

Inteligencia empresarial adaptativa y su aplicación predictiva sobre patrones taxonómicos de deserción

Adaptive business intelligence and its Predictive application on Taxonomic patterns of desertion

Recibido: 15 Ago 2017 – Revisado: 30 Sep 2017

Aceptado: 30 Oct 2017 – Publicado: 30 Dic 2017

Marcelo Fabio Roldán

*Departamento Académico de Ciencias Exactas, Físicas y Naturales.
Universidad Nacional de La Rioja, La Rioja, Argentina.
marcelofabio01@yahoo.com.ar*



Resumen: Las variables que surgen de las bases de datos de una Universidad Nacional de Argentina, favorecen el análisis y aplicación de una metodología de Inteligencia de Negocio Adaptativa, mediante la cual, aplicando los filtros adecuados de manera metodológica, como se expone en el presente trabajo, permitirán encontrar los alumnos en riesgo de desertar y realizar un trabajo preventivo sobre ellos. Existen datos socio-psicométricos, económicos y culturales, además de elementos físicos que han de ser tomados en consideración relacionándolos con la terminalidad de estudios de los sujetos evaluados. Surge a partir de allí, un modelo predictivo de aprendizaje que clasifica las causalidades de deserción. Finalmente, se puede explotar este conocimiento para desarrollar la aplicación en cualquier lenguaje de programación, de manera que aproveche tal conocimiento, con uso en nuevos sujetos y sus datos. En virtud de ello, se exponen las principales ventajas obtenidas de las capacidades predictivas sobre los datos del caso.

Palabras clave: minería de datos; inteligencia de negocios; predictivo; adaptativo.

Abstract: The variables that arise from the databases of a National University of Argentina, favor the analysis and application of a methodology of Adaptive Business Intelligence, by which, applying the appropriate filters in a methodological way, as it is exposed in the present work, will allow Finding students at risk of defecting and doing preventive work on them. There are socio-psychometric, economic and cultural data, in addition to physical elements that have to be taken into consideration by relating them to the finality of studies of the subjects evaluated. Emerges from there, a predictive learning model that classifies the causalities of desertion. Finally, this knowledge can be exploited to develop the application in any programming language, so that it takes advantage of such knowledge, with use in new subjects and their data. By virtue of this, the main advantages obtained from the predictive capabilities on the data of the case are exposed.

Key words data mining; business intelligence; predictive; adaptive.



1. INTRODUCCIÓN

En los últimos años se han presentado grandes avances en la generación y colección de datos, debido básicamente al poder de procesamiento de las máquinas, propiciando la existencia de enormes volúmenes de datos en donde 'se oculta' información que podría ser de suma importancia estratégica para cualquier proceso, pero a la que no se puede acceder por las técnicas clásicas de recuperación de información.

El presente estudio tiene como propósito exponer las ventajas aplicativas de una nueva metodología de *Inteligencia de Negocio Adaptiva*, que utiliza la base de datos de la Universidad Nacional de La Rioja, Argentina como datos de entrada, información respecto a diferentes aspectos de los alumnos, tales como datos personales, económicos, sociales, culturales, que favorecerán un análisis multivariado.

En el análisis se utilizaron diversas herramientas que permitieron el pre-procesamiento de los datos, su selección acorde a la sensibilidad con los resultados, clasificadores, discriminantes, normalizadores y técnicas de inteligencia artificial, para derivar en un árbol de clasificación que precisa aquellas causalidades principales de la deserción de los alumnos universitarios, en la institución mencionada.

Mediante una revisión sintética se exponen, además, aquellos conceptos que resultan más relevantes para la aplicación posterior de las tecnologías en cuestión.

La aplicación de las reglas obtenidas en la base de datos del Sistema de Autogestión de Alumnos, SIU Guarani, mediante simples sentencias facilita la detección de aquellos alumnos que actualmente se encuentran en riesgo de deserción, con la consecuente posibilidad de adopción de medidas correctivas y preventivas en la institución.

El trabajo se lleva a cabo a partir de casos de estudio y de bibliografía de referencia, que facilita el uso de técnicas de *Inteligencia de Negocio Adaptiva*.

Surge como una de las tantas aplicaciones posibles de la Ingeniería de Software, siendo sus resultados relevantes y formativos para los alumnos de las carreras de Licenciatura e Ingeniería de Sistemas de la Universidad Nacional de La Rioja y, de manera complementaria, para los alumnos y docentes de las carreras de Educación de la misma Universidad.

