

# First steps towards the development of an academic system to follow the trajectories of primary and secondary uruguayan students

Virginia Rodes<sup>1</sup>, Cristian Cechinel<sup>2</sup>, Henrique Lemos dos Santos<sup>3</sup>, Xavier Ochoa<sup>4</sup>, and Guillermo Ettlín Alonso<sup>5</sup>

<sup>1</sup> Universidad de la República (UDELAR), Montevideo, Uruguay

<sup>2</sup> Universidade Federal de Santa Catarina (UFSC), Araranguá, Brasil

<sup>3</sup> Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brasil

<sup>4</sup> Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Ecuador

<sup>5</sup> Administración Nacional de Educación Pública (ANEP), Montevideo, Uruguay  
virginia.rodes@cse.edu.uy, cristian.cechinel@ufsc.br,  
hlsantos@inf.ufrgs.br, xavier@cti.espol.edu.ec, gettlinal@anep.edu.uy

**Resumen** The present work describes the first steps towards the development of an Academic System focused on tracking the trajectories of uruguayan students from primary and secondary education. Since it is a large-scale project which covers an entire national educational system, several challenges and constraints (both technical and legal) were taken into account during its development. This paper presents some considerations regarding these issues along with the current state of the project. Early results point out the feasibility of finding meaningful patterns in the available data (using data mining techniques) which can be embedded into a prototype for tracking the students scholar trajectory.

**Keywords:** Academic trajectory, Early warning system, Primary and secondary education, Learning analytics, Educational Data Mining

## 1. Introducción

El sistema educativo uruguayo ha experimentado, a lo largo de las últimas décadas, importantes problemas asociados al rezago y la desvinculación educativa. Caracterizado por la cobertura universal y gratuita en el nivel primario, el rezago, abandono y no matriculación de los estudiantes se incrementa durante el avance en el sistema educativo, a la vez que desciende la cobertura en edad oportuna. Como consecuencia en las trayectorias educativas se observa que una parte muy importante de los estudiantes uruguayos tiene dificultades para permanecer en el sistema educativo [1].

Especialmente en el tránsito entre educación primaria y media se registra una caída cercana al 10 % de los estudiantes, siendo que a la edad de 13 años ya un 26 % tiene rezago y un 3 % abandonó el sistema. En secundaria, en el tránsito entre educación media básica (1ero, 2do y 3ero) y educación media superior (4to, 5to y 6to), entre los 15 y los 17 años, se constata un descenso de

20 % en la proporción que asiste al grado esperado para su edad (54 % a 34 %, respectivamente), a la vez que aumenta a 27 % la proporción que no asiste al sistema educativo. En 2015, el sistema educativo uruguayo presentaba una baja tasa de egreso de su tramo obligatorio, el cual se extiende desde el nivel inicial al ciclo medio superior (en total 12 años de escolarización obligatoria). El 31 % de los egresos se da a los 19 años, mientras que el 40 % egresa a los 24 años [1].

Existe un importante corpus de investigación que da cuenta de los aspectos sociales, económicos, históricos y políticos asociados a esta problemática en el Uruguay, muy bien documentado en [2]. En este estudio además avanzan hacia una conceptualización del fenómeno que denominan "*desafiliación educativa*", definido como "*una forma de realizar la transición al mundo adulto, un tipo de trayectoria, en la cual el joven decide que concluye su membresía a una organización y anticipa la finalización de las protecciones sociales, modificando su posición en el espacio social y quedando en una posición vulnerable o directamente excluido de las fuentes de bienestar (mercado y estado)*"[2], siendo el rezago, abandono y ausentismo tres factores de explicación desde la perspectiva temporal de la desafiliación educativa.

