

Impacto del rendimiento académico de primer semestre en el abandono de primer año

Alfonsa García, Ana I. Lías, Luis M. Pozo
Departamento de Matemática Aplicada a las TIC
Universidad Politécnica de Madrid

alfonsa.garcia@upm.es, anaisabel.lias@upm.es, lm.pozo@upm.es

Resumen

El objetivo de este trabajo es presentar una propuesta sostenible para el seguimiento y evaluación sistemática del abandono de primer curso en titulaciones relacionadas con la Ingeniería Informática, basada en el tratamiento automático de la información disponible en la universidad. En trabajos previos se ha desarrollado una aplicación informática online, ASIA^{1a} (Aplicación para el Seguimiento Institucional del Abandono de primer año). Usando esta aplicación, se puede analizar el impacto sobre el abandono de diferentes variables. En particular, se ha constatado que uno de los mejores predictores del abandono es el rendimiento académico de primer año. Pero esta variable se conoce demasiado tarde para poder adoptar medidas preventivas. El estudio presentado ahora demuestra que se puede predecir el abandono de primer año a partir de los resultados académicos de las asignaturas de primer semestre, con la ventaja de que esta información se conoce antes. Se concluye que es posible, mediante el tratamiento automático de la información, hacer un seguimiento del proceso de integración de los estudiantes de primer año, evaluar la eficacia de los programas de acogida e identificar alumnos en riesgo de abandono, con el fin de proponer acciones específicas.

Abstract

The goal in this work is to present a sustainable proposal to assess first-year drop-out, in Computer Engineering degrees, based on the systems for automatic treatment of information. An online computer application, ASIA^{1a} (Aplicación para el Seguimiento Institucional del Abandono de primer año), was previously developed, and it has been used to analyze the impact of several different variables on the first-year drop-out rate. It has been observed, in particular, that one of the best predictors of drop-out is academic performance throughout the first year. Unfortunately,

this information gets to be known too late to be able to take remedial actions. The present study shows that first-year drop-out can be reliably predicted using only the academic results of courses in the first semester, which are known much earlier. One may thus conclude that it is possible to keep track of the integration process of first-year students by using the automatic means of treatment of the information available on them, thus evaluating the efficiency of the welcoming programs, and identifying the students most likely to drop out in order to propose measures to avoid it.

Palabras clave

Integración en la universidad, estudiantes de primer año, rendimiento académico, tratamiento automático de la información.

1. Introducción

La necesidad de titulados cualificados en Tecnologías de la Informática y las Comunicaciones (TIC) ha crecido espectacularmente en los últimos años, en todos los sectores. El crecimiento de la economía depende, más que nunca, de formación avanzada en TIC. Según previsiones de la Unión Europea, se crearán casi 900 000 empleos tecnológicos de aquí a 2020. Sin embargo actualmente la demanda de ingenieros (con perfiles como big data, ciberseguridad o desarrollo software) es difícil de cubrir. Uno de los factores influyentes en esta situación es que las titulaciones relacionadas con la Ingeniería Informática tienen altas tasas de abandono de primer año, como se puede ver en el estudio realizado por García et al. [11], y mayores que en otras ingenierías [6]. Esto no es solamente un problema de España, sino que también se produce en otros países de Europa [14] y América [7]. Algo falla cuando en carreras en las que

quedan candidatos no admitidos, más del 15% de los estudiantes admitidos abandonan tras el primer curso.

La preocupación por la información previa y por la acogida e integración los estudiantes de nuevo ingreso (ENI) empieza a extenderse en ámbitos institucionales y por ello las universidades suelen diseñar actividades encaminadas a mejorar las tasas de retención y a facilitar la integración de los ENI. Sin embargo, es preciso acompañar estas acciones con protocolos de seguimiento y evaluación, que permitan seleccionar las más eficaces. En [16], F. Michavila recoge un informe exhaustivo sobre programas de acogida de las universidades españolas. Un análisis del proceso de integración puede facilitar el diseño e implementación de medidas específicas. En cualquier caso, los protocolos deben contemplar el seguimiento sistemático del impacto de las actuaciones.