Fomenta, además, el trabajo en equipo e interdisciplinar, fundamental en el área del desarrollo informático y la participación de diferentes estamentos sociales, como sindicatos,

universidades y gobierno.

2. ESTADO DEL ARTE

El portal educativo del estado de La Argentina, educ.ar, cita textualmente que "la pobreza, la exclusión y la escasa capacidad de las escuelas para contener serán las causas principales de la deserción" (2017). Así mismo, La universidad de Salta (Farías et al. 2007) informa que causas como la deficiente articulación del sistema educativo argentino; dificultades económicas o falta de políticas institucionales de retención representan las causas de deserción. Por su parte Georgina Binstock (2005) realiza un estudio basado en técnicas cualitativas y cuantitativas plasmándolas en su libro Carreras Truncadas. En el mismo identifica cómo en las vidas de los jóvenes –de hogares socialmente excluidos, hogares vulnerables y hogares de clase media-baja y clase media– confluyen y se entrelazan los distintos aspectos que promueven el proceso de abandono escolar.

Otros investigadores como Karina B Eckert y Roberto Suénaga (2015) han realizado varios trabajos en argentina a partir de diferentes tipos de software, dominios de datos y metodologías, particularmente KDD y CRISP-DM o SEMMA. Mediante este proceso fue posible identificar los atributos que caracterizan a los casos de deserción y su relación con el desempeño académico, especialmente en el primer año de la carrera de Ingeniería en Informática de la Universidad Gastón Dachary.

El presente estudio arroja nuevos conocimientos, a la luz de los resultados obtenidos, utilizando tecnologías de la información y particularmente la nueva metodología propuesta en Una Metodología para el Desarrollo de Aplicaciones Autoadaptativas basada en Business Intelligence. Aplicación en Medicina (2012). Estos resultados precisan y cuantifican la relevancia mediante métricas de calidad definidas, lo cual particulariza los resultados obtenidos.

3. ADAPTIVE BUSINESS INTELLIGENCE

El descubrimiento de conocimiento "oculto" es posible gracias a la minería de datos -data mining-, que brinda un conjunto de técnicas sofisticadas para encontrar patrones y relaciones dentro de los datos.

Esto permite la creación de modelos o representaciones abstractas de la realidad como parte del proceso de descubrimiento de

conocimiento, KDDP -por su sigla en inglés- que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan significado a estos patrones encontrados.

Por otro lado, Business Intelligence, también propuesto por Watson y Wixom (2007), provee beneficios que no sólo se aplican en los ámbitos empresariales, donde estaría demás decir las ventajas que reditúa. Esta alternativa para la toma de decisiones se ve potenciada cuando se complementa con la característica autoadaptativa de las aplicaciones.

Adaptive Business Intelligence, ABI, es la disciplina que combina la predicción, la optimización, y la capacidad de adaptación en un sistema capaz de responder a dos preguntas fundamentales: ¿qué es probable que ocurra en el futuro? y ¿cuál es la mejor decisión en este momento? (Michalewicz et al. 2007)

3.1 La Metodología de Adaptive Business Intelligence

A partir del estudio de numerosas variables que tienen correlación, en mayor o menor medida, con el indicador objetivo estudiado: deserción estudiantil, se ha aplicado una metodología para desarrollar un sistema de Inteligencia de Negocio Adaptiva. Ésta surge de los conceptos relacionados entre disciplinas afines tecnológicamente como: OLAP, Acceso a datos multiplataformas, Business

Intelligence, Data mining y Adaptabilidad. Para su desarrollo, se han analizado las interrelaciones existentes entre ellas, y de este modo, se ha conformado un entorno de soporte para las aplicaciones predictivas.

La metodología propone el uso de las técnicas implementadas en la herramienta de minería de datos utilizada, buscando aquellos resultados que proporcionen la información necesaria para acercarse a los objetivos del proyecto (Slater et al 2016) Esto involucra las etapas de Búsqueda de patrones, reglas o grupos, la etapa de Modelizado predictivo y la Validación del modelo, basadas en las etapas metodológicas para el desarrollo de aplicaciones basadas en Inteligencia de Negocio Adaptiva.

Es en este punto donde los aportes de la metodología permiten una interacción dinámica con los patrones que surgen de los datos, a través de las diferentes fases, los cuales tienen significancia como nuevo conocimiento (Cios et al. 2007)

3.1 Ciclo de Vida de la Metodología

Cada fase es estructurada en varias tareas generales de segundo nivel, las cuales se proyectan a tareas específicas donde finalmente se describen las acciones que deben ser desarrolladas para situaciones particulares, pero en ningún momento se propone como realizarlas.

