

Evaluación y monitoreo de plataformas educativas

Federico Molina
Natalia da Silva
Ignacio Alvarez-Castro
Juan José Goyeneche

Serie Documentos de Trabajo

Nº2/21

Feb. 2021

ISSN : 1688-6453

Universidad de la República.
Facultad de Ciencias Económicas y de Administración,
Instituto de Estadística (IESTA)

Montevideo, Uruguay.



Esta obra está bajo una Licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional.

Forma de citación sugerida para este documento:

Molina, Federico, da Silva, Natalia, Alvarez-Castro, Ignacio y Goyeneche, Juan José. (2021). *Evaluación y monitoreo de plataformas educativas*. (Serie Documentos de trabajo; Nº2/21). Montevideo: Universidad de la República. Facultad de Ciencias Económicas y de Administración, Instituto de Estadística.

<https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/10518>

Evaluación y monitoreo de plataformas educativas

Federico Molina ¹ Natalia da Silva ² Ignacio Alvarez-Castro ³ Juan José Goyeneche ⁴;
Universidad de la República,
Facultad de Ciencias Económicas y de Administración
Departamento de Métodos Cuantitativos
Instituto de Estadística

Resumen

En este trabajo se presentan los principales resultados del proyecto de la Agencia Nacional de Investigación e Innovación (ANII): “Nuevas Tecnologías Estadísticas para la Evaluación y Monitoreo de Plataformas Educativa” desarrollado en el marco del fondo sectorial de investigación a partir de datos 2017. Se evalúa el uso de la plataforma CREA, donde el objetivo específico es analizar los patrones de comportamiento de los estudiantes. Para ello se desarrolla una aplicación web e indicadores de resumen para monitorear el uso de la plataforma CREA a distintos niveles de análisis (clase, grado, escuela, departamento) y distintos momentos del tiempo. Estos indicadores y resúmenes de información están implementados para diferentes actores del sistema educativo como mentores, docentes y directores, entre otros. Adicionalmente, la solución propuesta genera de forma sencilla y sistematizada reportes dinámicos en los distintos niveles de análisis para los distintos actores. Este proyecto es completamente desarrollado en R (software libre para realizar estadística computacional y gráficos) combinando `shiny` y `rmarkdown` entre otros paquetes. El monitor educativo tiene como objetivo contribuir a la calidad de la enseñanza posibilitando la identificación de alumnos que estén en riesgo de rezago académico. A su vez se analiza la capacidad predictiva de la información generada en CREA respecto de los resultados de la prueba adaptativa del programa Ceibal en inglés mediante métodos de aprendizaje estadístico. Para obtener modelos predictivos más precisos es necesario contar con información del uso de la plataforma CREA específica para las actividades de Ceibal en Inglés.

Palabras clave: Monitor educativo, visualización estadística, aprendizaje estadístico, ciencia de datos.

Códigos JEL: C88, I29

Clasificación MSC2010: 62-07, 68N99

¹ *email:*fmolina@iesta.edu.uy, ORCID: <https://orcid.org/0000-0001-7613-1179>

² *email:*natalia@iesta.edu.uy, ORCID: <https://orcid.org/0000-0002-6031-7451>

³ *email:*nachalca@iesta.edu.uy, ORCID: <https://orcid.org/0000-0003-1633-2432>

⁴ *email:*jjgoye@iesta.edu.uy, ORCID: <https://orcid.org/0000-0002-4628-2731>

Abstract

This paper presents the main results from ANII project: “New Statistical Technologies for the Evaluation and Monitoring of Educational Platforms” founded by fondo sectorial de investigación a partir de datos 2017 grant. CREA platform use is evaluated, where the specific objective is to analyze student’s pattern behavior. A web application and summary indicators are developed to monitoring the use of CREA at different levels of analysis (class, grade, school, department) and different moments in time. These indicators and information summaries are implemented for different roles in the education system such as mentors, teachers, and principals, among others. Additionally, the proposed solution generates dynamic reports in a simple and systematized way at the different levels of analysis for the different roles. This project is completely developed using R (free software environment for statistical computing and graphics) combining `shiny` and `rmarkdown` among other packages. The educational monitor objective is to contribute to the teaching quality making it possible to identify students at risk of academic lag. Additionally, predictive models for adaptative English test results are fitted using CREA information as predictor variables. To get more accurate predictive models it is necessary to have specific CREA use the information for Ceibal en Inglés.

Key words: Educational data monitor, statistical visualization, statistical learning, data science

JEL Codes: C88, I29

Mathematics Subject Classification MSC2010: 62-07, 68N99

1. Introducción

En los últimos años Plan Ceibal ha implementado varias herramientas educativas que apuntan a complementar las clases presenciales con la educación virtual.

Por un lado, se han puesto a disposición dos plataformas educativas operativas desde 2013: la Plataforma Adaptativa de Matemática (PAM) y CREA. El entorno virtual de aprendizaje (CREA) permite gestionar cursos, crear o compartir materiales didácticos y trabajar colaborativamente en grupos. Por otro lado, Ceibal en Inglés es una iniciativa que surge en 2012 con el objetivo de apoyar la enseñanza de inglés. En primaria esta iniciativa intenta cubrir la falta de docentes mientras que en educación media se enfoca en la destrezas de oralidad mediante videoconferencias con un profesor remoto. Para la evaluación de aprendizaje de la lengua inglesa se han implementado evaluaciones anuales adaptativas en línea desde el 2014 para los niños de 4°, 5° y 6° año.

El uso de las computadoras de Ceibal y la introducción de las mismas en el proceso de enseñanza depende mucho de la propuesta pedagógica de los educadores, lo cual se evidencia en que no todos los docentes han incorporado la herramienta en su trabajo en el aula. En este sentido, es necesario para evaluar correctamente el plan la información sobre el uso de las computadores en el aula.

Plan Ceibal genera información a nivel de cada computadora individual que, en su mayor parte, ha estado enfocado en la gestión y no en la generación de información relevante para docentes con fines educativos. Elaborar herramientas que permitan resumir la información, ponerla a disposición en tiempo real o al menos utilizable y aplicar métodos estadísticos para la evaluación y monitoreo son fundamentales para el éxito del plan y el diseño de estrategias educativas.

El objetivo general del trabajo es desarrollar herramientas estadísticas que permitan evaluar y monitorear el uso de las plataformas educativas del Plan Ceibal en vínculo con los aprendizajes. Entender de qué forma el uso de las plataformas educativas afectan o influyen en el aprendizaje es crucial para mejorar su desarrollo e incentivar su uso. Así mismo desarrollar indicadores de uso de las plataformas que estén disponibles rápidamente es importante para el monitoreo de las mismas. Este trabajo se centrará en la información generada por la plataforma CREA y su vínculo con el resultado de los aprendizajes de las pruebas adaptativas de inglés.

En primer lugar, se busca desarrollar indicadores para monitorear el uso de la plataforma a distintos niveles de análisis (clase, grado, escuela, departamento) y en diferentes momentos del tiempo. Estos indicadores y resúmenes de información estarán implementados para diferentes actores del sistema educativo como mentores, docentes y directores (entre otros). La solución propuesta busca generar en tiempo real, sencilla y sistematizadamente la información a través de un reporte dinámico en los distintos niveles de análisis que incluyan visualización estadística así como medidas de resumen e indicadores para el monitoreo y modelos predictivos de aprendizaje. Las herramientas serán generadas con el software libre

R combinando `shiny` (Chang *et al.*, 2020) y `rmarkdown` (Xie *et al.*, 2018) para cumplir con este objetivo entre otros. El producto final se basa en la visualización estadística de datos que permitan resumir de forma sencilla para el usuario final patrones complejos en altas dimensiones. Para la visualización estadística se trabaja con la librería `ggplot2` (Wickham, 2016) ya que permite visualizaciones de alta calidad y su base teórica es la gramática de gráfica desarrollada por Wilkinson (1999).

Por otro lado, se busca responder cómo se relaciona el uso de la plataforma CREA con la performance en las pruebas adaptativas de inglés. Marconi *et al.* (2017) sugieren que el uso de CREA se correlaciona con los desempeños. En dicho trabajo, estudian esta relación comparando la distribución de performance en la prueba para distintos niveles del uso de la plataforma. Tomando como punto de partida este trabajo, se propone identificar de qué forma la información de uso de CREA puede ser utilizada para obtener nuevos indicadores de uso de la plataforma educativa. Los indicadores de uso en combinación con otras variables, son los insumos para implementar buenos modelos predictivos para los resultados de las pruebas adaptativas de inglés.

Para diseñar las herramientas y modelos mencionados anteriormente, se utilizaron datos de uso de la plataforma CREA generadas durante el año 2015 y 2017 y los resultados de la prueba adaptativa de Ceibal en Inglés de 2017 en ambos casos anonimizadas (disponibles al momento de realizar este proyecto).

En la siguiente sección se presentan algunos antecedentes relevantes para este trabajo, en la Sección 3 se describen los datos su estructura y el procesamiento de los mismos, en la Sección 4 se presentan los principales resultados del trabajo y finalmente en la Sección 5 se realizan algunos comentarios finales.

2. Antecedentes

Uruguay en los últimos 30 años, a partir de la creación de distintos programas e iniciativas (PEDECIBA, ANII y CEIBAL), ha dado señales de su apuesta al desarrollo del sector de Ciencia e Innovación. Entendiendo que la generación de conocimientos y de capital humano son la base para el desarrollo del país.

En esa línea, se implementa en Uruguay como política pública de carácter universal el Plan Ceibal que forma parte de la iniciativa mundial One Laptop per Child (OLPC). Según lo establecido en la ley número 18.640 promulgada en el año 2010 la cual genera la creación del Centro Ceibal, dos de sus principales cometidos son “Contribuir al ejercicio del derecho a la educación y a la inclusión social mediante acciones que permitan la igualdad de acceso al conocimiento” y “Promover, coordinar y desarrollar planes y programas de apoyo a las políticas educativas para niños y adolescentes elaboradas por los organismos competentes” (Uruguay, 2010).

Plan Ceibal ha implementado el “modelo uno a uno” que consiste en otorgar un dispositivo

(laptop o tablet) de su propiedad a cada alumno y docente de la enseñanza pública básica (educación inicial y primaria, y educación media básica). De esta forma ha promovido la igualdad de acceso a la tecnología, así como ha asegurado el acceso a internet en todos los centros educativos públicos en sus 10 años de implementación en el país.

Una de las innovaciones más importantes en el sistema educativo de las últimas décadas a nivel mundial es la introducción de nuevas tecnologías que modifican la forma de enseñar y aprender. En el sistema educativo el cambio tecnológico se plasma en cursos basados en la web, sistemas de gestión de contenidos para el aprendizaje, sistemas inteligentes y adaptativos del aprendizaje basados en la web. En este contexto el Plan Ceibal se empalma con el nuevo paradigma de aprendizaje con un fuerte componente tecnológico.

Estas nuevas tecnologías en el aula no solamente presentan un desafío en cuanto a el paradigma de aprendizaje mediado por tecnología, sino también un desafío para analizar los datos generados por el sistema. La información generada son datos en altas dimensiones los cuales tienen una estructura compleja, puesto que se cuenta con registros individuales de uso de las distintas tecnologías a lo largo del tiempo y con variables a nivel individual contenidas en distintas fuentes de información.

Estas dos características hacen que los métodos estadísticos clásicos presenten dificultades para capturar estructuras complejas por lo que es recomendable implementar técnicas estadísticas modernas que permitan mayor flexibilidad.

Los métodos de aprendizaje estadístico consisten en el entrenamiento de un modelo de forma que aprenda diversos comportamientos usando información de un subconjunto de observaciones. Los métodos de aprendizaje estadístico son ampliamente utilizado en una variedad de problemas de aplicación en diversas disciplinas, tales como la economía, informática, biología, medicina, etc. La complejidad de los problemas de aplicación requieren de algoritmos automáticos que capturen las características fundamentales de los datos de forma de tener una adecuada performance predictiva para el problema de interés. El aprendizaje estadístico comprende un amplio conjunto de métodos que se pueden dividir en tres enfoques, aprendizaje supervisado, aprendizaje no supervisado y aprendizaje semi-supervisado. El aprendizaje supervisado implica que la variable de respuesta es conocida y el objetivo es construir un predictor para la misma usando información de las variables explicativas. Cuando la respuesta es categórica el problema es llamado de clasificación y cuando la respuesta toma valores continuos el problema es llamado de regresión.

En segundo lugar el aprendizaje no-supervisado implica que la variable de respuesta no es conocida, es decir no contamos con etiquetas para la misma. En este caso el problema se centra en el estudio de la estructura de los datos tratando de identificar grupos homogéneos y la clasificación automática de los datos dentro de tales grupos, la posibilidad de identificar el mecanismo generador de los datos, por ejemplo la estimación de la densidad de una muestra.

Por último, el aprendizaje semi-supervisado se conoce la variable de respuesta solo para algunas observaciones. En su versión simple, en la etapa inicial se entrena un modelo

con las observaciones donde se conoce la respuesta, en una segunda etapa se predice la respuesta de las observaciones donde se desconoce la misma y las observaciones con mejor predicción se añaden al conjunto de entrenamiento, posteriormente se vuelve a entrenar el modelo. Este proceso requiere algún supuestos sobre la distribución de los datos, como por ejemplo el supuestos de continuidad, supuesto de agrupamiento entre otros.

Los métodos que se utilizarán en este proyecto se enmarcan en el análisis supervisado, donde se encuentran una gran variedad de métodos muchos de los cuales han tenido gran éxito en cuanto a la performance predictiva en una amplia variedad de aplicaciones. Uno de los métodos ampliamente utilizados en este contexto son los árboles de decisión como CART (Breiman *et al.*, 1984), utilizados tanto para la clasificación como para la regresión. Este método presenta un gran número de ventajas: presentación gráfica del modelo bajo la forma de un árbol binario, capacidad de manejar datos heterogéneos sin codificación, categóricos o continuos, el aporte de un índice de importancia para cada variable y el no ser afectado mayormente por la existencia de datos faltantes. Esta técnica ha sido objeto de numerosas extensiones como Loh (2014); Loh y Zheng (2013). Sin embargo CART presenta el inconveniente mayor de ser inestable, es decir un pequeño cambio en la muestra de entrenamiento puede conducir a modelos predictivos totalmente distintos. Los métodos de agregación de predictores aparecieron como solución al problema de inestabilidad de CART teniendo una gran importancia en el aprendizaje estadístico. Estos métodos combinan varios predictores estadísticos con la finalidad de obtener un predictor con mayor performance predictiva. Dentro de estas técnicas se puede citar a bagging (Breiman, 1996a), boosting (Schapire (1990), Freund (1995), Freund y Schapire (1997)) y los bosques aleatorios (Breiman, 2001). Asimismo existen muchas extensiones a los métodos clásicos de agregación que varían tanto en los predictores a agregar así como la forma en que son agregados en el modelo final.

