Final Report: COVID-19 Effect on Student Performance

Anubhav Dhakal[1], Nick Li[1], Sihan Lyu[1], Hung Anh Vu[1], and Yiqing Wang[1]

Duke University, Durham NC 27708, USA

Abstract. The COVID-19 pandemic profoundly affected education systems worldwide, reshaping student performance across multiple dimensions. This study examines the impact of both pandemic-related and family-related factors on student test performance using a dataset of 1,400 students across six subjects and six semesters—three before and three after the pandemic lockdowns. We applied a range of predictive methods, including linear regression, logistic regression, decision trees, and random forest models, to identify key influences on academic outcomes. Our findings consistently highlight the significant role of socio-economic factors, particularly school type, household income, and parental education, which outweighed the direct effects of COVID-19 infection. Exploratory analyses using interaction effect models, fixed-effect models, clustering, and causal inference further reveal the critical role of institutional environments and household resources in shaping outcomes. While household income demonstrated a robust impact on performance, this effect was notably weaker in poorer schools, underscoring the mediating influence of school environments. These results emphasize the need for long-term interventions targeting structural and economic disparities, including subsidies for low-income families, increased funding for underserved schools, and investments in teacher training and educational resources to foster equity and reduce achievement gaps. [Code[1]] [Video[2]]

Keywords: Covid-19 · Student performance · Machine learning

1 Introduction

Problem Statement The COVID-19 pandemic triggered global lockdowns and shifted education to remote platforms, significantly altering daily life. Our project aims to evaluate its impact on student academic performance and investigate correlations with family-related factors. We will conduct statistical analyses to discern initial patterns and develop various Predictive Methods, including decision trees, random forests, linear regression, and logistic regression to predict student outcomes. Additionally, some exploratory methods, such as fixed effect models, interactive effect model, causal inference models, and clustering, will be employed to reveal causal relationships among these factors and enhance our understanding of the pandemic's broader effects.

Challenge A primary challenge with our dataset is its limited diversity, presenting potential biases. While it categorizes students by gender and economic status, it lacks racial and ethnic diversity. This limitation, coupled with the data's origin from Portland, Oregon, might restrict the applicability of our findings to other regions. Furthermore, the scarcity of comparable pre- and post-COVID test score data complicates broader analyses.

Literature Review COVID-19 has notably disrupted educational progress, particularly in math and reading, as detailed in [2] and [1], which describe the significant "COVID slide" affecting primarily low-income and minority students due to unequal access to resources. [3] emphasized that socio-economic factors have played a crucial role in students' adaptation to remote learning. Our research leverages these findings to assess the influence of family conditions on academic outcomes during the pandemic.

2 Method

2.1 Dataset Description

We obtained our dataset from Kaggle[3], comprising a panel dataset that documents scores across six subjects over six semesters/trimesters. The first three semesters occurred prior to the COVID-19 lockdowns, with the subsequent three following the onset of the pandemic. The dataset includes ten variables designed to mirror real-world trends and the local demographics of Portland, Oregon. These variables provide insights

[1] https://github.com/yqwang01/COMPSCI526Project4-CovidEffects
[2] https://www.youtube.com/watch?v=a-Ya439eS8Y
[3] https://www.kaggle.com/datasets/dylanbollard/covid19-effect-on-grades-constructed-dataset?resource=download

into family background, including parental education levels, family income, household size, and the number of computers at home. Additionally, they capture individual student characteristics such as grade, gender, COVID-19 infection history, and school type.

2.2 Statistical Analysis

Student T-test For binary factors such as COVID-19 infection status, we conducted a two-sample t-test to compare the average test scores of students who had contracted COVID-19 versus those who had not. The null hypothesis posited that the mean test scores between these groups were equal.
ANOVA We also conducted ANOVA tests to examine the statistical relationships between student performance and multi-class categorical factors, such as parental education levels. Our analysis extends beyond the significance of p-values; we also consider the average scores, variances, and distributions across different groups to provide a more comprehensive understanding of these factors.

