



UNIVERSIDAD DE LA REPÚBLICA Facultad de Ciencias Económicas y de  
Administración Instituto de Estadística

## **Evaluación y monitoreo de plataformas educativas**

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

**Septiembre, 2020**

## **Serie Documentos de Trabajo**

DT (19/2) - ISSN : 1688-6453

## Evaluación y monitoreo de plataformas educativas

Federico Molina <sup>1</sup> Natalia da Silva <sup>2</sup> Ignacio Alvarez-Castro <sup>3</sup> Juan José Goyeneche <sup>4</sup>;

*Universidad de la República,  
Facultad de Ciencias Económicas y de Administración  
Departamento de Métodos Cuantitativos  
Instituto de Estadística*

En este trabajo se presentan los principales resultados basados en el proyecto fondo sectorial de investigación a partir de datos 2017 titulado nuevas tecnologías estadísticas para la evaluación y monitoreo de plataformas educativas. En cuanto a la evaluación de la plataforma, el objetivo específico es analizar la capacidad predictiva de la información generada en CREA respecto de los resultados de la prueba adaptativa del programa Ceibal en inglés. Métodos de aprendizaje automático serán utilizados para cumplir con este objetivo. En segundo lugar, se busca desarrollar indicadores para monitorear el uso de la plataforma a distintos niveles de análisis (clase, grado, escuela, departamento, etc) y a distintas ventanas temporales. Estos indicadores y resúmenes de información estarán implementados para diferentes actores del sistema educativo, mentores, docentes, directores, entre otros. La solución propuesta busca generar en tiempo real, de forma sencilla y sistematizada la información a través de un reporte dinámico en los distintos niveles de análisis. Las herramientas serán generadas con el software libre R combinando shiny y Rmarkdown. 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.

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

Códigos JEL: C14, C16, C41, G23, J11

Clasificación MSC2010: 62P05, 97M30

---

<sup>1</sup>email:fmolina@iesta.edu.uy, ORCID: 0000-0001-7613-1179

<sup>2</sup>email:natalia@iesta.edu.uy, ORCID: 0000-0002-6031-7451

<sup>3</sup>email:nachalca@iesta.edu.uy, ORCID: 0000-0003-1633-2432

<sup>4</sup>email:jjgoye@iesta.edu.uy

---

This paper presents the main results based on the project sector research fund from 2017 data entitled new statistical technologies for the evaluation and monitoring of educational platforms. Regarding the evaluation of the platform, the specific objective is to analyze the predictive capacity of the information generated in CREA regarding the results of the adaptive test of the Ceibal program in English. Machine learning methods will be used to accomplish this goal. Second, it seeks to develop indicators to monitor the use of the platform at different levels of analysis (class, grade, school, department, etc.) and at different time windows. These indicators and information summaries will be implemented for different actors of the educational system, mentors, teachers, principals, among others. The proposed solution seeks to generate information in real time, in a simple and systematized way through a dynamic report at the different levels of analysis. The tools will be generated with free R software combining shiny and Rmarkdown. The objective of the educational monitor is to contribute to the quality of teaching by enabling the identification of students who are at risk of academic lag.

JEL Code: C14, C16, C41, G23, J11

Mathematics Subject Classification MSC2010: 62P05, 97M30

## 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. No todos los docentes han incorporado la herramienta en su trabajo en el aula. En este sentido es importante para evaluar el plan usar la información disponible sobre el uso de las computadoras 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, etc) y a distintas ventanas temporales. Estos indicadores y resúmenes de información estarán implementados para diferentes actores del sistema educativo, mentores, docentes, directores, entre otros. La solución propuesta busca generar en tiempo real, de forma sencilla y sistematizada 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 y Rmarkdown para cumplir con este objetivo. El producto final se basa fuertemente en visualización estadística de datos que permitan resumir de forma sencilla para el usuario final patrones complejos en altas dimensiones. Para obtener visualización de alta calidad se utilizará ggplot2 en R ya que permite mayor flexibilidad que cualquier otro paquete en R y se respalda teóricamente en la gramática de gráficos 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. ? 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 variables crudas, son los insumos para implementar buenos modelos predictivos para los resultados de las pruebas adaptativas de inglés. Será de interés también identificar umbrales de uso de la plataforma para los cuales se obtienen distintos niveles de aprendizaje de las pruebas adaptativas.

## 2. Antecedentes

En los últimos 30 años en Uruguay, a partir de la creación de distintos programas e iniciativas (PEDECIBA, ANII y CEIBAL) el país 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 No18.640 de creación del Centro Ceibal, uno de sus principales cometidos es “contribuir al ejercicio del derecho a la educación y a la inclusión social mediante acciones que permitan la igualdad de acceso al conocimiento y al desarrollo saludable de la infancia y de la adolescencia” (Marconi, 2017).

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 logrado generar igualdad de acceso a la tecnología, así como se asegura 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 educativos de las últimas décadas a nivel mundial es la introducción de nuevas tecnologías que modifican la forma de enseñar y aprender conocimiento. 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 de aprendizaje basados en la web entre otros. En este contexto el Plan Ceibal se empalma con el nuevo paradigma de aprendizaje con 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. Los datos generados son datos en altas dimensiones, se cuentan con registros individuales de uso de las distintas tecnologías en el tiempo con variables a nivel individual contenidas en distintos sistemas de información. La estructura de los datos a su vez es compleja. Estas dos características hacen que los métodos estadísticos clásicos presenten dificultades por lo que es recomendable implementar técnicas estadísticas modernas que permitan descubrir patrones en datos con esta complejidad.

Los métodos de aprendizaje estadístico o machine learning consisten en el entrenamiento de un modelo de forma que aprenda diversos comportamientos usando información de un subconjunto de observaciones. Los métodos de aprendizaje automático son ampliamente utilizados 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 dos enfoques, aprendizaje supervisado y aprendizaje no supervisado. El aprendizaje supervisado implica que la variable de respuesta es conocida y el objetivo es construir un predicto automático para la misma usando información de las variables explicativas. Cuando la respuesta es discreta 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.

Los métodos que se utilizarán en este proyecto se enmarcan en el análisis supervisado, donde se encuentran una amplia variedad de métodos que han tenido gran éxito en sus aplicaciones. Algunos ejemplos son los métodos basados en 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 (Loh, Wei (2008); Yin, (2014)). 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, 1996), Boosting (Freund & Schapire, 1997) y los Bosques Aleatorios (Random Forests, 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.

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, et.al., (2010)). Thakar y Mehta (2015) presentan 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 automático (Machine Learning) 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, et.al., (2010), identifican 5 aproximaciones dentro de la minería de datos educativos: predicción, agrupamiento, minería de relaciones, descubrimiento con modelos y destilación 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 la destilación de datos para el juicio humano como lo especifica Baker, et.al., (2010).

Severin, E., Capota, C. (2011) mencionan que a pesar de la popularidad de la iniciativa de OLPC en Latin America 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 la tecnología que tiene 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, M. et al, 2017).

Por otra parte, el estudio realizado sobre el efecto de la modalidad de enseñanza de inglés (virtual vs presencial) en el nivel de engagement 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 mayores niveles de engagement y mejores performance por parte de los alumnos en la evaluación de aprendizajes (Marconi, C., et al. 2017)

### 3. Datos

Se utilizan 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 Listening y, Vocabulary, Reading and Grammar (VRG) en los años 2016 y 2017.

En el primer caso, los logs registran cualquier actividad que el alumno haya realizado, desde envío de trabajos, acceso a la plataforma, respuestas en foros, entre otras. Para cada alumno  $i$  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  $\sum_i^T N_i$  registros. En el año 2015 se cuenta con 13.745.973 millones de registros, los cuales son transformados en dos tablas finales. Una de corte transversal con una línea (registro) por alumno y otra de formato panel o longitudinales con una observación por alumno-mes, en el caso de no tener actividad en un mes se imputa el valor cero. Se cuenta con un total de 81310 alumnos.

Los datos de 2017 son facilitados por Plan Ceibal procesados parcialmente con un formato de una observación por cada día de ingreso del alumno. En este conjunto fueron añadidos los puntajes de los alumnos en las pruebas adaptativas de inglés de los años 2016 y 2017 (repitiendo el puntaje en todas las observaciones). Al igual que los datos de 2015, se transforman en dos tablas, una con formato transversal y otra longitudinal. Se cuenta con un total de 120276 alumnos, de los cuales 15776 tomaron la prueba adaptativa de Listening tanto en 2016 como en 2017 y 19352 tomaron la prueba de VRG en ambos



periodos.

A modo de ejemplo, se utilizaron variables del nivel de actividad como días de ingreso, de la actividad en foros como los comentarios posteados, del envío de tareas como los envíos calificables y medidas generales como total de logins. Información del dispositivo como tipo de dispositivo, del alumno como el sexo, fecha de nacimiento, curso, grado y grupo. Y del centro de estudio como el contexto socioeconómico (por quintiles, según definición de CEIBAL), la zona (urbana o rural), departamento y localidad entre otras.

## 4. Resultados

### 4.1. Monitor de plataformas educativas

En esta sección se describe las principales características del monitor para plataformas educativas desarrollado en el proyecto. Se desarrolla una aplicación web en lenguaje R usando la librería `shiny`, compuesta 4 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.

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. Los mentores pueden analizar distintos indicadores y dimensiones de la escuela de interés previo a reuniones con directoras y maestros, para ver en qué aspecto el centro de estudio puede mejorar. Este aspecto es relevante en la medida que los mentores no cuentan con esta información al momento de recorrer escuelas. Con la aplicación podrán consultar en tiempo real o con rezago de una semana o un mes.

