

112 UN ANÁLISIS DESCRIPTIVO DEL AUSENTISMO EN UN CURSO DE MATEMÁTICA PARA ECONOMISTAS UTILIZANDO HERRAMIENTAS DE LEARNING ANALYTICS

Nicolás Harari, Marco Spinelli, Javier Garcia Fronti

Facultad de Ciencias Económicas, Universidad de Buenos Aires

ndharari@gmail.com, marco.spinelli@economicas.uba.ar, javier.garciafronti@economicas.uba.ar

Especialidad: Educación Matemática

Palabras Claves: Learning Analytics, Ausentismo, Visualización de Datos

Resumen

El análisis de desgaste relacionado al abandono estudiantil durante la cursada de una materia es una técnica de *Learning Analytics* muy utilizada. Este trabajo muestra una descripción de este fenómeno en un curso de Matemática para Economistas en la Universidad de Buenos Aires, identificando perfiles de estudiantes en riesgo. Lo realizado permite entender y anticipar causas de abandono dentro de un proceso de enseñanza-aprendizaje, de forma de eficientizar la intervención docente y prevenir el ausentismo estudiantil.

Los resultados aquí presentados representan una primera aproximación de un proyecto general de *Learning Analytics* aplicado en la cátedra Javier García Fronti de Matemática para Economistas de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires y enmarcado en el proyecto UBATIC 2018-2019 “Desde el conocimiento matemático hacia la adquisición de técnicas cuantitativas aplicadas a las Ciencias Económicas” del Departamento de Matemática de la Facultad de Ciencias Económicas de la UBA.

INTRODUCCIÓN

Las instituciones educativas, especialmente las de altos estudios, generan y disponen de grandes cantidades de datos respecto a sus alumnos -y docentes- producto del propio funcionamiento de la institución. Algunas de las variables a las que se tienen acceso son resultados de diversos exámenes, género, edad. No obstante, su recopilación suele ser fragmentaria y su uso para facilitar la práctica docente -desde el diagnóstico de dificultades o problemáticas en el proceso de aprendizaje, a la evaluación de resultados de las distintas alternativas pedagógicas propuestas- no suele ser sistemático. En circunstancias habituales, esta información es raramente utilizada a su máximo potencial: usualmente los docentes carecen de la infraestructura necesaria para almacenar, organizar y analizar los datos obtenidos de los alumnos.

Sin embargo, durante los últimos años han ocurrido importantes avances en las técnicas de almacenamientos y análisis de datos que luego se han enfocado a distintas y diversas disciplinas. Comúnmente conocido como *Data Mining* la técnica para “extraer” conocimiento de grandes cantidades de datos en con el objetivo de lograr una toma de decisiones informada se conoce como *Knowledge Discovery in Databases* (KDD) (Pandey & Pal, 2011)

Es producto de las dificultades antes mencionadas -y a la luz del crecimiento y disponibilidad de estas herramientas- que surgen iniciativas de KDD especialmente aplicadas al ámbito educativo, conocidas por el nombre de *Educational Data Mining* (Kumar Baradwaj & Pal, 2011), interesadas en desarrollar métodos de extracción de conocimiento de los datos educativos.

Simultáneamente surgen en la literatura experiencias de “*Academic Analytics*” o “*Learning Analytics*” (Baepler & Murdoch, 2010) que buscan revolucionar la tarea docente al sumar al análisis de datos el desarrollo y aplicación de sistemas de gestión de datos con el objetivo de mejorar la experiencia estudiantil y el proceso de aprendizaje. Dentro de los grandes áreas de acción de esta metodología se encuentra el *Attrition Risk Detection* (Sin & Muthu, 2015) que intenta minimizar el número de alumnos que abandonan la materia durante su cursada.

