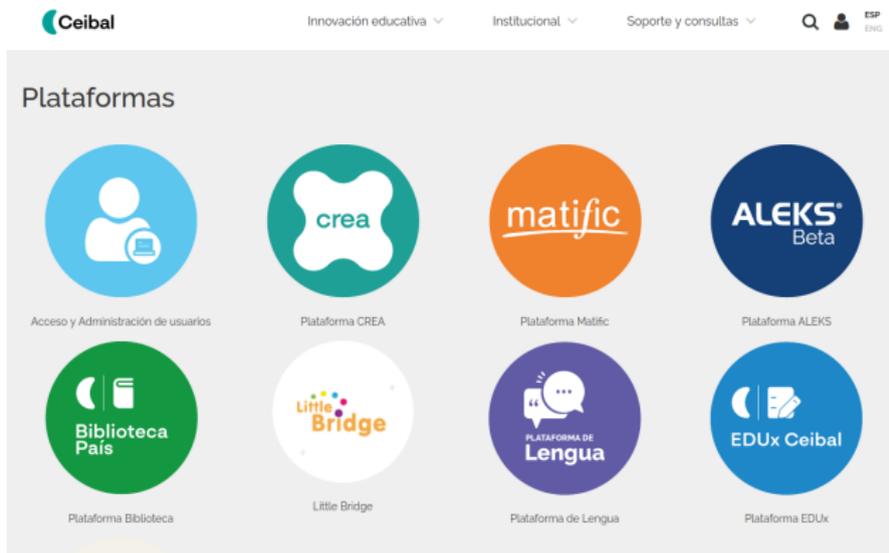


Ceibal en inglés: Modelos Predictivos para el resultado del pruebas adaptativas de inglés

Ciencia de datos en plataformas educativas



- ▶ Generan gran volumen de datos
- ▶ Desarrollar herramientas para generar información
- ▶ Fondo Sectorial de Educación - ANII

Hay 3 líneas principales de trabajo

- ▶ Monitor de Plataforma
 - ▶ Aplicación interactiva con `Shiny`
 - ▶ Mejorar su performance computacional para producción
 - ▶ Disfrutaron el taller ?
- ▶ Determinantes de uso
 - ▶ Medir el compromiso con la plataforma
 - ▶ Entender fuentes de vairabilidad de ese compromiso
 - ▶ Vean el poster !
- ▶ Modelos predictivos
 - ▶ Usar datos de plataforma Little Bridge
 - ▶ Predecir resultado Prueba Adaptativa de Inglés

Ciencia de datos en plataformas educativas

Los datos

Modelos predictivos

Datos de uso

Plataforma Little Bridge

- ▶ plataforma educativa e interactiva de inglés
- ▶ sistema de evaluación automático
- ▶ para 4°, 5° y 6° año de Primaria

Datos para 2021

- ▶ \approx 70 mil estudiantes
- ▶ Actividad en LB por día
- ▶ Información de efecto docente