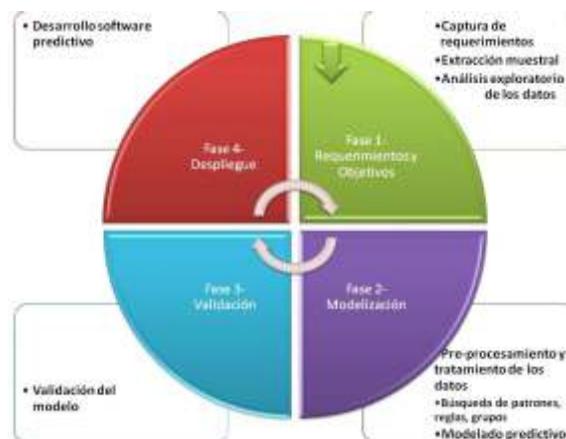


Fig 1. Ciclo de vida de la nueva metodología Adaptive Business Intelligence y su correspondencia con las etapas metodológicas propuestas

El ciclo de vida de la metodología representa aquellos aspectos que describen los pasos a seguir de modo tal que permitan una evolución coherente y progresiva, que facilite, partiendo del problema

planteado, llegar a una solución aceptable en los resultados esperados. Tal y como lo indican Michaeliewicz y otros: “Los sistemas de Adaptive Business Intelligence incluyen elementos de la



minería de datos, modelos predictivos, predicción, optimización, y adaptabilidad, y son utilizados... para tomar mejores decisiones.” (2007). La lógica estructural de la metodología contempla la implementación de la característica de autoadaptatividad y lo realiza semánticamente.

3.2 Alcances de la Metodología

Alcance de cada una de las etapas que se identificaron en la metodología ABI utilizada.

Fase 1- Requerimientos y Objetivos

- (a) Captura de requerimientos. Se trabaja en el dominio de los expertos para definir el problema y determinar los objetivos del proyecto que, finalmente, se traducen en hipótesis acerca de la selección inicial de las técnicas de data mining.
- (b) Extracción muestral. Se hace la recopilación de datos de la muestra y se toman decisiones respecto a cuáles serán necesarios para el análisis, teniendo en cuenta el formato y tamaño, entre otros aspectos. En este punto se comprueba la integridad de los datos, la redundancia, los valores que faltan, etc. Además, el paso incluye la verificación de la utilidad de los datos con respecto a los objetivos de data mining.
- (c) Análisis exploratorio de los datos. Se identifican los campos más importantes relacionados con el problema y se determinan cuáles valores derivados pueden ser útiles.

Fase 2 - Modelización

- (d) Pre-procesamiento y tratamiento de los datos. Se considera la limpieza de datos, que incluye la comprobación de la integridad de los registros de datos, la eliminación o la corrección de ruido y los valores faltantes, las remociones y duplicaciones, las adecuaciones de formato, el agregado de expresiones numéricas y fórmulas que mejoren o equilibren datos faltantes con mayor significancia o relevancia.
- (e) Búsqueda de patrones, reglas o grupos. Aquí se elige la técnica de minería de datos que permite encontrar patrones en los datos.
- (f) Modelado predictivo. Se hace el refinamiento del método seleccionado, testeando diferentes técnicas para luego decidir cuál algoritmo y qué parámetros pueden ser utilizados, de acuerdo a los requerimientos del experto.

Fase 3 - Validación

- (g) Validación del modelo. Se hace la comprensión de los resultados, comprobando si el conocimiento descubierto es novedoso e interesante.

Fase 4 - Despliegue

- (h) Desarrollo software predictivo. Una vez que se ha conseguido el modelo, para que éste sea útil, deberá ser desplegado. Esto significa salir del entorno de desarrollo en una forma que pueda ser usada externamente.

4. GESTIÓN DE LOS DATOS DE LOS ALUMNOS – SIU GUARANÍ

Las universidades utilizan sistemas de información adecuados a sus necesidades y características, pero es posible mantener criterios comunes para el desarrollo de sistemas informáticos que permitan la gestión de los datos de los alumnos desde sus primeros días en la universidad hasta que culminan sus estudios. La iniciativa del SIU Guaraní es un ejemplo de esta tendencia a la unicidad de elementos que favorezcan los desarrollos informáticos.