Los métodos utilizados en este proyecto fueron implementados en base a H2O (plataforma de código abierto para aprendizaje automático escalable) disponible en el paquete `h2o` (LeDell *et al.*, 2020) en R. A continuación se describen brevemente algunas de las técnicas implementadas en H2O y que fueron seleccionadas como posibles modelos predictivos para el problema de interés.

Los bosques aleatorios combinan árboles de clasificación o regresión (predictores) con el objetivo de reducir la varianza del modelo. En la construcción de los bosques aleatorios se introducen dos fuentes de aleatoriedad, en primer lugar se extraen muestras bootstraps y para cada una de ellas se ajusta un árbol de clasificación o regresión. La segunda fuente de aleatoriedad es introducida en la construcción de cada árbol individual donde se selecciona aleatoriamente un subconjunto de variables utilizados en la partición de los nodos. Para problemas de regresión la predicción del bosque es el promedio de las predicciones de los árboles individuales y en problemas de clasificación la predicción se obtiene por voto mayoritario.

Una extensión son los árboles extremadamente aleatorios (Extremely randomized trees,

XRT, [Geurts et al. \(2006\)](#)), donde la aleatoriedad va un paso más allá en la forma en que se calculan las divisiones. Al igual que en los bosques aleatorios, se utiliza un subconjunto aleatorio de variables, pero en lugar de buscar los umbrales que más discriminan, los umbrales se generan aleatoriamente para cada variable candidata y el mejor de estos umbrales se elige como regla de división. Lo cual puede reducir aún más la varianza del modelo, a expensas de un leve aumento del sesgo.

Gradient boosting machine (GBM, [Friedman \(2001\)](#), [Friedman \(2002\)](#), [Tibshirani et al. \(2009\)](#)) es una técnica de aprendizaje estadístico para problemas de regresión y clasificación, que produce un modelo de predicción en forma de un conjunto de predictores débiles, generalmente árboles de decisión. Construye el modelo por etapas como lo hacen otros métodos de boosting, y los generaliza al permitir la optimización de una función de pérdida diferenciable. Pueden verse como algoritmos iterativos de descenso de gradiente funcional, es decir, algoritmos que optimizan una función de costo sobre el espacio funcional eligiendo iterativamente una función (hipótesis débil) que apunta en la dirección negativa del gradiente. La idea principal es que se pueden obtener buenos resultados predictivos mediante aproximaciones cada vez más finas en la construcción de los predictores débiles. En el caso particular del algoritmo utilizado mediante H2O ([LeDell et al., 2020](#)), se construyen secuencialmente árboles de clasificación (regresión) en todas las características del conjunto de datos de una manera completamente distribuida, es decir, cada árbol se construye en paralelo aumentando la eficiencia computacional.

Una extensión de GBM, es XGBoost ([Chen y Guestrin, 2016](#)), el cual es un algoritmo novedoso capaz de manejar conjuntos de datos dispersos, con una justificación teórica basada en un esquema ponderado por cuantiles. Específicamente, XGBoost utiliza una forma de modelo más regularizada para controlar el sobreajuste, lo que le suele dar un mejor rendimiento que GBM y puede ser visto como aumento de gradiente regularizado ([Chen y Guestrin, 2016](#)).

Por último, el método de stacking (también conocido como stack generalization [Wolpert \(1992\)](#), stacked regression, [Breiman \(1996b\)](#) o superlearner ?) es una técnica de agregación de predictores al igual que bagging (bosques aleatorios) y boosting (GBM). Por tanto, son una clase de algoritmos que implica entrenar a un agregador (metalearner) en una etapa posterior para encontrar la combinación óptima de los predictores iniciales. Aunque a diferencia de bagging y boosting, el objetivo en stacking es lograr conjuntos de predictores diversos y potentes.

En este trabajo se utiliza SuperLearner, el cual representa un sistema asintóticamente óptimo para el aprendizaje ([Van Der Laan et al., 2007](#)), que es una variación del método de stacking dada la utilización de validación cruzada para formar el conjunto de datos (datos de nivel uno) sobre el cual se entrena al agregador.

El aprendizaje profundo (deeplearning, [Lecun et al. \(2015\)](#)) es parte de una familia de métodos de aprendizaje estadístico basados en redes neuronales artificiales (ANN) con aprendizaje de representación ([Bengio et al. \(2013\)](#)). Existen múltiples arquitecturas de

aprendizaje profundo como las redes neuronales profundas, las redes de creencias profundas, las redes neuronales recurrentes y las redes neuronales convolucionales. En este trabajo se utiliza una red neuronal artificial totalmente conectada de retroalimentación de múltiples capas (Fully-connected multi-layer feedforward artificial neural network) entrenada mediante descenso de gradiente estocástico con retropropagación (backpropagation). La red puede contener una gran cantidad de capas ocultas que consisten en neuronas con funciones de activación y características avanzadas como la tasa de aprendizaje adaptativo, el entrenamiento de impulso, deserción, el tipo de regularización (L1 o L2), los puntos de control y la búsqueda mediante una grilla permiten una alta capacidad predictiva.

Existen a su vez antecedentes específicos en el área de la educación ya que estas nuevas formas de aprender y enseñar basadas en tecnología requieren a su vez formas distintas de evaluar y monitorear las estrategias educativas. Recientemente, se ha identificado como “Educational Data Mining” a una sub-disciplina dedicada al estudio de las metodologías estadísticas para analizar los datos generados en sistemas educativos con el objetivo de mejorar la performance de dichos sistemas (Baker, 2011). Thakar (2015) presenta una revisión de artículos en este campo. Los objetivos de las investigaciones de este tipo son la creación de tipologías de estudiante (por ejemplo con técnicas de clustering) y más comúnmente la predicción de performance y/o de abandono en un curso o un programa de estudio. Los modelos predictivos pueden servir como indicadores tempranos que sirvan para detectar a estudiantes en riesgo. Rovira *et al.* (2017) realizan una comparación de varios métodos de aprendizaje estadístico para predecir la nota final y el abandono en estudiantes universitarios.

Bakhshinategh *et al.* (2018) estudian tareas y aplicaciones existentes en el área de minería de datos educativos identificando 13 categorías de aplicaciones categorizadas según su objetivo (modelado de estudiantes, sistema de soporte para la toma de decisiones, sistema adaptativo, evaluación e investigación científica). Por su parte Baker (2011), identifican 5 aproximaciones dentro de la minería de datos educativos: predicción, agrupamiento, minería de relaciones, descubrimiento con modelos y procesamiento de datos para el juicio humano. Estas dos referencias enmarcan nuestro proyecto dentro de los objetivos de modelado de estudiantes y soporte para la toma de decisiones. Dentro del primer objetivo la aproximación será basada en predicción y para el segundo enfocado en el procesamiento de datos para el juicio humano como lo especifica Baker (2011).

Severín y Capota (2011) mencionan que a pesar de la popularidad de la iniciativa de OLPC en América Latina y el Caribe, el monitoreo del impacto de las mismas no están debidamente cuantificado o cuentan con análisis parciales de la realidad. Mencionan a su vez la importancia fundamental de medir el impacto de la tecnología en el aprendizaje curricular. Las medidas de impacto pueden ser usadas para identificar estrategias pedagógicas y usar las tecnologías que tienen mayor impacto en el aprendizaje.

En este contexto, es relevante el uso de técnicas estadísticas que permitan obtener mejor provecho de la enorme cantidad de información generada por CEIBAL para el monitoreo

y el diseño de políticas educativas.

A nivel nacional, existen estudios que dan indicios sobre el impacto que tiene el uso de las plataformas educativas en el aprendizaje. En particular, la evaluación de impacto de la Plataforma Adaptativa de Matemática en los resultados de los aprendizajes realizado por el Centro de Investigaciones Económicas (CINVE) da cuenta de la ganancia de aprendizajes en matemática a partir de información longitudinal de una muestra de alumnos de la educación primaria. Esta primera evidencia a escala país, del impacto de una herramienta pedagógica de este tipo, muestra que la posibilidad de mejorar la calidad de la educación a través del uso de la tecnología es una alternativa real (Perera y Aboal, 2019).

Por otra parte, el estudio realizado sobre el efecto de la modalidad de enseñanza de inglés (virtual vs presencial) en el nivel de compromiso de los alumnos con la plataforma CREA, y su asociación con los logros de los estudiantes dan cuenta de la importancia del rol docente, así como del contexto sociocultural, y del grado escolar en el mayor nivel de apropiación de la plataforma. Además se evidencia la asociación positiva entre niveles de compromiso y desempeño por parte de los alumnos en la evaluación de aprendizajes (Marconi *et al.*, 2017).

3. Datos

En esta sección se describen los datos utilizados en el proyecto. Se trabaja con dos conjuntos de datos anonimizados facilitados por Plan Ceibal bajo acuerdo de confidencialidad. El primero corresponde a los logs de las actividades realizadas por los alumnos en la plataforma CREA durante los años 2015 y 2017. El segundo refiere a los puntajes de los alumnos en las pruebas adaptativas de inglés de comprensión auditiva (listening) y, vocabulario, lectura y gramática (VRG) en los años 2016 y 2017.

Los datos de 2015 son logs, los cuales registran cualquier actividad que el alumno haya realizado, desde envío de trabajos, acceso a la plataforma, respuestas en foros, entre otras. Para cada alumno i (de un total de n) hay N_i registros que corresponden a las N_i interacciones que realizan con la plataforma, por lo cual se cuenta con un total de $\sum_i^n N_i$ registros. En el año 2015 se tienen 13.745.973 millones de registros, en torno a cuatro millones de observaciones que corresponden a cuarto y quinto grado y más de cinco millones a sexto grado. Se cuenta con 24 variables con la fecha del evento, el identificador de alumno y escuela, el grado y el grupo del alumno, el departamento de la escuela, el tipo de dispositivo utilizado donde las variables más relevantes son el nombre del evento y la categoría del evento. El formato de los logs se observa en la Tabla 1, donde la primera columna es la variable que refiere al evento que se registra y la segunda a que categoría corresponde. Las variables se generan a partir de la categoría del evento (`event_category`). Los logs del año 2017 son procesados por Plan Ceibal. Se recibe un conjunto de datos con formato longitudinal donde cada observación (cada fila) contiene la información del total

Tabla 1: Formato de logs año 2015

Nombre del evento	Categoría del evento
visit:authenticated:new	visit
login:school_subscription:full	login
visit:authenticated:old	visit
login:school_subscription:full	login
file_upload:jpg	file_upload
comment_posted:course:assignment	comment_posted

de actividad realizada por el alumno en cada día que tiene alguna actividad. Son 1.355.972 registros correspondientes a 120276 alumnos de los cuales 87915 tienen alguna actividad y 32361 no registran ninguna interacción con CREA. Adicionalmente, los datos de 2017 contienen los puntajes de los alumnos en las pruebas adaptativas de inglés de los años 2016 y 2017, en ambos años un total de 15776 estudiantes tomaron la prueba adaptativa de listening y, 19352 rindieron la prueba de VRG. El total de variables es 77 (incluyendo 22 variables generadas en una etapa posterior para mejorar la performance del modelo utilizada en las predicciones de las categorías en las pruebas adaptativas de inglés), entorno a 30 refieren a alguna interacción del estudiante con la plataforma CREA que pueden ser catalogadas como: nivel de actividad (ejemplo, días de ingreso), actividad en los foros (comentarios posteados), envío de tareas (envíos calificables) y medidas generales como el total de ingresos a la plataforma. El resto se divide en variables de características del alumno, por ejemplo, el sexo, la fecha de nacimiento, el curso, grado y grupo, relacionadas al centro de estudio, como el contexto socioeconómico (por quintiles, según definición de CEIBAL), la zona geográfica (urbana o rural), el departamento y la localidad. A su vez se cuenta con información del dispositivo y el tipo de conexión a internet, variables de puntajes en las pruebas de inglés, los puntajes transformados a categorías según el marco común europeo y finalmente si el alumno toma clases extracurriculares de inglés o no. En la Tabla 3 se muestran las variables utilizadas en el análisis.

3.1. Procesamiento

En el caso de 2015 se generan variables de actividad de los alumnos a partir de la variable categoría del evento (`event_category`) y fecha del evento (`event_date`), ver Tabla 1. Como cada registro es una interacción del alumno con la plataforma se cuentan los registros de cada categoría, adicionalmente se combina con la fecha del evento para realizar un recuento de comentarios por día y el número de días en cada categoría. De la combinación de ambas surgen las variables de comportamiento de los estudiantes. Las mismas están disponibles en 2017 aunque en dicho caso son procesadas por CEIBAL. Algunas de estas variables son utilizadas posteriormente en la construcción del indicador de compromiso (`engagement`) y son analizadas en detalle en el monitor de plataformas educativas.

Como se observa en la Tabla 3 las variables utilizadas se han dividido en categorías prin-

cipales, como variables de comportamiento, otras variables de los estudiantes, variables de la escuela o puntajes en las pruebas de inglés. Se crean variables de comportamiento en base a los logs generados por la actividad del alumno en la plataforma. Como ejemplo, `cr_comentariosposteados` se refiere a todos los comentarios publicados por el alumno en todas las materias utilizadas en CREA. El número de comentarios está entre 0 y 842, el máximo se corresponde aproximadamente dos comentarios diarios. Además, hay información de la actividad general de los estudiantes en la plataforma relacionada a la cantidad de veces que inician sesión, comentan una actividad, cargan actividades, crean eventos, buscan recursos, envían tareas, entre otras. Si el alumno no tuvo ninguna actividad entonces no posee registros en la plataforma, sin embargo, puede pertenecer al marco de alumnos de primaria por ello esta presente en el conjunto de datos. En dicho caso, se cuenta con una observación por alumno la cual no tiene fecha de actividad, por lo tanto, todos los registros que cumplen dicha condición tienen actividad cero.