El grupo de Innovación Educativa GIEMATIC de la Universidad Politécnica de Madrid (UPM) ha creado la aplicación on-line ASIA^{1a} (disponible en <http://permanencia.eui.upm.es/>), para hacer un seguimiento institucional sistemático del abandono de primer año. Esta herramienta, presentada en [11], permite analizar la incidencia en la permanencia a un año de variables como el género, la edad, la nota de acceso, la opción de acceso o el rendimiento académico. Actualmente dispone de los datos de cuatro cohortes de estudiantes de nuevo ingreso de las titulaciones de grado de la UPM, las correspondientes a ENI de 2010-11, 2011-12, 2012-13 y 2013-14 y está en proceso la incorporación de la cohorte de estudiantes que accedieron el curso 2014-15. En [4], usando una copia de ASIA^{1a} realizada para la Universidad de Oviedo, se ha constatado que el rendimiento académico del primer curso es una variable con gran influencia en el abandono de primer año. El problema es que esta variable se conoce demasiado tarde para tomar medidas preventivas. El objetivo de este trabajo es, por una parte demostrar que es posible obtener, con técnicas de minería de datos, un modelo de predicción del abandono de primer año para los estudiantes de Ingeniería del Software (IW) e Ingeniería de Computadores (IC) de la UPM, a partir de los datos de acceso y del rendimiento académico y por otra ver si es plausible sustituir en el modelo el rendimiento académico de primer año por las calificaciones de primer semestre, con el fin de proponer protocolos de actuación con seguimiento sistemático, basado en el tratamiento automático de la información.

El resto del presente documento está organizado como sigue: En la sección 2 se comentan algunos trabajos relacionados, en la 3 se describe la población de estudio, en la 4 la metodología seguida y en la 5 los resultados obtenidos. El trabajo se completa presentando algunas conclusiones.

2. Trabajos relacionados

El problema del abandono universitario ha sido abordado por numerosos autores en las últimas décadas. Ya en 1975, V. Tinto [18] relaciona el abandono con los problemas de integración tanto académica como social. La investigación realizada por Álvarez et al. [1] pone de manifiesto la importancia de la transición a la universidad y el carácter multifactorial del proceso de integración, que se produce durante el primer año. Otros estudios [9, 17, 20] confirman que el mayor porcentaje de abandono universitario se produce en el primer curso. Las tasas de abandono son sin duda un indicador de la calidad del proceso de integración, que no es solamente responsabilidad del estudiante, sino que también es, en buena medida, responsabilidad de la Institución, como se pone de manifiesto en los trabajos de Berger y Braxton [3] y en las reflexiones de V. Tinto [19], quien indica que apenas se ha abordado la tarea de evaluar y mejorar la práctica institucional en relación con la acogida e integración de los estudiantes.

Para hacer un seguimiento de los programas de acogida y analizar el perfil del estudiante que abandona, diversos autores proponen la explotación automática de la información disponible en las universidades. Por ejemplo, usando información de la Universidad de Granada, F. Araque et al. [2] desarrollan modelos para obtener el riesgo de abandono en diferentes titulaciones y concluyen que el perfil del estudiante que abandona Ingeniería del Software es diferente del que abandona estudios de humanidades o economía. C. Díaz [10] también utiliza bases de datos de su universidad para analizar la retención, con un modelo de riesgos proporcionales de Cox y concluye que los estudiantes de ingeniería presentan altos riesgos de deserción entre el primer y tercer semestre e identifica factores de riesgo de carácter económico. Por su parte H. Kuma et al. [15], usando procesos de explotación de la información, demuestran que la superación de más de una materia del primer año es clave para la permanencia.

Tanto el trabajo de Christie [3], como el de Kori et al. [4] señalan que, en carreras TIC, las actividades sociales y la relación con los compañeros fuera del aula, es menos determinante para la permanencia que en otro tipo de carreras.

3. Población de estudio

La población de estudio la constituyen los estudiantes de nuevo ingreso (ENI) que accedieron al sistema universitario entre 2010 y 2014, matriculándose en una de las dos titulaciones de grado IW e IC de la UPM. El colectivo lo forman un total de 990 estudiantes (873 hombres y 117 mujeres). La media de edad de acceso a la universidad es de 19.7 años, la

nota media de acceso a la universidad es 6.78 (en un rango de 5 a 14), el valor mínimo de las notas del colectivo es 5 y el máximo 12.25. El número medio de créditos aprobados en el primer año ha sido 24.9 (en un rango de 0 a 60). Accediendo a ASIA^{1a} se puede ver la evolución de estos datos por año y titulación y compararlos con los de otros grados de la universidad.

3.1. Abandono de primer año

De los 990 alumnos del colectivo estudiado, 245 (24.7%) no se volvieron a matricular el curso siguiente en la misma titulación, lo que supone una tasa de abandono de primer año por encima de la media de la universidad. La evolución de las tasas por cohortes y titulaciones se recoge en la figura 1.