El SIU Guaraní, como sistema de información, permite el seguimiento de todas las actividades que realiza un estudiante: inscripción a exámenes y cursadas, reinscripción a carreras, consulta de inscripciones, consulta de plan de estudios e historia académica, consulta de cronograma de evaluaciones parciales, consulta de créditos, notas de evaluaciones parciales, materias regulares, actualización de datos censales y recepción de mensajes.

Se presenta como un sistema informático de autogestión académica por Internet, que le permite a los alumnos hacer uso de un sistema más acorde con las actuales tecnologías de la información y a la institución una recopilación de mayor cantidad de datos, debido a que no se requieren usos de equipamientos dedicados en la institución.

Con los datos recopilados por el SIU Guaraní se construye la base de datos utilizada por la Universidad Nacional de La Rioja, una universidad argentina, aportando datos de carácter social, económico y cultural de cada uno de los estudiantes.

Esta base de datos constituye la referencia inicial para el procesamiento de los datos mediante herramientas ETL -Extract-Transform-Load- que constituirán los datos de entrada para el análisis de

5. IMPLEMENTACIÓN ESTRATÉGICA DEL ABI

La posibilidad de entregar al decisor información precisa y confiable respecto a indicadores como el porcentaje de alumnos que han desertado es una medida que se comporta como una medición pasiva, toda vez que cualquier acción correctiva adoptada dependerá de una política y a su vez tendrá un destino más amplio que preciso. Es allí donde el aporte de las tecnologías de Inteligencia de Negocio Adaptativa provee información oportuna y selectiva que permita un proceso de toma de decisiones óptimas mejorando la celeridad y la precisión (Villegas y Luján-Mora. 2017).

5.1 Aplicación de la Metodología para la Predicción de Casos de Deserción Potencial de los Alumnos Ingresantes

Los datos empleados provienen de una Universidad Nacional de Argentina. Previo a la utilización de los datos aportados por la base de datos del SIU Guaraní fue necesario realizar una selección empírica de aquellos atributos cuya incidencia puede ser relevante, dejando de lado algunos como: datos de tarjetas de crédito, identificadores de tipos de documentos, fechas irrelevantes, códigos, nombres, entre otros.

Se ha definido que el modelo realice la clasificación de los alumnos con riesgo de deserción, con una precisión cuyo error absoluto medio sea $\leq 2\%$ y su precisión de al menos 80%.

Se trabajó con un banco de datos cuya información contiene un elevado número de muestras, 4707 registros, para ser exactos, a partir de los datos provistos por el SIU Guaraní de una Universidad Nacional de Argentina.

Esta base de datos contiene 45 atributos. 2 atributos adicionales han sido calculados con base en los valores contenidos en otros atributos. Se determinó la condición de Cursantes o Abandonos a partir de aquellos alumnos que no han rendido asignaturas desde hace 2 años. Los atributos restantes constan de valores numéricos en 6 atributos, y los restantes son nominales o discretos.

5.2 Análisis exploratorio de los datos, Información de los atributos, Distribución de Clases.

Base de datos facilitada por el Sistema Unificado Universitario de Gestión de Alumnos – SIU Guaraní de una Universidad Nacional de Argentina.

Tabla 1: Distribuciones por clases

| Clase | Número de muestras |
|----------|--------------------|
| Abandono | 2456 |
| Cursante | 2251 |

Esta etapa se orienta mayormente hacia la visualización de los datos, con la finalidad de simplificar el problema, detectando aquellos datos con poca o ninguna incidencia estadística hacia los objetivos predefinidos.

5.3 Búsqueda de patrones, reglas o grupos

La metodología propuesta aprovecha en este caso las bondades de un algoritmo de clasificación en árbol SimpleCart, J48 y Decision Table, los que proveen un método supervisado para la clasificación (Shmueli 2017)

Tabla 2: Resultados obtenidos en la clasificación de los subconjuntos de datos.

| Paso | Algoritmo utilizado | Porcentaje de aciertos |
|------|-----------------------|--|
| 1 | SimpleCart (Subset 4) | Correctly Classified Instances 2858 87.24 % Incorrectly Classified Instances 418 12.75% |

Con esta aplicación de la clasificación, queda demostrado el correcto uso de la técnica de selección de datos basados en correlatividad. Es decir, los porcentuales de error y las precisiones alcanzadas son mejores con los atributos recomendados por el evaluador de atributos, con el método de ranking (Peral et al. 2017) Para mayor información respecto a estos algoritmos, revisar Classification and Regression Trees, de Leo Breiman (2017).