Identificar estos objetos en las trayectorias educativas de los sujetos permite establecer acciones tempranas para mitigar los riesgos de la desafiliación. A partir de la abundante cantidad de datos generados por la progresiva digitalización de los sistemas de gestión académica y de la enseñanza, se han desarrollado nuevas perspectivas para el análisis de estas problemáticas, permitiendo desarrollar modelos que permiten establecer sistemas de alerta temprana para los individuos en situación de riesgo de abandono. Estos grandes volúmenes de datos son una fuente de conocimiento y pueden ser aplicados en diversos contextos. Los datos recogidos de diversas fuentes requieren de un método apropiado a fin de que sea posible extraer conocimiento de los mismos y auxiliar en la toma de decisiones [3]. Esto es porque los seres humanos tienen una capacidad limitada para extraer conocimiento de datos no tratados [4]. Con el fin de encontrar informaciones útiles dentro de estas grandes colecciones de datos se han utilizado técnicas de *Data Mining*[5].

La aplicación de *Data Mining* en el contexto educativo, conocido como *Educational Data Mining* (EDM), combinado con técnicas de *Learning Analytics* se presenta como un área de investigación muy importante. Esta apunta a encontrar conocimiento en las bases de datos educativas [6], tales como reglas de asociación, clasificación y clusterización [3]. Desde un punto de vista práctico, técnicas de EDM permiten el descubrimiento de conocimiento basado en datos provenientes de los propios estudiantes, con el objetivo de evaluar los sistemas educativos en sus diferentes niveles. Algunos aspectos de calidad de la educación pueden ser evaluados, contribuyendo a una mejora de los aprendizajes. Por este motivo, en los últimos años se ha presenciado un aumento en el número de investigaciones con técnicas de EDM contribuyendo a identificar factores que influyen el desempeño de los alumnos [7].

La habilidad de predecir el rendimiento de los estudiantes es beneficiosa para los sistemas educativos modernos. Sin embargo, esta no es una tarea fácil [8].

Recientemente, técnicas de *Data Mining* se han empleado para proporcionar nuevos *insights* para este problema, ya que muchos factores pueden influir en el rendimiento del estudiante [6]. Las técnicas de predicción ayudan en el momento de realizar intervenciones, con el fin de evitar una posible reprobación, rezago, abandono y desafiliación educativa. Esto es porque generalmente los alumnos presentan señales antes de abandonar formalmente un curso [9].

En este artículo se presentan los avances de un proyecto de investigación y desarrollo orientado a la creación de un sistema de alerta temprana de riesgo académico en los estudiantes de educación primaria y media públicas de Uruguay. Este sistema utilizará la gran cantidad de datos históricos y actuales capturados por los sistemas que se vienen utilizando en el marco de la Administración Nacional de Educación Pública (ANEP). En particular, el artículo da cuenta de los procesos que condujeron a la puesta en marcha de un proyecto de analítica del aprendizaje de alcance nacional, contribuyendo a la comprensión de las diversas dimensiones a tener en cuenta para la implementación de iniciativas similares con impacto en todo el sistema educativo de un país.

El presente artículo está estructurado de la siguiente forma. La sección 2 describe la literatura relacionada mientras la sección 3 presenta los principales aspectos de la metodología adoptada en el proyecto. La sección 4 discute los resultados iniciales obtenidos hasta la fecha, y la sección 5 presenta cuáles son las próximas etapas en el desarrollo.

## 2. Trabajos relacionados

La variedad de las investigaciones actuales destinadas a la predicción de los estudiantes de riesgo difieren en muchos aspectos: las técnicas de modelado que se utilizan para entrenar y probar los modelos de predicción (redes neuronales, regresión logística, análisis discriminante, árboles de decisión, máquinas de vectores soporte), las fuentes de datos utilizadas (diferentes Entornos Virtuales de Aprendizaje, sistemas académicos, pruebas y exámenes, formularios), las cantidades de dichos datos y la combinación de atributos utilizados en diferentes modelos (las interacciones de los estudiantes, el género, la edad, el origen étnico, la experiencia laboral, nivel de estudios, acceso a la tecnología). Una revisión más extensa de la literatura se puede encontrar en [10], pero aquí se mencionan algunas de las obras más recientes de una breve ilustración.

