Student’s academic performance based on parental education

Manasvi Shobanaboina

06/28/2023

## Summary

This document aids in our comprehension of the connection between parental education and student academic performance. We are aware that parental education has some bearing on their children's academic success, but we are unsure of the exact magnitude of this influence. The study aims to determine the significance of parental education in influencing educational results by looking at the variables of Mother's Education (Medu), Father's Education (Fedu), and Grade (G3). The investigation tries to shed light on how parental education affects student performance using statistical modelling techniques like linear regression, ridge regression, and random forest. Policymakers, educators, and parents can benefit from the findings of this investigation because they provide important insights that can guide choices and strategies for improving student academic achievements.

## Introduction

The power of education to influence one's future and opportunities in life. Academic achievement is reportedly influenced by a wide range of variables, including socioeconomic background, school resources, and family involvement. Among these characteristics, parental education has been recognised as a key factor in predicting student success. Parents with more education typically have more knowledge, abilities, and resources to support their children's academic success. For policymakers, educators, and parents looking to promote and improve educational outcomes, it is critical to comprehend the link between parental education and student academic accomplishment.

## Literature Review

It is widely believed that parental education has a significant impact on a child's academic development. The manner in which parental education affects educational results have been made clear by the numerous research that have been done to study the link between parental education and children's academic achievement.

Nathalie Tamayo Martinez et al. (2022) conducted an important study that looked at the connections between parental education, parenting styles, child intelligence, and academic achievement. The study found a correlation between parental education and a child's IQ and academic success using data from a population-based birth cohort. The study emphasised both the direct and indirect effects of parenting practises and family routines on a child's cognitive development. These findings emphasise the importance of considering parental education when creating interventions and regulations to raise kids' academic performance.

Numerous more studies have examined the connection between parental education and students' academic success in addition to the one mentioned above. For instance, a 2005 meta-analysis of several research by Davis-Kean found a consistent link between parental education and a range of educational outcomes, including school performance, academic success, and cognitive ability. The meta-analysis found that parental education both shields kids from academic issues and encourages better educational achievements for them.

Furthermore, Reardon (2011) investigated disparities in academic attainment between socioeconomic level (SES) groups in relation to parental education. The findings revealed that variations in parental education had a significant influence on the achievement gap between high- and low-income students. This underlines the importance of parental education in decreasing educational gaps and providing equal opportunities for all students.

## It is plainly clear from these results and the wealth of past research that parental education has a considerable impact on a child's academic success. Understanding the mechanisms underlying this connection can help policymakers, educators, and parents create successful interventions and methods to boost children's academic performance. The current study aims to extend this body of knowledge by investigating the connection between parental education and student academic achievement using the Student accomplishment dataset. To provide further details and insights into the importance of parental education in determining educational outcomes, this research examines the variables Mother's Education (Medu), Father's Education (Fedu), and Grade (G3).

## Theory

The primary purpose of this study is to investigate the relationship between parental education and academic accomplishment. According to the theory underpinning this study, parental education has a major impact on how well students succeed academically.

It is considered that parental education and academic achievement in children are favourably associated. According to the hypothesis, parents with higher formal education provide better learning environments for their children. They are more likely to provide their children with educational materials, school support, and intellectual stimulation that can increase their academic progress.

Higher-educated parents frequently have a broader knowledge base, improved cognitive skills, and a greater awareness for the value of education. They are more inclined to participate in educational activities with them.With their kids, they are more likely to participate in educational activities like reading aloud, discussing academic subjects, and helping with homework. These parents also frequently have higher expectations for their kids' academic performance and give them the necessary coaching and assistance to help them succeed.

According to the notion, children who have parents with higher levels of education have more opportunity for intellectual development and have access to helpful learning environments. Therefore, it is anticipated that these youngsters will perform academically better than those whose parents have less-educated backgrounds.