5.4 Modelado predictivo

Tabla 3: Resultado del modelado predictivo

| === Summary === | | |
|----------------------------------|------|-----------|
| Correctly Classified Instances | 2858 | 87.2405 % |
| Incorrectly Classified Instances | 418 | 12.7595 % |
| Kappa statistic | | 0.7434 |
| Mean absolute error | | 0.2025 |
| Root mean squared error | | 0.324 |
| Relative absolute error | | 40.8993 % |
| Root relative squared error | | 65.1173 % |
| Total Number of Instances | 3276 | |

Para el proceso de entrenamiento del modelo se utilizó la técnica de análisis denominada validación cruzada-cross validation-. Habiendo aplicado los filtros 'CFS -Subset Evaluator-' con los métodos



'Best first' y 'Genetic search', además del filtro 'Info Gain Attribute'" con el filtro 'Eval Attribute Ranking'".

5.5 Aplicación del aprendizaje

A este nivel ya se han logrado las características de los conocimientos necesarios, que facilitan la implementación del aprendizaje obtenido, mediante las fases metodológicas desarrolladas previamente.

Las reglas de inferencia obtenidas corresponden al algoritmo SimpleCART, debido a que durante las pruebas comparativas entre algoritmos se han presentado mejores porcentajes de instancias correctamente clasificadas y se describen a continuación:

Tabla 4: Resultado del algoritmo SimpleCART Decision Tree

| | |
|---|--|
| Edad < 21.5 | |
| Edad < 20.5: Cursante(551.0/8.0) | |
| Edad >= 20.5 | |
| mat_aprob < 18.5: Cursante(278.0/61.0) | |
| mat_aprob >= 18.5 | |
| Promedio < 6.5: Abandono(16.0/2.0) | |
| Promedio >= 6.5 | |
| mat_aprob < 28.5: Cursante (11.0/1.0) | |
| mat_aprob >= 28.5: Abandono (6.0/0.0) | |
| Edad >= 21.5 | |
| mat_aprob < 0.5 | |
| orientacion_recibida= "Si": Cursante (20.0/0.0) | |
| orientacion_recibida!= "No": Abandono (817.0/130.0) | |
| mat_aprob >= 0.5 | |
| mat_aprob < 11.5: Cursante(643.0/96.0) | |
| mat_aprob >= 11.5 | |
| Promedio < 5.5: Abandono(247.0/37.0) | |
| Promedio >= 5.5 | |
| mat_aprob < 24.5 | |
| Promedio < 6.5 | |
| mat_aprob < 15.5: Cursante (25.0/9.0) | |
| mat_aprob >= 15.5: Abandono (62.0/25.0) | |
| Promedio >= 6.5: Cursante (51.0/13.0) | |
| mat_aprob >= 24.5: Abandono (142.0/25.0) | |

Esta formulación de requerimientos facilita al desarrollador de software la realización de un programa o de una secuencia de sentencias por ejemplo en SQL y MySQL, que accedan a la Base de Datos a partir del modelo para nuevos datos proporcionados como entrada.

```
SELECT * FROM alumnos WHERE (Edad>=21)
AND (mat_aprob >=11.5) AND (Promedio<5.5)
```

Código fuente: Ejemplo aplicación con MySQL en BD SIU

6 CONCLUSIONES

Como se ha demostrado, la aplicación de la metodología ABI en su fase final culmina en una etapa en donde surge un modelo que, aprovechado para generar sentencias, se podrá implementar en diferentes lenguajes a los gestores de bases de datos. A partir de estos datos quedan expuestas las capacidades predictivas sobre los datos de los alumnos. Además de esto, se ha concluido en un modelo predictivo que facilita la construcción de filtros aplicables a bases de datos para la detección temprana de alumnos con riesgo de deserción estudiantil en el contexto planteado.

Este conocimiento obtenido se potencia al aplicar esta metodología, pudiendo responder preguntas de dificultosa resolución o que demandan excesivo tiempo ante tal cantidad de variables (Alom y Courtney 2018). Es así que a partir de los registros de datos que surgen desde las bases de datos de los estudiantes, ha sido posible encontrar patrones ocultos e información predictiva útil para los decisores. Por su parte, el proceso metodológico ha resultado exitoso, permitiendo autoevaluar los resultados alcanzados mediante la aplicación de las técnicas y algoritmos de Data mining en el contexto de ABI.