El segundo público objetivo son los analistas de CEIBAL quienes sintetizan y arman reportes manuales de cada institución, dichas tareas pueden ser automatizadas parcial o completamente. Con esta aplicación, la información queda sintetizada a distintos niveles de análisis espacio-temporal permitiendo flexibilidad a la hora de identificar el comportamiento de diferentes departamentos, centros de estudios y clases. Además, existe la opción de generar reportes automáticos parametrizados al seleccionar la escuela de interés.

#### 4.1.1. Nivel Nacional

El primer nivel de la aplicación, es Nacional. Aquí el usuario puede elegir que año desea visualizar, en conjunto con el tipo de indicador y la medida de resumen. Hay dos visualizaciones, la primera 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.

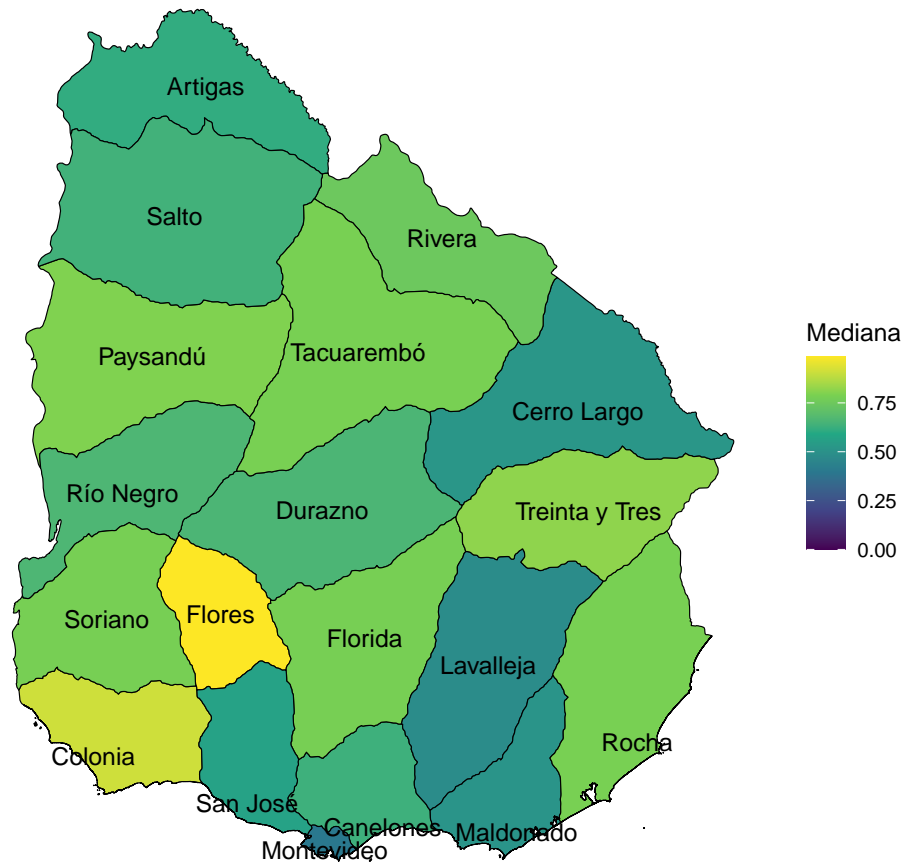


Figura 1: Mapa de Uruguay por departamentos. El color denota el valor de la mediana del índice de engagement, colores de tonalidad amarilla indican un mayor valor. Color de tonalidad oscura un menor valor. El indicador esta limitado entre cero y uno.

La Figura ?? 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 posibles<sup>5</sup>. Por ejemplo, en el caso del IE se observa que la distribución es bimodal, es decir, los alumnos se concentran en dos

<sup>5</sup>Por construcción el rango de valores es mayor al rango de la variable de interés, esto puede ser

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.

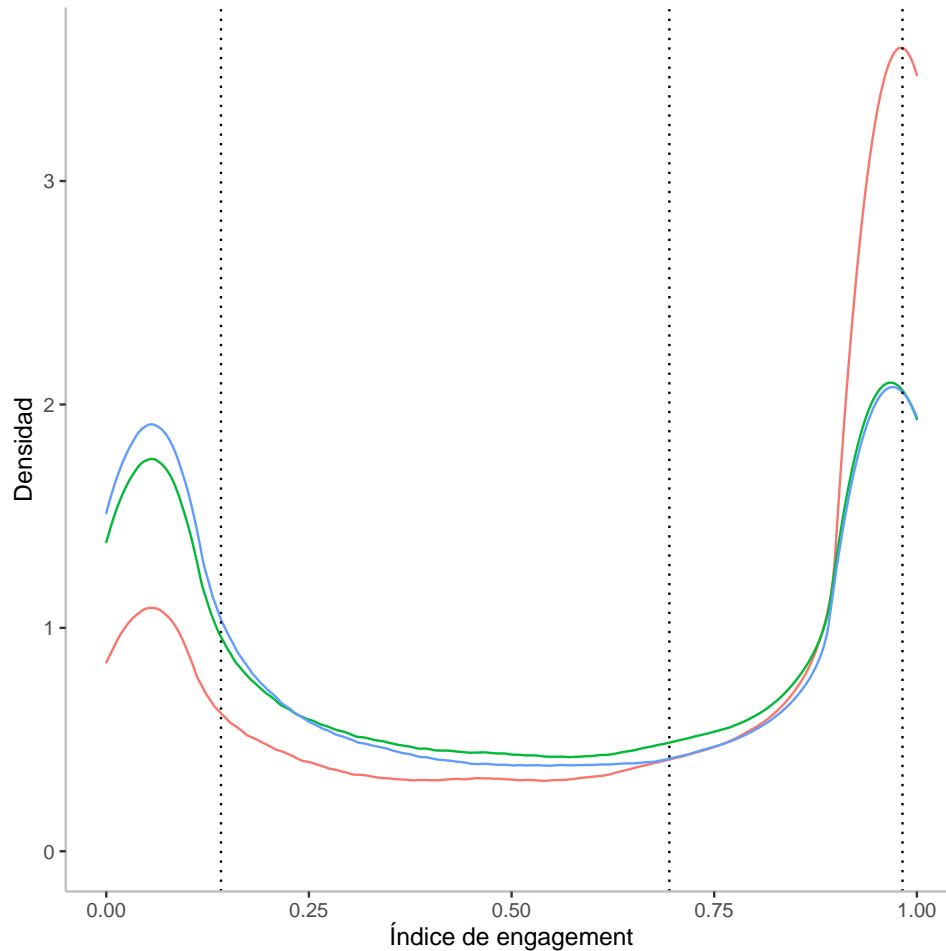


Figura 2: Gráfico de estimación de densidad de kernel. Usando el kernel de Epanechnikov. Las líneas punteadas denotan los cuantiles 25, 50 y 75. El color referencia a los grados.

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.

#### 4.1.2. Escuela

El segundo nivel de análisis (segunda pestaña de la aplicación) se denota Escuela, permite elegir: año, indicador, departamento y escuela. Existen cuatro tipos de visualizaciones que comparan a una escuela de interés con el resto de centros de su mismo departamento. Elegir el departamento de Montevideo implica que el conjunto de visualizaciones solo contiene datos de Montevideo, mientras al seleccionar un departamento diferente, los datos son filtrados para contener únicamente observaciones del Interior. Es decir, al observar una escuela de Salto y ver comparaciones respecto a un promedio de escuelas, el mismo hará referencia al conjunto de escuelas del Interior. Lo mismo se cumple para todos los restantes niveles de análisis.

En la Figura ?? se presenta una gráfica de violin. 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 (de Montevideo o Interior, respectivamente). Complementariamente, se agrega un punto de color rojo especificando la mediana de la escuela seleccionada por el usuario. La interpretación es que para el interior del país, las escuelas de 4to grado tienden a concentrarse mayoritariamente en los límites superiores del IE, las escuelas de 5to y 6to grado se agrupan en torno a los límites del indicador, con una leve superioridad en torno al cero, mostrando una mediana de IE inferior. La escuela 1169 tiene una mediana de IE para cada grado en torno a uno, es decir, alto engagement con la plataforma.

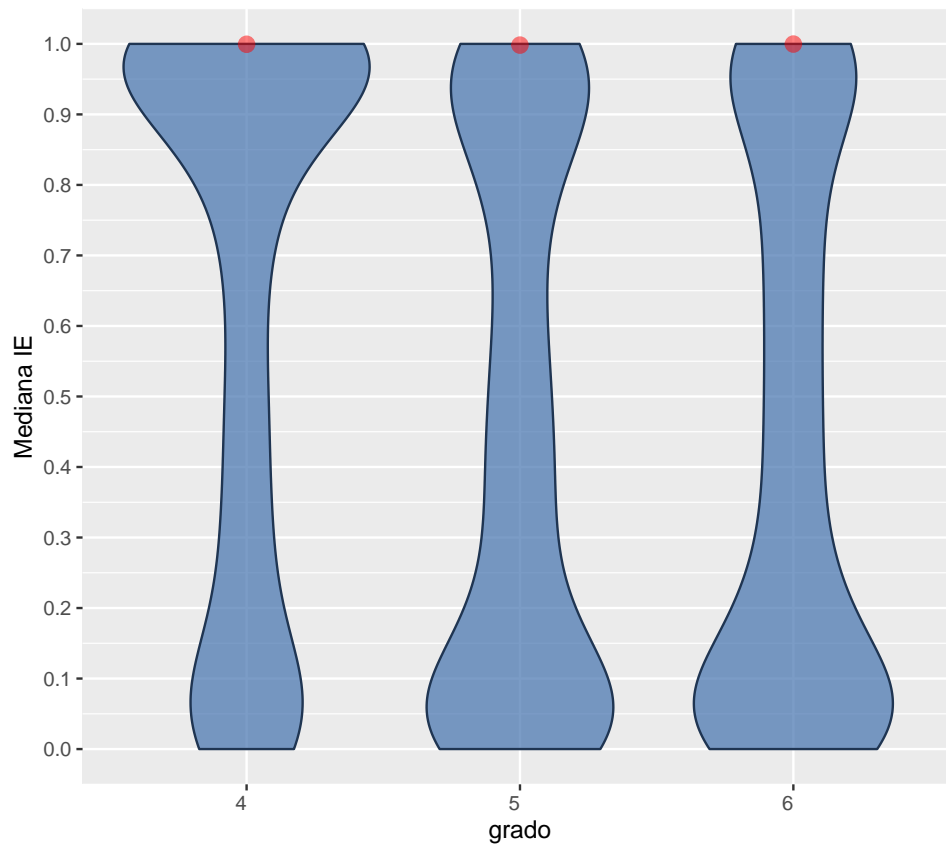


Figura 3: Gráfico de violín, representando la mediana del IE para cada grado del Interior. El color rojo refiere a la escuela 1169.