2.3 Predictive Methods

Linear Regression & Logistic Regression We performed linear and logistic regression analyses to explore the relationships between various factors and student scores, assessing the influence of each factor through coefficient hypothesis testing. To account for the wide range of household income values, we normalize the data by dividing each value by the maximum income in the dataset. To minimize uncertainty and noise in our data and to examine the impacts of the COVID-19 lockdowns on student performance, we calculated the average scores across six subjects for the three semesters before and after the lockdowns. Additionally, we established the median of these target scores as the threshold for logistic regression analyses.
Decision Tree & Random Forest We further enhanced our analysis by utilizing decision trees and random forests (both as regressors and classifiers) to model our data. Also, we aimed to assess the impact of lockdown on test scores. We also evaluated the performance of random forests in predicting the average score before and after the lockdown, Additionally, we implemented a classification task to predict whether scores would be higher or lower than the mean score across all samples.

2.4 Exploratory Methods

Interaction & Fixed Effect Models We employed interaction regression models to investigate how the impact of COVID-19 on student grade scores varies by gender, school type, and family income. In one model, "female" was a binary variable (0 = male, 1 = female), and "COVID" indicated whether a student was affected by the pandemic. By including an interaction term (the product of "female" and "COVID"), we evaluated whether COVID-19's effect on grade scores differed between male and female students. This approach captures how variables interact with COVID-19 in influencing grades, rather than assuming their effects are independent. To account for unmeasured student characteristics, we also conducted a fixed-effects linear regression using 1,400 dummy variables—one for each student—to control for variation in the dependent variable. Additionally, we examined the effects of COVID-19 lockdowns by incorporating a dummy variable for semesters conducted online into the model.
Clustering We employed the K-Means algorithm to categorize students based on various factors. To ascertain the optimal number of clusters, we utilized the elbow method, which indicated that two clusters provided the best fit for our data. Subsequently, we applied Principal Component Analysis (PCA) to reduce the dimensionality of our feature space, facilitating visualization. We then examined the distinct characteristics of each cluster, particularly analyzing variations in student backgrounds and their performance across related features.
Casual Inference We conducted causal inference by constructing a Directed Acyclic Graph (DAG). To assess the individual impact of each factor while accounting for the effects of the others, we iteratively treated each factor as a potential cause, with the others serving as confounders. We applied the backdoor criterion to identify individual causal effects by blocking all backdoor paths between the treatment and the outcome. A linear regression model was then used to quantify the causal effect of each treatment factor on the outcome. Finally, we employed a placebo treatment to evaluate the robustness of the estimated causal effect and determine whether the observed effect was influenced by noise or bias.

3 Result

3.1 Statistical Analysis

Student T-test Our statistical analysis showed significant disparities in test scores, with COVID-19 positive students underperforming compared to their non-infected peers. Similar trends were evident across various demographics, including students from wealthier versus poorer schools, males versus females, and those receiving free lunch versus those who did not. Figure 1 A. illustrates the change in average scores based on two COVID-related factors: lockdown and covidpos. The data clearly shows that both factors lead to a significant decrease in average scores.

ANOVA Our dataset, which includes 8.4k items, allows us to readily detect significant differences in ANOVA tests. However, upon examining the means, variances, and distributions across various groups, we find that the differences are quite minimal. Figure 1 B. shows the average score based on different parents education levels. Interestingly, the educational level of parents is not linearly correlated with the academic performance of their children.

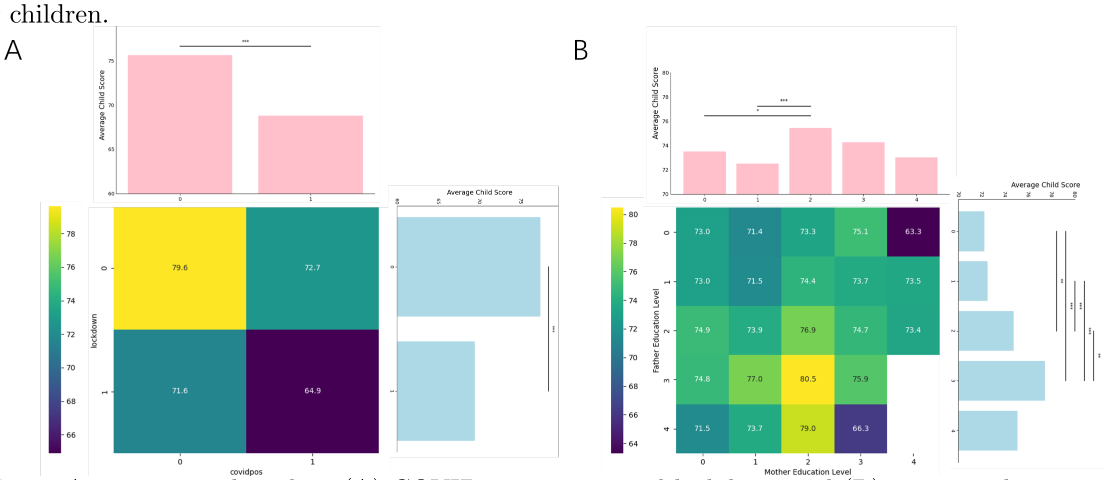


Fig. 1: Average score based on (A) COVID-19 positive and lockdown and (B) parents education.

