having the students on 11^{th} or 10^{th} grade a lower reprove rate (10%) when compared to students enrolled on 12th grade (30%). Regarding students on 10^{th} or 11^{th} grade we can observe that there is an average approval rate of 90% when compared with the 10% average reprove rate. The variable that presents the highest impact on this group of students on 11^{th} or 10^{th} grade is N_Enrollments. Here we can see that students with two or more enrollments have a much higher probability of reproving at year level (31%) when comparing to those with just one enrollment (8%). Regarding students with two or more enrollments the variable that better justifies their performance is N_Reprov. Here the two ranges formed are students that have never reprove before and students that reprove at least one year before, having the group of students that never reprove before a higher probability of reproving at year level (67%), comparing with those that have reprove at least one year before (21%). Like the previous model, such an outcome might not be so expected to happen.

Secondly, we have the set of students enrolled on 12^{th} grade. Here we find out that 70% of them present an average approval rate in the academical year 2014/2015, while 30% an average reprove rate. The variable with more impact on these results is Age. In this case, there was a split formed by students under 19 years old and students with 19 years old or more. The group of students with less than 19 years showed a much lower probability of reproving at year level (23%) comparing to students with 19 years or more (45%). To what respects on students that have less than 19 years old the variable mostly impacting the results is N_Enrollments. Here the two groups formed are students with one enrollment and students with two or more enrollments. The ones with more than two enrollments a much higher probability of reproving at year level (73%) when compared with their peers with only one enrollment (20%). For the range of students on the 12^{th} grade, under 19 years old, the variable that had a higher explanatory power is N_Reprov. Here, students that reproved at least two years before, have a higher probability of reproving at year level (51%) comparing to those with less than two reproved years (42%).

Until this point our focus has shed some light on the antecedents of academic achievement. For this purpose, we've developed models with characteristics that facilitate this objective. In other words, the parameters in the trees have been set up not with the goal of minimizing prediction error, but rather to generate simple and easy to interpret model trees. If our goal were to maximize prediction performance, even at the expenses of interpretation (i.e., clearly understand the predicted level of academic achievement based on simple rules), the models would be estimated differently. To have an idea of to what extent could we predict academic achievement with the data we have available, we have estimated several alternative and more complex models with this specific purpose.

We have estimated decision trees and gradient boost trees. Contrarily to what we have done previously, we have not limited the number of parent nodes to two, i.e., trees were not necessarily binary. Moreover, we allow trees to grow behind three levels. Additionally, different training algorithms (e.g., CHAID and CRT) were employed as well as different error measures. We have then select the best models (one for course-level and other for year-level) of the several alternatives.

Although we will not interpret these model trees (the models are in Appendix), our results would be the following:

