Student performance predictive models using LMS data in Primary Schools

Ignacio Alvarez-Castro

Instituto de Estadística (IESTA)
Universidad de la República, Uruguay.

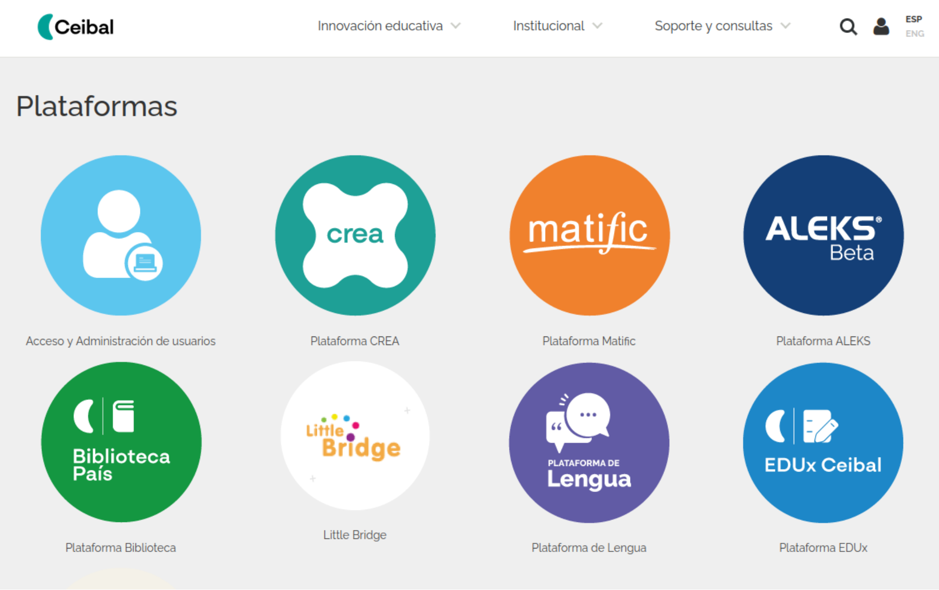
International Data Science Conference,
Santiago de Chile, Nnovember 2023.

Fondo Sectorial de Educación Inclusión digital 2020 - Modalidad financiado por ANIII. Proyecto "Monitor y evaluación de uso de las plataformas educativas"

Ceibal Program

- “One laptop per child"model in primary education (2007)
- Extended to secondary schools
- Key role during COVID-19 pandemic
- webpage: https://ceibal.edu.uy

Learning managment system (LMS)



3 lines of work

- LMS Monitor: `Shiny` app, draft version: http://164.73.240.157:3838/App-Ceibal/
- Key drivers of LMS use: measure student engagement
- Predictive modeling
  - Little Bridge data (LMS)
  - Predict English results

Project Team: Natalia da Silva, Oscar Montañez, Bruno Tancredi, jimena Padín

Introduction

Data sources

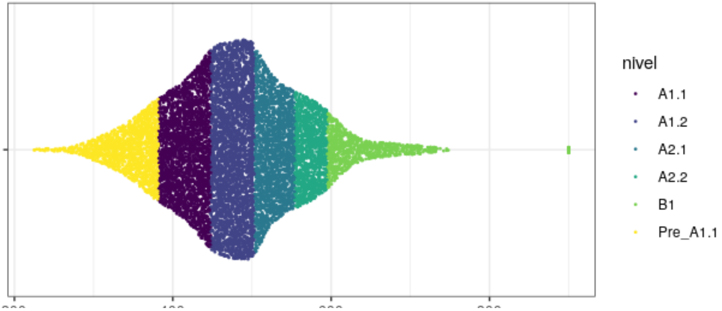
Predictive modeling

Performance data

English adaptive test

- 2 components: Vocabulary-Grammar (VG) and Reading (R)
- End of academic year (November-December)
- $\approx$ 35000 students, randomly selected