Es en este espíritu que el presente trabajo se propone presentar los resultados del análisis de datos en el contexto de la experiencia de *Learning Analytics* en un curso de Matemática para Economistas de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires. La siguiente propuesta se inspira y enmarca en el proyecto UBATIC 2018-2019 “Desde el conocimiento matemático hacia la adquisición de técnicas cuantitativas aplicadas a las Ciencias Económicas” del Departamento de Matemática de la Facultad de Ciencias Económicas de la UBA, como también de experiencias previas de uso de datos en el contexto de educación universitaria, entre las cuales cabe mencionar los esfuerzos de Bianco, Fraquelli y Gache tanto en el desarrollo de sistemas de autoevaluación de contenidos previos (Bianco, Fraquelli, & Gache, 2018), como en el análisis de la experiencia (Bianco, Fraquelli, & Gache, 2019).

MARCO TEÓRICO

Según uno de los equipos pioneros en el área, Kumar Baradwaj, B y Pal S (2011) se hace referencia a *Educational Data Mining* (EDM) como la aplicación de técnicas del paradigma de *Knowledge Discovery in Databases* (KDD) al ámbito educativo. Este paradigma, también conocido como minería de datos o *Data Mining* busca extraer conocimiento de grandes colecciones y volúmenes de datos, desde reglas de asociación, clasificación y agrupamiento. Sin embargo, otros autores refieren al proceso de análisis de datos con el término general de *Analytics* (análisis) acompañado de distintos términos específicos. En su trabajo de 2012, van Barneveld, Arnold, y Campbell (van Barneveld, Arnold, & Campbell, 2012) distinguen al menos 7 terminologías distintas con niveles de focos diferenciados -Institución, Departamento, Instructor, Alumno-, cuya aplicación se extiende a áreas muy disímiles, desde el rendimiento estudiantil hasta la gestión de negocios de la educación superior. En concordancia con nuestro rol como docentes de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires y desde una perspectiva de universidad pública, tomaremos de estas caracterizaciones los términos *Analytics*, *Learning Analytics*. El primero, refiere a la concepción más genera de la “toma de decisiones basada en el uso de datos”, mientras que el segundo tiene como foco de análisis el estudiante, su recorrido académico y el uso de las herramientas de datos para mejorar su experiencia educativa (van Barneveld et al., 2012). De esta forma, tomamos distancia de conceptualizaciones como *Bussines Analytics* que según los autores focalizan en la maximización de beneficios por parte de las instituciones educativas, desde posibilidades empresariales hasta la maximización de posibles donaciones de exalumnos. Parte fundamental de nuestro rol como docentes de Universidad Pública es un esfuerzo constante por la democratización del conocimiento y la igualdad de oportunidades, especialmente a la luz de los principios de la Reforma Universitaria, cuyo objetivo puede ser interpretado como democratizar sociedades profundamente desiguales (Arocena, 2018). Es desde esta perspectiva que consideramos que al ausentismo -que puede luego resultar en el abandono de la cursada en general- como una problemática de interés.

Según el estudio bibliográfico realizado por Sin y Muthu (2015) se reconocen el análisis del riesgo de desgaste y la visualización de datos como dos de las principales áreas donde es posible aplicar técnicas de *Big Data* en contextos de *Learning Analytics*.

En el caso del análisis de riesgo de desgaste -*Attrition Risk Detection*- se busca lograr identificar los grupos de riesgo de ausentismo para, en un futuro, aplicar políticas preventivas en el comienzo de cada cuatrimestre. De esta forma se espera retener la mayor cantidad de alumnos dentro del proceso educativo. En su trabajo, los autores marcan este

tópico como uno de los menos explorados en la literatura, lo que presenta una oportunidad de análisis en el contexto propio de la Universidad de Buenos Aires.

Por otro lado, la visualización de datos *-Data Visualization-* refiere a un conjunto de técnicas que, frente al creciente volumen de datos e información disponibles, logran resumir y exponer la información para facilitar la detección de tendencias y relaciones presentes en los datos.

LOS ESTUDIANTES -Y SUS RECORRIDOS- COMO UNIDADES DE ANÁLISIS

En concordancia con lo expuesto anteriormente respecto del rol de la Universidad Pública y de las posibilidades de aplicación de herramientas de *Learning Analytics*, tomamos a los estudiantes -y sus recorridos en la materia- como unidades de análisis. De esta forma, los consideramos como sujetos interpelados por una multiplicidad de variables que surgen tanto de sus condiciones identitarias como de la relación con la materia y sus docentes. Su rendimiento y crecimiento dentro de la asignatura refleja y cuestiona la efectividad de la labor docente y su perfeccionamiento. Reconocemos que los aspectos presentados a continuación no logran reflejar en su totalidad la individualidad de los estudiantes, su relación con la asignatura y el aprendizaje, pero creemos que pueden resultar útiles para identificar ciertos grupos de riesgo.