3.2 Predictive Methods

Linear Regression, Logistic Regression Tables 1 and 2 illustrate the coefficients and hypothesis testing results from linear and logistic regression models analyzing average scores before and after the COVID-19 lockdown. Key findings indicate significant impacts of factors such as school type, COVID-19 infection history, household income, and parental education levels on student performance. The lockdown notably enhanced the influence of household income and father's education level but reduced the impacts of COVID-19 infection and mother's education level. The average score of students in lower-income schools is over 6 points lower compared to those in higher-income schools. COVID-19 infection decreases the average score by approximately 2.8 points, though the lockdown mitigated this impact. An increase of around 18 thousand in annual income results in a 1-point increase in the average score, and this effect is further amplified by the lockdown.

Table 1: Linear Regression Between Various Factors and Average Course Scores.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Factor	Pre-lockdown			Post-lockdown		
	Coef	Std Err	P-value	Coef	Std Err	P-value
const	73.4175	0.934	0.000***	65.3653	0.932	0.000***
school	-6.5033	0.435	0.000***	-6.2920	0.436	0.000***
gradelevel	0.0371	0.055	0.501	-0.0218	0.055	0.692
gender	0.4624	0.220	0.036**	0.2522	0.220	0.252
covidpos	-2.7809	0.254	0.000***	-2.5538	0.254	0.000***
householdincome	9.8695	0.981	0.000***	10.1047	0.982	0.000***
freelunch	0.5032	0.348	0.149	0.8220	0.349	0.019**
numcomputers	0.0703	0.079	0.374	0.0745	0.079	0.347
familysize	-0.4493	0.105	0.000***	-0.3729	0.105	0.000***
fathereduc	1.4546	0.135	0.000***	1.7225	0.140	0.000***
mothereduc	1.0320	0.134	0.000***	0.7025	0.131	0.000***

Table 2: Logistic Regression Between Various Factors and Average Course Scores.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Factor	Pre-lockdown			Post-lockdown		
	Coef	Std Err	P-value	Coef	Std Err	P-value
const	-4.2208	0.978	0.000***	-2.4253	0.952	0.011**
school	-2.7584	0.378	0.000***	-2.7925	0.377	0.000***
gradelevel	0.0798	0.058	0.172	-0.0424	0.057	0.455
gender	0.3660	0.230	0.111	-0.0221	0.223	0.921
covidpos	-1.6358	0.256	0.000***	-1.4687	0.250	0.000***
householdincome	6.4237	1.123	0.000***	6.6083	1.110	0.000***
freelunch	0.4038	0.422	0.339	0.2675	0.401	0.505
numcomputers	0.0510	0.082	0.535	-0.0002	0.081	0.998
familysize	-0.0181	0.110	0.870	-0.2675	0.109	0.014**
fathereduc	0.8125	0.147	0.000***	1.0420	0.150	0.000***
mothereduc	0.5351	0.141	0.000***	0.4047	0.133	0.002***

Decision Tree & Random Forest Using 5-fold cross-validation, we optimized our decision tree and random forest models, setting the best tree depth at 4 and criterion as 'squared error'. Despite initial challenges in fitting the decision tree regressor to predict potential scores, as evidenced by Table 3 and Figure 5, performance significantly improved when predicting average scores. Also, the MSE dramatically reduced from over 100 to approximately 13, and r^2 increased to 0.74. Previously, we looked at feature importance of our linear and logistic regression models based on the magnitude of the features coefficients. For our decision tree and random forest models, we measured features' importance based on their contribution to the target variable, through reduction in entropy. Figure 4 & 6 both agree that the students' school, family income, and covid test results play an important factor in predicting their test scores. This underscores the strong impact of economic and social conditions on student performance.