A continuación, Figura ?? se observa un gráfico de serie de tiempo donde cada punto refiere a la mediana del indicador seleccionado en cada mes del año. La línea punteada de color rojo es la escuela seleccionada, mientras la línea de color negro es la mediana de las escuelas del interior y las bandas de color azul representan el intervalo de confianza de la mediana al 95 %. De esta forma se observa el comportamiento de la escuela de interés a lo largo del año, identificando meses con alta y baja utilización. Adicionalmente se compara con el resto de las escuelas del interior. En el caso de la escuela 1169, a lo largo del año 2017 muestra un engagement sistemáticamente superior al resto de escuela del interior.

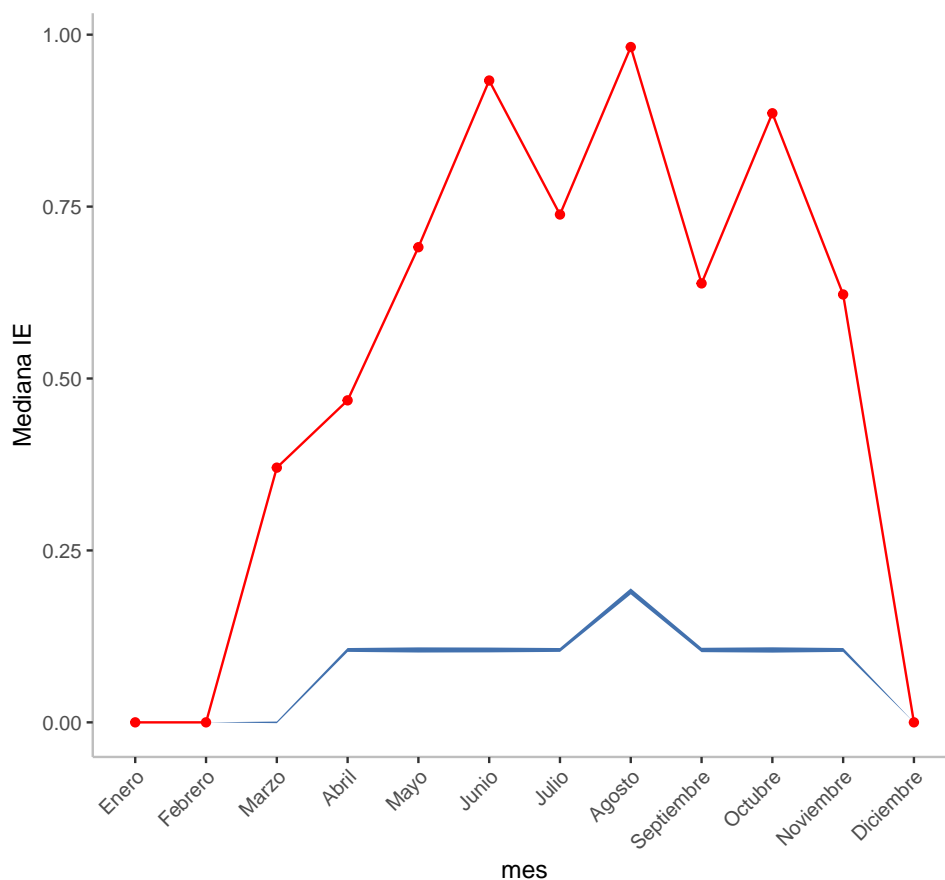


Figura 4: Gráfico de evolución temporal mediana del Índice de Engagement (IE) año 2017. Con color azul la mediana correspondiente a las escuelas del interior del país. Con color rojo la escuela número 1169 la cual se encuentra sistemáticamente por sobre la mediana del IE de escuelas del interior.

La tercera gráfica de la aplicación es una tabla 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. Hay dos columnas si se eligió Montevideo, con el ranking del grado a nivel del departamento y la cantidad de escuelas. Seleccionar un departamento del interior genera cuatro columnas finales, 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 ?? se presenta un gráfico de serie temporal para cada dimensión que compone el indicador. Cada rectángulo es una dimensión (variable) donde el eje vertical representa el promedio normalizado (restando la media y dividiendo por el desvío estándar) y es compartido de forma que las escalas sean comparables entre dimensiones. Las abcisas representan los meses del año. La línea punteada roja es la escuela seleccionada mientras la línea azul es el promedio de Montevideo o interior según corresponda. El área gris corresponde a más o menos dos desvíos estándar. 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 1169 muestra que las dimensiones con mayor actividad son el total de envíos calificables (`cr_totalenvioscalificables`) y total de días de ingreso (`cr_totaldiasingreso`).

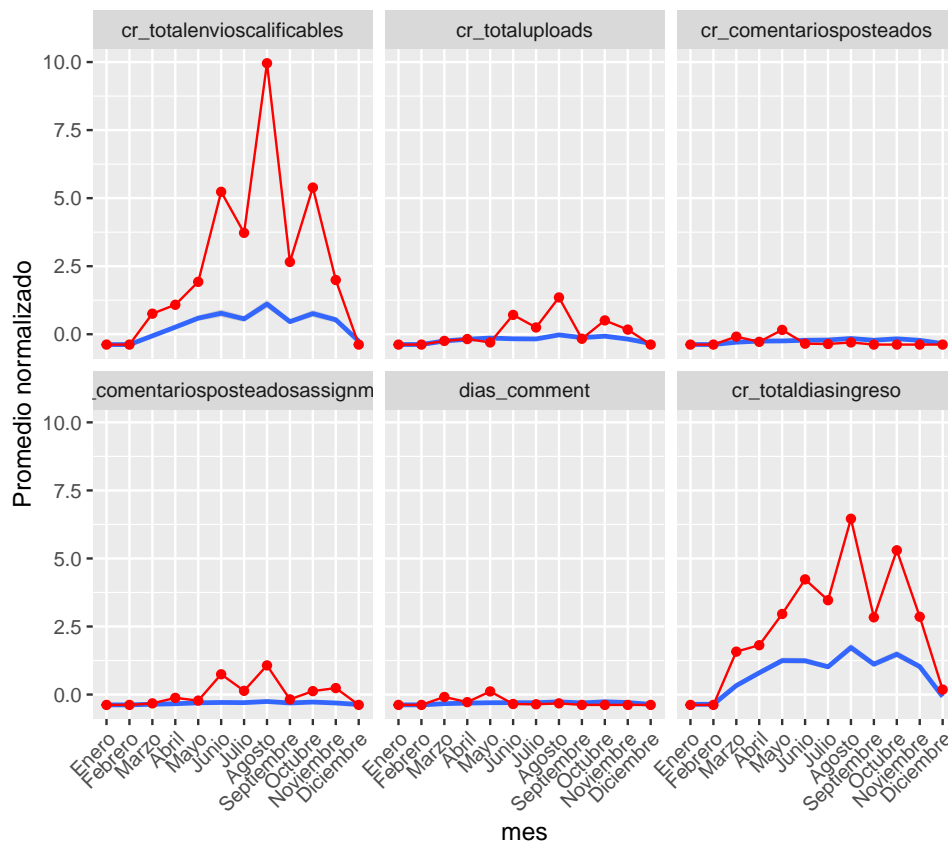


Figura 5: Gráfica de evolución temporal de las dimensiones del Índice de Engagement (IE). Con color rojo el promedio normalizado en cada dimensión de la escuela 1169 y con azul el promedio normalizado de las escuelas correspondientes al interior.

En resumen, en el ejemplo del departamento de Paysandú, la distribución de 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 hay una conglomeración mayor en el extremo superior. Esto nos marca que los cuartos obtienen en mediana mejores puntajes que quintos y sextos. La escuela 1073, en todos los grados obtiene los puntajes máximos en mediana. Al analizarla en su desempeño a lo largo del año se observa que es una escuela destacada que progresa mes tras mes y en todos los meses a partir de marzo se sitúa por encima de la mediana de 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 escuelas del interior. El resto de dimensiones son iguales o levemente mejores. Se concluye, de este nivel de análisis que, en mediana, la escuela 1169, en promedio envía mayor cantidad de envíos calificables que las escuelas del interior e ingresa mayor cantidad de días.

