

Influence of demographic, social and school factors on Mathematics performance

Erick Calderon-Morales

Fall 2021

Introduction

General overview of secondary education in europe

Education is one of the main key areas of investment that dominate the political agenda in the European union (EU). Despite the recent global economic crisis, policy makers has set ambitious goals like reducing school dropouts, and increase the percentage of the population with an educational degree. With this kind of strategies policy makers are trying to address the increasing social inequalities that affect the citizens of the EU (Busemeyer et al, 2018).

Achieving equality in education is recognized by the politicians as key, but among the EU members there is variation in the socioeconomic factors that influence the citizens of each country. This kind of variation among countries makes difficult to pass general bills that reduce education inequality in the EU. Because of this, it is important to investigate each country separately in order to draw conclusions that can potentially be generalized to the the whole EU. (Hippe, Araujo and da Costa, 2016)

Portugal educational system is characterized for providing universal access to its citizens. This has lead to reducing the number of schools dropouts, (having the lowest out of school rate at the high school level among OCDE members). Despite of this, Portugal has one of the highest unemployment rates of recent graduates with upper secondary education and one of the lowest medium and high educational attainment (O.E.C.D, 2021)

In this report, I investigate the demographic, social and school factors that influence the performance of students in the mathematics test. Specifically I address the following questions, which factors determine the score in mathematics (Q1) and which factors can be use to predict if a student fail or pass the test (Q2). The main goal of this report is to develop predictive models for the final year grade (G3) using statistical learning methods.

Methods

For addressing which factors determine the score in mathematics (Q1) I used multiple regression analysis, Partial least squares regression, Principal components regression, LASSO regression an Ridge regression. Each model was evaluated using the mean squared error of the testing data set.

For addressing which factors can be use to predict if a student fail or pass the test (Q2) I used decision trees, Bagging, Random forests, Random forest with gradient boosting and k-Nearest Neighbor. For evaluating these models I calculated the accuracy, the sensitivity and specificity of each model using the test data set.

Results

a brief description of the dataset including the number of observations and attributes

References

Busemeyer, M. R., Garritzmman, J. L., Neimanns, E., & Nezi, R. (2018). Investing in education in Europe: Evidence from a new survey of public opinion. *Journal of European Social Policy*, 28(1), 34-54.

Hippe, R., Araujo, L., & da Costa, P. D. (2016). Equity in education in Europe. Luxembourg: Publications Office of the European Union.

Indicators, O.E.C.D. (2021). Education at a Glance 2016. Editions OECD, 90.