Table 3: Performance metrics for decision tree and random forest

Model	Decision Tree								Random Forest	
Metric	Read	Write	Math	RSL	WSL	MSL	AVGB	AVGA	AVGB	AVGA
MSE	111.13	128.86	131.19	123.38	120.88	153.11	14.34	14.86	13.71	13.05
R^2	0.37	0.34	0.27	0.33	0.28	0.21	0.73	0.70	0.74	0.74

Comparison Tables 4 and 5 present the results for various regression and classification models. Among all the regression methods, the linear regression model demonstrates the best performance both before and after the lockdown period. This suggests a linear relationship between the predictors and the average test scores. For classification models, logistic regression performs the best overall. During the modeling process, we also observed that decision trees tend to overfit, although they offer good interpretability. Random forests, on the other hand, improve performance through ensemble methods but sacrifice interpretability. The simple linear model, however, remains both effective and interpretable, making it the most suitable for this project.

Table 4: The comparison of different regression methods.

Method	Pre-lockdown				Post-lockdown			
	MSE	R2	RSS	RSE	MSE	R2	RSS	RSE
Linear Regression	12.79	0.7388	3580	0.2224	12.33	0.7319	3451	0.2184
DecisionTree	14.34	0.7258	4014	1.6945	14.86	0.7039	4160	1.7251
RandomForest	13.71	0.7378	3839	1.6571	13.05	0.7398	3655	1.6169

Table 5: The comparison of different classification methods.

Method	Pre-lockdown				Post-lockdown			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Logistic Regression	0.9393	0.94	0.94	0.9416	0.9107	0.91	0.91	0.9141
DecisionTree	0.8893	0.89	0.88	0.8864	0.9036	0.92	0.90	0.9078
RandomForest	0.9214	0.96	0.89	0.9225	0.9071	0.92	0.90	0.9110

3.3 Exploratory Methods

Interaction & Fixed Effect Models Our interaction models in Table 9 reveal several key findings. Female students infected with COVID-19 show no significant difference in performance compared to their peers, which is an encouraging result. Students in poorer schools who contracted COVID-19 perform worse in reading but better in math. Additionally, female students in poorer schools outperform their peers in both reading and writing. While higher household income generally improves grades, this effect is notably weaker for students in poorer schools, emphasizing the critical role of school environments in shaping the influence of household income on academic performance. The fixed-effects model analysis in Table 6 and Figure 2 explores how lockdowns affected student scores while accounting for unobserved characteristics such as race/ethnicity. By incorporating dummy variables for 1,400 students and online learning semesters, the analysis identifies a significant decline in scores for online learning compared to in-person learning. The inclusion of an online learning indicator enhances the model's explanatory power, reflected in the higher R^2 value and increased coefficient significance between Model 1 and Model 2. Figure 2 presents a box plot of predicted scores from Model 2.

Table 6: Fixed Effect regression on Average Course Scores.

$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Model 1	Model 2
Online (=1)		-8.008***(0.141)
Constant	76.199***(2.693)	80.203***(2.389)
Individual Fixed Effect	✓	✓
R2	.47	.637
Observations	8,400	8,400

Table 7: Confusion matrix for clusters of six scores with the following labels.

Predicted Cluster	Learning Mode		School		Income	
	In person	Online	0	1	High	Low
Cluster 0	3071	1808	3593	1286	3484	1395
Cluster 1	1129	2392	607	2914	716	2805

Clustering Figure 3 illustrates the clustering outcomes when school data was removed. Despite the absence of explicit school labels, the clustering results demonstrate a strong alignment with students' actual school affiliations. This finding highlights the implicit influence of school-related factors on student groupings, suggesting that the clustering algorithm effectively captures patterns reflective of the institutional environments and resources. Table 7 compares clustering accuracy across different labels, using six pairs of academic scores as inputs. The results indicate that clusters align most closely with students' school affiliations and family economic status. Specifically, the clusters effectively differentiate between higher and lower-income groups, as well as between students from resource-rich and resource-limited schools. In contrast, clustering based on learning methods (in-person vs. online) exhibits less alignment, suggesting that learning mode alone is insufficient to define distinct clusters.