#### **4.1.3. Grado**

El tercer nivel de estudio es Grado. Se pueden elegir los mismos parámetros que en el caso anterior: año, indicador, departamento y escuela. Aplican las mismas consideraciones que en la pestaña anterior, si se selecciona Montevideo el análisis es condicional Montevideo, si se selecciona cualquier otro departamento la comparación es condicional al interior. A diferencia del caso anterior, la agrupación para examinar es por grado de escuela (4to, 5to y 6to).

El primer gráfico, Figura ?? es una serie temporal con la mediana del IE en las ordenadas y los meses del año en las abcisas. Se analiza la mediana del indicador en cada mes del año y se compara con el resto del interior. La línea de color violeta representa la mediana de escuelas del interior o Montevideo, según corresponda, y el área gris dos desvíos estándar. Con color rojo, verde y celeste se representa a cuarto, quinto y sexto grado de la escuela escogida. En el caso de la escuela 1169, todos los grados tienen una mediana del IE superior al resto del interior de forma sistemática.

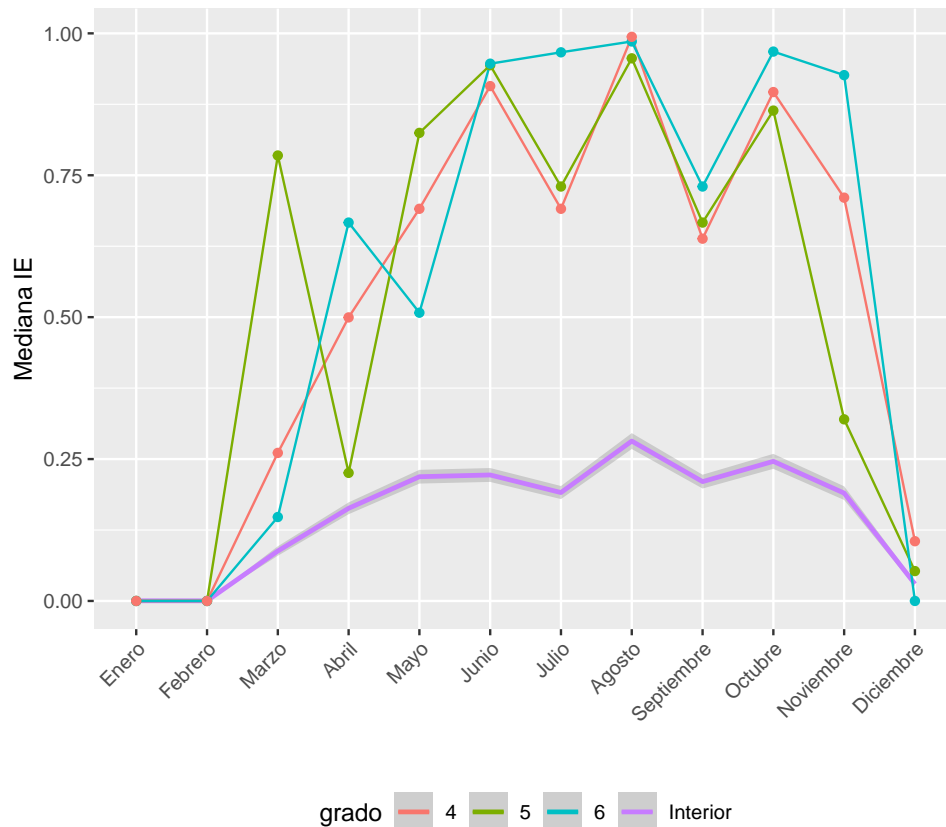


Figura 6: Serie temporal por grado.

La Figura ?? es una diagrama de cajas con el IE en el eje vertical y los grados de la escuela en el eje de las abscisas. Sirve para ver 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 %. La línea del medio en cada caja es la mediana del grado y cada punto es el valor del índice de cada alumno en cada grado de la escuela. Es decir, los puntos observados es el IE de cada alumno en cada grado (sin discriminar por clase). Los estudiantes de la escuela 1169 tienen un IE de entre 0.9 y 1, salvo cuatro estudiantes rezagados respecto a sus compañeros con un IE entre 0.5 y 0.7.

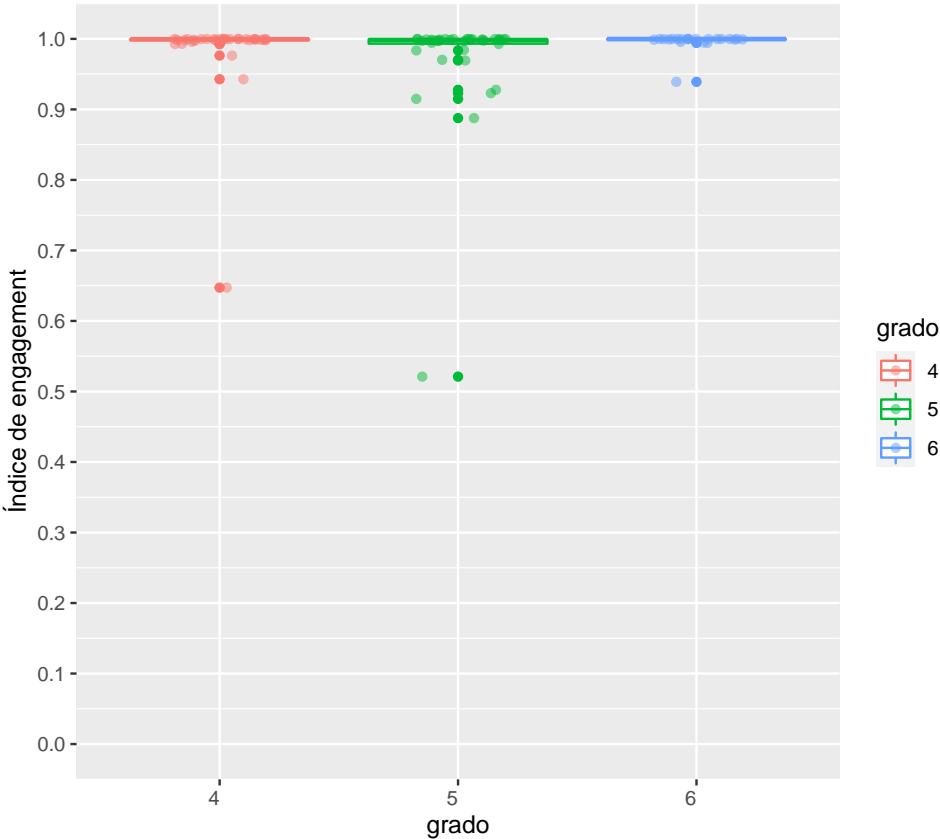


Figura 7:

La tercera representación, Figura ?? es una serie de tiempo para cada dimensión del indicador. El eje vectical es el promedio normalizado en cada dimensión, el cual es compartido en cada rectangulo de forma que las escalas sean comparables. Con color rojo tenemos a cuarto, verde a quinto y celeste a sexto grado. 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 ver si el mismo se da de forma sistemática o en algún mes aislado. Como vimos en la pestaña anterior, la escuela 1169 repite el mismo comportamiento con mayor actividad en total de envíos calificables y total de días de ingreso.

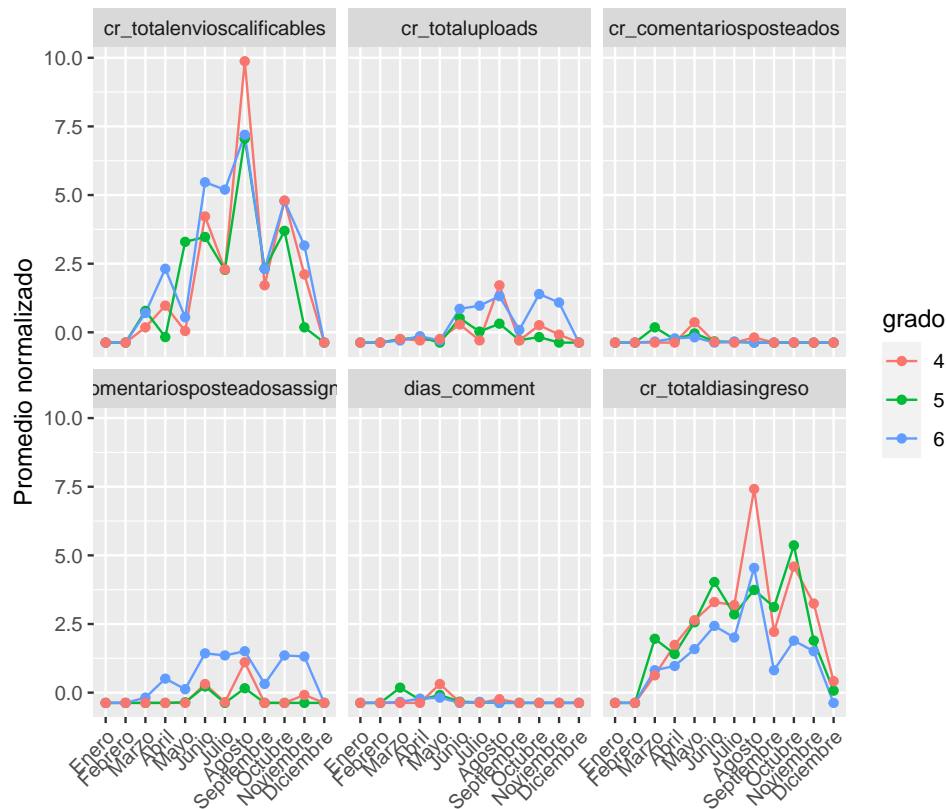


Figura 8: .