Signals [11] puede ser considerado uno de los más conocidos ejemplos de sistemas de Analítica de Aprendizaje capaz de predecir el riesgo académico. Utilizando informaciones históricas y actuales sobre el comportamiento del estudiante dentro de un curso, *Signals* es capaz de predecir la probabilidad que el estudiante posee de fallar en un curso. Un sistema similar es presentado por [12] para predecir la probabilidad de éxito de un estudiante, pero basado en un modelo que compara el compromiso (*engagement*) actual y anterior del estudiante en un Entorno Virtual de Aprendizaje (EVA) con el compromiso de los demás estudiantes. Hay también muchos otros predictores modernos de riesgo, de ellos el trabajo de [13] puede ser considerado un buen ejemplo representati-

vo. Este sistema utiliza un árbol de clasificación entrenado con datos históricos para obtener reglas para evaluar el riesgo de desafiliación de un estudiante de un programa universitario.

Entre los antecedentes a nivel del Uruguay se encuentran las iniciativas que el Plan Ceibal ha impulsado para la incorporación de esta perspectiva con la creación de una línea e investigación específica. Algunas experiencias se están implementando como el caso del Sistema de Evaluación en línea de ANEP, o el proyecto de implementación de un sistema de analítica del aprendizaje que integre los datos a gran escala de las diferentes bases de los sistemas que integra el Ceibal[14]. Otro ejemplo es el proyecto Descubrimiento de Interacciones que Impactan en el Aprendizaje (DIIA) [15].

### **3. Metodología**

El presente proyecto tiene como objetivo principal la generación, testeo y validación de modelos predictivos de seguimiento de trayectorias educativas para el apoyo a la toma de decisiones en la creación de políticas educativas. Los objetivos específicos se orientan a: 1) obtener e integrar datos de los estudiantes las plataformas digitales de la ANEP; 2) utilizar distintas técnicas de generación de reglas y/o modelos para detección de estudiantes en riesgo académico y observar posibles indicadores relacionados con los factores que inciden en las trayectorias educativas; 3) aportar conocimiento que contribuya a la creación de políticas educativas así como la detección de forma temprana de estudiantes en situación de riesgo académico.

#### **3.1. Etapas**

La creación del sistema de predicción de riesgo académico sigue un proceso metodológico estructurado en seis etapas: 1) recolección de datos históricos sobre los estudiantes; 2) estos datos, juntamente con los desempeños de los estudiantes son utilizados para entrenar uno o más modelos de predicción basados en métodos estadísticos o en minería de datos; 3) el desempeño de la predicción de cada modelo (o de lo mismo modelo con parámetros diferentes) es medido con un grupo de datos todavía no utilizados; 4) basado en las medidas de desempeño, el mejor modelo (o parámetros) son seleccionados y un sistema es construido alrededor de los mismos; 5) a continuación, la información de un estudiante actual es pasada al sistema, y basado en el modelo, ese sistema va a predecir su riesgo académico; 6) la predicción es usualmente presentada por medio de algún tipo de técnica de visualización.

Una atención especial necesita ser dedicada a la etapa 4, la selección del modelo (o parámetros) para el sistema, donde usualmente hay un conflicto entre los intereses del usuario final y los del desarrollador del sistema. El objetivo principal del desarrollador del sistema en esa etapa es obtener un modelo que pueda producir una predicción adecuada para la parte más grande de la población. Dependiendo del tipo de modelo, diferentes métricas de desempeño de predicción

son utilizadas para seleccionar el mejor modelo para los datos testeados. Por otro lado, el principal objetivo del usuario del sistema, es que el sistema prediga para un caso individual, independientemente de cómo el sistema funciona. La selección de un modelo que favorezca un desempeño promedio para toda una población no necesariamente es la mejor solución para determinados grupos en que otros modelos puedan producir resultados más favorables.