Los datos utilizados en el contexto del siguiente trabajo fueron recopilados durante el segundo cuatrimestre de 2018 y el primer cuatrimestre de 2019 en un curso de Matemática para Economistas de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires orientado a alumnos de las carreras de Economía, Actuario en Administración y Actuario en Economía y se exponen a continuación:

Número de registro:

Este es un identificador numérico administrativo que la facultad le otorga a los alumnos al completar el Primer Tramo (primer año) de la carrera. Como su valor es creciente conforme aumenta el número de inscriptos, la diferencia entre el valor del número de registro de los alumnos y sus compañeros muestra, de alguna forma, una medida de "retraso" comparativos. (Dato disponible para períodos 2018-2019).

Género:

(Dato disponible para períodos 2018-2019).

Rendimiento en los exámenes:

En la materia y según la normativa vigente se realizan dos exámenes parciales por cuatrimestre. Por decisión de la cátedra estos exámenes se encuentran estandarizados, divididos en cinco unidades temáticas individuales con cuatro ejercicios cada una con dificultad creciente. Esta organización posibilita un seguimiento ejercicio a ejercicio al realizar el análisis. Se recopila para cada punto si se respondió de forma correcta o incorrecta. (Dato disponible para períodos 2018-2019).

Participaciones en el Campus on-line de la asignatura:

La asignatura cuenta con un Campus on-line proporcionado por la Facultad de Ciencias Económicas mediante el sistema Sharepoint de Microsoft. Dentro de las funcionalidades se encuentra un foro de consultas que donde los alumnos pueden consultar dudas entre ellos y con el equipo docente. Se recopilaron el número de participaciones que los alumnos realizaron antes de cada período de examen (Dato disponible sólo para el primer cuatrimestre de 2019).

Asistencias a clases prácticas

Conforme al vigente plan de estudios de la materia la asignatura cuenta con clases teórico-prácticas cuya distribución es a discreción del docente a cargo. Es costumbre del curso en el que se realizó la experiencia la realización de una clase práctica por semana, relacionada a los contenidos de la unidad correspondiente. Al recolectar información sobre la asistencia a las clases prácticas, es posible recolectar información respecto al desempeño de cada unidad. (Dato parcialmente disponible para el primer cuatrimestre de 2019).

Resultados del test inicial

El presente trabajo continúa la tarea realizada por Bianco, Gache y Fraquelli (Bianco et al., 2019) respecto al uso de datos en la gestión matemática universitaria. En aquel trabajo, las autoras presentan un modelo de autoevaluación realizado con el objetivo de facilitarle a los alumnos la recuperación de aquellos contenidos aprendidos pero olvidados, conocido como conocimiento frágil. La implementación de este test inicial se realizó durante los cuatrimestres en los cuales el presente trabajo recopiló información. (Dato disponible para periodos 2018-2019).