El conjunto de datos cuenta con aproximadamente 44 mil niños y niñas y 32000 estudiantes sin ninguna información sobre su sexo, estos últimos son los mismos que no tiene actividad en CREA. La edad y la fecha de nacimiento nos dan la posibilidad de identificar a los alumnos repetidores o adelantados. Contamos con información de 4^o, 5^o y 6^o grado en donde hay escuelas con grados que tienen un grupo o hasta seis grupos inclusive. Además, se sabe a qué turno asisten los alumnos, casi 50.000 van a doble turno, qué tipo de dispositivo y marca están utilizando lo cual puede ser relevante en caso de una falla sistemática en alguna marca de computadora.

Se tiene información geográfica relacionada al departamento y localidad de la escuela pero no del estudiante. Si bien es plausible suponer que el estudiante es del mismo departamento que su centro de estudios, no es evidente que sea de la misma localidad, lo cual dificulta poder inferir la situación económica familiar. Es posible aproximarse por el contexto socioeconómico de la escuela, ya que, la variable contexto brinda una categorización hecha por CEIBAL sobre el contexto económico de las escuelas las cuales se dividen en quintiles, sin embargo, puede existir heterogeneidad en los contextos socioeconómicos de las familias que pertenecen a un centro educativo.

Tanto los datos de 2015 y 2017 son transformados en dos tablas finales que se utilizan en el monitor. La primer tabla es de corte transversal donde cada fila presenta un registro de un alumno (hay tantos registros por alumno como eventos realice) y otra de formato panel o longitudinal con una observación por alumno-mes. En el caso de no tener actividad en un mes se imputa el valor cero, ya que, si bien no se observa que hizo el alumno, se sabe con certeza que no tuvo actividad en la plataforma.

La elección de los formatos de datos a utilizar tienen como objetivo analizar la variabilidad de la información en el tiempo y en el espacio. De esta forma con un conjunto de datos longitudinal es posible analizar la evolución de los alumnos a lo largo del año mientras con los datos en formato de sección cruzada el foco es en un momento particular (por ejemplo, fin de año) en el cual es posible estudiar la variabilidad entre alumnos. De esta forma,

es posible identificar dinámicas generales como la caída de la actividad en la plataforma durante los meses con vacaciones o bien comportamientos individuales con caídas abruptas de la actividad que quedan solapadas en el indicador al final del periodo de análisis. Los datos de sección cruzada fueron agregados con sus valores a final de año mientras los datos longitudinales se agrupan con formato alumno-mes, esto se realiza para tener un volumen de datos manejables puesto que inicialmente se trabaja con el paquete `dplyr` (Wickham *et al.*, 2019), sin embargo, posteriormente se reescribe el código utilizando la librería de alta performance `data.table` (Dowle y Srinivasan, 2020) escrita en C y multithread. Este cambio posibilita realizar cálculos en tiempo real en la aplicación sin necesidad de tener que generar múltiples tablas intermedias pre-computadas.

Por último, con los datos de CEIBAL en inglés se obtienen los puntajes y la categoría de las pruebas de inglés realizadas en comprensión auditiva (`thotfin_list`), vocabulario, lectura y gramática (`thotfin_vrg`). Sin embargo, es necesario filtrar las puntuaciones debido a puntajes que se encuentran por fuera del rango de valores posibles de la variable de interés. Esto se realiza utilizando las variables `fuera_rank_list17` y `fuera_rank_vrg17`. Las mismas variables y problemas están disponibles para 2016. Al momento de analizar los puntajes de inglés, solo se utilizan los alumnos con calificación en ambos años, dado que es de interés analizar la evolución de los alumnos y ver si en promedio su desempeño se modifica.

4. Resultados

En esta sección se presentan los principales resultados del proyecto, el cual incluye el índice de engagement (IE) o compromiso utilizado, un análisis exploratorio del comportamiento de los alumnos en la plataforma CREA, sus resultados en las pruebas adaptativas de inglés, el monitor de plataformas educativas y los modelos de aprendizaje estadístico utilizados para predecir los puntajes en las pruebas de inglés. Tanto los resultados que se presentan sobre el monitor educativo como sobre los modelos de aprendizaje son realizados utilizando datos para el año 2017.

4.1. Índice de Engagement

Para medir el uso de la plataforma CREA por parte de los estudiantes se calcula el índice de engagement o compromiso (IE) planteado inicialmente por Marconi *et al.* (2017). Según Marconi *et al.* (2017) el IE tiene propiedades deseables, por ejemplo: a mayor intensidad y constancia de uso, mayor es el valor del índice, su valor está acotado inferior y superiormente (cero y uno), plantea un rendimiento marginal decreciente lo cual es coincidente con la curva de aprendizaje, es comparable entre alumnos y, por último, el valor del IE de un alumno no se ve influenciado por la actividad de otros alumnos. Su

forma funcional se presenta en la Ecuación 1.

$$IE_{it} = \frac{\delta + 1}{\delta} \left(\frac{e^{\pi_{it}}}{\delta + e^{\pi_{it}}} - \frac{1}{\delta + 1} \right) \quad (1)$$

i refiere al alumno y t refiere al momento del tiempo en este caso algún mes particular. El parámetro δ representa un coeficiente de exigencia elegido de forma tal que la distribución acumulada del índice sea *suave*. En el análisis anual el valor de δ es igual a 90, mientras en el cálculo mensual el valor de δ se modifica a 90/12, es decir, se ponderan todos los meses por igual. El valor $e^{\pi_{it}}$ representa la exponencial de la suma de los logaritmos naturales de las variables seleccionadas como dimensiones del índice, las cuales cumplían tener una correlación mayor o igual a 0.8 con la mayor cantidad de variables (2 a 2). En la sumatoria de la exponencial todas las variables tienen la misma ponderación. Las variables utilizadas fueron: cr.comentariosposteados, cr.comentariosposteadosassignmen, cr.totalcomentariosposteados, cr.totalsubmissions, cr.totalvisitas, cr.totaldiasingreso, cr.comentariosdiscusiones (ver definiciones en Tabla 3).

Adicionalmente se elaboró un índice de engagement de Ceibal en Inglés (IE_{CEI}), el mismo sigue exactamente los mismos pasos con la diferencia que las variables utilizadas son seleccionadas en base a reglas definidas por CEIBAL en base a su trabajo diario y experiencia docente. Las dimensiones utilizadas son: cr.comentariosposteados, cr.totalvisitas, cr.totalcomentariosposteados, cr.totalsubmissions, cr.totaldiasingreso, cr.comentariosdiscusiones, cr.comentariosposteadosassignmen (ver definiciones en Tabla 3).

Finalmente se implementaron formas alternativas para construir un indicador de compromiso mediante componentes principales pero los resultados no permitieron continuar por ese camino ya que para acumular una variabilidad de al menos el 80 % era necesario trabajar con 7 componentes.

4.2. Resumen de resultados

En esta sección se presentan algunos resultados del análisis exploratorio del comportamiento de los alumnos en la plataforma CREA durante el año 2017. Inicialmente se examina la cantidad de centros de estudio, cantidad de alumnos y el IE por departamento. Es importante mencionar que el análisis realizado solo toma en consideración las escuelas del estrato urbano dejando por fuera cerca de 6000 alumnos del ámbito rural.

En la Figura 1 se presenta un gráfico de barras con la cantidad de centros por departamento donde la intensidad de color indica la cantidad de alumnos. Se observa que la mayor cantidad de centros se encuentran en Montevideo (303), seguido por Canelones (186) y Maldonado (58). La menor cantidad de escuelas está en el departamento de Flores (11), seguido por Río Negro (27), mientras la media de escuelas es 61 y la mediana es de 44.

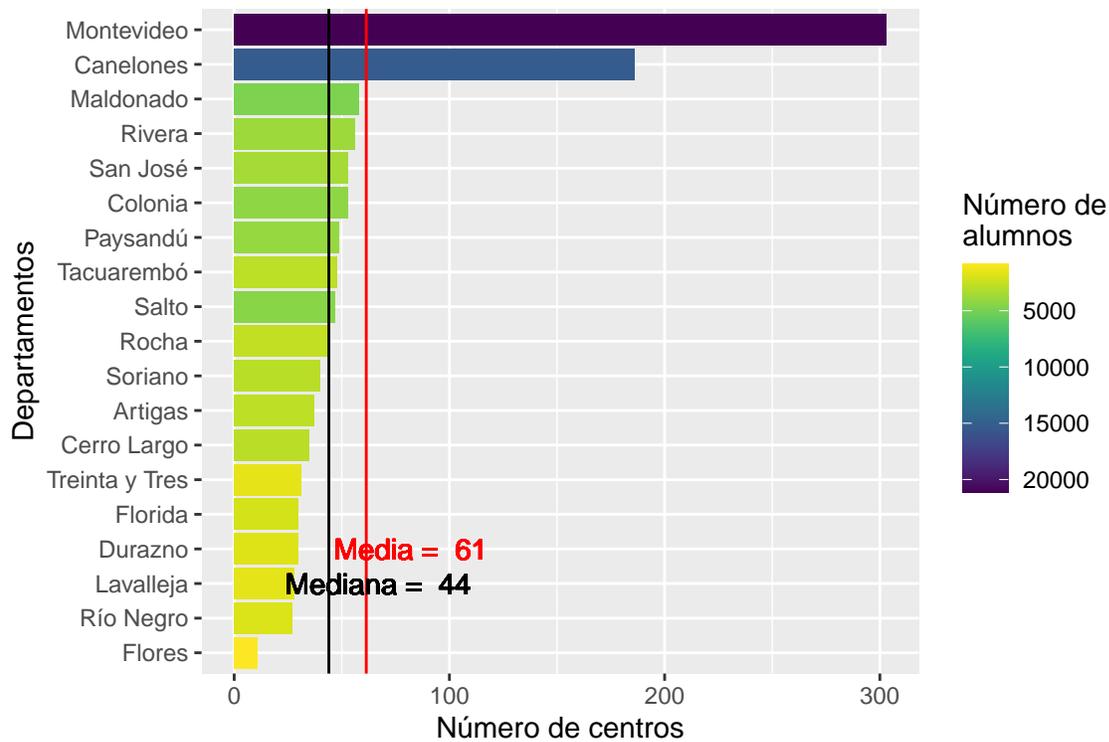


Figura 1: Gráfico de barras de la cantidad de centros por departamento en el año 2017. La intensidad de color indica el número de alumnos siendo Montevideo y Canelones los que tienen mayor cantidad de escuelas y estudiantes.

Similarmente, el departamento con mayor cantidad de alumnos es Montevideo, seguido de Canelones, Maldonado y Salto, mientras el departamento con menor cantidad de alumnos es Flores, seguido de Treinta y Tres, Lavalleja y Durazno.

En la Figura 2 se observa una gráfica de violín con el IE para cada departamento de Uruguay. La gráfica de violín se utiliza para mostrar datos numéricos, si bien es similar a un gráfico de cajas se diferencia en que presenta la densidad de probabilidad de los datos estimados y suavizados mediante un estimador de densidad de kernel (rotado y en ambos lados) (Hintze y Nelson, 1998). Permite observar la distribución completa de los datos y a diferencia del gráfico de cajas se podría ver por ejemplo si la distribución de los datos es unimodal o multimodal. En el caso del IE por departamento, la Figura 2 muestra que en la mayoría de los casos la distribución del IE en cada departamento es bimodal, con valores en torno al máximo (1) y el mínimo (0) del indicador, esto se cumple especialmente en los departamentos donde la mediana del IE es menor a la mediana del índice a nivel nacional (0.62), como Montevideo, Lavalleja, Maldonado, Cerro Largo, San José y Artigas. Distinto es el caso de Flores, departamento con mayor actividad en CREA, según el IE, el cual se

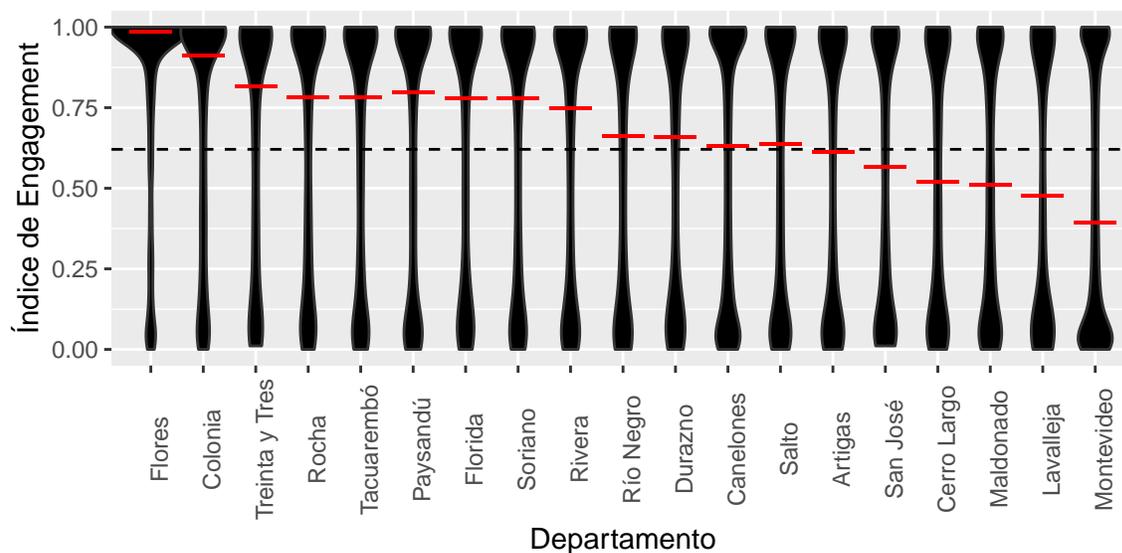


Figura 2: Gráfica de violín del Índice de Engagement (IE) por departamento en el año 2017. Con color rojo se hace referencia a la mediana del IE en cada departamento. La línea punteada de color negro denota la mediana a nivel nacional. Hay seis departamentos que tienen una mediana del IE menor a la mediana del IE nacional, entre los que se encuentran los de peor desempeño como Montevideo, Lavalleja y Maldonado.

concentra en torno al máximo (mediana de 0.99), con una menor amplitud que el resto de los departamentos y muestra una distribución unimodal. En contraposición, Montevideo evidencia el peor desempeño en términos de engagement, con una mediana de 0.39 y una distribución bimodal con mayor concentración en torno al mínimo. Vale destacar que la mediana del IE que se muestra en la Figura 2 es calculada a nivel de alumno (0.62), si fuese calculada como la mediana de las medianas de los departamentos su valor sería mayor (0.66), en ambos casos no son considerados los alumnos sin departamento asignado los cuales tienen un IE de cero, puesto que no tienen actividad en CREA.