Performance data



12 % of students below A1.1 level

LSM data

Little Bridge

- Interactive LSM to learn English
- Automatic evaluation
- In children from 4°, 5° y 6° grades (9-11 years old)

2021 data

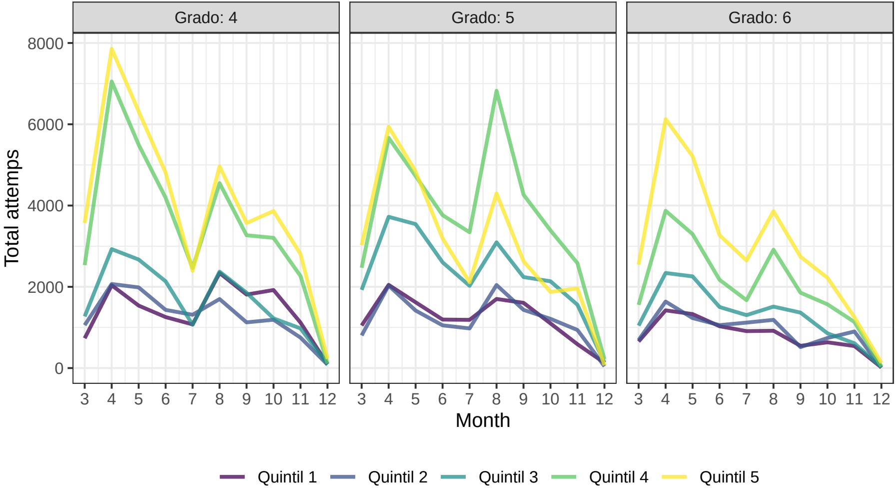
- $\approx$ 70000 students
- LB activity per child-day
- Some information about teachers

LB snapshot

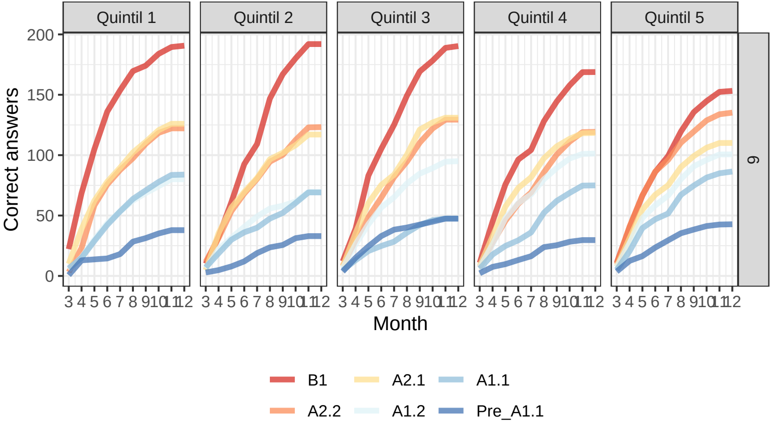
```
##          Act min.pts max.pts ActTot Preguntas Correctas
## 1   act_32    0.50    0.50      1        10         5
## 2   act_32    0.50    0.50      1        10         5
## 3   act_33    1.00    1.00      1         2         2
## 4  act_402    1.00    1.00      1         1         1
## 5               NA      NA     NA        NA        NA
## 6   act_16    0.30    0.60      2        20         9
## 7   act_18    1.00    1.00      1        12        12
## 8   act_19    1.00    1.00      1         5         5
## 9   act_20    0.88    0.88      1         8         7
## 10  act_21    1.00    1.00      1         5         5
```

Other variables: school, socioconomic level ...

Monthly attemps



Right answers and English level



Clasification problem

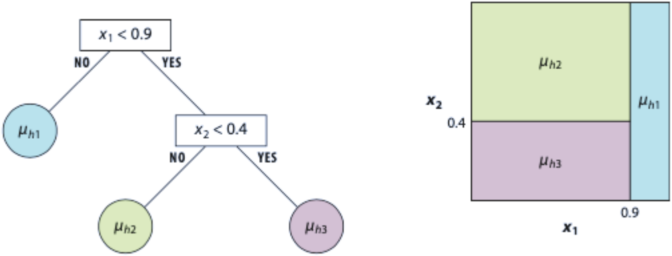
Children in 6th grade are *expected* to reach **A2.1** level.

