

Student performance predictive models using LMS data in Primary Schools

Ignacio Alvarez-Castro

Joint Statistical Meetings. Toronto, Canada

August 2023



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

IESTA

INSTITUTO
DE ESTADÍSTICA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

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 management system (LMS)



Innovación educativa ▾

Institucional ▾

Soporte y consultas ▾



Plataformas



Acceso y Administración de usuarios



Plataforma CREA



Plataforma Matific



Plataforma ALEKS



Biblioteca País

Plataforma Biblioteca



Little Bridge



PLATAFORMA DE Lengua

Plataforma de Lengua



EDUx Ceibal

Plataforma EDUx

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

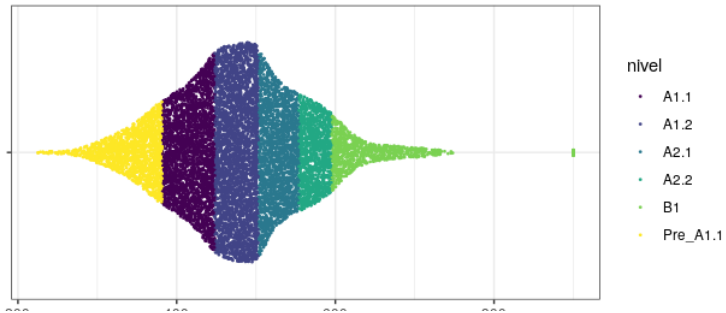
Predictive modeling

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^o, 5^o y 6^o grades (9-11 years old)

2021 data

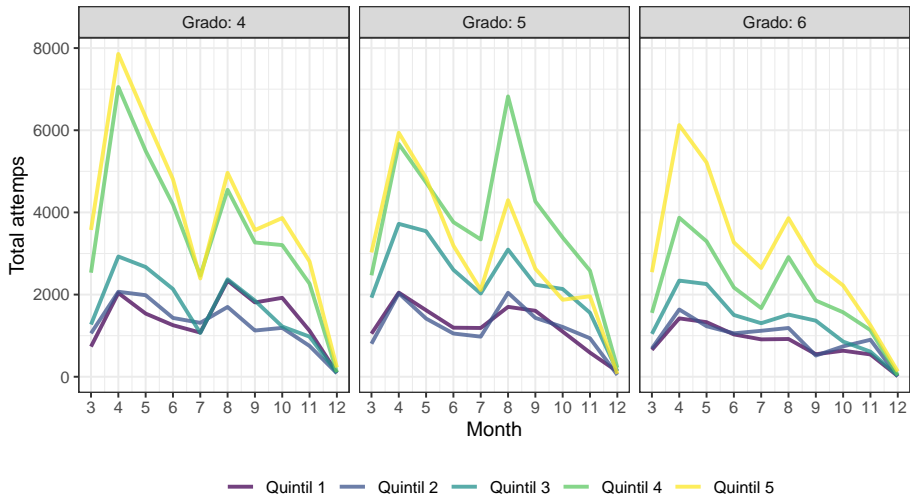
- ▶ \approx 70000 students
- ▶ LB activity per child-day
- ▶ Some information about teachers

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, socioeconomic level ...

Monthly attempts

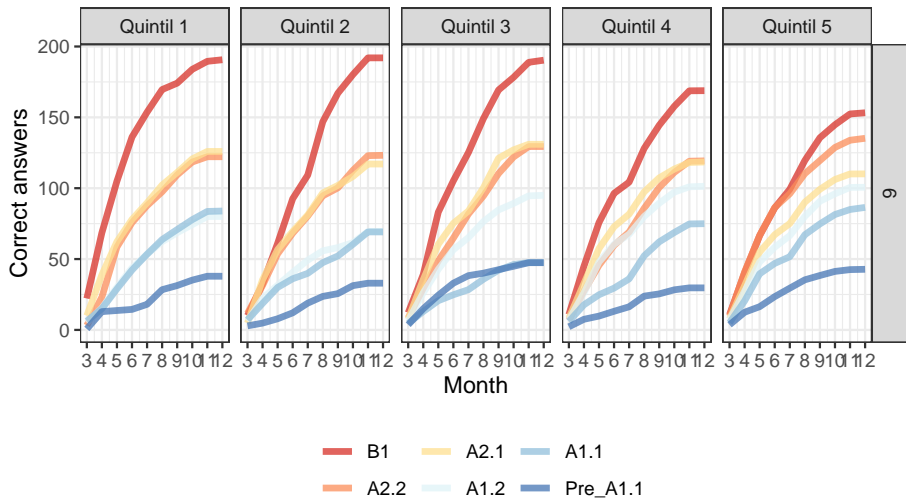


Introduction

Data sources

Predictive modeling

Right answers and English level



Classification problem

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

- ▶ Sample size: ≈ 3000 students
- ▶ Response:

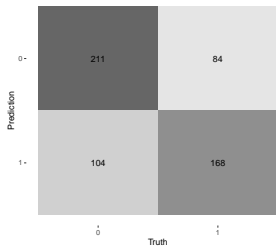
$$Y_i = \begin{cases} 1 & \text{reaches A2 level or higher} \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Use LB accumulated work up to July
- ▶ Fit several statistical learning methods

Classification results

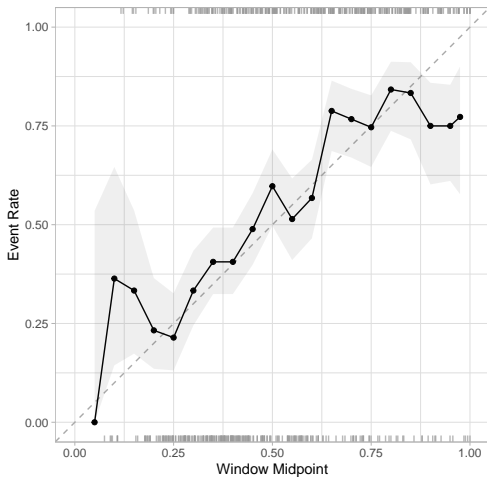
Showing some results from a random forest (`ranger`)

Confusion matrix



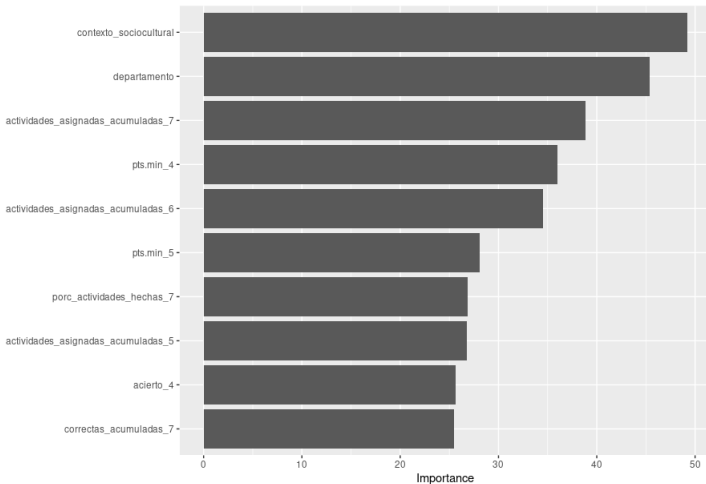
Accuracy $\approx 66\%$

Calibration plot



Variable importance

Most relevant variables are not individual specific



Future (present?) steps

- ▶ Include class effect: learning occurs in class environment, so

$$E(Y_{ij}) = class_j + f(x_{ij})$$

for instance: BART with random effects

- ▶ Data exploration suggests separation occurs in linear combinations. We plan using Projection pursuit methods.

Thank you!