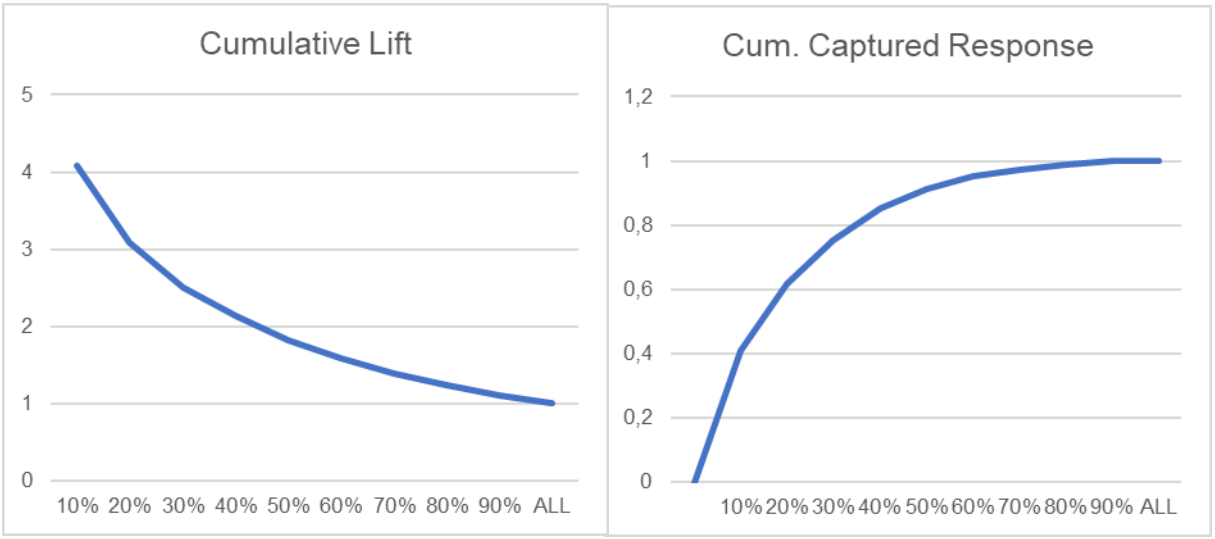

Figure 4.4 – Cumulative Lift and Cumulative Captured Response for Model 1

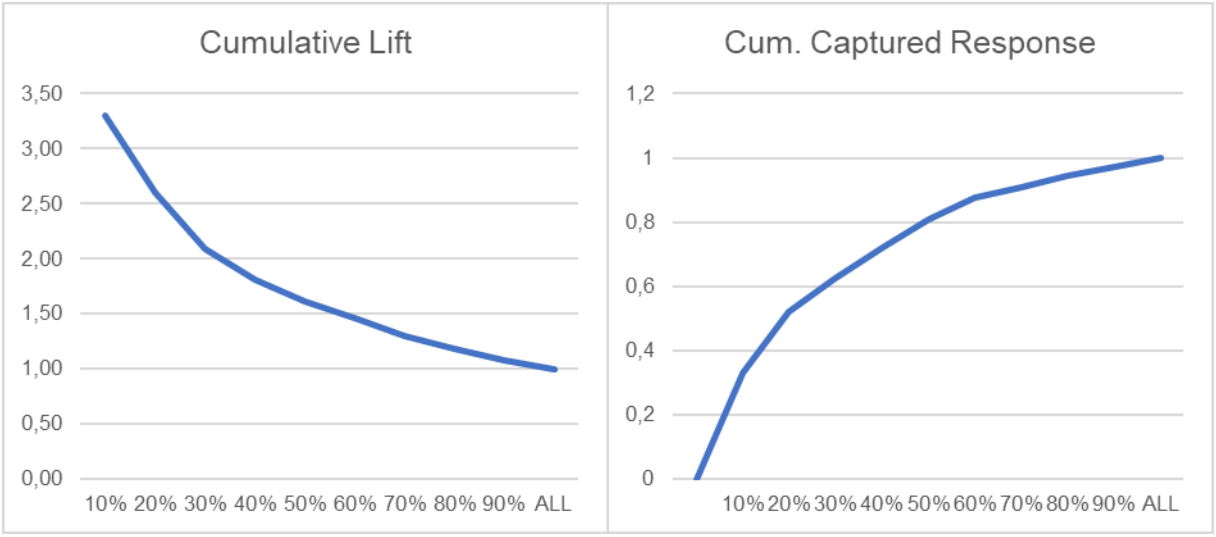

Figure 4.5 – Cumulative Lift and Cumulative Captured Response for Model 2

5. DISCUSSION

5.1. DISCUSSION OF FINDINGS

To get a better understanding of academic achievement at course level, we must first realize that the variable that better explains alone academic achievement is the course itself. In other words, the specific course is the most important antecedent of academic achievement at course level. Hence, different courses have different antecedents, at least in importance. In this case, given that the tree is binary, the best option is to isolate quantitative and qualitative courses and to estimate kind of "different models" for each one and define different antecedents for each type of courses instead of trying to fit a model to all without distinction. So, comparing the two type of courses, quantitative courses have a higher reprove rate comparing to qualitative courses.

Concerning the knowledge of academic achievement at year level, the first conclusion made, to understand if the student will reprove or not in 2014/2015, was that the most important thing to know is the Year in which the student is enrolled. This means that it is not so efficient having a possible good explainable model for all courses as much as, from the start, to split them in two groups: students from 10^{th} or 11^{th} grade and students from 12^{th} grade.

After the results of non-parametric tests and decision tree, it is possible to conclude that the following hypothesis, presented before are verified: H1, H2, H3, H4, H5, H6, H7, H9, H10, as they are all are statistically significant to students' academic achievement. For the hypothesis H8 we could not infer any conclusions as we did not collect sufficient data.

Regarding hypothesis H1 we can conclude that students from female gender obtain better results than their peers from the opposite gender. This conclusion is reinforced when we find that female students outperform male students in almost all six core courses from the secondary degree (presented in table 4), being the only exception Foreign Language – English. For hypothesis H2, the results show that Portuguese native students present higher academic results, comparing to foreign students. In hypothesis H3 and H4 we verify that students with access to both computer and internet can achieve better results. However, the results from the decision trees show that for quantitative courses, the use of computer should be more moderate since they have slightly lower approval rates. This reinforces the idea that, being Portugal a developed country, we should investment in providing better conditions so that our students could have access to such tools, particularly in schools where students spend much of their time. For hypothesis H5 the decision tree results confirm that the number of years that the student reproved before is an important factor to explain academic achievement, mainly for quantitative courses and for students enrolled on 11^{th} or 12^{th} grade from qualitative courses if we are referring to the academic achievement course level. This is also a key factor to explain academic achievement at year level, especially for students enrolled on 10^{th} or 11^{th} grade with two or more enrollments and for students on 12^{th} with 19 years old or more. This tells us that, for example, teachers should be more observant to the historic of reprove rates. Based on these results, it would be interesting to understand which are the phycological implications a reprove has on the students academical path. For H6 and H7, regarding SASE support and family financial support, we can conclude that students that receive one or both supports, have a worse academic achievement. However, being students registered in public institutions, with equal learning opportunities, this situation should not occur. For hypothesis H9 and H10, as it was already