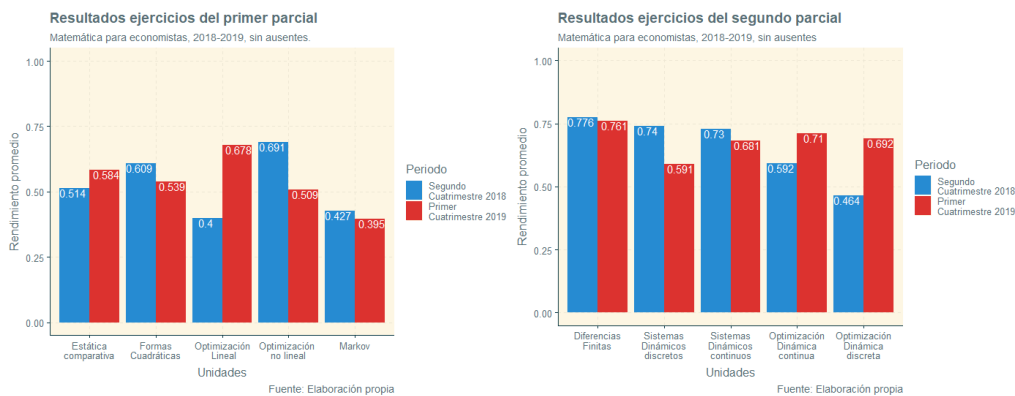
Nivel

En el caso del segundo cuatrimestre de 2018, las autoras (Bianco et al., 2019) dividen a los alumnos en tres grupos (Alto, Medio, Bajo) según las condiciones de aprobación de las materias anteriores (Álgebra y Análisis Matemático II) para identificar aquellos alumnos que podrían necesitar un mayor nivel de apoyo. Esta información es utilizada como a priori en nuestra investigación.

RESULTADOS

En esta sección se presentan resultados preliminares del análisis de datos realizado respecto a la información recopilada. Mientras que la gestión de datos se realizó utilizando el software Excel, el análisis posterior se realizó íntegramente en los lenguajes de programación R y Python.

Como se mencionó anteriormente, los exámenes parciales de la materia se encuentran divididos en cinco unidades temáticas. Como una primera aproximación, es posible armar el siguiente gráfico de barras mostrando el rendimiento promedio por ejercicio para cada una de ellas, donde se presentan en el orden que aparecen en sus respectivos exámenes.



Resultan interesantes ciertos aspectos de las representaciones anteriores. En primer lugar, la aprobación en el primer parcial suele ser menor y con cierta tendencia decreciente (más presente en el primer cuatrimestre de 2019) a lo largo de las unidades, donde se nota un rendimiento promedio menor en la unidad sobre Cadenas de Markov. En segundo lugar, es posible apreciar que el rendimiento general es superior durante los segundos parciales, encontrando una tendencia decreciente sólo en el caso del segundo cuatrimestre de 2018.

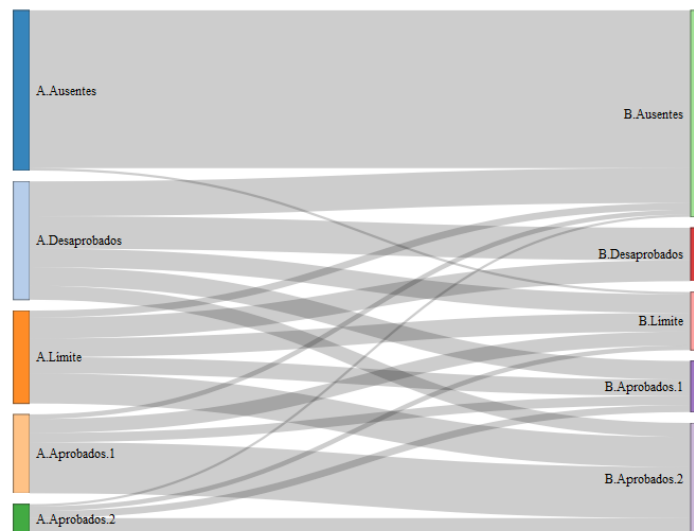
A continuación, resulta relevante comprender como es el recorrido del rendimiento de los alumnos a través de los exámenes. Para eso, se clasifica los resultados de cada evaluación mediante la escala expuesta a continuación. El caso de los alumnos catalogados como "límite" refiere a aquellos desaprobados que se encuentran cerca de una posible aprobación. Se diferencia estos estudiantes del resto de los desaprobados para luego intentar apreciar si estos lograron cambiar su condición respecto al segundo examen parcial.

Grupos	Ausentes	Desaprobados.1	Desaprobados.2	Límite	Aprobados.1	Aprobados.2
Límites	0	[1:5]	[6:9]	[10:12]	[13:15]	[15:20]

En la siguiente tabla se puede encontrar el número de alumnos presente en cada una de las clasificaciones antes mencionadas. El caso del primer parcial, para ambos periodos analizados se muestra a continuación, seguido por los segundos exámenes.