Por otro lado, es de interés conocer el uso de la plataforma CREA según el estrato socioeconómico. Para esto se utiliza la variable *cont_urb2015* que divide a las escuelas según su estrato socioeconómico en cinco quintiles. Donde el primer quintil (1) es de contexto crítico y el quinto (5) es el de mejor contexto socioeconómico. En la Figura 3 se observa una gráfica de cajas del IE según quintiles de contexto socioeconómico y sexo. Estos gráficos sirven para analizar la variabilidad de los datos e identificar posibles valores atípicos. Para su construcción se ordenan los datos, donde la caja representa el 50% de la distribución conteniendo entre el 25% y 75% de las observaciones y, la línea al interior de la caja representa la mediana de la distribución de los datos. En casi todos los casos las distri-

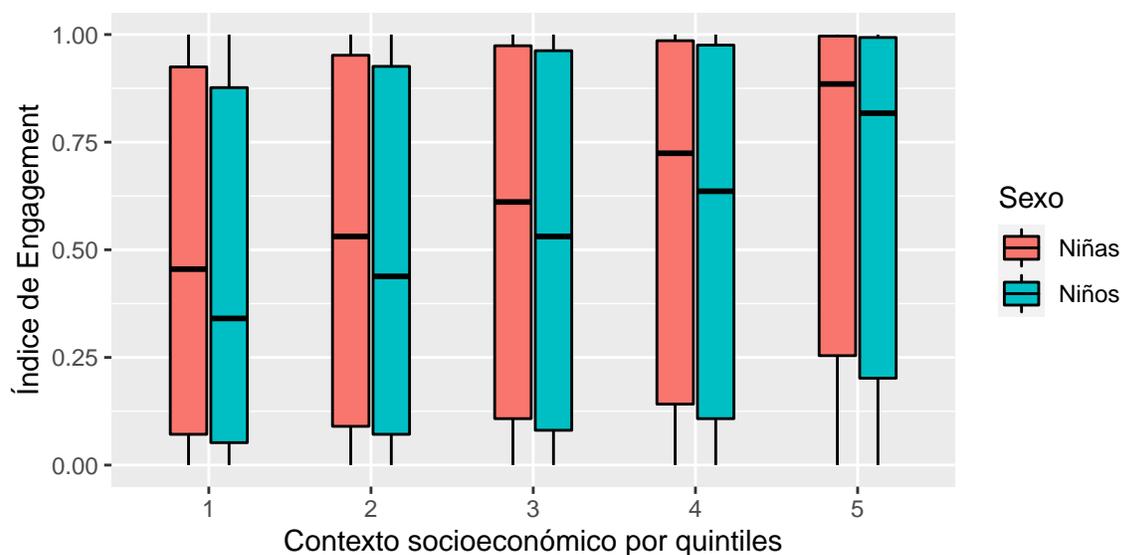


Figura 3: Gráfico de caja del IE según quintiles de contexto socioeconómicos y sexo e el año 2017. La mediana de engagement en todos los quintiles es siempre menor para niños que para niñas. La mediana del IE aumenta de forma sistemática desde el contexto crítico (quintil 1) hasta los quintiles más altos (quintil 5).

buciones son asimétricas, en especial en el quintil más alto. . Notar que la mediana del IE por estrato es sistemáticamente menor a medida que el contexto se vuelve más crítico, desagregando o no por sexo. Los valores de la mediana del IE desde el estrato más crítico hasta el de mejor contexto, sin desagregar por sexo, son respectivamente: 0.08, 0.17, 0.28, 0.39, 0.66. Adicionalmente, la Figura 3 muestra que las niñas tienen una mediana del IE permanentemente superior al IE de los niños y dicha superioridad se mantiene en todos los estratos socioeconómicos.

4.3. CEI

Se analizan las categorías de los puntajes de las pruebas de Ceibal en Inglés en los años 2016 y 2017 respecto a vocabulario, lectura y gramática (VRG) y comprensión auditiva (listening), aunque aquí se presentan solo los resultados del año 2017. En todos los casos se excluyen del análisis los alumnos con puntajes incorrectos y sin actividad. Para detectar los puntajes incorrectos se cuenta con una variable dicotómica generada por CEIBAL que especifica si dicho alumno tiene o no valores correctos, la cual esta basada en que el rango de valores posibles esta acotado a un intervalo de puntajes admisibles. Los puntajes según el Marco Común Europeo (MCE), por lo cual el menor puntaje (visto por categorías) es A0 y el mayor es B1 (la clasificación máxima es C2 pero en alumnos de primaria de

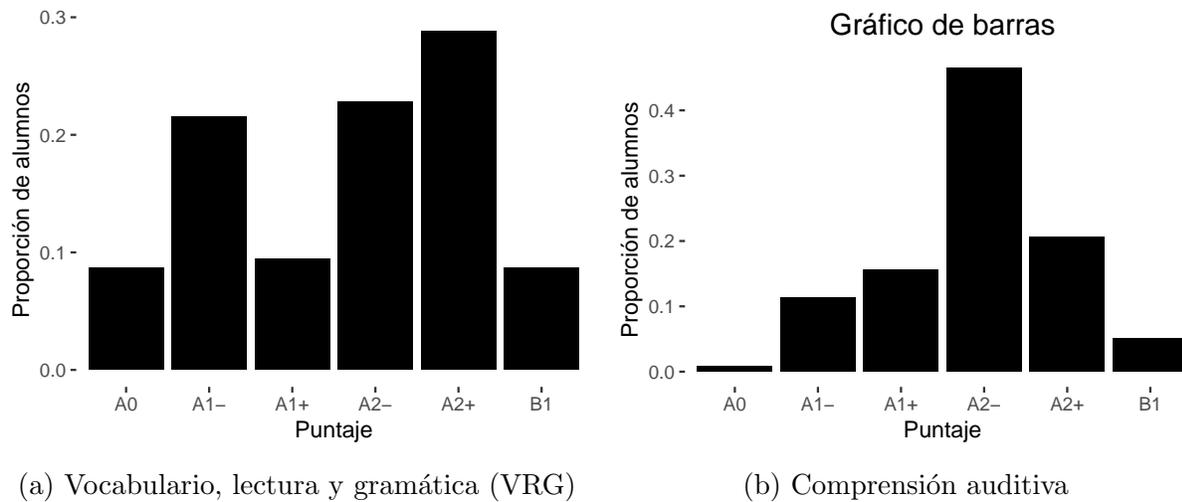


Figura 4: Gráfico de barras del resultado en las pruebas de vocabulario, lectura y gramática (VRG) y comprensión auditiva del año 2017 solo considerando los alumnos que tomaron las pruebas ambos años (2016 y 2017)

CEIBAL el máximo que se considera es B1).

En la Figura 4a se presenta un gráfico de barras con la proporción de alumnos en cada categoría de resultados posibles para las pruebas VRG (panel izquierdo) excluidos quienes no tienen puntajes. La mayor proporción de alumnos está en el nivel A2+, el cual debe considerarse alto al tomar en cuenta que el máximo es B1. El nivel A2+ es seguido por la categoría A2- (nivel medio alto) y luego A1-, clasificación baja. Las menores proporciones de alumnos se encuentran en las categorías de menor y mayor desempeño, A0 y B1.

En el caso de los puntajes en la prueba de comprensión auditiva (panel derecho), presentado en la Figura 4b se observa que la mayor proporción de alumnos se encuentran en la categoría A2-, el cual es un nivel medio, seguido de lejos por el nivel A2+ y A1+ niveles medio alto y medio bajo respectivamente. Al igual que en la prueba de VRG la menor proporción de estudiantes están en las categorías de las puntas, A0 y B1.

4.4. Monitor de plataformas educativas

El monitor de plataformas educativas es una herramienta cuyo objetivo principal es observar y analizar el uso de plataformas educativas de forma sistemática, estandarizada y sencilla, ayudando a distintos actores del sistema educativo a responder preguntas que requieren distintos niveles de análisis y redundan en un fortalecimiento de la educación tanto a distancia como presencial. El público objetivo es variado pero hay dos usuarios principales, analistas de CEIBAL y mentores. Los mentores, cuya labor consiste en asistir

a los centros educativos a corroborar el uso de la plataforma CREA y obtener información específica del centro de estudio a través de reuniones con directores y maestros del establecimiento. Parte de su rol de control queda sujeto a la información que tienen disponible respecto al desempeño de la escuela de interés. En el caso de los analistas de CEIBAL su labor consiste en analizar y sintetizar información para generar reportes que sean insumo para controlar el desempeño de los centros de estudio e insumo en una posterior toma de decisiones respecto a la ejecución de políticas.

El monitor sirve a los mentores como una fuente de información consolidada y procesada con datos actualizados que permite saber con detalle la evolución en el uso de la plataforma de la escuela, grado, clase o alumno de interés y la información agregada en un momento del tiempo.

En el caso de los analistas, el monitor permite consultar información consolidada para cada alumno en un momento del tiempo, la misma esta sistematizada y procesada de forma que es automático conocer el desempeño de los distintos departamentos, centros de estudio y clases de una escuela. Se brindan información en distintos niveles de análisis espacio-temporal permitiendo flexibilidad a la hora de identificar el comportamiento a nivel agregado de regiones o a nivel micro de centros de estudio, sus clases y específicamente los alumnos. Adicionalmente, la información que se va consultando en distintos niveles se guarda y permite generar reportes dinámicos automatizados lo cual redundo en una disminución del tiempo de análisis requerido.

A continuación se describen las principales características del monitor para plataformas educativas desarrollado en el proyecto. El monitor es desarrollado como una aplicación web usando la librería *shiny* en R. Si bien esta diseñada para crear aplicaciones complejas sin ser necesario conocimiento previo de programación en html, los componentes referidos a la interfaz del usuario se pueden personalizar ampliamente. Además, su servidor utiliza programación reactiva que permite crear cualquier lógica de backend que se necesite (Chang *et al.*, 2020).

La aplicación esta compuesta de cuatro pestañas principales con información correspondiente a distintos niveles de análisis. Se suma una pestaña inicial donde se explica cada uno de los niveles de análisis y sus respectivas visualizaciones, y una pestaña final que permite documentar la exploración mediante la generación de un reporte dinámico, parametrizado y automático. El monitor está disponible en el servidor del IESTA en **IESTA-CEIBAL-APP**.

En la pestaña **Inicio** se describe el contenido del monitor en cada nivel de análisis. Luego hay cuatro niveles de análisis: **Nacional**, **Escuela**, **Grado** y **Clase**. El monitor está diseñado para que los distintos usuarios puedan realizar una exploración de los datos de forma guiada seleccionando distintos parámetros de interés. Los distintos niveles que componen la aplicación están encadenados, es decir, la selección de parámetros (año, indicador, escuela, grado, clase) se mantiene de un nivel a otro (aunque es posible modificarlos) de forma tal que se va analizando desde una perspectiva macro (departamental)

hasta una visión micro (la clase), en donde se observa el desempeño de los alumnos a nivel particular.

A modo de ejemplo, el análisis comienza a nivel nacional con la elección de un indicador de desempeño en la plataforma y un año de interés, posteriormente se elige una escuela condicional a la selección previa de año e indicador, luego se pasa al nivel de grado donde se comparan los 4to, 5to y 6to y finalmente se elige la clase de un 4to, 5to o 6to donde se observa el desempeño de los alumnos, pudiendo identificar quienes tienen una participación menor al promedio de la clase. A continuación se describen con detalle los distintos niveles de análisis y sus respectivas visualizaciones estadísticas.

4.4.1. Nivel Nacional

El primer nivel de la aplicación permite un análisis **Nacional** del uso de la plataforma. En la Figura 5 se muestran las dos visualizaciones que se presentan en este nivel. La aplicación permite la interacción con el usuario que puede seleccionar que año desea visualizar, en conjunto con el tipo de indicador y la medida de resumen. Hay dos visualizaciones, la primera, Figura 5a es un mapa del territorio uruguayo dividido por departamentos y coloreado por el indicador seleccionado (las dos variaciones del IE). Puntajes con tonalidad amarilla indican un mejor desempeño, mientras colores oscuros denotan peor performance. Montevideo es el departamento con peor puntaje independiente del indicador seleccionado, mientras Flores se ubica siempre en primer lugar.

La Figura 5b es una densidad de kernel (se utiliza el kernel de Epanechnikov) que muestra la distribución de los puntajes del indicador. Permite identificar si existen concentraciones de puntajes en algún intervalo del rango de valores posibles⁵. Por ejemplo, en el caso del IE se observa que la distribución es bimodal, es decir, los alumnos se concentran en dos intervalos del rango posible de puntajes. En torno al cero y en torno al 1, mínimo y máximo respectivamente. Por lo tanto, hay alumnos con una alta y baja participación en la plataforma según el IE.

4.4.2. Escuela

El segundo nivel de análisis se muestra en la Figura 6, se denota **Escuela** y permite elegir los siguientes parámetros: año, indicador, departamento y escuela. Existen cuatro tipos de visualizaciones que comparan la escuela seleccionada con el resto de centros de su misma región. Definimos región como dos subconjuntos que engloban al total de

⁵Por construcción el rango de valores es mayor al rango de la variable de interés, esto puede ser solucionado restringiendo los valores máximos y mínimos. Sin embargo, el objetivo en nuestro caso es observar si la distribución es unimodal, bimodal o multimodal. Es evidente que los puntajes siguen una distribución bimodal, siendo innecesario para el objetivo del trabajo realizar una corrección.