mentioned on literature review, our study reinforces, even more, the importance of reducing the number of students per class, and consequently the number of students per school, as that would allow a much better accompaniment by teachers to the needs of each of their students.

5.2. PRACTICAL IMPLICATIONS

Several practical implications can be drawn from this study. First this study reinforces the idea of the importance on investing on a society and especially on an education with better digital and technological networks, which can be stimulated by financing schools and classes with computers, technical manuals for IT support, school programs or even proposing programming classes as a mandatory or even optional as we could withdraw from this paper that students with computer and internet access are capable of better academical results. This enhances the urgency of promoting and supporting increasingly science and innovation programs.

Secondly, it is equally important to continue to bet and aid students with less income, as we have found and concluded from this work that, these continue to be the group of students with lower academic achievement results. This support may start, for example, by subsidizing the distribution of materials and school meals.

Thirdly, it continues to be crucial to invest on the reduction of the number of students per school and consequently the number of students per class. As so, it would be much easier for teachers to engage and better understand students constrains and needs. In addition, an effort should be made to recover and create better conditions and comfort in public Portuguese schools as these are the places where students spend most of their time. These implications are supported by the fact that, after reaching our conclusions and results, classes between 11 and 30 students are those with better academic achievement

Lastly, our findings point out and reinforce the importance of knowing students' scholar background to what refers to the number of reproves or previous good performance, as the path one takes may possibly be a key factor to explain their academic achievement.

5.3. THEORETICAL IMPLICATIONS

Concerning the different theoretical implications, it is first important to note that data mining techniques proved to be yield good results, especially the use of non-parametric methods as decision trees given the characteristics of our data. This result makes us think that data mining methods are an eligible and much valid alternative to the classical econometric techniques used on most of the studies conducted on the area of students' academic achievement. These are techniques that provide good results as they are specific tools to handle large quantities of data and with highly detailed analysis capable of answering to different errors that big data bases might contain (outliers, missing values, variables transformation, statistical analysis). With this in mind, it is recommendable that more researchers use these new data analysis techniques.

Secondly the non-parametric tests appear to be equally reliable and a good alternative to parametric tests as they can handle variables that do not follow a normal statistical distribution, which in most cases limits most of classical approaches and techniques. Besides, they are tests easy to implement providing good results.

6. CONCLUSIONS

Understanding the factors that have greater impact on academic achievement is a topic far from being resolved, in fact, there is still a lot to improve. However, the biggest surplus of this study was the possibility of working with data that correctly translated the reality of where we stand in terms of the educational level at secondary degree, although we only focused on academic year 2014/2015. Our findings suggest that there are still differences and gaps on academic achievement among female and male genders, as female students obtain better results on academic achievement. We can also point out that access to computers and Internet, when well used for school purposes are a powerful mean to help students achieve better results. Students coming from less wealthy households obtain lower scholar performances and is crucial to urgently act on this topic that still stains our educational paradigm. Finally, we can assess that the student reprove background still has a great emotional weight on his academic achievement. The current research provides a drilldown analysis, which allows us to discover findings not only at course level but also at year level.

7. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

Despite our best efforts, some limitations must be acknowledged. The first one is regarding the data quality. Although MISI database comprises every student enrolled in the Portuguese high school system, being a source of tremendous potential in education data, some further developments are needed in the way data is recorded and stored. In fact, data pre-processing took a very meaningful part of the efforts conducted in this study. Missing values and data inconsistency are aspects to improve. Secondly, the study is in respect to a specific point in time, i.e., data used is cross-sectional. In the future, it would be interesting to do analysis on academic achievement for multiple points in time, where each student is "*tracked*" through his/her high school experience. This approach would further shed light on academic achievement antecedents. In third place, as we have used secondary data, we couldn't include other potential antecedents of academic achievement, such as teachers' characteristics or even (textual) notes on students' behavior by assiduity. Finally, we also acknowledge some limitations in terms of methods employed, which is related with the previous limitation. Have we had the opportunity to include additional variables, specifically interval ones, and the methods used would be different. We probably have used neural networks and regression analysis to improve our predictions, and explanation, on academic achievement. Moreover, the high sample size of our data implies that non-parametric tests will be much more likely to reject the null hypotheses, as mentioned earlier.