Periodo	Examen	Ausentes	Desaprobados.1	Límite	Aprobados.1	Aprobados.2
2018	Primero	32	19	19	13	4
2018	Segundo	38	7	15	12	15
2019	Primero	37	32	21	21	9
2019	Segundo	51	16	10	10	33

Aquí se puede apreciar como el número de estudiantes que logran aprobar el examen es mayor en el caso del segundo parcial respecto del primero para el caso de ambos cuatrimestres. Es posible construir un Diagrama de Sankey que muestre el cambio de condición de un examen al siguiente:



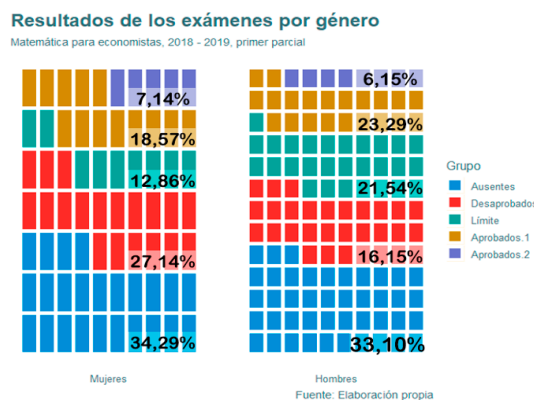
En primer lugar, es posible apreciar que el gran número de ausentes en el primer parcial abandonan definitivamente la asignatura y no se vuelven a presentar en la segunda instancia de examinación: en ambos cuatrimestres sólo se contabiliza un caso de lo contrario, un alumno que aun estando ausente el primer parcial haya rendido el segundo (pasando al grupo “límite”). Por otro lado, como era de esperarse, el número de ausentes aumenta en el segundo parcial alimentado especialmente por desaprobados. Sin embargo, el grupo de alumnos que abandonan la materia luego de un resultado negativo es minoritario, 17 de 51 desaprobados (33%), mientras sólo 3 de 40 (7,5%) de aquellos “límite”. Continuando por estos últimos, se puede ver que el 50% (20 de 40) mejora su rendimiento para el segundo parcial, caso similar se encuentra en los aprobados de nivel bajo, que en la gran mayoría aumentan su rendimiento. En forma general, se vuelve a apreciar que el rendimiento general mejora en el segundo examen a lo largo de las categorías, mostrando que el crecimiento apreciado en gráficos anteriores no es producto del cambio en la composición de los alumnos (ausentes). Es en base a estos resultados que se considera importante remarcar la importancia de la integración *antes del primer parcial* para evitar el abandono definitivo de la materia. Parte fundamental para lograr este cometido es identificar aquellas características que puedan indicar pertenencia a grupos de riesgo. Las consideradas a continuación son género, registro y nivel.

GÉNERO

Comenzamos por ver el número de estudiantes en la materia en cada uno de los cuatrimestres, acompañado del número de mujeres para cada periodo (la definición de género fue hecha *ex-post* mediante los datos de los alumnos, por este motivo hay estudiantes no clasificados):

Periodo	Estudiantes	Mujer	N/A
2018	87	32	7
2019	120	38	-

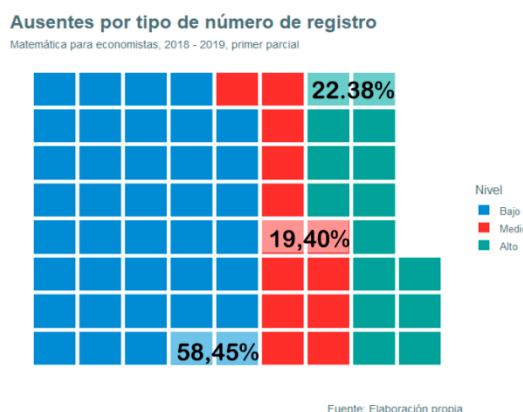
Se puede apreciar que el número de mujeres es menor que el de hombres en ambos cuatrimestres analizados, un efecto que se encuentra más pronunciado en el primer cuatrimestre de 2019. Continuando la clasificación por rendimiento considerada anteriormente, se grafica su proporción en ambos géneros considerando los resultados del primer parcial en ambas instancias de evaluación.