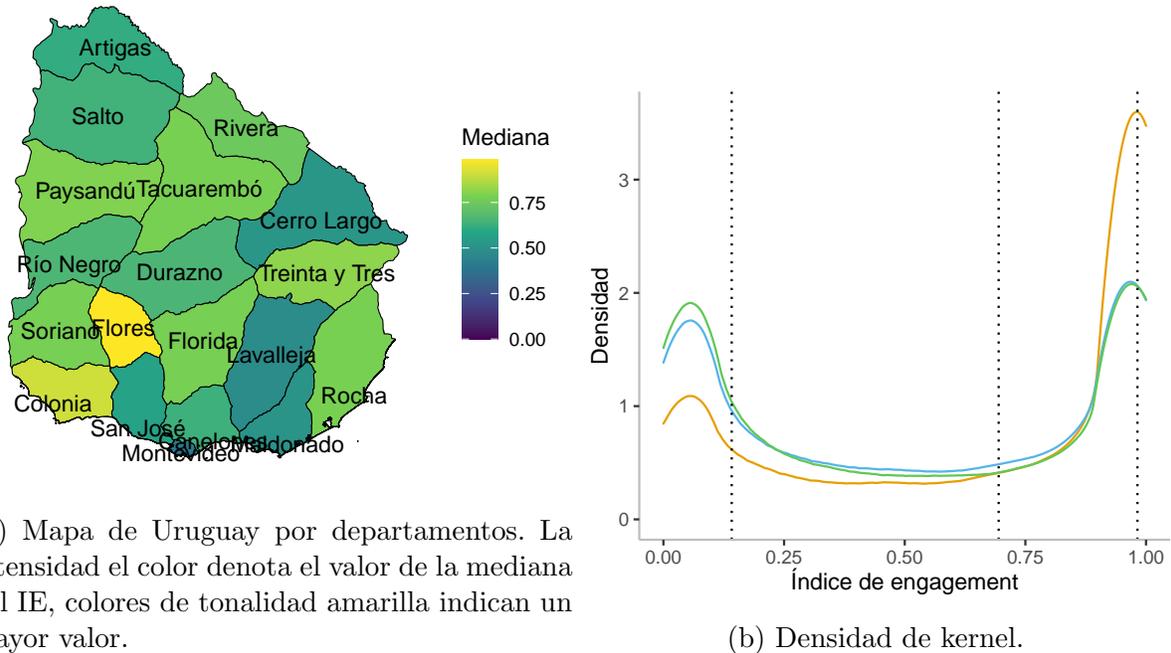


Figura 5: Panel izquierdo presenta el mapa de Uruguay coloreado con la mediana del IE y el panel derecho presenta la estimación de densidad por kernel (Epanechnikov) del IE donde las líneas punteadas denotan los cuantiles 25, 50 y 75. El color referencia a los grados.

departamentos. Por un lado, el subconjunto interior contiene todos los departamentos distintos de Montevideo, siendo Montevideo su propio subconjunto.⁶

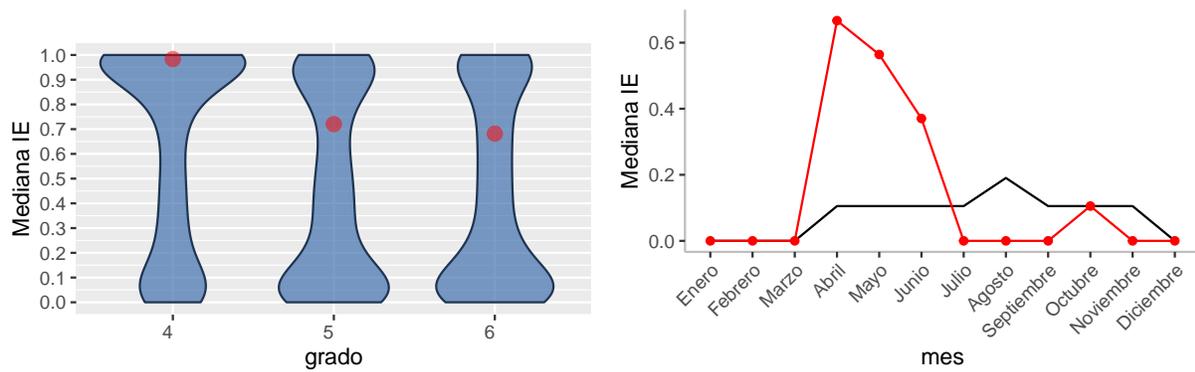
A modo de ejemplo, en la Figura 6 se elige el año 2017, departamento de Salto, IE y escuela número 1119. Estas selecciones se mantienen y profundizan en los siguientes niveles de análisis del monitor. En la Figura 6a se presenta una gráfica de violín, la misma tiene en el eje vertical las medianas de cada escuela, mientras el eje horizontal muestra una apertura por grado (4to, 5to y 6to), con lo cual se obtiene la forma de la distribución de las medianas para cada grado del interior. Complementariamente, se agrega un punto de color rojo especificando la mediana de la escuela seleccionada. De esta forma, se interpreta que la mediana de la escuela 1119 en los distintos grados está en la cola superior de la distribución del IE comparado a los mismos grados de todo el interior del país. Adicionalmente las distribuciones de probabilidad muestran ser bimodales, en el caso de cuarto grado con mayor concentración en el límite superior, mientras quinto y sexto grado en el límite inferior.

La Figura 6b muestra una serie de tiempo donde cada punto refiere a la mediana del indicador seleccionado (IE) en cada mes del año para una escuela particular (en color rojo) y para las escuelas de Montevideo o Interior dependiendo el caso a comparar (color negro). Se observa el comportamiento de la escuela 1119 a lo largo del año, identificando una participación nula en los meses correspondientes a vacaciones de verano (diciembre, enero y febrero), luego un aumento sostenido hasta junio, que se mantiene aunque con caída en los meses de julio (vacaciones de invierno) y septiembre (vacaciones de primavera). La mediana del IE para las escuelas del interior (color negro) cae en los meses de vacaciones de verano y luego se mantiene estable, salvo un aumento en agosto. De esta gráfica resulta evidente que la escuela 1119 tiene un IE en el año 2017 sistemáticamente superior al IE mediano de las escuelas del interior.

Esto se puede ver en la Figura 6c, tabla interactiva de resumen por grado. La información contenida refiere al grado (4to, 5to y 6to), el número de la escuela, el departamento en el cual se encuentra, la mediana de cada grado y el número de alumnos totales, lo cual engloba a todos los cuartos, quintos y sextos que existan en el centro de estudio. Si se elige Montevideo hay dos columnas que contienen el ranking del grado a nivel del departamento y la cantidad de escuelas. Si se selecciona un departamento del interior se generan cuatro columnas, el total de escuelas en el departamento y en el interior y, la cantidad de alumnos en el departamento e interior. De esta forma se puede comparar a la escuela-grado con el resto de centros de su departamento y también, con todo el interior.

Finalmente, en la Figura 6d se presenta una serie temporal para cada dimensión que com-

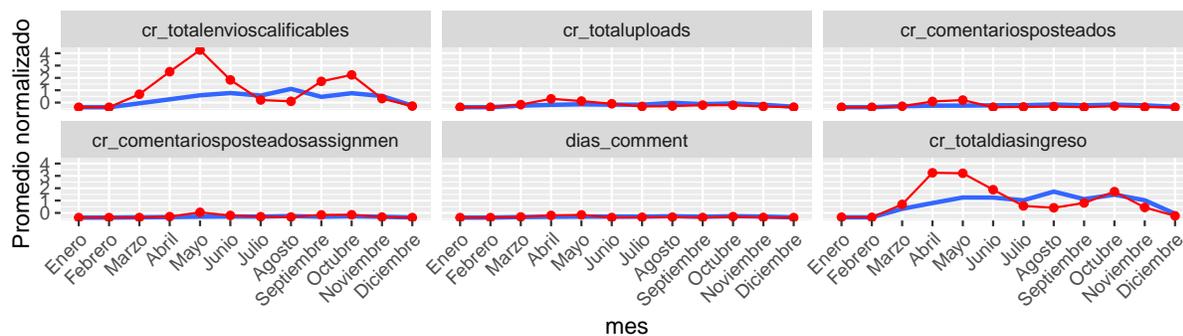
⁶Esto se realiza debido a una diferencia notoria en el comportamiento de los alumnos montevideanos en la plataforma CREA respecto al resto del país. Por lo tanto, elegir el departamento de Montevideo implica que el conjunto de visualizaciones solo contiene datos de Montevideo, mientras al seleccionar un departamento diferente los datos contienen únicamente observaciones del interior, por lo cual, el promedio o mediana de comparación sera referido al interior.



(a) Gráfico de violín de la mediana del IE para (b) Evolución mensual de la mediana del IE para cada grado del Interior. El punto rojo refiere a la escuela seleccionada (1119 del interior en rojo) y escuelas del interior (en negro).

Grado	Escuela	Departamento	Mediana	Alumnos	Ranking Salto	Total Escuelas Salto	Ranking Interior	Total Escuelas Interior
4	1119	Salto	0.98	88	10	38	208	679
5	1119	Salto	0.72	106	15	45	242	679
6	1119	Salto	0.68	75	10	39	231	665

(c) Ranking de escuelas según el IE en cada grado condicional a escuelas del interior



(d) Evolución temporal de las dimensiones del IE. Con color rojo el promedio normalizado (PN) en cada dimensión de la escuela 1119 y con azul el PN de las escuelas del interior.

Figura 6: Análisis a nivel de escuela del Índice de Engagement (IE), año 2017. Se muestra la escuela 1119.

pone el IE. Cada rectángulo es una dimensión (variable) donde el eje vertical representa el promedio normalizado (restando la media y dividiendo por el desvío estándar) el cual es compartido de forma que las escalas sean comparables entre dimensiones y las abscisas representan los meses del año. La línea roja es la escuela seleccionada mientras la línea azul es el promedio de Montevideo o interior según corresponda. El área gris corresponde al promedio más y menos dos desvíos estándar $\bar{x} \pm 2\sqrt{\text{var}(x)}$. El gráfico permite identificar que dimensiones son las que pesan más en el indicador seleccionado en cada mes del año. En el ejemplo, la escuela 1119 muestra que las dimensiones con mayor actividad son el total de envíos calificables (`cr_totalenvioscalificables`) y total de días de ingreso (`cr_totaldiasingreso`).

En resumen, en el departamento de Salto, la distribución del IE en cuarto año difiere de quinto y sexto. Se observa una mayor concentración de escuelas en torno al mínimo del puntaje del IE para quinto y sexto, mientras en cuarto el compromiso es mayor ya que la distribución se concentra en el extremo superior. Marcando que los cuartos presentan una mediana del IE mayores que quintos y sextos. La escuela 1119 obtiene en todos los grados una mediana del IE mayor que la del interior, de hecho, su lo largo del año progresa constantemente y a partir de marzo se sitúa por encima de la mediana del IE de las escuelas del interior. Las dimensiones con mejor desempeño son total de envíos calificables y total días ingreso, las cuales están por encima del promedio de IE de escuelas del interior. El resto de dimensiones son iguales o levemente mejores. Se concluye, de este nivel de análisis que, en mediana, la escuela 1119, en promedio envía mayor cantidad de envíos calificables que las escuelas del interior e ingresa mayor cantidad de días.

4.4.3. Grado

El tercer nivel de análisis es el **Grado** (4to, 5to y 6to), presentado en la Figura 7. Se pueden elegir los mismos parámetros que en el caso anterior: año, indicador, departamento y escuela. A su vez se aplican las mismas consideraciones, si se selecciona Montevideo el análisis es condicional a Montevideo, si se selecciona cualquier otro departamento la comparación es condicional al interior. Pero con una diferencia, la agrupación para examinar es por grado de escuela (4to, 5to y 6to).

La Figura 7a, es una serie temporal con la mediana del IE en las ordenadas y los meses del año en las abscisas que analiza la mediana del indicador en cada mes del año y lo compara con el resto del interior. Con color mostaza, celeste y verde se representa a cuarto, quinto y sexto grado respectivamente. En el caso de la escuela 1119, cuarto grado tiene una mediana del IE superior al de quinto y sexto, especialmente en el segundo semestre.

La Figura 7b es un diagrama de cajas con el IE en el eje vertical y los grados de la escuela en el eje de las abscisas. La línea al interior de la caja es la mediana del IE en el grado y cada punto es el valor del IE de cada alumno en cada grado de la escuela seleccionada. Es decir, los puntos observados son el IE de cada alumno en cada grado (sin discriminar

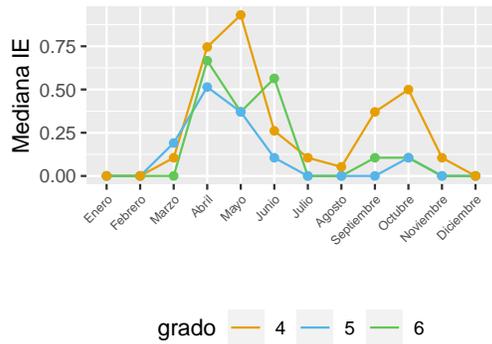
por clase). Los estudiantes de la escuela 1119 en 4to grado se concentran en un IE entre 0.9 y 1, mientras 5to y 6to grado muestran una variabilidad mayor.

La tercera representación, Figura 7c, es una serie de tiempo para cada dimensión del indicador con el eje vertical representando el promedio normalizado (bajo una escala común). El color mostaza representa a cuarto grado, el celeste a quinto y el verde a sexto. Es posible comparar el desempeño por grado para cada mes del año y en cada dimensión del indicador, de esta forma se identifica si algún grado tiene en promedio un desempeño mejor o peor que el resto, a la vez que se observa si ese comportamiento se da permanente o aisladamente en el tiempo. La escuela 1119 repite el mismo comportamiento que en el nivel previo, con una mayor actividad en total de envíos calificables y total de días de ingreso.