Para que diferentes modelos puedan ser aplicados a distintos grupos, el presente proyecto adopta el modelo de agrupamiento adaptativo con varios niveles [16]. Es decir, un modelo de agrupamiento de varios niveles será creado para identificar relaciones jerárquicas y semi jerárquicas entre las diferentes características de los estudiantes. Ejemplos de niveles de características que pueden ser utilizadas son: 1) desempeños pasados de los estudiantes (nivel 1: promedios de las evaluaciones, nivel 2: dominio de diferentes tópicos); 2) informaciones del programa educativo (nivel 1: número de cursos, nivel 2: número de cursos por tópico, nivel 3: códigos de los cursos) y 3) situación económica (nivel 1: gastos totales; nivel 2: porcentaje de sueldo gasto por categoría). La selección de un determinado nivel para cada unas de las características de varios niveles permitirá la creación de diferentes modelos de agrupamiento.

### **3.2. Acceso a las bases de datos**

La implementación de los modelos de predicción están siendo realizados utilizando datos de diferentes plataformas de gestión académica de la ANEP. Este aspecto constituye un factor determinante del éxito de la iniciativa. En proyectos con alcance nacional e impacto en todo el sistema educativo la disponibilidad y autorización para acceder a los datos necesarios para la ejecución de la propuesta es clave, no solamente para poder contar con los datos, sino, fundamentalmente, para su posterior adopción en el marco del sistema educativo.

En nuestro caso, el desarrollo de las estrategias de adopción implicó la realización, durante el segundo semestre de 2017, de varias reuniones de trabajo con integrantes de equipos técnicos y autoridades del Consejo Directivo Central (CODICEN) de la ANEP. A estas reuniones concurrieron integrantes y autoridades de la Dirección Sectorial de Información para la Gestión y la Comunicación, la Dirección Sectorial de Planificación Educativa y la Dirección Sectorial de Integración Educativa de dicho Consejo. Esta estrategia permitió el trabajo conjunto y articulado con los responsables y técnicos de ANEP, así como encaminar la futura adopción de los productos del proyecto. Como resultado de estas instancias de trabajo, y a sugerencia de los representantes de ANEP, se entendió oportuno un cambio de enfoque en los objetivos y resultados del Proyecto. Se decide desarrollar un prototipo de herramienta de seguimiento de trayectorias educativas que proporcione indicadores generales para la toma de decisiones, en vez de centrar la herramienta únicamente en el riesgo académico. De esta forma se ampliará el horizonte permitiendo sumar la detección de factores de éxito y buenas prácticas. La herramienta a desarrollar presentará la información evitando la estigmatización y los posibles sesgos hacia factores negativos. La ANEP también solicitó que el prototipo presente algún tipo de visualización de las reglas

y/o modelos generados permitiendo acompañar de manera visual determinados aspectos de la trayectoria de los estudiantes. El producto que se desarrolle en el marco del proyecto se integrará a las estrategias de seguimiento de trayectorias educativas que se están implementando en el marco de la ANEP.

La ANEP no entregó los datos hasta no acordar una serie de garantías y lograr acuerdos de trabajo conjunto. Esto se debe a que el proyecto es innovador, implica el tratamiento de datos de menores de edad e involucra la transferencia de información sobre las trayectorias educativas de todos los estudiantes del país, por lo que se consideró como un tema sensible. Este proceso de negociación concluyó con una Resolución del CODICEN en la que se establecen las condiciones de acceso a las bases de datos y designa una comisión de seguimiento institucional para el proyecto. Este trabajo conjunto y el modo en que se condujo derivó en una confianza mutua, el establecimiento de instancias de colaboración y el desarrollo de estrategias de adopción desde el propio inicio del proyecto. Luego de la autorización de la transferencia de datos, fueron necesarias varias reuniones de trabajo para instrumentar el acceso efectivo a las mismas. Las reuniones se realizaron con autoridades y miembros de equipos técnicos de ANEP-CODICEN, de las Direcciones Sectoriales de Integración Educativa, Planificación Educativa e Información para la Gestión y la Comunicación; así como con responsables del área de Tecnologías de la Información del Plan Ceibal. Finalmente, solo se accedió a las bases de datos administradas por ANEP, no así las administradas por el Plan Ceibal.