This study tries to offer empirical evidence that either supports or contradicts the notion by analysing the data and looking at the connection between parental education and student academic achievement. The results of this study will provide important new understandings regarding the importance of parental education in influencing student outcomes.However, this research focuses specifically on investigating the impact of parental education as a key determinant of student academic success.

Therefore, a potential hypothesis is that there is a relationship between the educational background of the parents and the academic achievement of students. The hypothesis assumes that higher levels of education for both parents (represented by larger values for "Medu" and "Fedu") will positively correlate with higher final grades (G3) of students

## Data

For this research project, we utilized the Student Performance dataset obtained from the UCI Machine Learning Repository. The dataset includes information on students' demographics, parental education levels, and academic performance indicators. It provides a comprehensive collection of variables required to examine the relationship between parental education and student performance.

# Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:

1 school - student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)

2 sex - student's sex (binary: "F" - female or "M" - male)

3 age - student's age (numeric: from 15 to 22)

4 address - student's home address type (binary: "U" - urban or "R" - rural)

5 famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)

6 Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart)

7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)

8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)

9 Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at\_home" or "other")

10 Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at\_home" or "other")

11 reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")

12 guardian - student's guardian (nominal: "mother", "father" or "other")

13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)

16 schoolsup - extra educational support (binary: yes or no)

17 famsup - family educational support (binary: yes or no)

18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)

19 activities - extra-curricular activities (binary: yes or no)

20 nursery - attended nursery school (binary: yes or no)

21 higher - wants to take higher education (binary: yes or no)

22 internet - Internet access at home (binary: yes or no)

23 romantic - with a romantic relationship (binary: yes or no)

24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)

26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)

27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

29 health - current health status (numeric: from 1 - very bad to 5 - very good)

30 absences - number of school absences (numeric: from 0 to 93)

# these grades are related with the course subject, Math or Portuguese:

31 G1 - first period grade (numeric: from 0 to 20)

31 G2 - second period grade (numeric: from 0 to 20)

32 G3 - final grade (numeric: from 0 to 20, output target)

## Methodology

The following steps were performed:

Data preprocessing: The datasets for the mathematics course and the Portuguese course were combined. Relevant columns, including "Fedu" (father's education), "Medu" (mother's education), and "G3" (final grade), were selected for analysis.

The analysis utilized two datasets: one for the Mathematics course and another for the Portuguese course. The datasets were combined to create a unified dataset that captures a broader student population. The selected variables for analysis were father's education (Fedu), mother's education (Medu), and the final grade (G3).

*# Read the data files*

Port\_course <- read.csv("/kaggle/input/student-performance-data-set/student-por.csv", sep = ",", header = TRUE)

Math\_course <- read.csv("/kaggle/input/math-students/student-mat.csv", sep = ",", header = TRUE)

*# Combine the datasets*

dataset <- rbind(Math\_course, Port\_course)

To gain initial insights into the data, scatter plots were created to visualize the relationship between father's education, mother's education, and the final grade.

*# Visualize the initial data*

ggplot(data = data, aes(x = Fedu, y = Medu, color = G3)) +

geom\_point() +

labs(x = "Father's Education", y = "Mother's Education", color = "Grade (G3)") +

ggtitle("Initial Data")

A picture containing text, screenshot, diagram, square

Description automatically generated

Visualization: The ggplot2 library was used to create visualizations, such as scatter plots, box plots, histograms, and bar plots, to explore the relationships between parental education and student performance.

A picture containing diagram, screenshot, rectangle, plan

Description automatically generated

Modeling:

Linear Regression: The lm function was used to build a linear regression model to predict the final grades of students based on parental education.

Ridge Regression: The glmnet package was employed to perform ridge regression and assess its predictive power.

Random Forest: The randomForest package was utilized to construct a random forest model and evaluate its effectiveness in predicting final grades.

The model aimed to predict the final grade based on father's education and mother's education.