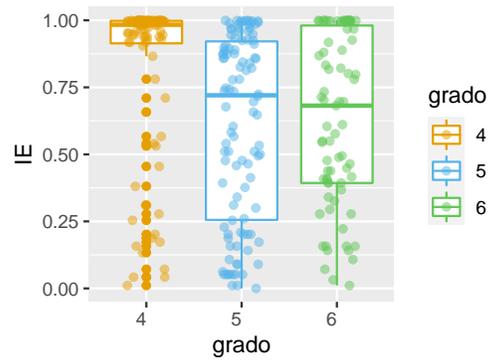
La Figura 7d, tiene en el eje vertical el promedio normalizado de todas las dimensiones del indicador, a nivel horizontal están las variables que componen el índice y el color que diferencia los valores por grado. El gráfico compara que grados se destacan en cada dimensión a nivel anual sin tomar en cuenta el factor temporal. En la escuela 1119, el comportamiento de los grados sigue un patrón similar aunque cuarto tiene mayor amplitud en sus puntajes.

Por último, la Figura 7e exhibe la mediana del IE en cada grado y grupo de la escuela seleccionada. La apertura por clase, permite hacer una comparación intra grado (grupo A y grupo B de cuarto grado, por ejemplo) y extra grado (cuarto A versus quinto B). A nivel intra grado es posible diferenciar comportamientos diferentes en el mismo nivel y focalizar el análisis en los cursos más rezagados. Mientras que la visión extra grado permite notar posibles heterogeneidades entre niveles que puedan tener una causa común.

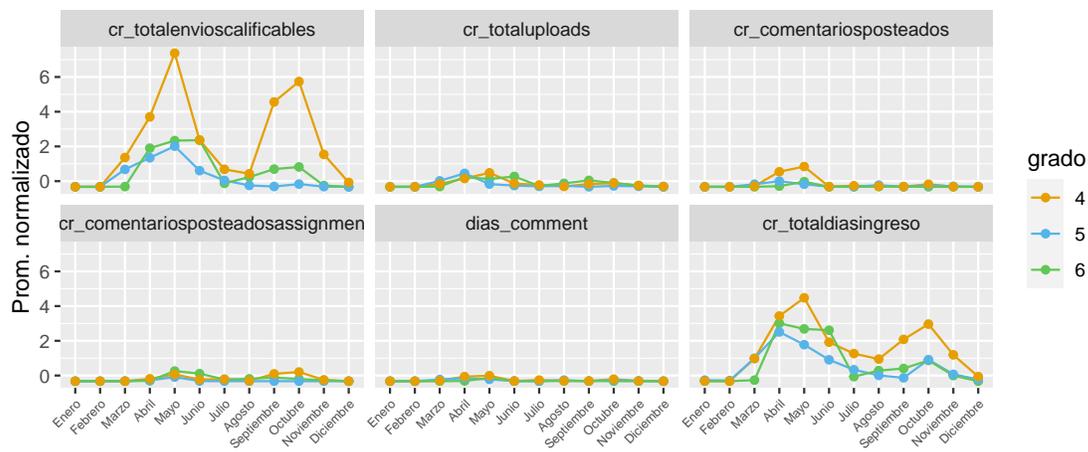
En síntesis, en el año 2017 en el departamento de Salto, seleccionando la escuela número 1119 y usando el IE se observa según la Figura 7a que cuarto grado es el de mejor desempeño. En Figura 7b los datos de cuarto grado están concentrados en torno al máximo valor del IE mientras quinto y sexto evidencian mayor variabilidad. La comparación de las dimensiones en la Figura 7c a través del año muestra que el total de envíos calificables y total días de ingreso son las variables de mayor actividad y en ambos casos cuarto grado obtiene los mejores puntajes. En el resto de las dimensiones no se observan grandes diferencias. En la Figura 7d se observa que cuarto grado tiene un valor más elevado tanto en total envíos calificables como en días de ingreso, a la vez que estas dos dimensiones son las de mayor actividad. Finalmente, desagregando los grados por grupo, se observa que la mediana del IE en todos los cuarto grado se ubica en torno al máximo valor del IE mientras quinto y sexto grado tienen una variabilidad mayor con medianas desde 0.4 hasta 1.0.



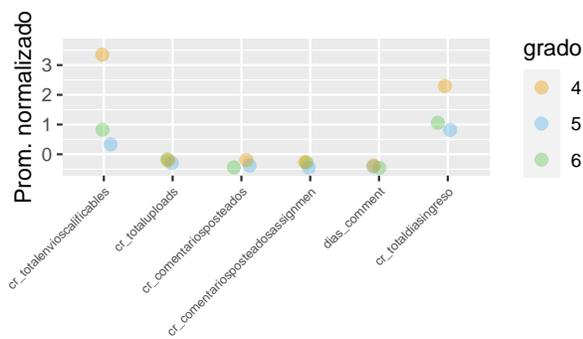
(a) Mediana del IE por grado.



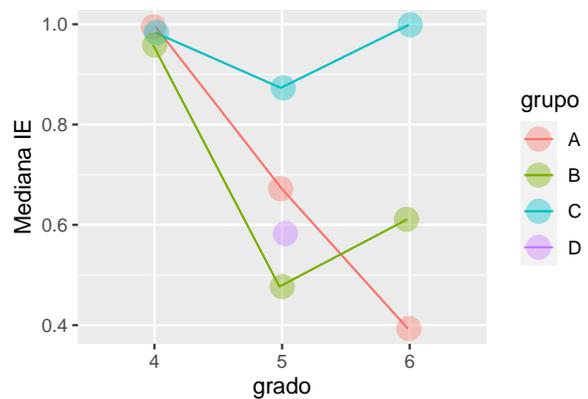
(b) Gráfico de cajas del IE por grado



(c) Serie temporal categorías del IE (normalizadas)



(d) Categorías del IE



(e) Mediana del IE por grado y grupo

Figura 7: Análisis del Índice de Engagement (IE) a nivel de grado en escuelas del interior DT (02/21)-Instituto de Estadística Molina, Federico *et al*

4.4.4. Clase

El cuarto nivel de análisis se denomina **Clase**, mantiene los parámetros seleccionados previamente (año, indicador, departamento y escuelas) y se agregan dos nuevas elecciones, el grado y la clase. Aplican las mismas consideraciones que en la pestaña anterior, si se selecciona Montevideo el análisis es condicional a Montevideo, si se selecciona cualquier otro departamento la comparación es condicional al interior. La diferencia es que se analizan los datos sin agregar, es decir, se analiza el IE de cada alumno.

La Figura 8a esta compuesta de dos gráficos superpuestos, un diagrama de cajas y un gráfico de violín, con los cuales podemos examinar la variabilidad de los datos y la forma de la distribución. Adicionalmente, se agregan puntos representando los valores del IE de cada alumno. En el ejemplo de la escuela 1119, cuarto grado, clase A, si bien la mediana del IE es cercana a uno para un grupo de alumnos con valores en torno al máximo, en el diagrama de cajas se observa una alta variabilidad y, que existe un grupo de alumnos rezagados con puntajes desde 0.50 hasta 0.

La Figura 8b, es una serie temporal donde cada línea punteada corresponde al IE mensual de un alumno particular. Mientras la línea azul es el promedio de la clase y el área gris denota el intervalo de confianza al 95 %. Es posible identificar el desempeño de los alumnos a lo largo del año y, comparar si estuvieron por encima o debajo del promedio de su clase. Finalmente, en la Figura 8c se presenta un diagrama de cajas para cada dimensión del indicador seleccionado, donde el eje vertical representa el promedio normalizado y los puntos denotan el valor de cada alumno para esa variable. De esta forma se observa la variabilidad de los datos en cada dimensión del indicador y a cada alumno en particular lo cual permite identificar, por ejemplo, valores atípicos que son alumnos o muy destacados o de bajo desempeño en comparación a su clase (por dimensión).

En el caso de la escuela 1119, cuarto grado C, se observa que la distribución de los datos es asimétrica con alta variabilidad y dos grupos de alumnos, la mayoría con puntajes en torno al máximo y una minoría rezagada con valores menores a 0.5.

La evolución de los alumnos a lo largo del año muestra gran variabilidad tanto por alumno como entre alumnos. En el diagrama de cajas es evidente que las variables que se encuentran por sobre el promedio son total envíos calificables y total días ingreso, en especial hay dos alumnos en la primer dimensión con un rendimiento muy superior a sus compañeros. Las restantes variables están en torno al valor promedio de las dimensiones, con poca variabilidad porque los gráficos de cajas están comprimidos en torno al cero.

4.4.5. Reporte dinámico

El último nivel sintetiza toda la información de los niveles anteriores y genera un reporte dinámico. En la medida que se fueron seleccionando los parámetros año, indicador, medida de resumen, departamento, escuela, grado y clase, estos permanecen con los mismos

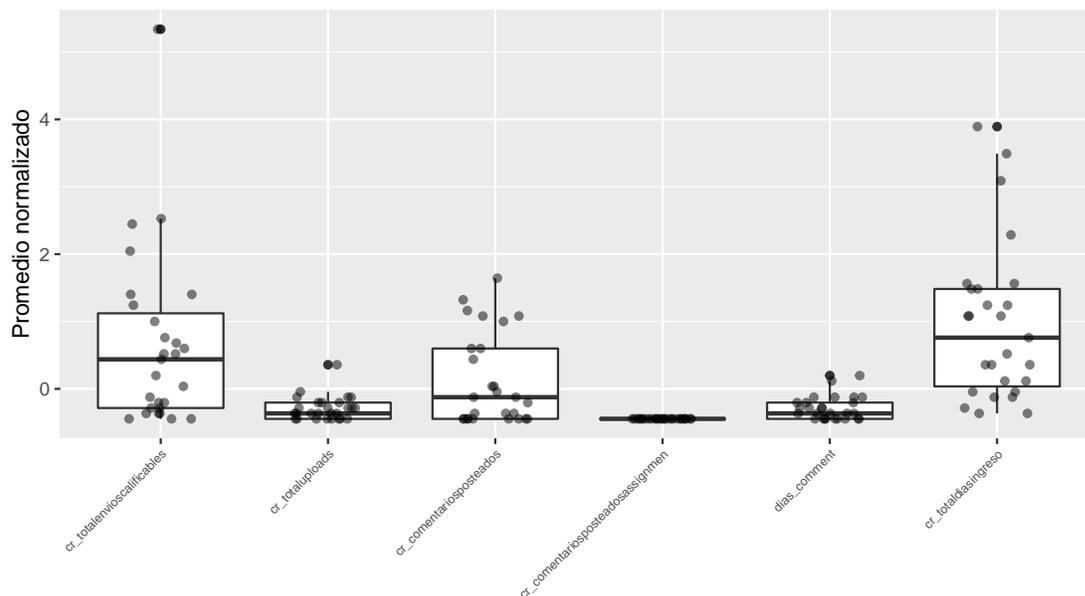
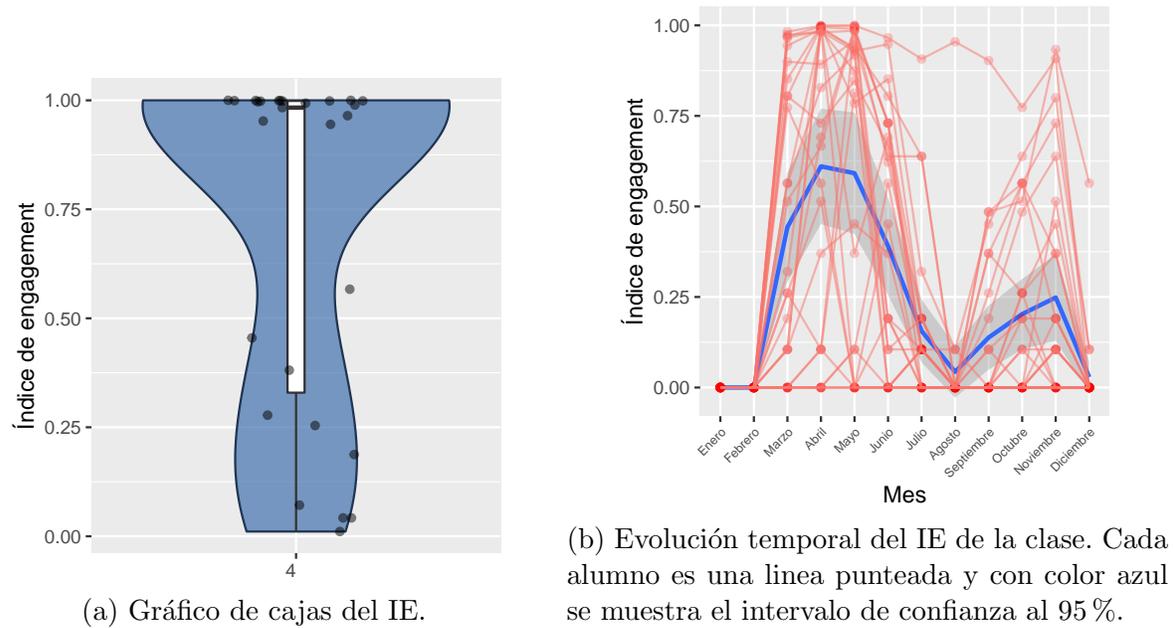


Figura 8: Análisis a nivel de clase del Índice de Engagement (IE) donde cada punto representa el valor del IE de cada alumno. Se observa que envíos calificables y total de ingresos son las dimensiones con mayor variabilidad y actividad.

valores. Si se desea modificar algún parámetro se puede realizar en este mismo nivel. Una vez definida la selección deseada, se presiona el botón *Mostrar reporte* el cual genera un reporte dinámico que puede ser visto en la aplicación.

Se puede elegir el formato: PDF, HTML y Word. Al seleccionar la opción descargar se genera el mismo reporte automático y se descarga en el formato deseado. El informe además de incluir las gráficas previamente mostradas, incluye una explicación de cada una y un breve análisis automatizado. En caso de querer utilizarlo como una prototipo a extender por el analista el documento puede ser descargado en formato Word y ser modificado.