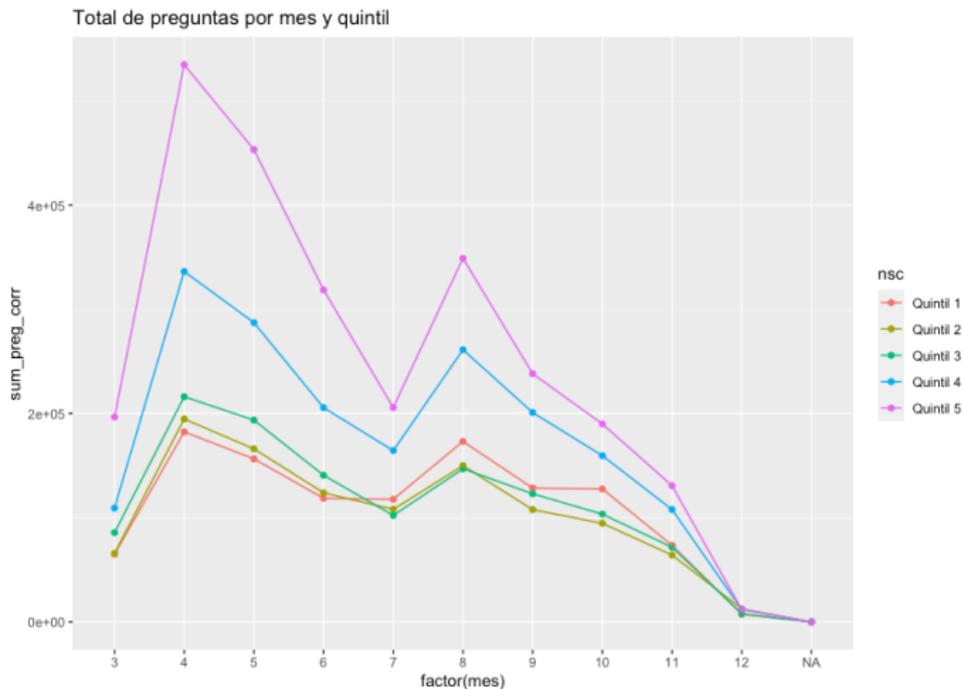
Datos de uso

Variables de uso de LB

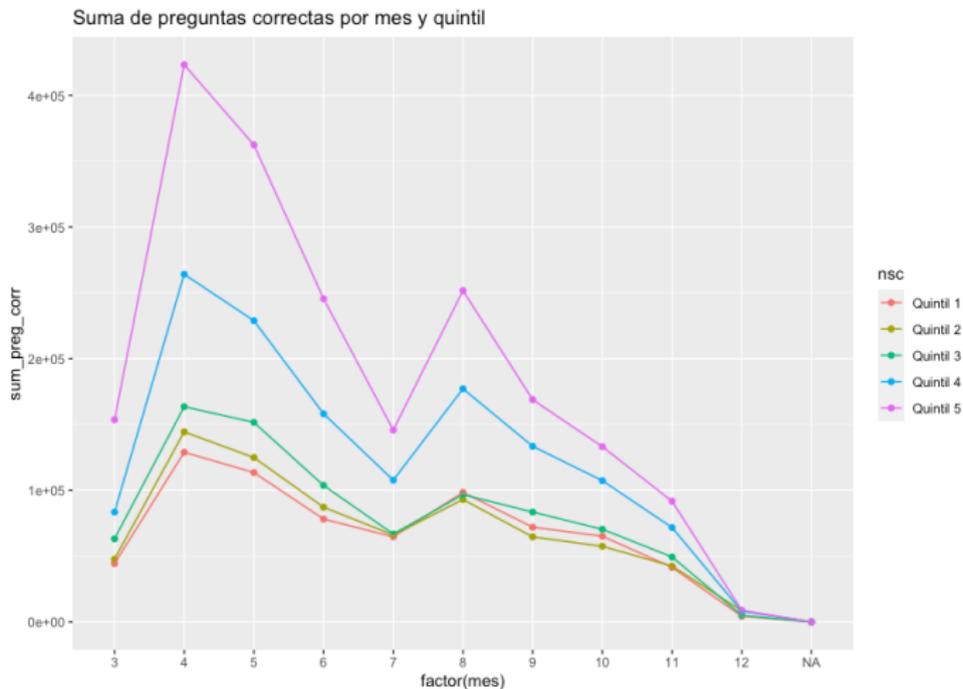
##	Cd.act	Min.act	Max.act	Tot.act	Tot.preg	Cor.preg
## 1	act_32	0.5	0.5	1	10	5
## 2	act_32	0.5	0.5	1	10	5
## 3	act_33	1.0	1.0	1	2	2
## 4	act_402	1.0	1.0	1	1	1
## 5		NA	NA	NA	NA	NA
## 6	act_16	0.3	0.6	2	20	9
## 7	act_18	1.0	1.0	1	12	12

Otras variables: estudiante, centro, contexto, fecha,

Intentos por mes



Preguntas correctas por mes

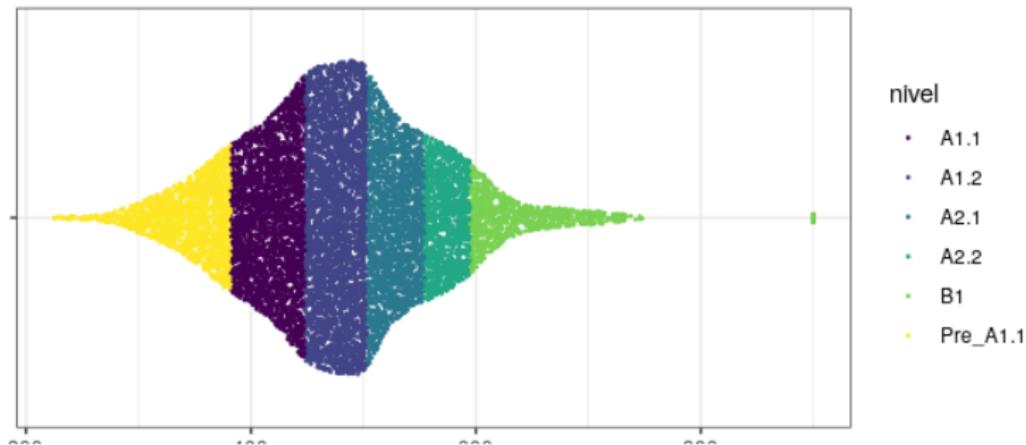


Datos de resultados

Prueba adaptativa de Inglés

- ▶ Componentes evaluados: Vocabulario-gramática (VG) y Lectura (R)
- ▶ entre 15 Noviembre y 15 Diciembre (aprox)
- ▶ \approx 12 mil estudiantes que usan LB

Datos de resultados



Hay 12% de estudiantes que no alcanzan el nivel A1.1

Ciencia de datos en plataformas educativas

Los datos

Modelos predictivos

Modelos predictivos

Es posible predecir el resultado de la prueba en base a la información de uso de plataformas ?