Como se ve en el gráfico, los rendimientos discriminados por género -especialmente al hacer referencia a ausentes- no presentan cambios significativos entre ellos ni respecto a la estadística general (69 ausentes considerando los 209 alumnos en los dos cursos representan un 33,33%)

REGISTRO

El número de registro de los alumnos, como antes mencionado, puede considerarse como una forma de medir la rapidez de cursada con respecto a sus compañeros; dado que los registros se obtienen cuando el alumno entra al segundo tramo de la carrera, un registro cercano al más alto de su camada puede interpretarse como una realización “en tiempo y forma” de la carrera. Sin embargo, es necesario aclarar que esta clasificación no sólo considera a aquellos alumnos “rezagados”, sino también a aquellos adultos mayores que, ya anotados a su primera carrera en su juventud, cursan su segunda carrera.



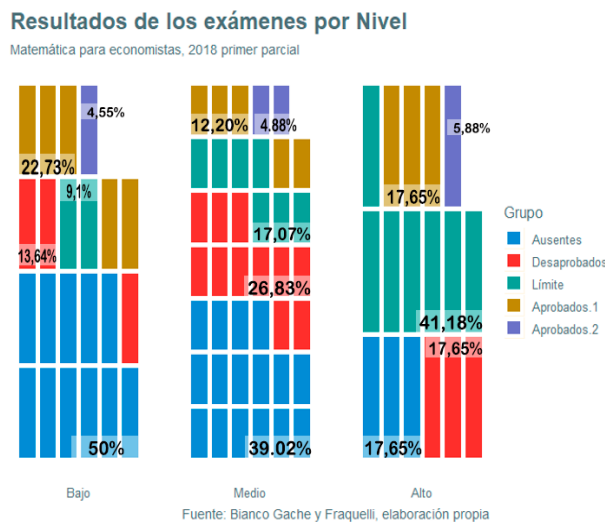
Para realizar el gráfico que se muestra a continuación se dividió a los alumnos en tres partes iguales por periodo dependiendo de su número de registro. De esta forma, los estudiantes clasificados como “Alto” pertenecen al tercio con número de registro elevado con relación a sus compañeros. De forma equivalente aquellos marcados como “Bajo” o “Medio”. Consecuentemente, se muestra los porcentajes de cada una de estas categorías en los ausentes durante ambos cuatrimestres. Se puede notar que la mayor cantidad de ausentes (58,45%) proviene de aquel grupo con número de registro relativamente mayor al de sus compañeros.

NIVEL

Utilizando los datos recopilados por Bianco, Gache y Fraquelli (Bianco et al., 2019), consideramos los niveles predefinidos para cada uno de los estudiantes en el caso del segundo cuatrimestre de 2018. La cantidad de estudiantes en cada una de las categorías se puede encontrar en la siguiente tabla

Periodo	Alto	Medio	Bajo
2018	17	41	22

Donde se encuentra un predominio de la clasificación de tipo “Medio”.

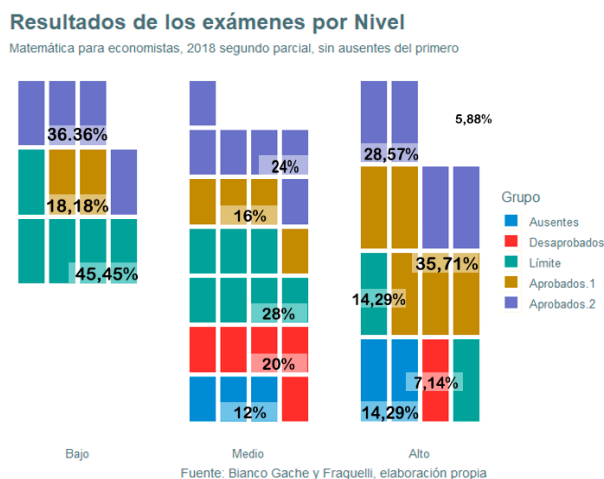


Continuando con los grupos antes expuestos, el siguiente gráfico discrimina por los niveles mencionados. De esta forma, se puede ver que el 50% de los alumnos de nivel bajo no se presentan al primer parcial (lo que significa, en la mayoría de los casos, abandonar la materia). Sin embargo, este grupo tiene el mayor porcentaje de aprobados con 27,28% y el menor de desaprobados (22,74%, considerando los “límites”). Esto pareciera indicar que, de presentarse al examen, los alumnos preclasificados como bajos tuvieron un rendimiento en promedio superior que sus compañeros. Con respecto al grupo de aquellos alumnos con niveles “altos”, su porcentaje de aprobación resulta comparativamente elevado (23,53%), mientras que la cantidad de estudiantes “límites” resulta elevada (41,18%).