La flexibilidad de los reportes dinámicos es amplia, en este caso se genera un reporte genérico, sin embargo, los mismos pueden personalizarse según el usuario de interés. Es decir, se pueden generar análisis automatizados si el rol es de analista o de mentor, con lo cual se modifican los cálculos, gráficas y el análisis, el cual no tiene porque restringirse a lo visto en la aplicación. De hecho, es posible agregar fuentes de datos adicionales y parametrizar el reporte. Parametrizar quiere decir que el reporte cambie en función de los parámetros dados. A modo de ejemplo, en la aplicación el reporte genérico esta parametrizado según el índice escogido, el año, el departamento, la escuela, grado y la clase. Entonces los cálculos y visualizaciones cambian de forma dinámica según dicha selección y la descripción de dichos análisis (el texto) también. La cantidad de parámetros no esta acotada y se pueden definir según las necesidades deseadas. Si se desea un reporte detallado es posible fijar la opción detallado y generar automáticamente un análisis más extenso sobre el mismo conjunto de datos.

También es posible que el reporte funcione de forma desagregada del monitor de plataformas educativas manteniendo (y ampliando) la selección de parámetros. De hecho, una forma de uso puede ser su ejecución automática y programada en ciertos momentos del tiempo mediante linea de comando.

4.5. Predicción

En esta sección se presentan los modelos de aprendizaje estadístico y las estrategias utilizadas con el objetivo de predecir las categorías correspondientes a los puntajes de los alumnos en las pruebas adaptativas de inglés de vocabulario, lectura y gramática (VRG) y comprensión auditiva (Listening). Los puntajes obtenidos por los alumnos corresponden a una variable cuantitativa numérica con un rango de valores no negativos, los cuales son categorizados de acuerdo al marco común europeo. Los niveles básicos son A, seguido de B y finalmente C corresponde a individuos con un nivel fluido, específicamente el nivel C2 es de una persona nativa. En este trabajo se cuenta con seis niveles (A0, A1-, A1+, A2-, A2+ y B1) que se buscan predecir.

Inicialmente se generan 22 covariables incluyendo los IE y, se restringe el conjunto de datos para mantener solo a los alumnos que realizaron la prueba en ambos año y el formato de

datos a utilizar es de una observación por cada alumno.

El primer paso es analizar la calidad de la información provista por CEIBAL (y las variables generadas a partir de dichos datos) con el fin de predecir las categorías de los puntajes de inglés. Para lograr dicho objetivo se sigue la metodología planteada por [Genuer *et al.* \(2015\)](#) y posteriormente se estiman los modelos utilizando la librería `h2o`. Se muestran los resultados predichos de las categorías de la prueba de comprensión auditiva en el año 2017 (variable de respuesta).

4.5.1. VI

A modo de tener una evaluación sobre cuan importantes son los datos facilitados dado el objetivo de predicción de las categorías en VRG y comprensión auditiva, se utiliza un procedimiento de detección de variables importantes siguiendo a [Genuer *et al.* \(2015\)](#). La propuesta se realiza en tres etapas, en la primera se identifican variables redundantes y se fija un umbral a partir del cual las variables son descartadas utilizando como métrica una medida de importancia de variables (promedio de importancia de variable). A continuación, en la etapa dos, se mantienen las covariables consideradas relevantes para explicar la variable de interés aunque permitiendo cierta redundancia. Se construyen bosques anidados (desde un modelo solamente con la variable más importante, hasta un modelo con todas las variables de la etapa uno) y se eliminan las variables que no superen el mínimo del promedio del error out-of-bar (OOB) adicionado su desvío. En la última etapa se restringe aún más la cantidad de variables, las mismas solo se añaden (de forma stepwise) al modelo si logran disminuir el error OOB por sobre un umbral (mayor al promedio de la diferencia de los errores OOB de un modelo con las variables de la etapa dos y otro sin ellas).

Los resultados obtenidos muestran que, bajo esta metodología, hay dos variables relevantes. Las mismas son: “aprende16” y “nivel_vrg16”. Las cuales hacen referencia a si el alumno contesta que en el año previo (2016) curso inglés por fuera del ámbito escolar, y a su nivel en la prueba de VRG en el año previo. Es decir, que los puntajes previos son un buen predictor así como la actividad extra fuera del aula de clase relacionada a la materia de interés. En la [Figura 9](#) se observan las tres etapas consideradas para identificar los posibles predictores más importantes. En la [Figuras 9a](#) y [9b](#) se observa la primer etapa del procedimiento donde se busca eliminar variables poco importantes que se pueden asociar a elementos puramente aleatorios, se muestra la importancia de variable promedio de cada predictor y su desvío estándar. La etapa dos se presenta en la [Figura 9c](#), donde se observa claramente que no hay mejora en el error OOB en los modelos anidados si se agregan más predictores. Por último, la [Figura 9d](#) denota la etapa tres, en la cual se eligen las variables más importantes con fines predictivos, en este caso son las mismas variables de la etapa dos.

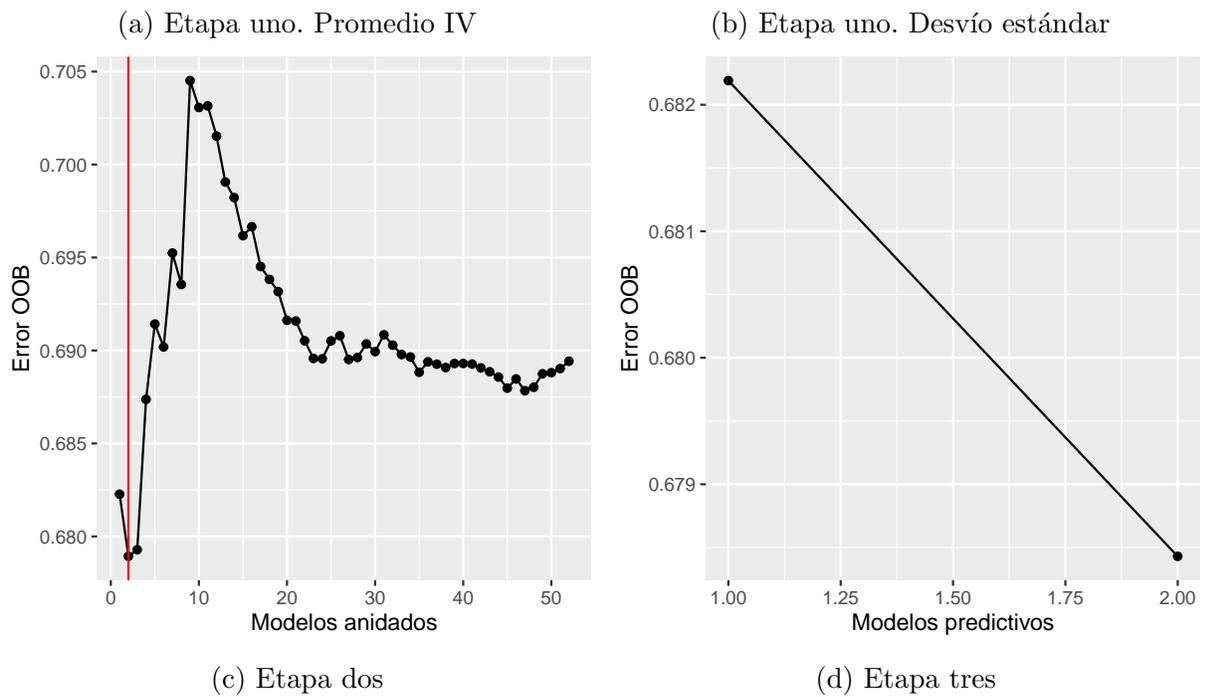


Figura 9: Análisis de importancia de variables (IV).

4.5.2. Modelos

La predicción se realiza utilizando el paquete `h2o` y `ranger`. La librería `ranger` (Wright y Ziegler, 2017) estima modelos de bosques aleatorios de forma eficiente (C++) en contraposición a librerías clásicas como `randomForest`. Por su parte, `h2o` se destaca tanto por su eficiencia al trabajar con grandes volúmenes de datos como por su facilidad para generar resultados comparables entre modelos, es paralelizable y utiliza de backend la librería `data.table`. Además, dispone de la modalidad `h2o.automl`, función que automatiza el proceso de entrenamiento de los modelos de aprendizaje estadístico supervisado (o no supervisado). En específico, entrena los modelos, utilizando validación cruzada y una grilla aleatoria con los siguientes algoritmos: bosques aleatorios, bosque extremadamente aleatorizado (Extremely-Randomized Forest), GBM y redes neuronales. Finalmente, entrena un modelo apilado (Stack Ensemble) usando todos los modelos utilizados previamente. Posterior a la etapa de estimación, uno de los productos de `automl` es un objeto de clase `H2OAutoML` el cual contiene una tabla de clasificación con el ranking de los modelos según la métrica escogida.

Inicialmente utilizamos los bosques aleatorios sobre un conjunto de combinaciones de datos, variables y agrupaciones de clases. Dado que existen seis niveles a clasificar, se corrieron modelos con seis clases, cuatro clases agrupando las categorías de las puntas y tres clases agrupando en nivel bajo, medio y alto. Se combinaron dichas clases utilizando todas las variables (originales y creadas) y sin incluir las variables creadas. Adicionalmente, en cada corrida se genera un modelo donde solo se utilizan las variables que tengan al menos un 50 % de la importancia de la variable más importante. Por último, se agrega a las combinaciones el filtro de alumnos con tipo de dispositivo sin dato. En total se utilizan 24 variaciones. En ningún caso se obtienen resultados satisfactorios. En la Tabla 4 puede observarse que la precisión tuvo un máximo de 43 % mediante una agrupación en tres clases (nivel bajo, medio y alto).

Posteriormente se trabaja con el paquete `h2o` utilizando la función `automl` para ver que modelos ajustan mejor utilizando validación cruzada y el error promedio entre clases (`mean_per_class_error`) como métrica de elección. Se sigue la metodología utilizada con los bosques aleatorios y los resultados, al igual que en el caso previo, no son satisfactorios. De allí surge que el mejor modelo se genera al utilizar el algoritmo de XGBoost, el cual se ejecuta a través de una grilla para elegir los hiperparámetros que mejor ajuste generen medido mediante validación cruzada. Se obtiene un error promedio entre clases en torno al 60 % cuando las clases se reducen a tres. Los resultados se pueden ver en la Tabla 2, donde se muestran los mejores cinco modelos cuando la variable de respuesta tiene seis, cuatro o tres categorías a predecir. En todos los casos se utilizan los mismos hiperparámetros, por ejemplo, `eta`, `colsample_bylevel`, `lear_rate`, `sample_rate`, `col_sample_rate`, `max_depth`, `gamma`, `ntrees`, `subsample`.

La performance predictiva de los distintos métodos aplicados no es buena. La principal

Tabla 2: Resultados de los 5 mejores modelos prediciendo 6, 4 y 3 niveles. Error promedio entre clases mediante validación cruzada. En todos los casos los modelos surgen de utilizar una grilla y usar XGBoost.

modelos	nro clases 3	nro clases 4	nro clases 6
Modelo 1	0.58	0.69	0.78
Modelo 2	0.58	0.69	0.79
Modelo 3	0.58	0.69	0.79
Modelo 4	0.58	0.69	0.79
Modelo 5	0.58	0.69	0.79

Figura 10: Error promedio entre clases utilizando distinta cantidad de categorías en la predicción de la prueba de audición de inglés.

debilidad que tiene el problema planteado es la información que se utiliza para predecir el resultado de la prueba de Inglés. Los datos de uso de la plataforma CREA brindados por CEIBAL no pueden ser filtrados para las actividades solamente relacionadas a Ceibal en Inglés por lo que la calidad de la información para el problema no es buena. En el futuro se está trabajando en una plataforma específica para Ceibal en Inglés que permitirá obtener mejor información para el problema planteado y es esperable que la performance predictiva de los métodos presentados mejore.

5. Consideraciones finales

En este trabajo se presentan los principales resultados del proyecto ANII: “Nuevas Tecnologías Estadísticas para la Evaluación y Monitoreo de Plataformas Educativa” desarrollado en el marco del fondo sectorial de investigación a partir de datos 2017. Parte importante del trabajo es la transformación de los datos brindados por CEIBAL en información relevante para la toma de decisiones informadas. Se combinaron distintas herramientas computacionales que permitieron realizar este trabajo de forma estandarizada y eficiente. Como uno de los productos del proyecto se desarrolló una aplicación web usando la librería `shiny` de R la cual consta de 6 pestañas, una introducción, cuatro niveles de análisis y un reporte dinámico. La introducción explica cada uno de los niveles de análisis y sus respectivas visualizaciones. Existen diferentes opciones (parámetros) que el usuario puede elegir para ver distintos años, departamentos, escuelas, grados, clases e indicadores.

El público objetivo es de dos tipos, los primeros son llamados mentores cuya labor consiste en asistir a diferentes escuelas a corroborar el uso de la plataforma y obtener información directamente a partir de la directora y maestros. Este aspecto es relevante en la medida que los mentores no cuentan con esta información (actualmente) al momento de recorrer escuelas mientras que al tener la aplicación podrán consultar datos actualizados. La fre-

cuencia de actualización dependerá de la estructura e infraestructura de CEIBAL pero dados los volúmenes de información y la capacidad de computo no sería de extrañar que la información este actualizada en tiempo real o con un rezago de un día o una semana. El segundo público objetivo son los analistas de CEIBAL quienes arman reportes referidos a distintos centros de estudios de forma manual. Esto genera que dediquen tiempo a tareas que pueden ser automatizadas, optimizando tiempo y recursos.

Con esta aplicación, la información queda sintetizada a distintos niveles de análisis tanto en el espacio como en el tiempo lo que permite una gran flexibilidad a la hora de identificar el comportamiento de diferentes departamentos, centros de estudios y clases. Además, se da la opción de generar reportes automáticos al seleccionar la escuela de interés.

Se puede elegir el formato: PDF, HTML y Word. Al seleccionar la opción descargar se genera el mismo reporte automático y se descarga en el formato deseado. El informe además de incluir las gráficas previamente mostradas, incluye una explicación de cada una y un breve análisis automatizado. Esto es posible puesto que los parámetros seleccionados son variables que comparte el reporte, lo que se denomina un reporte parametrizado.

De esta forma tanto analista o mentor de CEIBAL pueden obtener un reporte de forma instantánea con todos los niveles de análisis y flexibilidad que deseen, disminuyendo de forma drástica la cantidad de tiempo que dedican a analizar o generar reportes y puedan enfocarse en sus objetivos primordiales.