Previo a la entrega se detecta la existencia de gran cantidad de datos no integrados y dispersos en los diferentes subsistemas de la ANEP. Debido al volumen de datos, la falta de integración, la diversidad de programas y planes de estudio, así como la cantidad limitada de tiempo con que cuenta el proyecto, se decide seleccionar en esta etapa únicamente las bases de CES (Consejo de Educación Secundaria) y CEIP (Consejo de Educación Inicial y Primaria), descartando, en esta oportunidad, el trabajo con las bases de CETP (Consejo de Educación Técnico Profesional).

### **3.3. Tratamiento de datos personales**

Se considera dato personal la información de cualquier tipo referida a personas físicas o jurídicas determinadas o determinables, a modo enunciativo, cualquier información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo que refiera a ellas directa o indirectamente, conforme con lo dispuesto en el artículo 4 de la Ley N° 18.331 y artículos 1 y 4 del Decreto N° 414/009. La Udelar como institución responsable del proyecto y el equipo de investigadores se obligan a tratar los datos personales a los que tuvieren acceso, de conformidad con la Ley N° 18.331, de 11 de agosto de 2008 y Decreto N° 414/2009, de 31 de agosto de 2009, únicamente para la realización del Proyecto, no pudiendo utilizarlos para otra finalidad, ni en beneficio propio ya sea gratuito u oneroso, ni cederlos, comunicarlos o transferirlos a terceros, salvo previa autorización por escrito del CODICEN. La Institución y los Investigadores son

responsables del tratamiento (artículo 4 Ley N° 18.331) de los datos proporcionados por la ANEP o accedidos directamente en el marco de del proyecto, debiendo cumplir con los requerimientos previstos en la Ley N° 18.331, .

En este sentido, los Investigadores se obligan a adoptar las medidas de seguridad necesarias para garantizar la seguridad y confidencialidad de los datos personales y evitar su adulteración, pérdida, consulta o tratamiento no autorizado, así como detectar desviaciones de información.

Los datos proporcionados por ANEP se deben alojar en servidores dentro del territorio uruguayo, a los que únicamente tendrán acceso los investigadores del Proyecto (inclusive los investigadores extranjeros a los únicos efectos de la realización del Proyecto). Al término del Proyecto, la Institución y los Investigadores se obligan a suprimir de todos sus sistemas y archivos físicos y lógicos, sean propios o contratados a terceros, los datos personales accedidos, obtenidos o tratados en virtud de este contrato, así como los metadatos asociados, en caso de corresponder. En virtud de lo anterior, fue necesaria la instalación de un servidor para acceder al banco de datos y la realización de los experimentos de manera remota. Inicialmente, el servidor disponibilizado por la Universidad de la República Uruguay (UDELAR) fue configurado, optándose por colocarlo en una máquina virtual, con sistema Ubuntu 16.04 LAMP (Linux, Apache, MySQL,PHP). Además, se instaló PostgreSQL 9.5.11 y configurado para acceso externo. Se utilizó el software Pentaho Data Integration CE 8.0 para la extracción de las bases del servidor de la ANEP.

Una vez que la base de datos (anonimizada y disociada) fue puesta a disposición, se dio inicio al proceso de extracción de los datos. Estos se movieron a desde una base ubicada en el servidor de la ANEP para servidor de la Udelar, siendo almacenados en una base PostgreSQL. La extracción de estos fue hecha a través del desarrollo de ETLs con el Pentaho. Los datos que se está trabajando hasta la fecha se refieren a las bases de bases de CES (Consejo de Educación Secundaria).

#### 4. Estado técnico actual de la iniciativa

Inicialmente se intentó la generación de reglas para el análisis de las trayectorias utilizando los datos de ANEP y **algoritmos de asociación**. Para eso, inicialmente fue utilizada el lenguaje de programación Python con la IDE Spyder<sup>6</sup> y la librería apyori, que contiene reglas de asociación para el algoritmo Apriori. Sin embargo, incluso utilizando solamente un subset (10%) del dataset original (de aproximadamente 3 millones de registros), el *script* tomaba mucho tiempo para la carga y ejecución, y generaba un conjunto de reglas demasiado grande para su análisis (en formato poco legible y con poca documentación disponible para su comprensión). Teniendo eso en cuenta, el equipo de desarrollo comenzó a utilizar el lenguaje R 3.4.3 <sup>7</sup>.