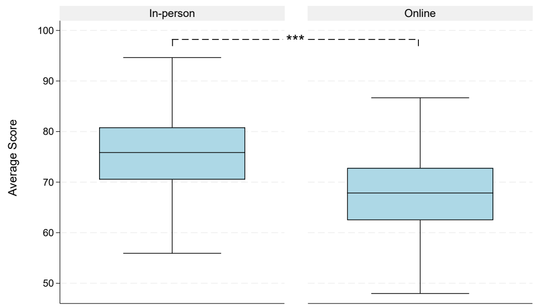


Fig. 2: Predicted scores from Fixed effect Model 2 (Table 6).

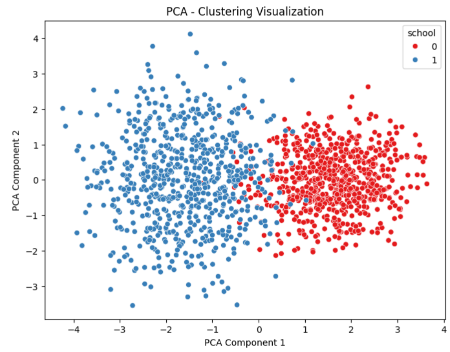


Fig. 3: PCA - Clustering Visualization by School

Casual Inference Tables 8 present the estimated causal effects and placebo treatment results for analyzing average scores before and after the COVID-19 lockdown. All ten factors show significant causal impacts on the final outcomes, with most exhibiting high p-values and near-zero placebo effects when the p-value is zero. The ratio of the estimated causal effect to the placebo effect for household income is significantly larger compared to other factors, indicating a robust causal relationship between household income and student performance. This effect is further amplified by the COVID-19 lockdown, highlighting social inequities during the crisis.

Table 8: The Estimated Causal Effects and Placebo Treatment Results.

Factor	Pre-lockdown			Post-lockdown		
	Estimated Effect	Placebo Effect	P-value	Estimated Effect	Placebo Effect	P-value
school	-6.7524	-0.0124	0.94	-6.3525	0.0019	1.00
gradelevel	0.0275	-0.0025	0.98	-0.0474	-0.0018	0.92
gender	0.3739	-0.0189	0.76	0.2626	0.0245	0.96
covidpos	-2.6148	-0.0338	0.94	-2.4209	-0.0059	0.98
householdincome	9.9288	-9.9e-14	0.00	10.2544	-1.9e-13	0.00
freelunch	0.1505	-0.0184	0.78	0.5106	-0.0291	0.90
numcomputers	0.0418	-0.0003	0.98	0.0676	-0.0037	0.98
familysize	-0.4077	0.0082	0.78	-0.4205	0.0001	0.88
fathereduc	1.4483	-0.0036	0.88	1.7163	-0.0071	0.98
mothereduc	0.9779	-0.0018	0.98	0.6525	-0.0098	0.86

4 Conclusion

We investigate the impact of the COVID-19 pandemic and family-related factors on students' test performance. We apply several predictive methods, including linear regression, logistic regression, decision trees, and random forest models, to analyze the significance of these factors. All methods consistently highlight the importance of school type, household income, parental education, and COVID-19 infection. Surprisingly, school type and household income have a greater influence than COVID-19 infection, indicating that economic and social conditions have a stronger impact on student performance.

We used exploratory methods, including interaction effect models, fixed-effect models, clustering, and causal inference, to examine factors influencing student performance. The analysis highlights socio-economic factors and school affiliations as central to differentiating student groups, with clusters aligning closely with school types and family economic status, emphasizing the impact of institutional environments and household resources on outcomes. A strong causal-to-placebo effect ratio for household income confirms its robust influence, though this effect weakens in poorer schools, highlighting the mediating role of school environments. In contrast, the weaker correlation with learning methods indicates that short-term changes, such as shifts between in-person and online learning, have limited impact.

These findings highlight the need for long-term interventions to address structural and economic disparities and reduce achievement gaps. Policymakers should focus on promoting social equity through subsidies for low-income families and increased funding for underserved public schools. Investments in teacher training, modern learning tools, and community engagement can further enhance educational opportunities, creating a more equitable system that benefits all students.

Limitations & Future directions A key limitation of our research is the dataset's lack of diversity, as it categorizes students by gender and economic status but excludes racial and ethnic representation. Future work will explore feature interactions beyond scores and COVID-19 impacts using methods like Difference-in-Differences (DID) regression and analyze more diverse datasets covering additional demographics and performance trends pre- and post-COVID.