La cuarta representación, Figura ?? tiene en el eje vertical el promedio normalizado de todas las dimensiones del indicador, mientras a nivel horizontal están las variables. El color diferencia los valores por grado de la escuela de interés. Este gráfico compara que grados se destacan más o menos en cada dimensión a nivel anualizado sin tomar en cuenta el factor temporal. En la escuela 1169, cuarto y quinto grado son similares, con puntuaciones cercanas en todas las dimensiones, mientras sexto tiene un patrón diferente.

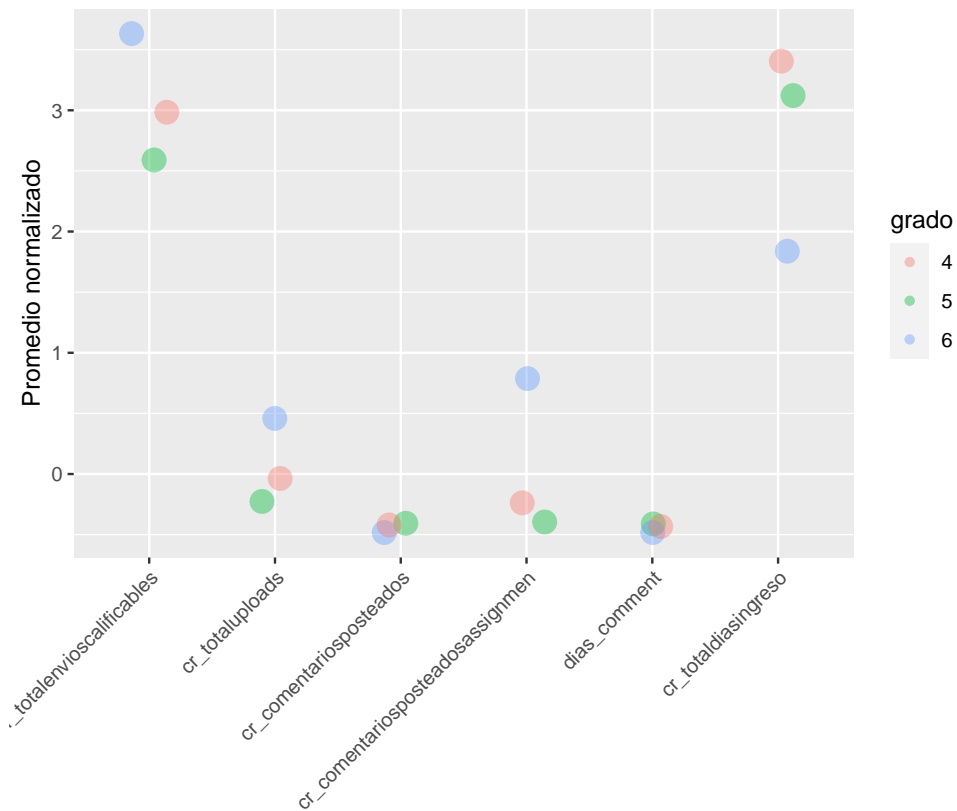


Figura 9: .

Por último, la Figura ?? muestra la mediana del indicador para cada grado y grupo de la escuela seleccionada. Permitiendo comparar en que grupo cada grado tiene un mejor desempeño, si bien se mantiene la agrupación por grado al desagregar por grupo podemos ver si existe variabilidad entre las clases de cada grado. Logrando observar si alguna clase esta rezagada respecto al resto.

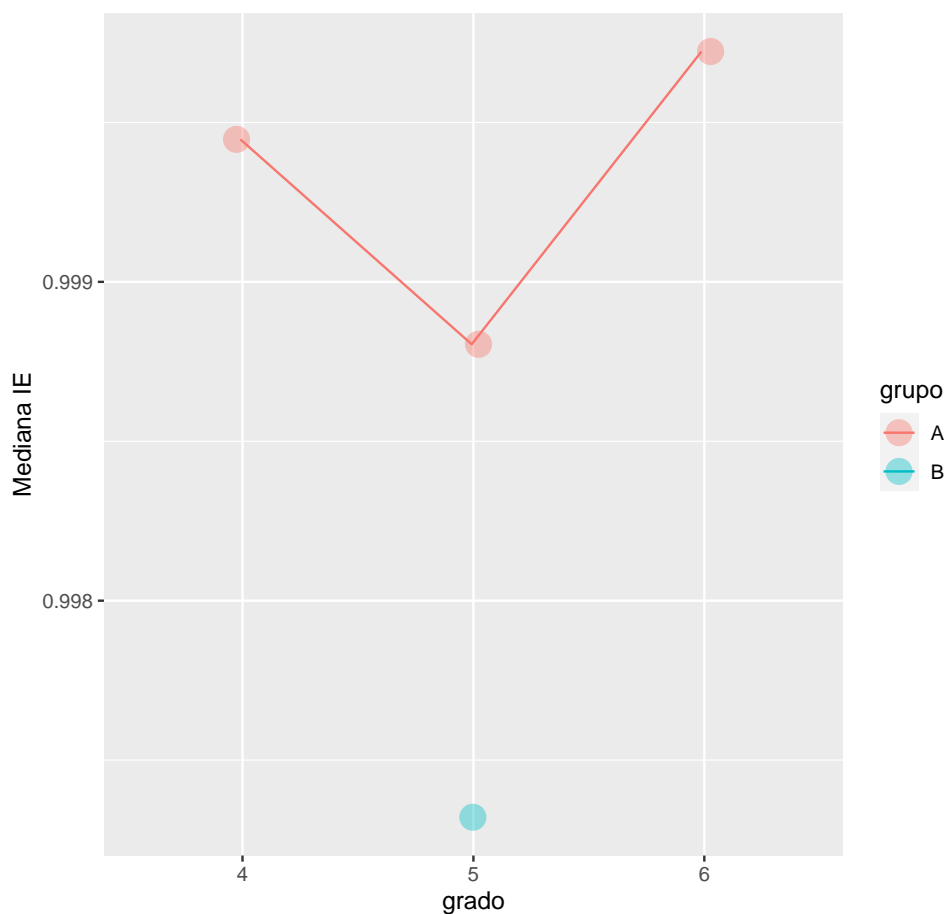


Figura 10: .

En síntesis en el año 2017, departamento de Paysandú escuela 1169, indicador de engagement se observa según el gráfico uno que cuarto y quinto están todos los meses sobre la mediana de escuelas del interior, mientras lo mismo sucede con sexto a partir de mayo. En el gráfico de cajas vemos que los datos están concentrados en torno al máximo valor del IE, en especial en quinto grado, hay pocos alumnos en cada grado con baja participación. La comparación de las dimensiones a través del año muestra diferencia para total de envíos calificables y total días ingreso, en el primer caso los tres grados tienen valores elevados y se mueven de forma conjunta. En el segundo, quinto grado muestra mejores resultados en especial los últimos meses del año, donde se diferencia notoriamente. En el resto de las dimensiones no se observan grandes diferencias. En la figura cuarta se examina que total envíos calificables tiene un valor elevado en todos los grados, muy superior al promedio por

dimensión y al resto de las dimensiones, a excepción de total días ingreso la cual muestra valores elevados pero con mayor dispersión, ya que, cuarto muestra un valor notoriamente inferior a quinto grado. El resto de las dimensiones no muestran diferencias por grado, las cuales están en torno al valor promedio. Finalmente, desagregando los grados por grupo, se advierte que en mediana todos grupos están en torno a los máximos del IE, a excepción de sexto A cuyo mediana se encuentra en torno a 0.8.

#### 4.1.4. Clase

El cuarto nivel de análisis es la Clase. Se mantienen 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 Montevideo, si se selecciona cualquier otro departamento la comparación es condicional al interior. A diferencia del caso anterior, en esta pestaña se miran los datos sin agregar, vemos directamente el IE de cada alumno. La primera visualización esta compuesta de dos gráficos superpuestos, un diagrama de cajas y un gráfico de violin como se observa en la Figura ???. 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 la clase seleccionada. En el ejemplo de la escuela 1169, cuarto grado, clase A, salvo dos alumnos el resto se concentra entre un IE de 0.9 a 1.

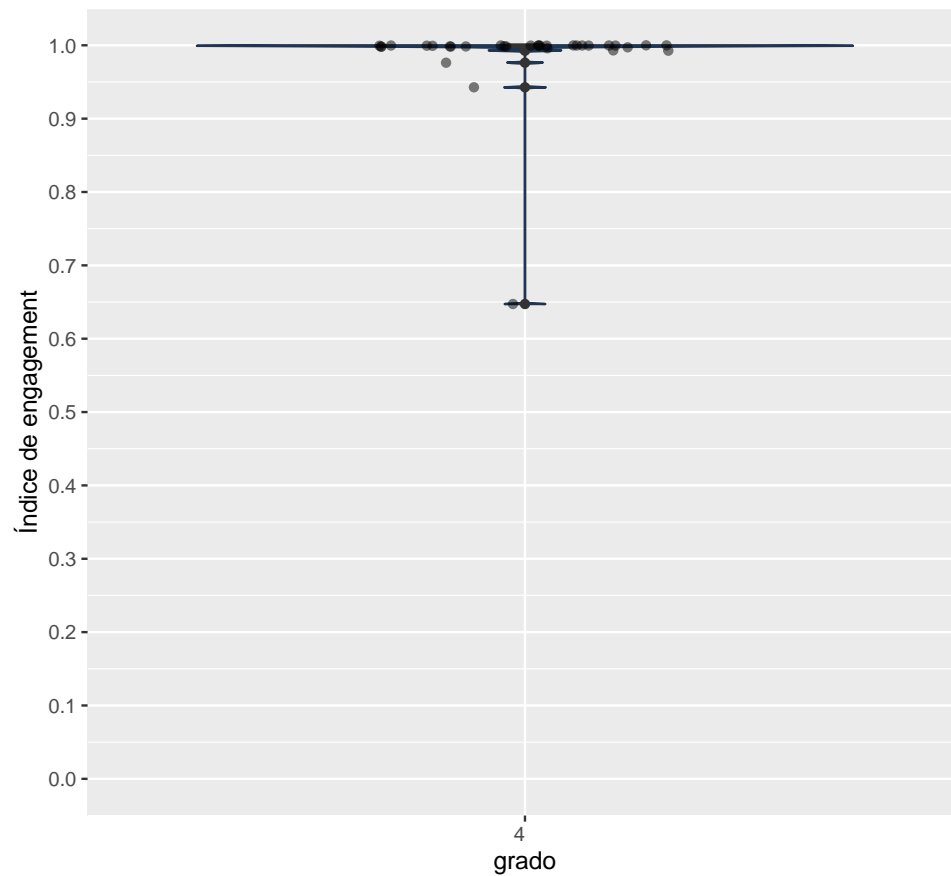


Figura 11: .

