

LatinR, Conferencia Latinoamericana sobre el Uso de R en Investigación + Desarrollo

Ordenando, “*tidy*”, la información sobre educación: una aplicación al Plan Ceibal en Uruguay

Juan José Goyeneche¹, Cecilia Marconi Ferrari², Ana Coimbra Marfetán¹,

Guillermina Costabel¹

¹ Instituto de Estadística, Facultad de Ciencias Económicas y de Administración,
Universidad de la República, Uruguay

² Plan Ceibal, Montevideo, Uruguay

Keywords/Palabras Claves: bases de datos en educación, edición y limpieza
de datos, modelos predictivos, Plan Ceibal

En la ya clásica propuesta de Hadley Wickham, trabajar con grandes cantidades de datos, especialmente si los mismos vienen de bases administrativas creadas para otros fines, puede ser una tarea desafiante y que ponga en peligro un correcto análisis y modelización.

“Tidy datasets are all alike, but every messy dataset is messy in its own way.” —
Hadley Wickham (R for Data Science, Wickham y Grolemund, 2017)

Se presentan dos proyectos desarrollados en el marco del convenio vigente desde 2008 entre Plan Ceibal y el Instituto de Estadística de la Facultad de Ciencias Económicas y de Administración, UDELAR, en los cuales la tarea de importar, limpiar y transformar los datos son claves para poder realizar un análisis robusto y confiable.

En ambos proyectos se desarrollan sistemas en R que leen la información directamente de bases de datos en Oracle, IBM Server y plataformas de Business Intelligence, los editan y limpian usando las bibliotecas tidyverse (como ggplot2, dplyr y tidyr, etc). y se presentan resultados semi-automáticamente mediante knitr y RMarkdown.

Datos de distintas bases son conectados usando merge():bases de personas, equipos, escuelas, entregas y recuperación de equipos, órdenes de trabajo de reparaciones, etc.

Plan Ceibal se creó en 2007 como un plan de inclusión e igualdad de oportunidades con el objetivo de apoyar con tecnología las políticas educativas uruguayas. Desde su implementación, cada niño que ingresa al sistema educativo público en todo el país accede a una computadora para su uso personal con conexión a Internet gratuita desde el centro educativo. (<https://www.ceibal.edu.uy/es/institucional>)

Uno de los proyectos refiere a Análisis Predictivo de Abandono en alumnos de primer año de Enseñanza Media. Se desarrolla un modelo para detectar aquellos alumnos, pertenecientes a primero de Enseñanza Media que tengan una alta probabilidad de abandonar el sistema educativo público en el año siguiente. Con este propósito se utilizó información de la transición de dichos alumnos entre los años 2016 y 2017.

El otro proyecto estima la función de Supervivencia de Tablets y Laptops. Se desarrollan modelos para varios modelos de tablets y de laptops que se entregan a los estudiantes de Enseñanza Primaria y de Enseñanza Media. Las funciones de supervivencia así estimadas brindan información sobre la durabilidad media de los equipos bajo distintos escenarios, la previsibilidad de repuestos en términos de demanda de cantidad y momento del consumo del repuesto, así como de los recursos humanos estimados para realizar reparaciones.

Referencias

1. Wickham, H., Grolemund, G., *R for Data Science*. O'Reilly (2017)