- Sample size: $\approx$ 3000 students
- Response:

$$Y_i = \begin{cases} 1 & \text{reaches A2 level or higher} \\ 0 & \text{otherwise} \end{cases}$$

- Use LB acumulated work up to July
- Fit several statistical learning methods
- Include school random effect

Bayesian additive regression trees



- Picture from: Hill, J., Linero, A., & Murray, J. (2020). *Bayesian additive regression trees: A review and look forward*. Annual Review of Statistics and Its Application, 7, 251-278.

BART model

A single tree is denoted as

$$g(x|T, M) = \sum_k \mu_k I(x \in R_k)$$

having two basic parameters: tree structure T and set of leaves values $M = (\mu_1, \ldots, \mu b)$.

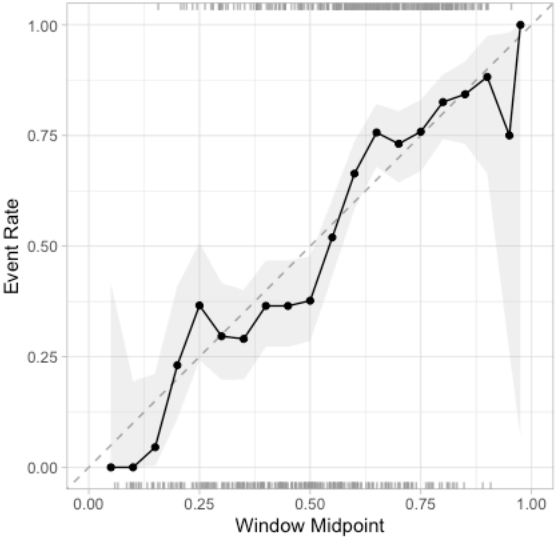
BART: sums of trees model

$$\begin{aligned} Y_i &= f(X_i) + \epsilon_i \\ &= \sum_j g_j(X_i|T_j, M_j) + \epsilon_i \\ &\epsilon_i \sim N(0, \sigma^2) \end{aligned}$$

Is possible to add random effects, $f(X_{ig}) + \alpha_g$.

Clasification results

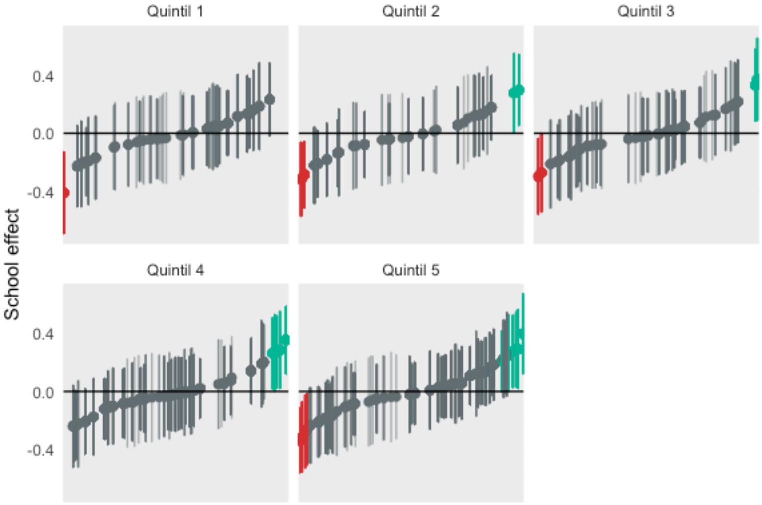
Calibration plot



Accuracy $\approx 70\,\%$, Specificity $\approx 58\,\%$

Random effects results

There are schools with positive effetcs in most quintile groups



Future (present?) steps

- Include effects for other levels (class)
- Extend school effects to slope for selected variables

Thank you!