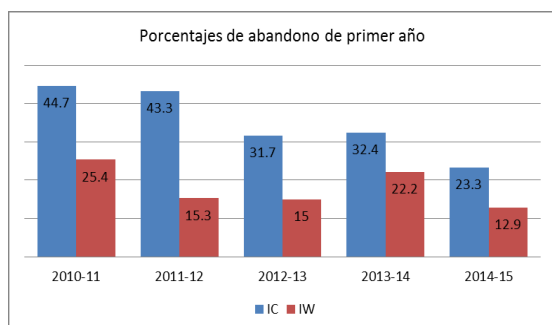


Figura 1: Tasas de abandono en IC e IW.

Las tasas de abandono son menores en IW (titulación con mejores notas de acceso y mejores resultados en rendimiento académico). Se puede apreciar una evolución positiva, tal vez explicada por la toma de conciencia del problema y la puesta en marcha de algunas medidas de acogida e integración. En el curso 2013-14 hubo un importante incremento del abandono en toda la universidad, coincidiendo con una subida de tasas académicas, ante la cual muchos estudiantes optaron por cambiar de carrera (lo que supone incremento del abandono), mejor que pagar segundas matrículas.

3.2. Peculiaridades de la población

Las titulaciones de grado en IC e IW de la UPM, como la mayoría de las titulaciones TIC, tienen pocas mujeres entre sus estudiantes.

Por ejemplo, mientras que, según datos de 2014 del Ministerio de Educación Cultura y Deportes, en España las mujeres suponen el 54.1 % de los estudiantes universitarios o el 30.6 % de los estudiantes del área de ingeniería o arquitectura, en las titulaciones IC e IW el porcentaje de mujeres en cada cohorte (ver figura 2) apenas supera el 10 %.

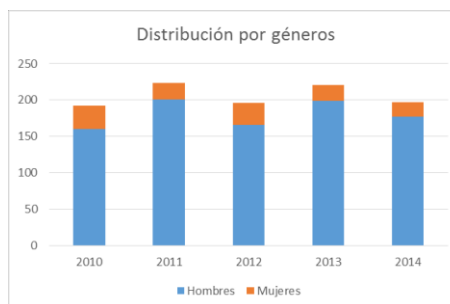


Figura 2: Evolución del porcentaje de mujeres.

En investigaciones cualitativas previas [12,13], con encuestas realizadas a los estudiantes y grupos de discusión, hemos verificado que muchos ENI de IC e IW de la UPM admiten que su conocimiento previo de la realidad de la carrera era muy escaso antes de llegar a la universidad y las dificultades con las asignaturas básicas como las matemáticas, debidas habitualmente a carencias en la formación previa, se suelen incluir entre los factores del abandono, como es habitual en titulaciones de ingeniería [5, 8, 14].

Las distintas vías de acceso a la universidad implican distintas formas de afrontar el proceso de integración. En las titulaciones estudiadas, más de un 15 % de los ENI acceden tras realizar un ciclo superior de formación profesional (FP) relacionado con la informática. Prácticamente el 85 % restante acceden después de haber hecho bachillerato y superar la prueba de acceso a la universidad. En la mayoría de las titulaciones de la UPM, el abandono de primer año es superior entre los estudiantes que provienen de FP, como se puede ver en el trabajo de A. Casaravilla et al. [6]. Pero esto no suele ser así en IC e IW [12,13]. Además, las causas de abandono son claramente diferentes entre los estudiantes que provienen de FP y los de bachillerato. Los primeros tienen mejor conocimiento previo de la carrera, pero su dificultad para la integración reside en su mala preparación para afrontar asignaturas básicas. Los alumnos del segundo grupo apuntan como causas de su fracaso sus propias dificultades de planificación y organización. Por otro lado, casi todos los procedentes de bachillerato que abandonan se reubican en otra universidad (por conveniencia geográfica) o en otra carrera, porque la iniciada no ha respondido a sus expectativas, mientras que los ENI procedentes de FP que abandonan la titulación, renuncian a seguir otros estudios universitarios.