8. BIBLIOGRAPHY

Abosede, S., & Akintola, O. (2016). Mothers ' Employment , Marital Status and Educational Level on Students ' Academic Achievement in Business Studies. *Multidisciplinary Research*, *4*(2), 159–165.

Archibald, S. (2006). Narrowing in on Educational Resources That Do Affect Student Achievement. *Peabody Journal of Education*, *81*(4), 23–42. https://doi.org/10.1207/s15327930pje8104_2

B. Spinath. (2012). Academic Achievement. In V. S. Ramachandran (Ed.), *Encyclopedia of Human Behavior* (Second Edi, pp. 1–8).

Barnard, W. (2004). Parent involvement in elementary school and educational attainment. *Children & Youth Services Review*, *26*(1), 39. https://doi.org/10.1016/j.childyouth.2003.11.002

Barnett, R., Glass, J. C., Snowdon, R., & Stringer, K. (2002). Size, Performance and Effectiveness: Cost-Constrained Measures of Best-Practice Performance and Secondary-School Size. *Education Economics*, *10*(3), 291–311. https://doi.org/10.1080/09645290210127516

Barro, R., & Lee, J.-W. (2001). International data on educational attainment: updates and implications. *Oxford Economic Papers*, *53*(3), 541–563. https://doi.org/10.1093/oep/53.3.541

Berry, M. J. a., & Linoff, G. S. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. *Wiley Publising* (Third). Retrieved from http://portal.acm.org/citation.cfm?id=983642

Borzekowski, D., & Robinson, T. (2012). The Remote, the Mouse, and the No. 2 Pencil. *Archives of Pediatrics and Adolescent Medicine*, *159*(2), 607–613. https://doi.org/10.1001/archpedi.159.7.607

Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics*, *22*(2), 141–165. https://doi.org/10.1080/09645292.2011.568698

Bourdieu, P. (1973). Cultural reproduction and social reproduction. *Papers in the Soiology of Education*, 71–112.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgment of Taste*. *Cambridge, MA:Harvard University Press*. https://doi.org/10.1007/s13398-014-0173-7.2

Bowles, S., & Levin, H. (1968). The determinants of scholastic achievement--An appraisal of some recent evidence. *The Journal of Human Resources*, *3*(1), 3–24. https://doi.org/10.2307/144645

Brunner, M., Gogol, K., 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. https://doi.org/10.1016/j.intell.2013.05.009

Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, *36*(5), 403–421. https://doi.org/10.1016/j.intell.2007.11.002

Buddin, R., & Zamarro, G. (2009). Teacher Qualification and Student Achievement in urban elementary School. *Rand Education*, 1–48.

Cabot, J., De Virgilio, R., & Torlone, R. (2017). Web engineering: 17th international conference, ICWE 2017 Rome, Italy, June 5-8, 2017 proceedings. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 10360

LNCS). Springer. https://doi.org/10.1007/978-3-319-60131-1

Caro, D., Mcdonald, J., & Willms, J. D. (2009). Socioeconomic Status and Academic Achievement Trajectories from Childhood to Adolescence. *Canadian Journal of Education*, *32*(3), 558–590.

Chowa, G., Masa, R., Ramos, Y., & Ansong, D. (2015). International Journal of Educational Development 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. https://doi.org/10.1016/j.ijedudev.2015.09.009

Clotfelter, C., Ladd, H., & Vigdor, J. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, *41*(April 2005), 778–820. https://doi.org/10.2307/40057291

Codjoe, H. (2007). The importance of home environment and parental encouragement in the academic achievement of African-Canadian youth. *Canadian Journal of Education*, *30*(1), 137–156. https://doi.org/10.2307/20466629

Coleman, J. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*(1988), 95–120. https://doi.org/10.1086/228943

Coleman, J., & Hopkins, J. (1966). Equality of Educational Opportunity. *U.S. Department of Health, Education and Welfare*, 66–675. https://doi.org/10.2307/2091096

Croninger, R., Rice, J., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, *26*(3), 312–324. https://doi.org/10.1016/j.econedurev.2005.05.008

DGE. (2018). Ação Social Escolar | Direção-Geral da Educação. Retrieved May 11, 2018, from http://www.dge.mec.pt/acao-social-escolar

Driessen, G., Smit, F., & Sleegers, P. (2005). Parental Involvement and Educational Achievement. *British Educational Research Journal*, *31*(4), 509–532. https://doi.org/10.1080/01411920500148713