<sup>6</sup> <https://pythonhosted.org/spyder/>

<sup>7</sup> <https://www.r-project.org/>

Previo a la generación de las reglas de asociación, los datos fueron preprocesados utilizando diferentes técnicas, como: generación de atributos derivados, binarización de atributos, y categorización de variables numéricas. Esa etapa de minería no generó conocimiento que pudiera ser aprovechado en el prototipo del sistema de manera que se empezó a trabajar con técnicas de **clusterización**.

En esa etapa fue generada una tabla conteniendo 17 atributos, siendo que 16 de ellos fueron utilizados como entrada para la clusterización y la variable *fallo final* del estudiante fue utilizada para la evaluación de los *clusters*. El cuadro 1 muestra las variables utilizadas juntamente con sus respectivas descripciones. Para la clusterización se utilizó el algoritmo *k-means* implementado en *Java* e utilizando la API de la herramienta *Weka*. Fue utilizada una cantidad de *clusters*  $k = 3$ , que son los valores referentes a la cantidad de resultados posibles para el *fallo final* del estudiante (*Promovido*, *Repite por Rendimiento* e *Repite por Inasistencia*).

**Cuadro 1.** Variables utilizadas en la clusterización

Variable	Descripción
cant_materias	Cantidad de materias que el estudiante cursó
edad	Edad del estudiante
pmat_mx_ry	Porcentaje de materias (del total que estaba cursando) que el estudiante obtuvo la calificación general (promedio) menor que $X$ en la reunión $Y$ , donde $X$ varía de 3 hasta 7, y $Y$ varía de 1 hasta 2
inasinjust_rx	Cantidad de faltas injustificadas hasta la reunión $X$
inasjust_rx	Cantidad de faltas justificadas hasta la reunión $X$

Algunas de los patrones identificados por medio de la clusterización para la identificación de estudiantes en situación de riesgo académico son:

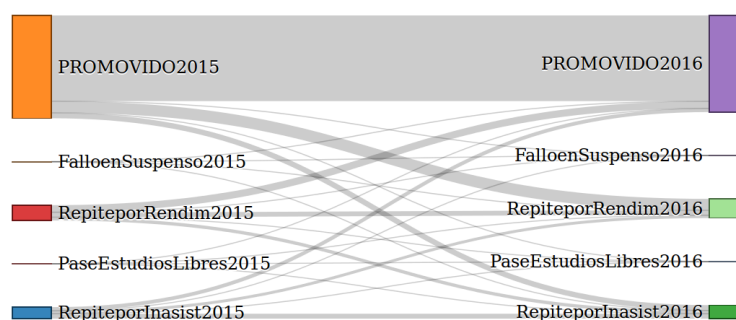
1. Estudiantes con cinco faltas injustificadas
2. 50% de las notas medias bajo cinco en el momento de la primera reunión
3. Cinco o más faltas injustificadas en el momento de la primera reunión y el doble de ese valor en el momento de la segunda reunión
4. Poseer edad mayor que la edad media en determinado grado (con fallos anteriores) y con 50% de las notas más bajas que seis en el momento de la primera reunión

## 5. Consideraciones finales

En las próximas etapas del proyecto se elaborará un prototipo de sistema de seguimiento de trayectorias educativas de estudiantes de educación primaria y media pública de Uruguay. La herramienta (prototipo) servirá de apoyo a los tomadores de decisión en la creación de políticas educativas.



La herramienta combinará visualizaciones estadísticas de las variables asociadas a la trayectoria de los estudiantes a nivel individual, regional y global, como es presentado en la figura 1. En esa figura, es presentado el flujo de los estudiantes entre los años 2015 y 2016 considerando las sus situaciones: Fallo en suspenso (en examen), repitió el año por rendimiento, promovido, repitió el año por inasistencias, o pase estudios libres (estudiantes dispensados por alguna razón). Además, los perfiles medianos encontrados por medio de los experimentos de minería de datos anteriormente mencionados, serán añadidos al prototipo con el objetivo de ayudar en la identificación temprana de los factores de riesgo y éxito en la trayectoria estudiantil.