Posteriormente se utiliza la información de uso de la plataforma CREA para predecir la performance en las pruebas adaptativas de inglés. Se lleva a cabo una estrategia de identificación de variables relevantes y se encuentra que los resultados de las pruebas de inglés en años previos y la actividad extra fuera del aula de clases son las variables más importantes para predecir los puntajes, no así la información de uso la plataforma CREA. Esto indica la necesidad de obtener información respecto a los puntajes históricos del alumno en años previos en la materia de interés, su actividad extracurricular y que la información de CREA debe estar desagregada por materia, es decir, se debe identificar a que tema corresponde cada actividad si el objetivo es utilizar dicha información con fines predictivos. Al momento de realizar el proyecto no era posible filtrar la información de uso de la plataforma CREA solamente para actividades relacionadas a Ceibal en Inglés. Se plantean distintos modelos buscando predecir las categorías de los puntajes en las pruebas siguiendo distintas metodologías como bosques aleatorios, XRT, redes neuronales, XGBoost y SuperLearner. En todos los casos los resultados obtenidos no son satisfactorios, tampoco lo son si se reducen las clases a predecir, pasando de seis a cuatro y finalmente trabajando con tres clases. Reducir la cantidad de categorías a predecir implica una mejora en promedio de diez puntos de precisión (seis a cuatro categorías y seis a tres categorías). De esta forma se corrobora el resultado previo y se concluye que si el objetivo deseado es predecir el resultado de las pruebas adaptativas de Inglés es necesario poder desagregar la información de uso de la plataforma según las distintas materias y utilizar solamente la información de uso que corresponda a actividades de Inglés.

Referencias Bibliográficas

- Baker, R. S. J. (2011). Data Mining for Education. En *International Encyclopedia of Education*. 3 edici
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S., e Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23(1):537–553.
- Bengio, Y., Courville, A., y Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.
- Breiman, L. (1996a). Bagging Predictors. *Machine Learning*, 24(421):123–140.
- Breiman, L. (1996b). Stacked regressions. *Machine Learning*, 24(1):49–64.
- Breiman, L. (2001). Random Forests. *Random Forests*, 45:5–32.
- Breiman, L., Friedman, J., Olshen, R., y Stone, C. (1984). *Classification and Regression Trees*.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., y McPherson, J. (2020). shiny: Web Application Framework for R.
- Chen, T. y Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Aug:785–794.
- Dowle, M. y Srinivasan, A. (2020). data.table: Extension of ‘data.frame’.
- Freund, Y. (1995). Boosting a Weak Learning Algorithm by Majority. En *Information and Computation*, volumen 121, pp. 256–285.
- Freund, Y. y Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119–139.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4):367–378.

- Genuer, R., Poggi, J., y Tuleau-Malot, C. (2015). VSURF : An R Package for Variable Selection Using Random Forests. *The R Journal*, 7(2):19–33.
- Geurts, P., Ernst, D., y Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.
- Hintze, J. L. y Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2):181–184.
- Lecun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyou, P., Kurka, M., y Malohlava, M. (2020). h2o: R Interface for the 'H2O' Scalable Machine Learning Platform.
- Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 82(3):329–348.
- Loh, W. Y. y Zheng, W. (2013). Regression trees for longitudinal and multiresponse data. *Annals of Applied Statistics*, 7(1):495–522.
- Marconi, C., Goyeneche, J. J., y Cobo, C. (2017). When teachers and machines achieve the best combination: A national comparative study of face-to-face and blended teaching and learning. *The Sixth International Conference on Data Analytics*, pp. 34–40.
- Perera, M. y Aboal, D. (2019). The impact of a mathematics computer-assisted learning platform on students' mathematics test scores. *UNU-MERIT Working Paper Series*, 7(January):1–24.
- Rovira, S., Puertas, E., e Igual, L. (2017). Data-driven system to predict academic grades and dropout. *PLoS ONE*, 12(2):1–21.
- Schapire, R. E. (1990). The Strength of Weak Learnability. *Machine Learning*, 5(2):197–227.
- Severín, E. y Capota, C. (2011). La computación uno a uno: nuevas perspectivas. *Revista Iberoamericana de Educación*, 56:31–48.
- Thakar, P. (2015). Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. 110(15):60–68.
- Tibshirani, R., Hastie, T., y Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* by Trevor Hastie, Robert Tibshirani, Jerome Friedman, volumen 2.

- Uruguay (2010). Ley Nro 18640 Declaración de interés nacional: Programas de apoyo a la promoción de la educación en la niñez y la adolescencia.
- Van Der Laan, M. J., Polley, E., y Hubbard, A. (2007). Super learner. Statistical applications in genetics and molecular biology. *Statistical Applications in Genetics and Molecular Biology*, 6(1).
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. springer.
- Wickham, H., François, R., Henry, L., y Müller, K. (2019). dplyr: A Grammar of Data Manipulation.
- Wilkinson, L. (1999). *Statistics and Computing: Editorial*.
- Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks*, 5(2):241–259.
- Wright, M. N. y Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1):1–17.
- Xie, Y., Allaire, J., y Grolemond, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida.

6. Apéndice

Tabla 3: Variables disponibles en 2015, 2017 sobre uso de plataforma CREA2, variables generadas a partir de actividad en CREA y puntajes en pruebas de inglés.

Nombre de variables	Significado	Valores
VARIABLES DE COMPORTAMIENTO		
cr_comentariosdiscusiones	Comentarios en discusiones	0 a 93
cr_comentariosposteados	Comentarios posteados	0 a 842
cr_comentariosposteadosassignmen	Comentarios posteados por tarea	0 a 29
cr_comentariosposteadosgradeddis	Comentarios posteados con calificación	0 a 93
cr_totalacciones	Total de acciones en la plataforma	0 a 2473
cr_totalautoregistros	Total de registros	0 a 3
cr_totalbusquedarecurso	Total de búsquedas de recursos	0 a 40
cr_totalcambiocalificacion	Total de veces que cambió la calificación	0 a 62
cr_totalcomentariosmediaalbum	Total de comentarios en el album	0 a 73
cr_totalcomentariosposteados	Total de comentarios posteados	0 a 307
cr_totaldiasingreso	Total de días de ingreso	0 a 31
cr_totalenvioscalificables	Total de post con calificación	0 a 153
cr_totaleventoscreados	Total de eventos creados	0 a 70
cr_totalfiltrocurso	Total de filtros del curso	0 a 100
cr_totallogins	Total de login	0 a 315
cr_totalmegusta	Total de comentarios con me gusta	0 a 1666
cr_totalrecursocreado	Total de recursos creados por el estudiante	0 a 180
cr_totalrecursos	Total de recursos	0 a 418
cr_totalrecursovisitado	Total de recursos visitados por el estudiante	0 a 418
cr_totalsubmissions	Tiempo total de subidas	0 a 151
cr_totaluploads	Total de archivos subidos	0 a 177
cr_totalusuariocreado	BUSCAR	0-1
cr_totalvisitas	Total de veces que visitó CREA	0 a 315

Tabla 3: Variables disponibles en 2015, 2017 sobre uso de plataforma CREA2, variables generadas a partir de actividad en CREA y puntajes en pruebas de inglés. (*continued*)

Nombre de variables	Significado	Valores
crusuarioactivo	Total de veces activo	No-Si
crusuarioactivofecha	Total de veces activo por fecha	2017-01-12: 2017-12-31
actividad	Tiene actividad o no	1-0
Otras variables del estudiante		
sexo	Genero	Femenino- Masculino
edad	Edad	9 a 13
fechanacimiento	Fecha de nacimiento	Fecha de nacimiento
turno	Turno	Diferentes horarios
Variabes de la escuela		
aprende16	Aprende 16	1-2
conectividad	Tipo de conexión a internet	Fibra Óptica; Otros
cont2015_urb	Contexto urbano de la escuela	1-5
contexto	Contexto	1-5
departamento	Región geográfica de la escuela	19 departamentos
localidad	Pueblo de la escuela	Nombre de localidades
id.centro	ID de la escuela	ID de los centros
grado	Grado de la escuela	4-6
grupo	Grupo-grado de la escuela	A-F
estrato	Estrato de la escuela	1 y 3
dias_comment	Total de comentarios por día	0 a 31
id	ID del estudiante	ID
mes	Mes	meses
modelodispositivo	Marca de la computadora	4 modelos
tipodispositivo	Tipo de aparato	Laptop-Tablet Ceibal

Puntaje/Categoría en Inglés

Tabla 3: Variables disponibles en 2015, 2017 sobre uso de plataforma CREA2, variables generadas a partir de actividad en CREA y puntajes en pruebas de inglés. (*continued*)

Nombre de variables	Significado	Valores
fuera_rank_list17	Variable para filtrar puntaje de inglés	0-2
fuera_rank_vrg17	Variable para filtrar puntaje de inglés	0-2
nivel_list16	Nivel en listening 2016	A0-B1
nivel_listening17	Nivel en listening 2017	A0-B1
nivel_vrg16	Nivel en vocabulario, lectura y gramática (VRG) 2016	A0-B1
nivel_vrg17	Nivel en vocabulario, lectura y gramática (VRG) 2017	A0-B1
nivelgrupo	Nivel grupo	1-3
thotfin_list17	Puntaje listening 2017	0 a VER valor al filtrar por thotfin
thotfin_vrg17	Puntaje en vocabulario, lectura y gramática (VRG) 2017	–
tthotfin_list16	Puntaje en Listening 2016	–
tthotfin_vrg16	Puntaje en vocabulario, lectura y gramática (VRG) 2016	–
Otras variables creadas		
ind_eng90	Índice de engagement	0-1
ind_eng90_cei	Índice de engagement CEI	0-1
extraEdad	Identificado alumno extra edad	0-1
ubicacion	Montevideo/Interior	Montevideo/Interior
cantActividadMes	Cantidad de actividad en el mes	1-12
sobreQ3cantActividadMes	Cantidad de actividad sobre 3er cuartil en el mes	1-12
sdActividadMesTotal	Desvío estándar actividad en el mes	R+
meanActividadMesTotal	media actividad en el mes	meanActividadMesTotal
sobreMedianMaxActividadMesTotal	Identificar sobre media en el mes	0-12
sobreMedianMinActividadMesTotal	Identificador sobre mediana en el mes	0-12

Tabla 3: Variables disponibles en 2015, 2017 sobre uso de plataforma CREA2, variables generadas a partir de actividad en CREA y puntajes en pruebas de inglés. (*continued*)

Nombre de variables	Significado	Valores
cantActividadMesNula	Cantidad de actividad nula	0-12
cantActividadMesBaja	Cantidad de meses con actividad baja	0-12
cantActividadMesMedia	Cantidad de meses con actividad media	0-12
cantActividadMesAlta	Cantidad de veces con actividad alta	0-12
cantSobreMeanEng	Cantidad de veces sobre la media del IE	0-12
cantSobreMeanUbicEng	Cantidad de veces sobre media del IE Montevideo/Interior	0-12
cantSobreMeanDptoEng	Cantidad de veces sobre media del IE por departamento	0-12
sdEng	Desvío estandar del IE	R+
cantSobreMeanEngCei	Cantidad de veces sobre media del IE CEI	0-12
cantSobreMeanUbicEngCei	Cantidad de veces sobre media por Montevideo/Interior IE CEI	0-12
cantSobreMeanDptoEngCei	Cantidad de veces sobre la media del IE por dpto	0-12
sdEngCei	Desvío del IE CEI	R+

Notas:

Los nombres de las variables son las especificaciones que se utilizaron en el estudio, el significado explica a que corresponden, mientras los valores refieren al rango de valores si la variable es numérica y a las categorías si la variable es categórica. Las variables de comportamiento se crearon en base al registro de la plataforma CREA. El puntaje y categoría de inglés son los puntajes que obtienen los estudiantes en la prueba de inglés.

Tabla 4: Precisión en el conjunto de testeo para el bosque aleatorio

precisión	modelos
0.32	1.Todos los alumnos sin agrupamiento
0.30	2.Todos los alumnos sin agrupamiento y usando IV
0.32	3.No todos los alumnos, sin agrupamiento
0.30	4.No todos los alumnos, sin agrupamiento y usando VI
0.31	5.Todos los alumnos, sin agrupamiento sin variables generadas
0.32	6.Todos los alumnos, sin agrupamiento y usando VI
0.30	7.No todos los alumnos, sin agrupamiento y sin variables generadas
0.32	8.No todos los alumnos, sin agrupamiento y usando VI
0.41	9.Todos los alumnos, con agrupamiento de clases y variables generadas
0.43	10.Todos los alumnos, con agrupamiento de clases y usando VI
0.42	11.No todos los alumnos, con agrupamiento de clases
0.42	12.No todos los alumnos, con agrupamiento de clases y usando VI
0.41	13.Todos los alumnos, con agrupamiento, sin variables generadas
0.42	14.Todos los alumnos, con agrupamiento y usando VI
0.41	15.No todos los alumnos, sin variables generadas
0.42	16.No todos los alumnos, sin variables generadas y usando VI
0.43	17.Todos los alumnos, con agrupamiento de clases intermedias
0.43	18.Todos los alumnos, con agrupamiento de clases intermedias y usando VI
0.43	19.No todos los alumnos, con agrupamiento de clases intermedias
0.43	20.No todos los alumnos, con agrupamiento de clases intermedias y usando VI
0.42	21.Todos los alumnos, con agrupamiento de clases intermedias, sin variables generadas
0.43	22.Todos los alumnos, con agrupamiento de clases intermedias, sin variables generadas y usando VI
0.42	23.No todos los alumnos, con agrupamiento de clases intermedias, sin variables generadas
0.43	24.No todos los alumnos, con agrupamiento de clases intermedias, sin variables generadas y usando VI

Instituto de Estadística

Serie Documentos de Trabajo



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

IESTA INSTITUTO
DE ESTADÍSTICA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Eduardo Acevedo 1139. CP 11200 Montevideo, Uruguay
Teléfonos y fax: (598) 2410 2564 - 2418 7381
Correo: ddt@iesta.edu.uy
<https://iesta.fcea.udelar.edu.uy/>
Área Publicaciones

Febrero, 2021

Nº2/21