Dronkers, J., Van Der Velden, R., & Dunne, A. (2012). Why are migrant students better off in certain types of educational systems or schools than in others? *European Educational Research Journal*, *11*(1), 11–44. https://doi.org/10.2304/eerj.2012.11.1.11

Egalite, A., & Kisida, B. (2016). School size and student achievement: a longitudinal analysis. *School Effectiveness and School Improvement*, *27*(3), 1–12. https://doi.org/10.1080/09243453.2016.1190385

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, *13*(1), 1–22. https://doi.org/10.1023/A:1009048817385

Gabinete do Secretário de Estado da Educação. (2017). Despacho normativo nº1-A/2017. Diário da República. Retrieved from https://dre.pt/application/conteudo/106436777

Ghazvini, S., & Khajehpour, M. (2011). Gender differences in factors affecting academic performance of high school students. *Procedia - Social and Behavioral Sciences*, *15*, 1040–1045. https://doi.org/10.1016/j.sbspro.2011.03.236

Gil-Flores, J. (2009). Computer use and students' academic achievement. *Research, Reflections and Innovations in Integrating ICT in Education*, 1291–1295.

Goldhaber, D., & Hansen, M. (2013). Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance. *Economica*, *80*(319), 589–612. https://doi.org/10.1111/ecca.12002

Greenwald, R., Hedges, L., & Laine, R. (1996). The Effect of School Resources on Student Achievement. *Review of Educational Resear*, *66*(3), 361–396. Retrieved from https://books.google.de/books?hl=de&lr=&id=N3UIwF9P1WUC&oi=fnd&pg=PA43&dq=gdp+spending+education+resource+school+effectiveness&ots=SVHxRAceof&sig=Zlkppzcg8fC7wtrlCzWUP6cUZ3Y#v=onepage&q=gdp spending education resource school effectiveness&f=fals

Guarino, C., Reckase, M., Stacy, B., & Wooldridge, J. (2015). A Comparison of Student Growth Percentile and Value-Added Models of Teacher Performance. *Statistics and Public Policy*, (May). https://doi.org/10.1080/2330443X.2015.1034820

Hanushek, E. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

Hanushek, E., & Kimko, D. (2000). Schooling, Labor Force Quality, and the Growth of Nations. *American Economic Review*, *90*(5), 1184–1208. https://doi.org/10.1257/aer.90.5.1184

Hanushek, E., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, *17*(4), 267–321. https://doi.org/10.1007/s10887-012-9081-x

Hanushek, E., & Wößmann, L. (2010). Education and Economic Growth Early Studies of Schooling Quantity and Economic Growth. *International Encyclopedia of Education*, *2*, 245–252.

Hartas, D. (2011). Families' social backgrounds matter : socio-economic factors, home learning and young children's language, literacy and social outcomes. *British Educational Research Journal*, *37*(6), 893–914. https://doi.org/10.1080/01411926.2010.506945

Hill, N., & Taylor, L. (2004). Parental and Children's Involvement Academic Achievement Pragmatics and Issues. *Curent Directions in Psychological Science*, *13*(4), 161–164. https://doi.org/10.1111/j.0963-7214.2004.00298.x

Hodis, F., Johnston, M., Meyer, L., McClure, J., Hodis, G., & Starkey, L. (2015). Maximal levels of aspiration, minimal boundary goals, and their relationships with academic achievement: The case of secondary-school students. *British Educational Research Journal*, *41*(6), 1125–1141. https://doi.org/10.1002/berj.3189

Hoxby, C. (2000). The effects of class Size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, *115*(4), 1239–1285. Retrieved from http://www.jstor.org.proxy.library.ucsb.edu:2048/stable/info/2586924

Jackson, L., von Eye, A., Biocca, F., Barbatsis, G., Zhao, Y., & Fitzgerald, H. (2006). Does home internet use influence the academic performance of low-income children? *Developmental Psychology*, *42*(3), 429–435. https://doi.org/10.1037/0012-1649.42.3.429

Jayanthi, S. V., Balakrishnan, S., Ching, A., Latiff, N., & Nasirudeen, A. M. A. (2014). Factors Contributing to Academic Performance of Students in a Tertiary Institution in Singapore. *American Journal of Educational Research*, *2*(9), 752–758. https://doi.org/10.12691/education-2-9-8

Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education*, *42*(1), 82–110. https://doi.org/10.1177/0042085906293818

Krassel, K., & Heinesen, E. (2014). Class-size effects in secondary school. *Education Economics*, *0*(0), 1–15. https://doi.org/10.1080/09645292.2014.902428

Kubey, R., Lavin, M., & Barrows, J. (2001). Internet use and collegiate academic performance decrements: Early findings. *Journal of Communication*, *51*(2), 366–382. https://doi.org/10.1093/joc/51.2.366

Lee, C. L., & Mallik, G. (2015). The impact of student characteristics on academic achievement: Findings from an online undergraduate property program. *Pacific Rim Property Research Journal*, *21*(1), 3–14. https://doi.org/10.1080/14445921.2015.1026128

Lee, H. (2007). The Effects of School Racial and Ethnic Composition on Academic Achievement During Adolescence. *Journal of Negro Education*, *76*(2), 154–172.