Variable de respuesta

- ▶ **Puntos en la prueba**
- ▶ Nivel alcanzado
- ▶ Alcanza nivel A1.1

Variable explicativas

- ▶ **Totales de uso en LB por mes**
- ▶ Uso de CREA (en inglés)
- ▶ Variables adicionales (departamento, contexto, clase, etc)

Un primer intento ...

- ▶ Utilizando la biblioteca `h2o`
- ▶ Problema de regresión
- ▶ Aplicamos una búsqueda automática

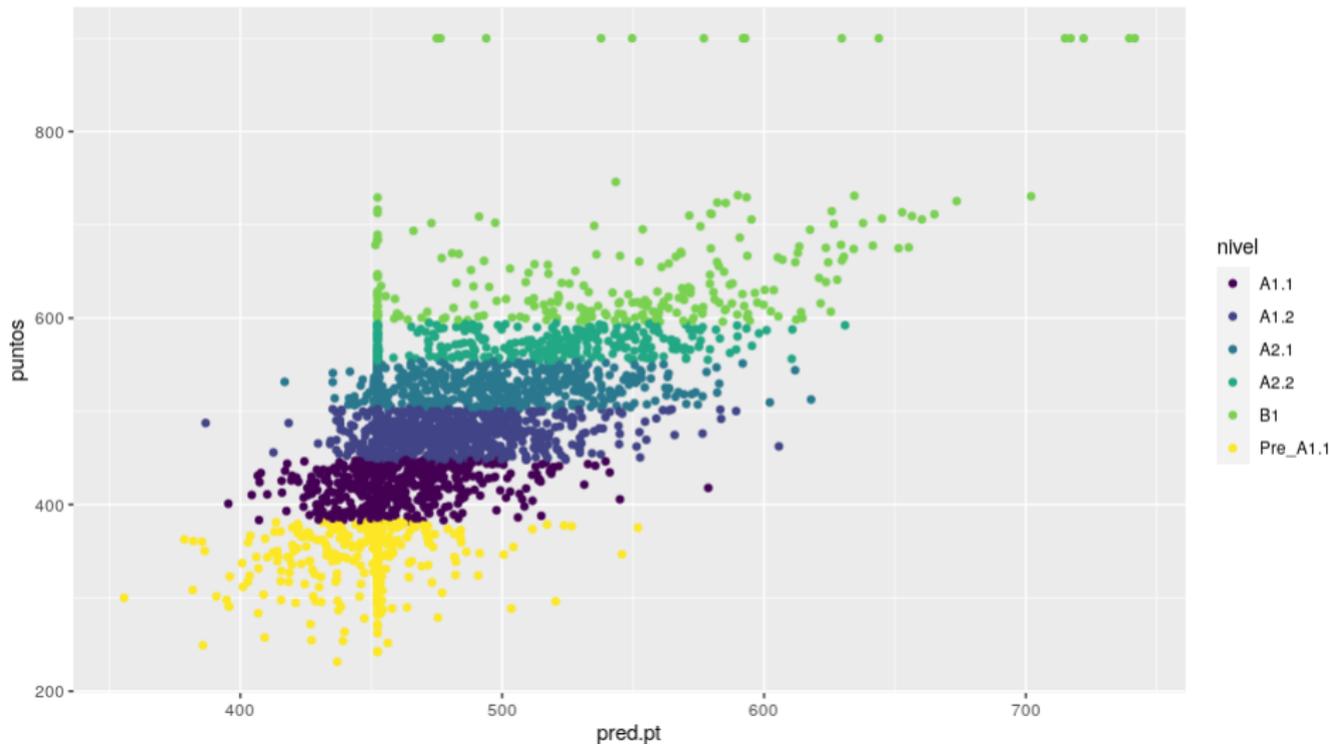
Un primer intento ...

Ranking de los modelos ajustados por h2o

##		m1	m2	rmsle
## 1	StackedEnsemble		AllModels	0.1615970
## 2	StackedEnsemble		AllModels	0.1620236
## 3	StackedEnsemble		Best1000	0.1619228
## 4	StackedEnsemble		AllModels	0.1620080
## 5	StackedEnsemble		AllModels	0.1620925
## 6	StackedEnsemble	BestOfFamily		0.1620567

Los primeros son todas combinaciones de otros modelos (*Stacked*)

Un primer intento ...



Siguientes pasos

- ▶ Incorporar otras fuentes de información
- ▶ Mejorar variables de uso
- ▶ Trabajar con modelos a medida tomando en cuenta la estructura de los datos.