**Figura 1.** Trayectoria de la promoción de los estudiantes entre los años 2015 y 2016

En los próximos pasos, nuevas variables derivadas serán incluidas y sus efectos en la tentativa de identificación de patrones serán evaluados. Otra posibilidad de experimentación futura será basada en la distinción de materias que llevaron los estudiantes a una situación de examen (y sus respectivos resultados finales) de manera a auxiliar en la creación de estrategias de apoyo especializadas en acuerdo con los resultados. Se espera también construir una noción más amplia de la trayectoria estudiantil por medio de la inclusión de los datos de la enseñanza primaria, es decir, identificar relaciones entre el perfil medio de un determinado estudiante y sus características de aprendizaje en la enseñanza media. Con la intersección entre las dos bases de datos, se espera contribuir a un mayor entendimiento de la evolución del estudiante uruguayo dentro del sistema de enseñanza pública en la transición hasta los primeros años de la enseñanza media.

## Agradecimientos

Ese trabajo fue financiado por el Fondo Sectorial "*Inclusión Digital: Educación con Nuevos Horizontes 2016*" de la *Agencia Nacional de Investigación e Innovación* (ANII) de Uruguay, por medio del proyecto "*Modelos de predicción para la determinación de riesgo académico*" (código FSED\_2\_2016\_1\_130897).

## Referencias

1. INEEd: Informe sobre el estado de la educación en Uruguay 2015-2016: Síntesis y desafíos. Imprenta Blueprint (2017)
2. Pereda, T.F.C.: Explicar/intervenir sobre la desafiliación educativa en la enseñanza media. EL URUGUAY DESDE LA SOCIOLOGÍA VIII 165
3. Baradwaj, B.K., Pal, S.: Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417 (2012)
4. Cortez, P., Silva, A.M.G.: Using data mining to predict secondary school student performance. (2008)
5. Mannila, H.: Data mining: machine learning, statistics, and databases. In: Scientific and Statistical Database Systems, 1996. Proceedings., Eighth International Conference on, IEEE (1996) 2–9
6. Daud, A., Aljohani, N.R., Abbasi, R.A., Lytras, M.D., Abbas, F., Alowibdi, J.S.: Predicting student performance using advanced learning analytics. In: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee (2017) 415–421
7. Marquez-Vera, C., Romero, C., Ventura, S.: Predicting school failure using data mining. In: EDM, ERIC (2011) 271–276
8. Ramaswami, M., Bhaskaran, R.: A chaid based performance prediction model in educational data mining. arXiv preprint arXiv:1002.1144 (2010)
9. Barber, R., Sharkey, M.: Course correction: Using analytics to predict course success. In: Proceedings of the 2nd international conference on learning analytics and knowledge, ACM (2012) 259–262
10. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **40**(6) (2010) 601–618
11. Arnold, K.E., Pistilli, M.D.: Course signals at purdue: Using learning analytics to increase student success. In: Proceedings of the 2nd international conference on learning analytics and knowledge, ACM (2012) 267–270
12. Wolff, A., Zdrahal, Z., Nikolov, A., Pantucek, M.: Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Proceedings of the third international conference on learning analytics and knowledge, ACM (2013) 145–149
13. Dekker, G.W., Pechenizkiy, M., Vleeshouwers, J.M.: Predicting students drop out: A case study. International Working Group on Educational Data Mining (2009)
14. Cobo, C.: Analíticas de aprendizaje - an overview of educational software and analytics (March 2016)
15. Techera, B., Rodríguez, C., Ferrero, T., Motz, R.: Learning analytics for teachers. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. Volume 6. (2017) 744
16. Ochoa, X.: Adaptive multilevel clustering model for the prediction of academic risk. In: Learning Objects and Technology (LACLO), Latin American Conference on, IEEE (2016) 1–8