La segunda gráfica, Figura ??, es una serie temporal donde cada línea punteada es un alumno y el valor del indicador a lo largo del año. Mientras la línea azul es el promedio de la clase y el área gris denota los intervalos 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.



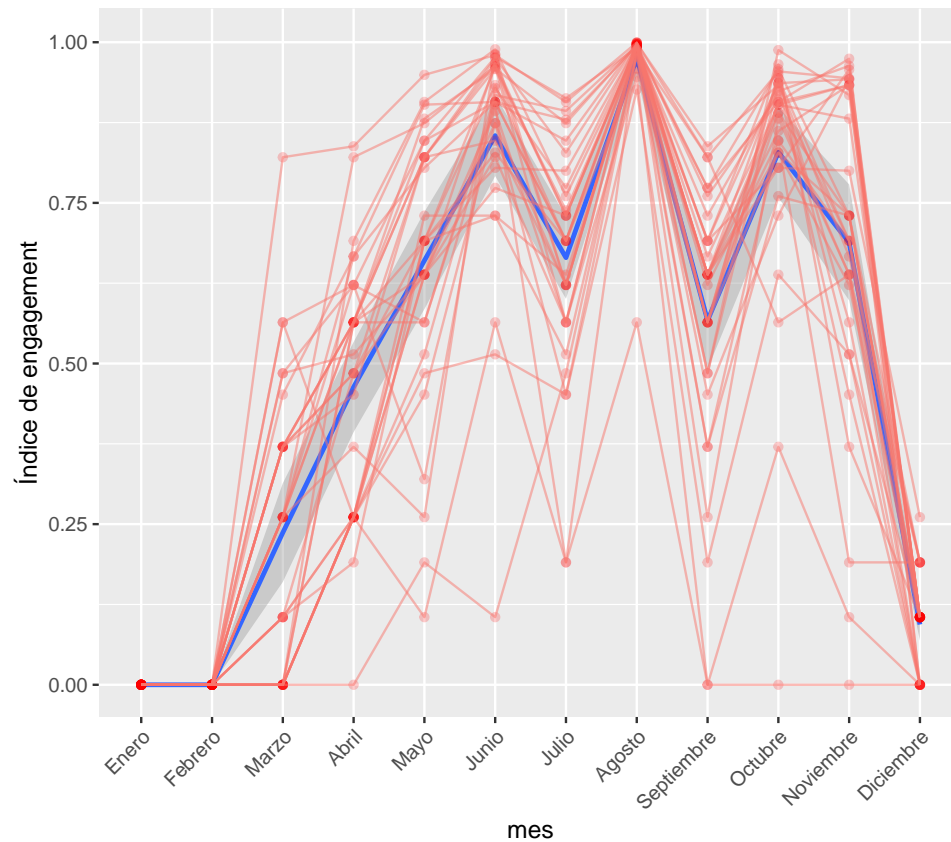


Figura 12: .

Finalmente, en la Figura ?? se presenta un diagrama de cajas para cada dimensión del indicador seleccionado, donde se observa el valor del promedio normalizado en el eje vertical. La explicación de las cajas es idéntica a la que ya se mencionó. Los puntos representan el valor de cada alumno de la clase para esa variable. De esta forma podemos observar la variabilidad de los datos por dimensión en la clase y donde se sitúa cada alumno. Pudiendo identificar si esta por encima o debajo de la mediana, el primer cuartil o el tercer cuartil, y notar posibles valores atípicos, es decir, alumnos o muy destacados o de bajo desempeño en comparación a su clase.

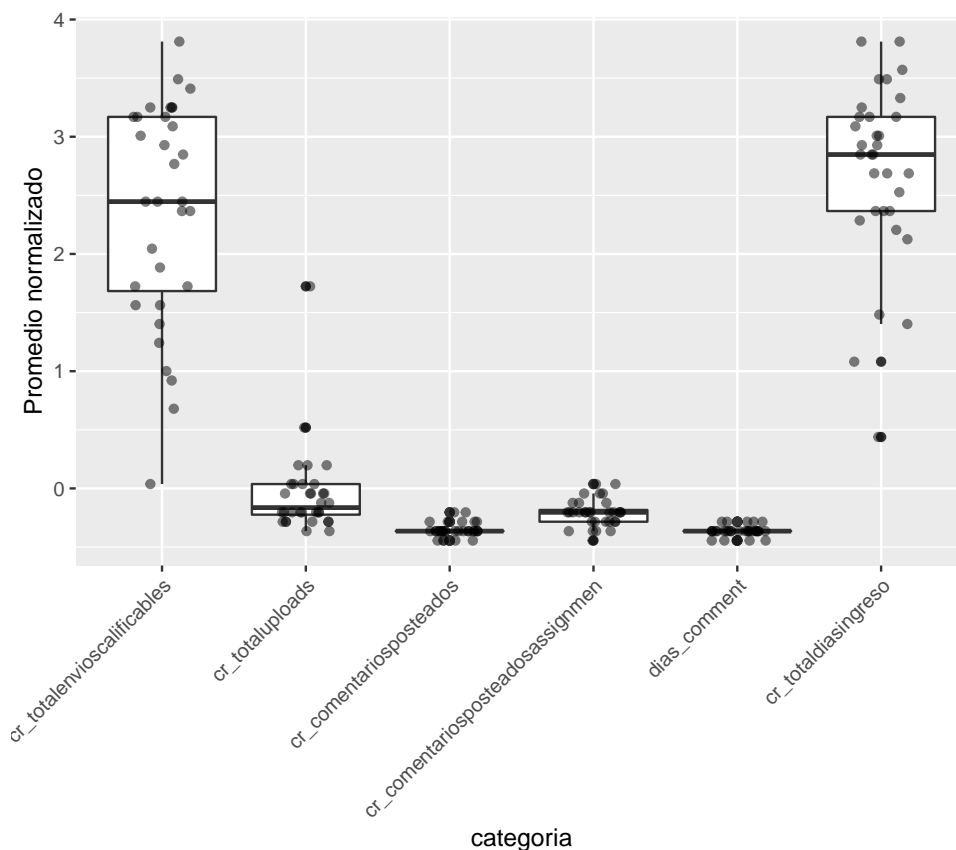


Figura 13: Diagrama de cajas para cada dimensión del índice de Engagement (IE).

En el caso de la escuela 1169, cuarto grado y clase A observamos que la distribución de los datos se concentra en torno al máximo del indicador con tres alumnos de muy bajo desempeño comparativamente a sus compañeros de clase, en torno a un valor del engagement de 0.6. 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 variable con un rendimiento muy superior a sus compañeros. Las restantes variables están en torno al valor promedio de las dimensiones, con poca variabilidad porque las cajas están comprimidas en torno al cero.

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, me-

dida de resúmen, departamento, escuela, grado y clase, estos permanecen con los mismos valores. Si se desea modificar algún parámetro se puede realizar en este mismo nivel. Una vez definida la selección que se desea, se apreta 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.

## 4.2. Resumen de resultados y modelo predictivo

Departamento	Escuelas	Alumnos	Engagement	Ranking
Artigas	37	2810	0.61	14
Canelones	186	15329	0.63	13
Cerro Largo	35	2919	0.52	16
Colonia	53	4220	0.91	2
Durazno	30	1831	0.66	10
Flores	11	834	0.99	1
Florida	30	2127	0.78	5
Lavalleja	28	1650	0.48	18
Maldonado	58	4773	0.51	17
Montevideo	303	21120	0.39	19
Rivera	56	3775	0.75	9
Rocha	44	2676	0.78	6
Salto	47	4430	0.64	12
Soriano	40	2930	0.78	7
sin_dato	1079	28850	0.00	20
Paysandú	49	4034	0.80	4
Treinta y Tres	31	1586	0.82	3
Tacuarembó	48	2854	0.78	8
Río Negro	27	1909	0.66	11
San José	53	3554	0.57	15

Tabla 1: Tabla resúmen

En la Tabla ?? hay 19 departamentos y una fila que hace referencia a los alumnos sin departamento asignado, quienes no tienen actividad en la plataforma CREA. La segunda y tercera columna muestran la cantidad de escuelas y alumnos que participan en la plataforma por departamento, respectivamente. Engagement y Ranking refieren a la mediana del índice de engagement por alumno agrupada por departamento y la posición que ocupa cada departamento. Montevideo es el departamento con mayor cantidad de escuelas y alumnos seguido por Canelones, mientras Flores es el departamento con menor cantidad a la vez que es el de mejor desempeño.