Lee, J.-S., & Bowen, N. (2006). Parent Involvement, Cultural Capital, and the Achievement Gap Among Elementary School Children. *American Educational Research Journal*, *43*(2), 193–218.

Lei, J., & Zhao, Y. (2007). Technology uses and student achievement: A longitudinal study. *Computers and Education*, *49*(2), 284–296. https://doi.org/10.1016/j.compedu.2005.06.013

Leithwood, K., & Jantzi, D. (2009). Review of Empirical Evidence about School Size Effects. *Review of Educational Research*, *79*(1), 464–490. https://doi.org/10.3102/0034654308326158

Liebert, M. A., & Chou, C. (2001). College Students : An Online Interview Study. *Cyber Psychology & Behavior*, *4*(5), 573–586. https://doi.org/doi:10.1089/109493101753235160

Maehr, M., & Zusho, A. (2009). *Achievement goal theory: The past, present, and future*. *Handbook of motivation at school*.

Marks, G., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation*, *12*(2), 105–128. https://doi.org/10.1080/13803610600587040

Mensah, F., & Kiernan, K. (2010). Gender differences in educational attainment: influences of the family environment. *British Educational Research Journal*, *36*(2), 239–260. https://doi.org/10.1080/01411920902802198

Miedel, W., & Reynolds, A. (1999). Parent Involvement in Early Intervention for Disadvantaged Children: Does It Matter? *Journal of School Psychology*, *37*(4), 379–402. https://doi.org/10.1016/S0022-4405(99)00023-0

Neamtu, D. (2015). Education, the economic development pillar. *Procedia - Social and Behavioral Sciences*, *180*(November 2014), 413–420. https://doi.org/10.1016/j.sbspro.2015.02.138

Oecd. (2012). Equity and Quality in Education - Supporting Disadvantaged Students and Schools. In *Equity and Quality in Education* (p. 165). https://doi.org/http://dx.doi.org/10.1787/9789264130852-en

Oecd. (2014). Are Boys and Girls Equally Prepared for Life?, 1–8. https://doi.org/10.1787/9789264064072-en

Okan, Z. (2003). Edutainment: Is learning at risk? *British Journal of Educational Technology*, *34*(3),

255–264. https://doi.org/10.1111/1467-8535.00325

Oksana, A., & Elena, Y. (2015). Edutainment as a modern technology of education. In *Procedia - Social and Behavioral Sciences* (Vol. 166, pp. 475–479). Elsevier B.V. https://doi.org/10.1016/j.sbspro.2014.12.558

Patterson, M., & Pahlke, E. (2011). Student Characteristics Associated with Girls' Success in a Single-sex School. *Sex Roles*, *65*(9–10), 737–750. https://doi.org/10.1007/s11199-010-9904-1

Perkins, R., Kleiner, B., Roey, S., & Brown, J. (2004). The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000. NCES 2004-455. *National Center for Education Statistics*, 131. Retrieved from http://131.211.208.19/login?auth=eng&url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=eric3&AN=ED483081

Portes, A., & Rumbaut, R. (2005). Introduction: The Second Generation and the Children of Immigrants Longitudinal Study. *Ethnic & Racial Studies*, *28*(6), 983–999. https://doi.org/10.1080/01419870500224109

Ramachandran, K. M., & Tsokos, C. P. (2015). *Mathematical Statistics with Applications*. *Mathematical Statistics with Applications*. https://doi.org/10.1016/B978-0-12-417113-8.00009-6

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *Vol. 73*(No. 2), 417–458. https://doi.org/10.1002/polq.12145

Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, *94*(2), 247–252.

Sandstrom, H., & Huerta, S. (2013). The Negative Effects of Instability on Child Development: A Research Synthesis. *Urban Institute*, (September).