Un resultado interesante se puede encontrar si se replica esta clasificación para el segundo parcial del segundo cuatrimestre de 2018, sin considerar los ausentes del primero (que en su totalidad no se presentan al segundo). A continuación, se presenta en la siguiente tabla los números de estudiantes en cada uno de los niveles

Periodo	Alto	Bajo	Medio
2018	14	13	25

Se puede apreciar de esta forma en el gráfico que la totalidad de los alumnos de nivel bajo se presentan al segundo examen. Esto se encuentra emparejado a su vez con una mejora del rendimiento en los tres niveles respecto al primer parcial. En el caso de los alumnos preclasificados “Bajos” el 54,54% de los estudiantes aprueban el examen, contabilizando ambos niveles mientras que el restante obtiene una calificación “Límite”. Por otro lado, el grupo con mayor nivel de ausentismo en esta etapa son los alumnos de nivel “Alto” con un 14,29% cercano al 12% de sus compañeros del grupo “Medio”.



CONCLUSIONES Y TRABAJO A FUTURO

En este trabajo se expusieron los resultados preliminares de una experiencia de implementación de *Learning Analytics* en un curso de Matemática para Economistas de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires. Se mostraron los resultados generales del curso, en cada uno de los ejercicios y luego un análisis exploratorio sobre la relación de determinadas variables, -género, número de registro y nivel según rendimiento en materias anteriores- con el ausentismo en la materia.

Uno de los desafíos a futuro es analizar la relación entre el ausentismo y distintas variables asociadas a los alumnos de forma más precisa, para luego realizar un modelo Probit que logre predecir la probabilidad individual de ausentismo de alumnos en la materia. Esto generaría una valiosa retroalimentación y lograría mejorar, en última instancia, la experiencia del alumno y la práctica docente.

REFERENCIAS BIBLIOGRÁFICAS

- Arocena, R. (2018). DE LOS PRINCIPIOS DE LA REFORMA DE CÓRDOBA A LA UNIVERSIDAD PARA EL DESARROLLO DE AMÉRICA LATINA EN EL SIGLO XXI. 1, 13.
- Baepler, P., & Murdoch, C. (2010). Academic Analytics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*, 4(2). <https://doi.org/10.20429/ijstl.2010.040217>
- Bianco, M. J., Fraquelli, A., & Gache, A. (2018). Mediación tecnológica para la recuperación del conocimiento frágil. En Libro de Actas XXXIII Jornadas de Docentes de Matemática de Facultades de Ciencias Económicas y Afines (p. pp 55-62).

Bianco, M. J., Fraquelli, A., & Gache, A. (2019). MEDIACIÓN TECNOLÓGICA PARA UN APRENDIZAJE CRÍTICO. ANÁLISIS DE LOS RESULTADOS EN MATEMÁTICA PARA ECONOMISTAS. 1, 19.

Kumar Baradwaj, B., & Pal, S. (2011). Mining Educational Data to Analyze Students Performance. International Journal of Advanced Computer Science and Applications, 2(6). <https://doi.org/10.14569/IJACSA.2011.020609>

Pandey, U. K., & Pal, S. (2011). Data Mining: A prediction of performer or underperformer using classification. arXiv:1104.4163 [cs]. Recuperado de <http://arxiv.org/abs/1104.4163>

Sin, K., & Muthu, L. (2015). APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS – A LITERATURE REVIEW . ICTACT Journal on Soft Computing, 05(04), 1035-1049. <https://doi.org/10.21917/ijsc.2015.0145>

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in Higher Education: Establishing a Common Language. 12.