Esta última información puede ser vista en el mapa de la figura ??, en la cual el color amarillo refiere a mayor puntaje mientras un color oscuro uno menor. Se observa un comportamiento diferente entre Montevideo y el Interior del país, mientras Montevideo

muestra el peor desempeño, Flores destaca siendo el de mayor involucramiento en terminos medianos.

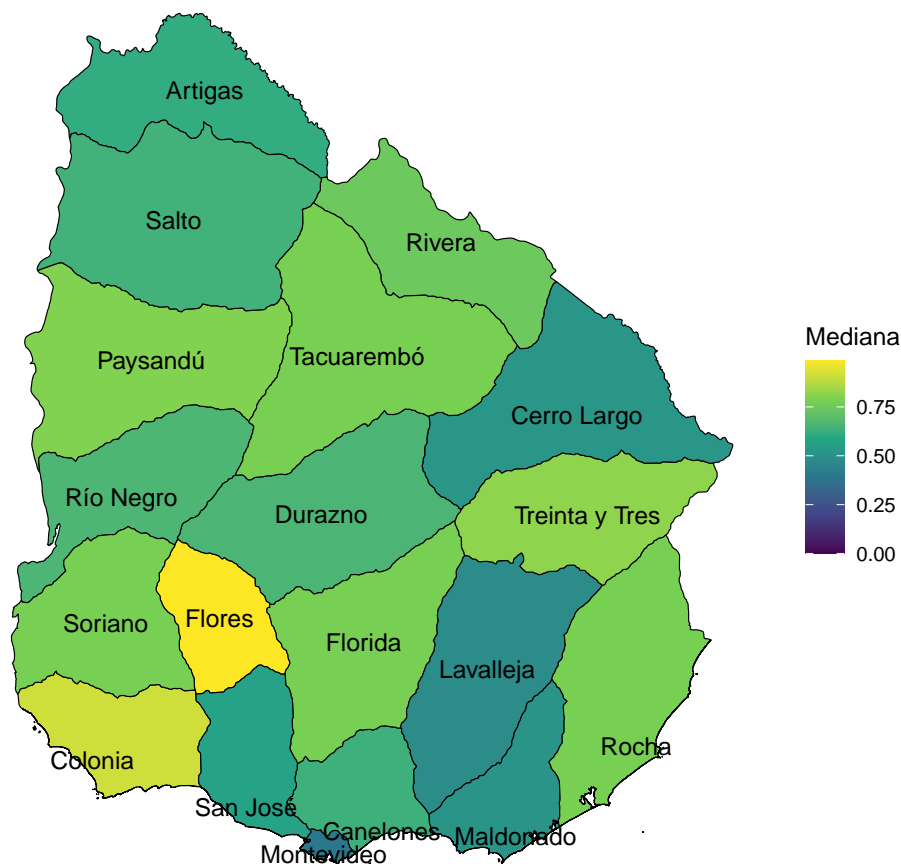


Figura 14: Mapa de Uruguay

En la Figura ?? se observa un gráfico de violin que muestra la distribución del IE por departamento. Se añaden tres líneas rojas por departamento, de menor a mayor hacen referencia al primer cuartil, mediana y tercer cuartil. Y una recta de color negro que muestra la mediana del IE a nivel Nacional.

Se contempla que Montevideo se encuentra por debajo de la mediana nacional, al igual que Maldonado, Lavalleja, Cerro Largo y, levemente Artigas y San José. Flores y Colonia muestran valores cercanos al máximo del índice, Salto, Durazno y Río Negro se encuentran levemente sobre la mediana nacional mientras Florida, Rivera, Rocha, Soriano, Paysandú,

Treinta y Tres y Tacuarembó obtienen un desempeño superior. Es llamativo el desempeño de Flores dado que el primer cuartil del IE se encuentra en un valor entorno al 0.75 superando la mediana de casi todos los restantes departamentos y sobrepasando por casi 0.10 puntos la mediana del IE a nivel nacional. Su distribución de puntajes esta concentrada entre 0.75 y 1.

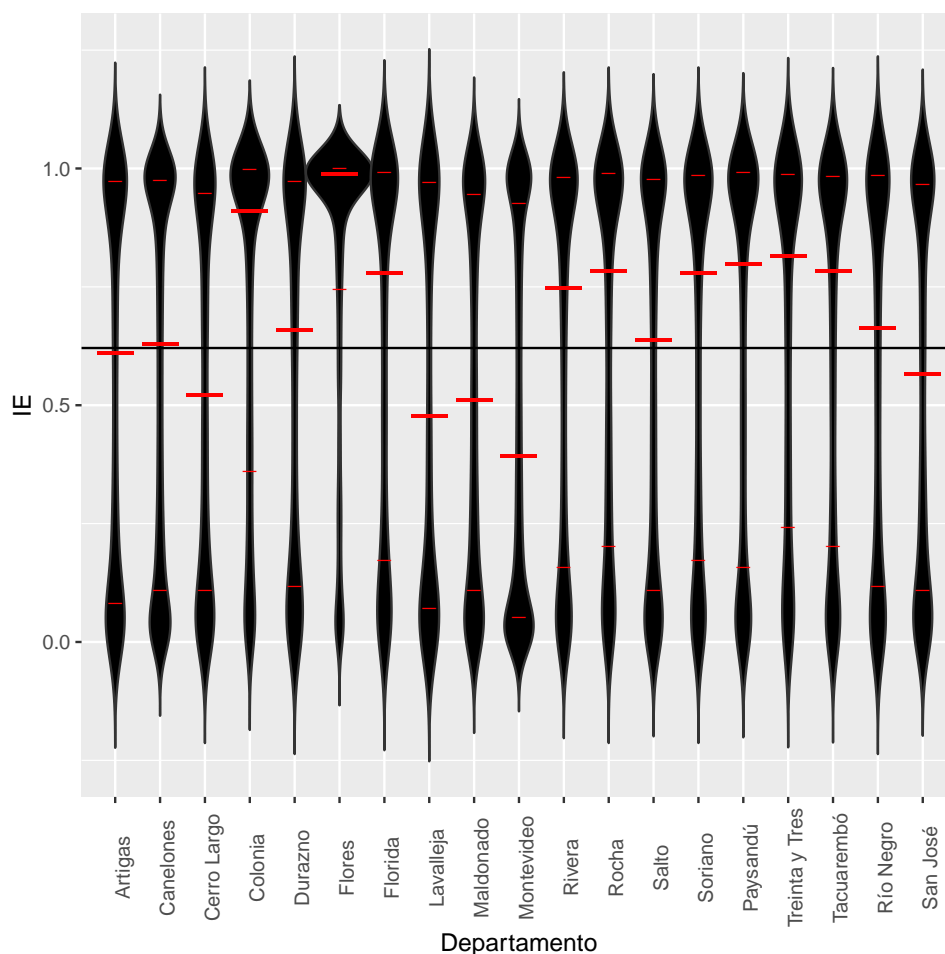


Figura 15: Gráfico de violín departamento de Uruguay.

### 4.3. CEI

Analizamos los puntajes de las pruebas de CEI de 2016 y 2017 respecto a VRG y Listening. En todos los casos se han excluido los alumnos con puntajes incorrectos y sin actividad. Para detectar los puntajes incorrectos se cuenta con una variable booleana 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 siguen el Marco Común Europeo, por lo cual el puntaje más bajo es A0 y el mayor es B1 (la clasificación máxima es C2 pero en alumnos de primaria de CEIBAL el máximo que se considera es B1).

En la Figura ?? se observa la proporción de alumnos en cada categoría de VRG excluidos quienes no tienen puntajes. La mayor cantidad de alumnos estan en el nivel A2+, el cual es un nivel alto. Seguido por la categoría A2- (nivel medio alto) y luego A1-, clasificación baja. Las menores proporciones de alumnos se encuentran en las categorías más bajas y altas, A0 y B1.

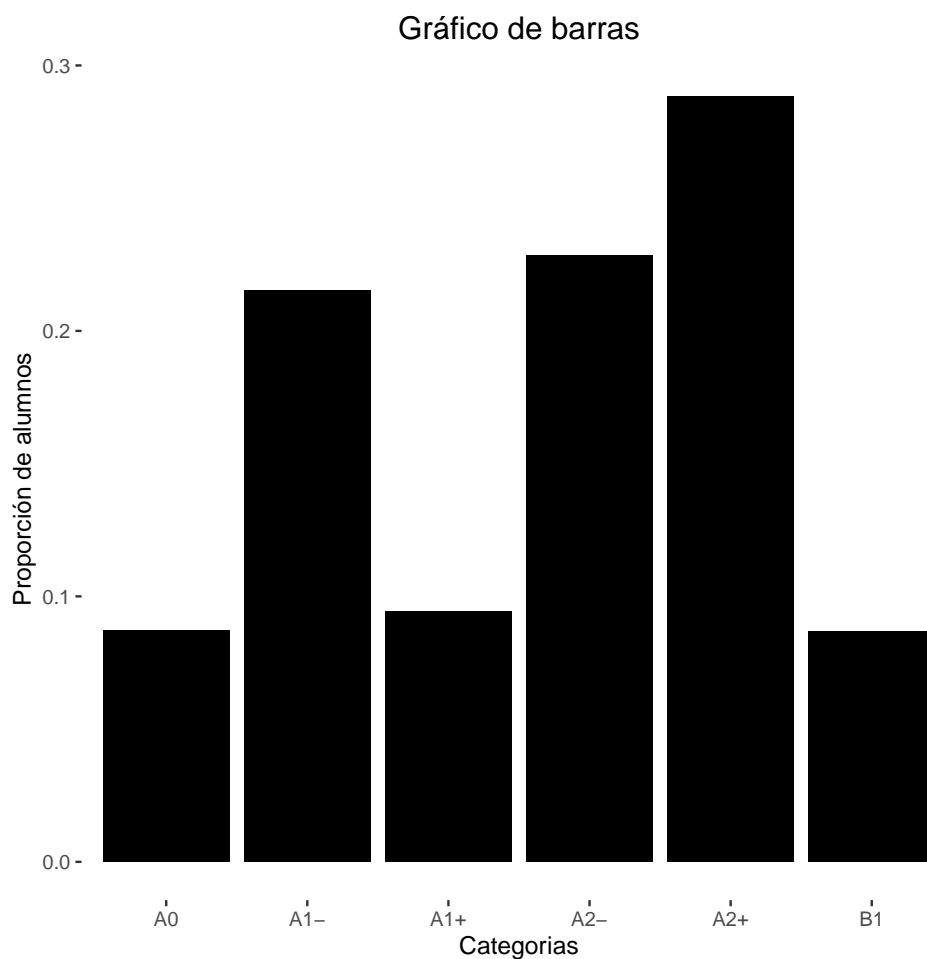


Figura 16: Resultados prueba Vocabulary, Reading and Grammar año 2017. Solo se consideran los alumnos que tomaron la prueba en 2016 y 2017