5 Contribution

All authors contributed to the statistical analysis and conclusions. AD developed the fixed-effect and interaction effect models, while NL implemented the clustering method and conducted the literature review. SL worked on decision tree and random forest models, with HV addressing challenges and contributing to the decision tree analysis. YW implemented the linear and logistic regression models, performed causal inference, and organized the report.

References

1. Dorn, E., Hancock, B., Sarakatsannis, J., Viruleg, E.: Covid-19 and learning loss—disparities grow and students need help. McKinsey & Company, December **8**, 6–7 (2020)
2. Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., Liu, J.: Projecting the potential impact of covid-19 school closures on academic achievement. Educational Researcher **49**(8), 549–565 (2020)
3. Maldonado, J.E., De Witte, K.: The effect of school closures on standardised student test outcomes. British Educational Research Journal **48**(1), 49–94 (2022)

6 Appendix

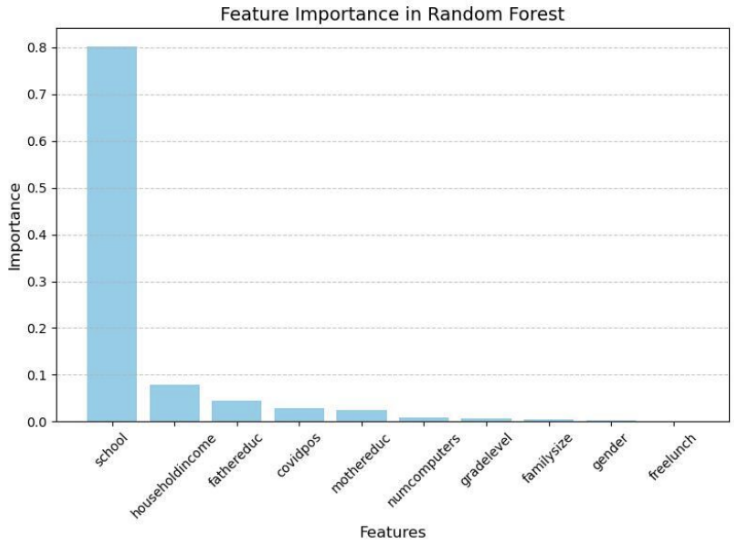
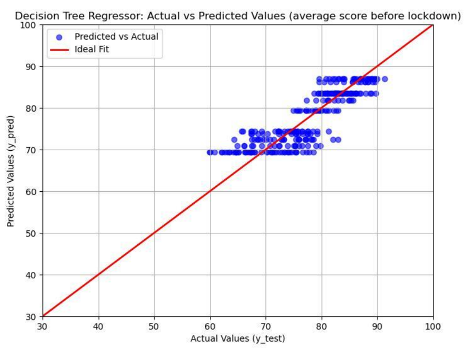
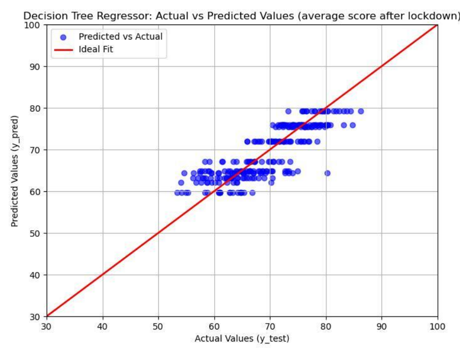


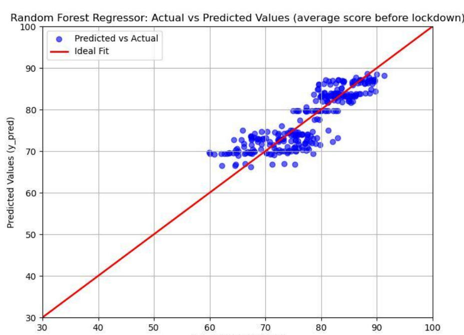
Fig. 4: Feature importance plot from our Random Forest model highlighting the contributions of various features to the target variable. 'School' is the most significant feature, followed by 'household income,' 'father's education,' and 'covid impacts,' with other features contributing marginally.