Con relación a las características del conocimiento adquirido, acorde a las reglas de clasificación, se ejemplifican algunas de las más relevantes conclusiones consultadas con miembros académicos institucionales:

Alumnos mayores de 21.5 años que no han recibido orientación vocacional al ingreso a las carreras y sin asignaturas aprobadas dentro de los 2 años, desertan en un porcentaje de 24.9%, lo que equivale a un cuarto del alumnado ingresante.

Alumnos mayores de 21.5 años, posiblemente en un segundo año de la carrera, con más de 11.5 asignaturas aprobadas y cuyo promedio no supera 5.5, abandonan las carreras en un 7.5%.

Mediante estos resultados es posible la aplicación de las reglas obtenidas directamente a la base de datos del SIU Guaraní, facilitando así la detección de aquellos alumnos que actualmente se encuentran en riesgo de deserción.

Estas conclusiones permiten la detección temprana de casos de riesgo y de esta manera facilitan la adopción de las decisiones necesarias para orientar los estudios y transformarlos en exitosos, en instituciones de similares características a la del presente estudio.

Al conocer estos datos, la Institución podrá implementar la Orientación Vocacional temprana y evitar la deserción posterior.

TRABAJO FUTURO Existen diferentes líneas de investigación factibles de abordar a partir de la propuesta de esta metodología. En lo referente a los análisis de extracción de conocimiento realizados tenemos que:

Una de las premisas de los trabajos a realizar se centra en la divulgación de los resultados obtenidos para dar a conocer esta tecnología disponible.

Adicionalmente, en el contexto de nuevas investigaciones, este trabajo podrá servir como base disciplinar para construir una aplicación con una base de datos incremental. La misma se utilizaría como una taxonomía de la enorme cantidad de herramientas de data mining, clasificándolas de acuerdo a sus características, aspectos estadísticos de sus datos de entrada, procesos a los que se aplicaría, así como la posibilidad de determinar asociaciones potencialmente superadoras. De esta manera, proveer campos de aplicación para determinados tipos de problemas de desarrollo, donde su aplicabilidad resulta óptima.

En el aspecto referido a Business Intelligence:

La generación de un banco de datos nacional para la adquisición y recopilación de datos relativos a problemas de educación de contexto nacional brindaría ventajas significativas.

Complementariamente al Banco de datos, un archivo de modelos de aprendizajes facilitará el crecimiento exponencial del trabajo de obtención del conocimiento en los ámbitos educativos.

REFERENCIAS

- Alom, B. M., & Courtney, M. (2018). Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia.
- Binstock Georgina. (2005). Carreras truncadas. UNICEF. Recuperado de [https://www.unicef.org/argentina/spanish/Carrerastruncadas\(1\).pdf](https://www.unicef.org/argentina/spanish/Carrerastruncadas(1).pdf)
- Breiman, L. (2017). *Classification and regression trees*. Routledge.
- Cios, K. J., Pedrycz, W., Swiniarski, R. W., & Kurgan, L. A. (2007). *Data mining: a knowledge discovery approach*. Springer Science & Business Media.
- Eckert, K. B., & Suénaga, R. (2015). Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. *Formación universitaria*, 8(5), 03-12.
- Educ.ar El portal educativo argentino. (2017). Recuperado de <http://portal.educ.ar/noticias/educacion-y-sociedad/desercion-escolar.php>
- Farias, M., Fiol, D, Kit, I. Melgar, S. (2007). Todos pueden aprender - Propuestas para superar el fracaso escolar. UNICEF - Oficina de Argentina.
- Michalewicz, Z., Schmidt, M., Michalewicz, M., & Chiriac, C. (2006). *Adaptive business intelligence* (pp. 37-46). Springer Berlin Heidelberg.
- Peral, J., Maté, A., & Marco, M. (2017). Application of data mining techniques to identify relevant key performance indicators. *Computer Standards & Interfaces*, 54, 76-85.
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. John Wiley & Sons.
- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, 42(1), 85-106.
- Una Metodología para el Desarrollo de Aplicaciones Autoadaptativas basada en Business Intelligence. Aplicación en Medicina. (2012). (Tesis para optar a la titulación de postgrado correspondiente a la Maestría en Ingeniería de Software).
- Villegas-Ch, W., & Luján-Mora, S. (2017, March). Analysis of data mining techniques applied to LMS for personalized education. In *World Engineering Education Conference (EDUNINE)*, IEEE (pp. 85-89). IEEE.
- Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer*, 40(9)