En el caso de los puntajes de Listening, en la Figura ?? se observa que la mayor cantidad 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. Nuevamente la menor proporción de estudiantes están en las categorías A0 y B1.

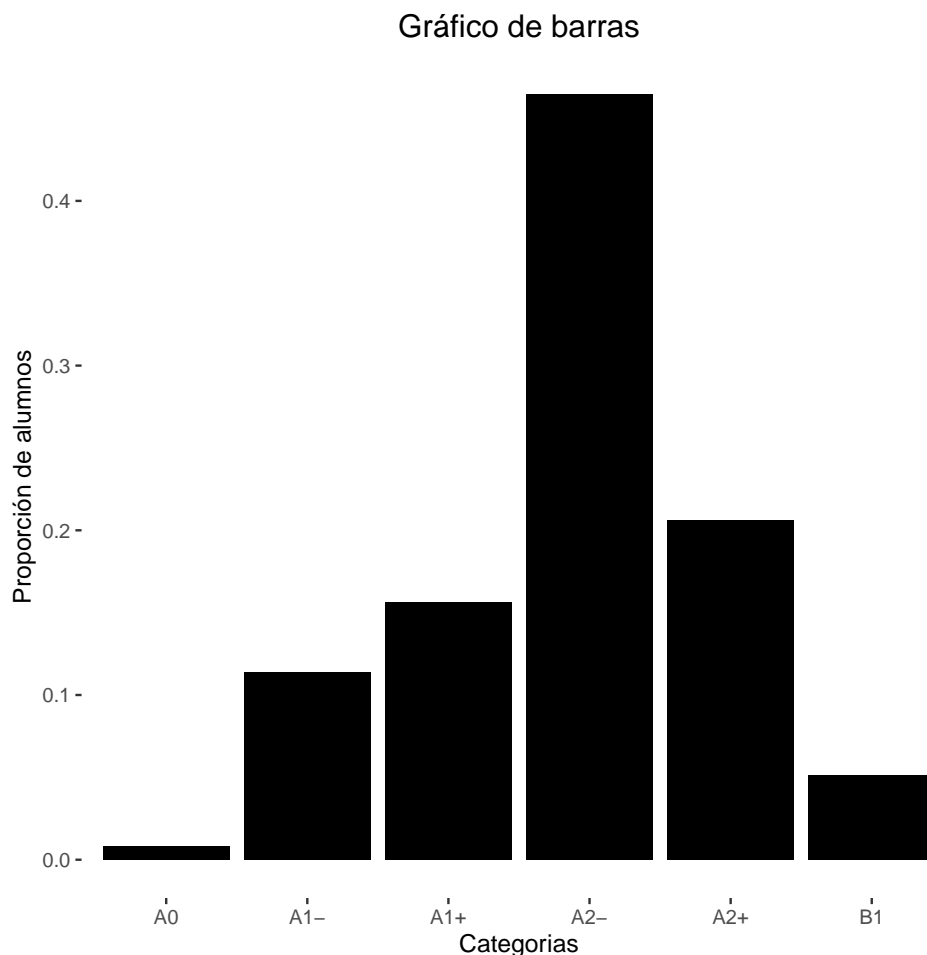


Figura 17: Resultados prueba Listening año 2017. Solo se consideran los alumnos que tomaron la prueba en 2016 y 2017

#### 4.4. Modelo

Se busca predecir la clasificación en las pruebas de inglés. Presentamos resultados para la prueba de Listening restringiendo los datos a estudiantes que tomaron la prueba durante ambos años. Se crean un total de 22 variables, incluyendo dos Índices de Engagement.

Inicialmente utilizamos Random Forest 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 generó un modelo donde solo se utilizaron las variables que tuviesen al menos un 50 % de la importancia de la variable más importante, esto a modo de utilizar pocas variables para remediar posibles problemas asociados a altas dimensiones (*curse of dimensionality*). Por último, se agregó a las combinaciones el filtro de alumnos con tipo de dispositivo sin dato. En total fueron 24 variaciones utilizadas. En ningún caso se obtuvieron resultados satisfactorios. En la Tabla ?? puede observarse que el accuracy tuvo un máximo de casi 44 % y se dio haciendo una agrupación en tres clases (nivel bajo, medio y alto).

accuracy_test	modelos
0.32	no_agrupo_clases_con_variables_alumnos_todos
0.30	no_agrupo_clases_con_variables_alumnos_todos_imp_var
0.32	no_agrupo_clases_con_variables_alumnos_subconjunto
0.30	no_agrupo_clases_con_variables_alumnos_subconjunto_imp_var
0.31	no_agrupo_clases_sin_variables_alumnos_todos
0.32	no_agrupo_clases_sin_variables_alumnos_todos_imp_var
0.30	no_agrupo_clases_sin_variables_alumnos_subconjunto
0.32	no_agrupo_clases_sin_variables_alumnos_subconjunto_imp_var
0.41	agrupo_clases_puntas_con_variables_alumnos_todos
0.43	agrupo_clases_puntas_con_variables_alumnos_todos_imp_var
0.42	agrupo_clases_puntas_con_variables_alumnos_subconjunto
0.42	agrupo_clases_puntas_con_variables_alumnos_subconjunto_imp_var
0.41	agrupo_clases_puntas_sin_variables_alumnos_todos
0.42	agrupo_clases_puntas_sin_variables_alumnos_todos_imp_var
0.41	agrupo_clases_puntas_sin_variables_alumnos_subconjunto
0.42	agrupo_clases_puntas_sin_variables_alumnos_subconjunto_imp_var
0.43	agrupo_clases_con_variables_alumnos_todos
0.43	agrupo_clases_con_variables_alumnos_todos_imp_var
0.43	agrupo_clases_con_variables_alumnos_subconjunto
0.43	agrupo_clases_con_variables_alumnos_subconjunto_imp_var
0.42	agrupo_clases_sin_variables_alumnos_todos
0.43	agrupo_clases_sin_variables_alumnos_todos_imp_var
0.42	agrupo_clases_sin_variables_alumnos_subconjunto
0.43	agrupo_clases_sin_variables_alumnos_subconjunto_imp_var

Tabla 2: Tabla accuracy test

Posteriormente se trabajó con el paquete `h2o` utilizando la función `automl` a modo explorativo para ver que modelos son los que ajustar mejor utilizando validación cruzada y el error promedio entre clases (`mean_per_class_error`) como métrica de elección. Se sigue la metodología utilizada con Random Forest (vía paquete `ranger`) y los resultados, al igual que en el caso previo, no son satisfactorios. De allí surge que el mejor modelo surge de utilizar el algoritmo de XGBoost. Dicho modelo se corre mediante una grilla de forma



de elegir los hiperpárametros que mejor ajuste generen utilizando validación cruzada. Se obtiene un error promedio entre clases en torno al 60% cuando las clases se reducen a tres, lo cual

## 5. Consideraciones finales

Se desarrolla una aplicación web en lenguaje R usando la librería Shiny la cual consta de una introducción, 4 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. El problema que soluciona la aplicación en este sentido es elegir a qué escuela-clase-grado debe asistir el mentor siendo posible elegir las escuelas de peor desempeño. Con esto es posible optimizar los recursos recorriendo los centros más necesitados en un tiempo menor, tomando acciones de forma temprana y ahorrando costos. 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 en tiempo real o con un rezago de una semana o un mes.

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 espacio-temporal lo que permite una flexibilidad total a la hora de identificar el comportamiento de diferentes departamentos, centros de estudios y clases. Además, damos 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. El nivel de automatización que se desee no tiene límite y puede ser total. Es decir, es posible definir los parámetros, variables y reglas que determinen que explicación e información debe aparecer. Más aún, se puede elegir que información mostrar dependiendo del usuario de interés. Por ejemplo, si se pretende enviar un informe al encargado del área, se puede parametrizar de forma tal que al elegir *Encargado* el informe se modifique automáticamente.

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.

Instituto de Estadística

---

Documentos de Trabajo



**Eduardo Acevedo 1139. CP 11200 Montevideo, Uruguay**

**Teléfonos y fax: (598) 2410 2564 - 2418 7381**

**Correo: [ddt@iesta.edu.uy](mailto:ddt@iesta.edu.uy)**

**[www.iesta.edu.uy](http://www.iesta.edu.uy)**

**Área Publicaciones**

**Setiembre, 2020**

**DT (19/2)**