Schanzenbach, D. W. (2014). Does class size matter? *Economics of Education Review*, (February), 15. Retrieved from http://www.sciencedirect.com/science/article/pii/0272775795000044%5Cnhttp://nepc.colorado.edu

Segurança Social. (2018). Abono de família para crianças e jovens. Retrieved May 11, 2018, from http://www.seg-social.pt/abono-de-familia-para-criancas-e-jovens

Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Sousa, S., Portela, M., & Sá, C. (2003). Teacher characteristics and student. *Review of Educational Research*, *73*(1), 89–122. https://doi.org/10.3102/00346543073001089

Steinmayr, R., Dinger, F., & Spinath, B. (2010). Parents' Education and Children's Achievement: The Role of Personality. *European Journal of Personality*, *24*(6), 535–550. https://doi.org/10.1002/per.755

Steinmayr, R., Meißner, A., Weidinger, A., & Wirthwein, L. (2016). Academic Achievement. *Oxford Bibliographies*, (June), 3–5. https://doi.org/10.1093/OBO/9780199756810

Steinmayr, R., & Spinath, B. (2008). Sex Differences in School Achievement: What Are the Roles of

Personality and Achievement Motivation? *European Journal of Personality*, *22*(3), 185–209. https://doi.org/10.1002/per.676

Stevens, P., Clycq, N., Timmerman, C., & Van Houtte, M. (2011). Researching race/ethnicity and educational inequality in the Netherlands: a critical review of the research literature between 1980 and 2008. *British Educational Research Journal*, *37*(1), 5–43. https://doi.org/10.1080/01411920903342053

Strand, S. (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal*, *37*(2), 197–229. https://doi.org/10.1080/01411920903540664

Torres-Díaz, J.-C., Duart, J., Gómez-Alvarado, H.-, Marín-Gutiérrez, I., & Segarra-Faggioni, V. (2016). Internet Use and Academic Success in University Students. *Comunicar*, *24*(48), 61–70. https://doi.org/10.3916/C48-2016-06

Vigdor, J., Ladd, H., & Martinez, E. (2014). Scaling the digital divide: Home computer technology and student achievement. *Economic Inquiry*, *52*(3), 1103–1119. https://doi.org/10.1111/ecin.12089

Walberg, H. (1984). Improving the Productivity of America's School. *Educational Leadership*, *41*, 19–27.

Wally-Dima, L., & Mbekomize, C. (2013). Causes of gender differences in accounting performance: Students' perspective. *International Education Studies*, *6*(10), 13–26. https://doi.org/10.5539/ies.v6n10p13

Welsch, D., & Zimmer, D. (2016). The Dynamic Relationship between School Size and Academic Performance: An Investigation of Elementary Schools in Wisconsin. *Research in Economics*, *70*(1), 158–169. https://doi.org/http://dx.doi.org/10.1016/j.rie.2015.07.006

Wentworth, D., & Middleton, J. (2014). Technology use and academic performance. *Computers and Education*, *78*(September 2014), 306–311. https://doi.org/10.1016/j.compedu.2014.06.012

Wilder, S. (2014). Effects of parental involvement on academic achievement: a meta-synthesis. *Educational Review*, *66*(3), 377–397. https://doi.org/10.1080/00131911.2013.780009

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics*, *65*(2), 117–170. https://doi.org/10.1111/1468-0084.00045

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. https://doi.org/10.1016/j.euroecorev.2004.11.005