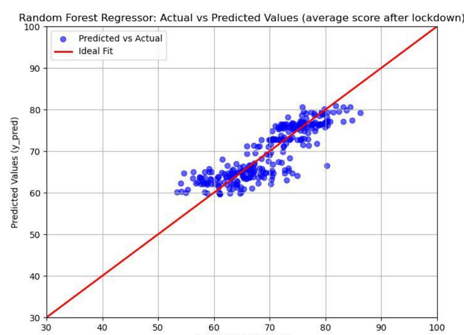
(a) Before Lockdown



(b) After Lockdown



(c) Before Lockdown



(d) After Lockdown

Fig. 5: Decision tree and random forest regressor performance on prediction the average score before and after the lockdown

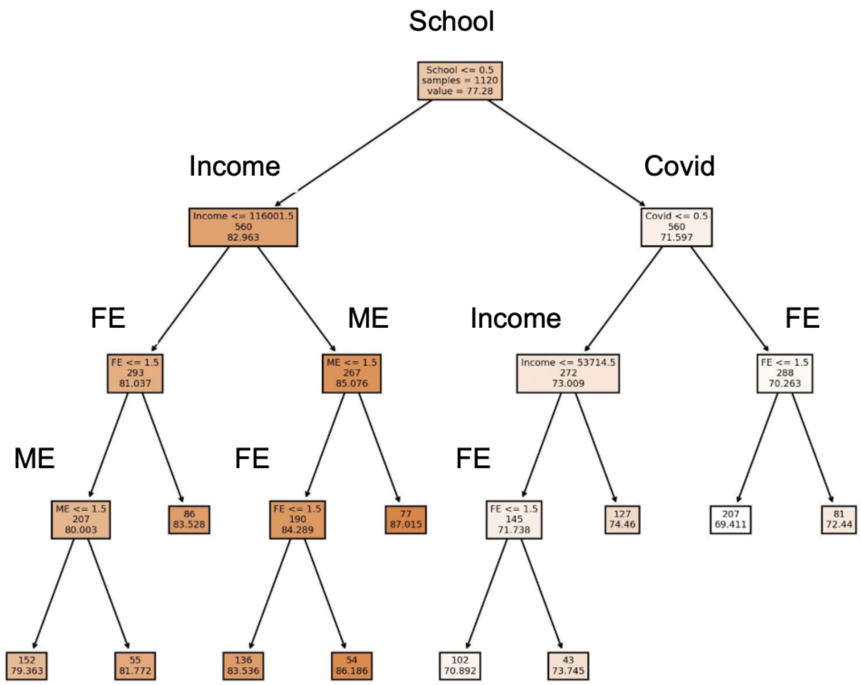


Fig. 6: Sample decision tree illustrating feature importance in determining splits based on impurity reduction. The root node splits on the feature 'school,' followed by 'income' and 'covid,' highlighting these as the most significant predictors of test scores.

Table 9: Four regression models with average, reading, writing, and math test scores as dependent variables. Each row corresponds to a regression variable, such as the interaction term *Female* X *COVID*, while columns display coefficients with standard errors in parentheses.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) scoreSL	(2) readingscoreSL	(3) writingscoreSL	(4) mathscoreSL
COVID (=1)	-3.238*	-2.947**	-3.307**	-3.460*
	(0.068)	(0.014)	(0.036)	(0.252)
Female (=1)	-3.335**	-3.301**	-3.188***	-3.517*
	(0.024)	(0.015)	(0.004)	(0.084)
Poor School	-0.844*	-1.309*	-2.079	0.856*
	(0.060)	(0.064)	(0.164)	(0.049)
HHIncome	0.077**	0.088**	0.071**	0.071***
	(0.001)	(0.001)	(0.001)	(0.000)
Female X Covid	0.323	0.292	0.145	0.531
	(0.141)	(0.061)	(0.051)	(0.537)
Poor School X Covid	0.095	-0.103*	0.193	0.193**
	(0.008)	(0.006)	(0.018)	(0.001)
Female X Poor School	0.537	0.545*	0.566*	0.499
	(0.044)	(0.032)	(0.037)	(0.202)
Poor School X HHIncome	-0.070**	-0.080**	-0.062*	-0.068**
	(0.001)	(0.001)	(0.001)	(0.000)
Constant	72.088**	70.581**	74.055**	71.630***
	(0.336)	(0.234)	(0.849)	(0.076)
R2	.551	.353	.306	.234
Observations	8,400	8,400	8,400	8,400
Time Fixed Effect	Yes	Yes	Yes	Yes
Father's Education FE	Yes	Yes	Yes	Yes
Mother's Education FE	Yes	Yes	Yes	Yes