*# Linear regression*

linear\_model <- lm(G3 ~ ., data = train\_data)

A graph with blue dots

Description automatically generated with low confidence

In this analysis, ridge regression was employed to predict the final grades. The optimal regularization parameter (lambda) was determined using cross-validation.

*# Ridge regression*

cv\_ridge <- cv.glmnet(x\_train, train\_data$G3, alpha = 0)

ridge\_model <- glmnet(x\_train, train\_data$G3, alpha = 0, lambda = cv\_ridge$lambda.min)

A graph with green dots

Description automatically generated with low confidence

In this analysis, a random forest model was utilized to predict the final grades.

*# Random Forest*

rf\_model <- randomForest(G3 ~ ., data = train\_data)

A graph with red dots

Description automatically generated with low confidence

The performance of each model was evaluated using metrics such as RMSE, R-squared, and MAE. Additionally, scatter plots can be inserted to visualize the predicted grades against the actual grades.

*# Evaluation and visualization*

linear\_pred <- predict(linear\_model, newdata = test\_data)

ridge\_pred <- predict(ridge\_model, newx = x\_test)

rf\_pred <- predict(rf\_model, newdata = test\_data)

Performance Evaluation: The root mean square error (RMSE) was calculated to compare the performance of the linear regression, ridge regression, and random forest models.

*# Calculate RMSE, R-squared, and MAE*

Linear\_rmse <- sqrt(mean((test\_data$G3 - linear\_pred)^2))

rmse\_ridge <- sqrt(mean((test\_data$G3 - ridge\_pred)^2))

rf\_rmse<- sqrt(mean((test\_data$G3 - rf\_pred)^2))

## Results

The investigation produced some interesting findings regarding the relationship between parental education and student academic achievement. Using scatter plots and box plots, the study examined the connections between parental education, earlier grades, and final grades. various visualisations allowed for a clear understanding of the relationships between various parameters.

Further research into the relationship between parental education and children's projected final grades involved the use of a linear regression model. It was feasible to measure the connection between parental education and student achievement thanks to this research. The results demonstrated a significant relationship between parental education levels and the anticipated final grades, highlighting the importance of family educational background in determining student outcomes.

A ridge regression model was applied in addition to the linear regression model to evaluate the predictive impact of parental education and prior grades on student performance.This study emphasised the importance of parental education in impacting academic achievement by demonstrating how both variables significantly impacted forecasting student grades.

A random forest model was utilised to obtain understanding of feature relevance and effectiveness in predicting final grades. This methodology delivered insightful information on the key elements influencing student success. The significance of many characteristics was evaluated by researchers in order to identify key components that affected academic accomplishment.

The performance of the models was evaluated using a comparison using the root mean square error (RMSE) as the assessment metric. It was feasible to assess how well the models foresaw final grades thanks to RMSE. The advantages and disadvantages of each model for outlining the relationship between parental education and academic achievement were more clearly understood after this comparison.

Overall, the analysis generated informative results about how parental education affects adolescents' grades and provided a solid foundation for the theoretical framework. The study emphasised the positive link between parental education and kids' academic success. By considering feature importance and model performance, the study defined the factors that significantly influenced student outcomes. student achievement.A picture containing text, screenshot, rectangle

Description automatically generated

## 

## Conclusion

## This study looked at the connection between parental education and kids' academic success. According to the research, parental education significantly affects how students turn out. The study deepens our understanding of educational inequities and offers guidance to educators and policymakers in developing solutions that support equity and improve student performance. The effectiveness of initiatives designed to improve parental involvement in education can be assessed by additional study that also looks at other variables affecting student accomplishment.

# References

Martinez, N. T., Alberola, C. O., Buil, J. M., León, C. F., & Roca, J. G. (2022). The influence of parental education, parenting practices, child intelligence and health behaviours on children's academic achievement: a longitudinal analysis. European Journal of Public Health, 32(1), 100-105.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19(2), 294-304.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances (pp. 91-116). New York: Russell Sage Foundation.