9. APPENDIX 1

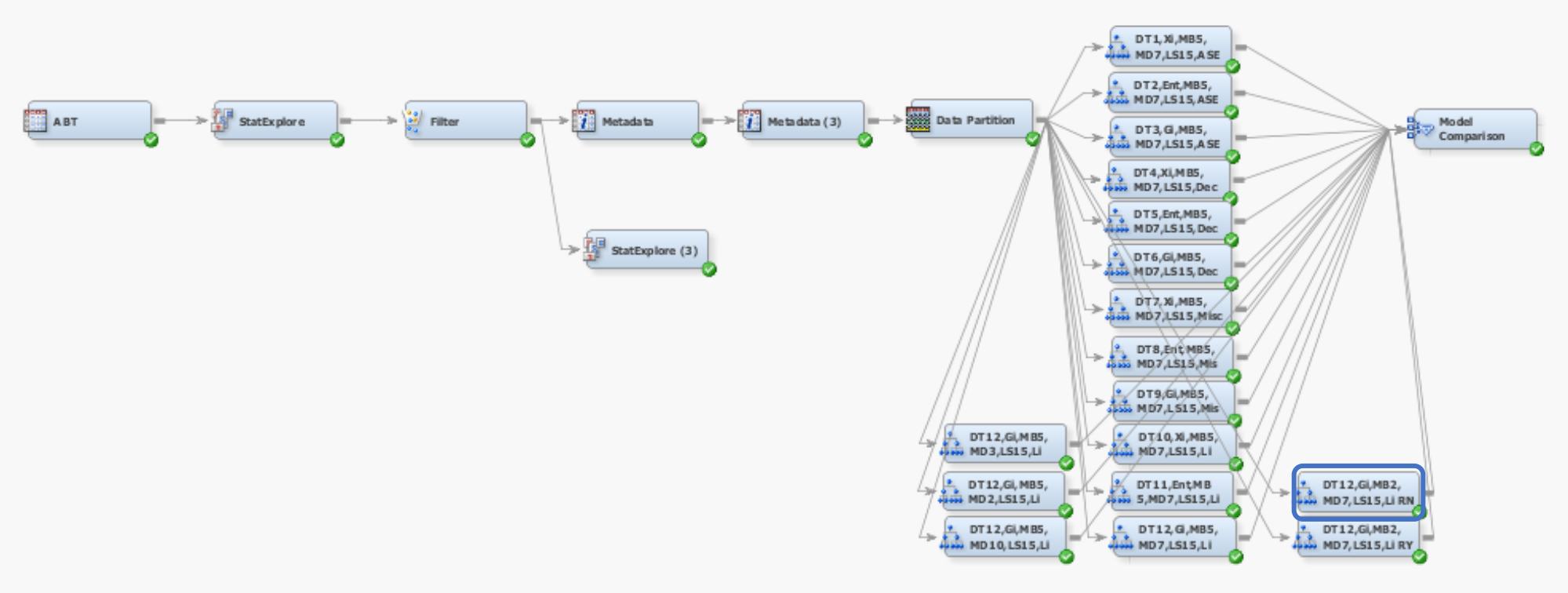

Figure 9.1 – SAS MINER MODEL 1

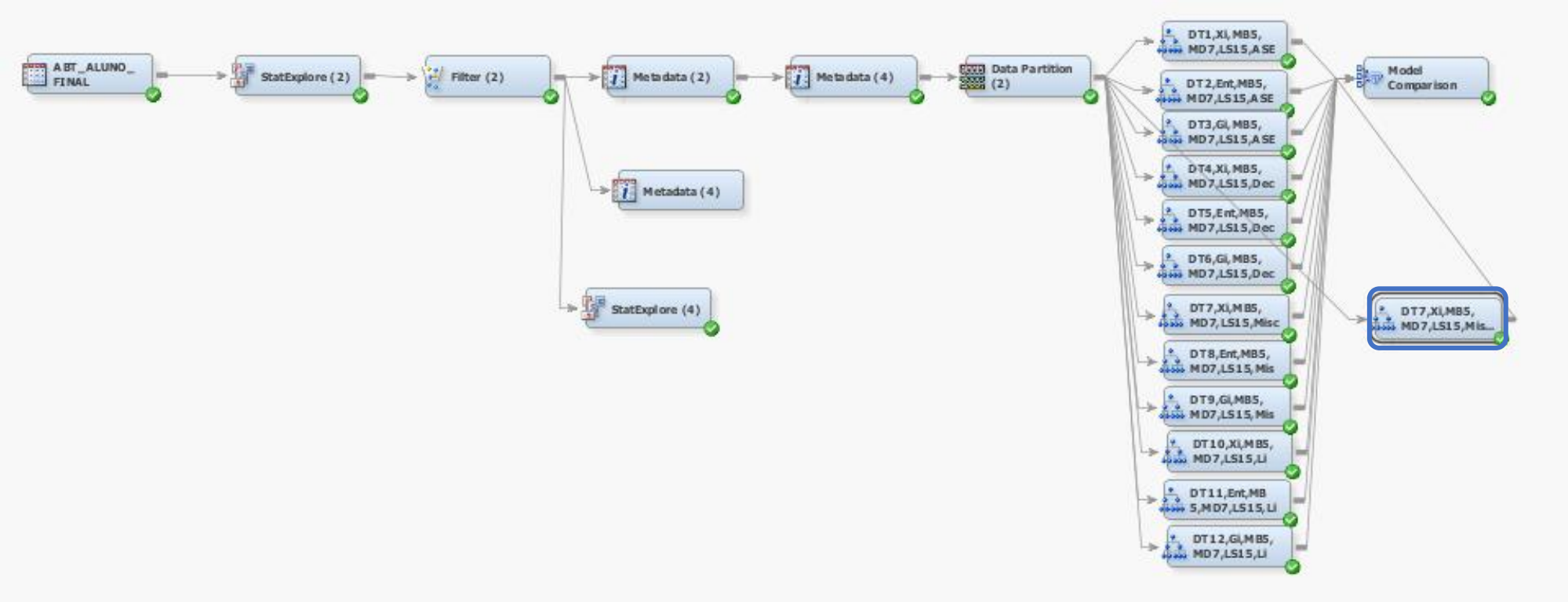

Figure 9.2 – SAS MINER MODEL 2