4. Metodología

En primer lugar hemos llevado a cabo un estudio descriptivo sobre el perfil del abandono en la totalidad del colectivo, usando la aplicación ASIA^{1a}. En este estudio hemos podido apreciar que las variables más influyentes en el abandono son la nota de acceso

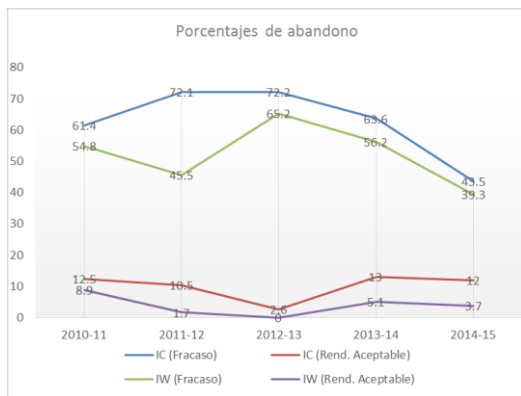


Figura 3: Abandono según rendimiento académico.

a la universidad (*NotAcc*) y el número de créditos aprobados el primer año (*CrSup*). Esta mayor influencia se ha corroborado con un análisis de correlaciones.

A continuación, usando estas variables como predictores, se han ensayado distintos modelos de regresión logística y clasificación, entrenados mediante técnicas de aprendizaje supervisado. Primero tomando como conjunto de entrenamiento el de los ENI de IC e IW de las cuatro cohortes de 2010 a 2013 (793 registros) y como conjunto de prueba la base de datos de los ENI que accedieron a estas carreras en 2014, que contiene 197 ENI, de los que 34 abandonaron su titulación tras el primer curso. Después se han ensayado modelos de regresión y clasificación con validación por resustitución usando la base de datos total, con los 990 registros de todo el colectivo de estudio, haciendo selecciones aleatorias para los conjuntos de entrenamiento y prueba.

Aunque el rendimiento de primer año pueda ser un buen predictor, el problema es que se conoce demasiado tarde. Por ello, se han recopilado datos de rendimiento académico de primer semestre del curso 2014-15 y se ha comparado la influencia en el abandono de la variable *CrSup* con la de la variable *CrSupIS*=Número de créditos aprobados en el primer semestre. Después se han analizado los datos de rendimiento académico en cada una de las asignaturas del primer semestre de la cohorte 2014-15 y su influencia en el proceso de integración.

5. Resultados

La figura 3 refleja la importancia del rendimiento académico en la integración. Las tasas de abandono de primer año son mucho mayores entre los estudiantes que fracasan (superando menos de 10 créditos de los 60 del primer curso) que entre los que tienen un rendimiento mínimamente aceptable (superando al menos 24 créditos).

Nuestro objetivo es elaborar modelos predictivos del abandono en función de las variables más significativas (nota de acceso, créditos superados, titulación). Abordamos el proceso desde dos alternativas distintas: la regresión logística y los modelos de clasificación.

Se han ensayado diversas opciones para obtener un modelo de regresión logística (modelos logit y probit y diversas combinaciones de variables), obteniéndose los mejores resultados con una fórmula $\text{logit}(y) \sim 1 + \text{NotAcc} + \text{CrSup} : \text{Titulación}$. Se ha utilizado resustitución con un subconjunto aleatorio de entrenamiento del 80 % de los datos y un subconjunto de validación del 30 %.

	Coefficiente	SE	tStat	pValue
(intercept)	-0.75688	0.62287	-1.215	0.2243
<i>NotAcc</i>	0.18623	0.10020	1.823	0.0683
<i>CrSup:IW</i>	-0.13538	0.01605	-8.436	<2e-16
<i>CrSup:IC</i>	-0.08272	0.01022	-8.090	<2e-16

Cuadro 1: Modelo de regresión logística.

Tras realizar 500 ensayos, el mejor modelo (ver cuadro 1) alcanza, con una probabilidad umbral de 0.4, un error de entrenamiento del 16.1% y proporciona un área bajo la curva ROC dado por $AUC=0.8755$.

Al comparar sobre el subconjunto de validación las predicciones de este modelo con el abandono real, se obtiene la matriz de confusión del modelo y los indicadores dados en el cuadro 2 (donde el signo + representa abandono y el signo - permanencia, sensibilidad es la fracción de abandono predicha correctamente y especificidad es la fracción de permanencia correctamente predicha).

Matriz de confusión	Valores predichos		Indicador		
	-	+	Tasa de éxito		
Valores reales	-	187	38	Sensitividad	79.5%
	+	20	53	Especificidad	72.6%
				MCC	83.1%
					0.52

Cuadro 2: Evaluación del modelo logit.

Respecto a los modelos de clasificación ensayados, los mejores resultados (ver cuadro 3) se han obtenido con un modelo de máquinas de soporte vectorial, usando resustitución 80/30 para la validación.

Matriz de confusión	Valores predichos		Indicador		
	-	+	Tasa de éxito		
Valores reales	-	211	18	Sensitividad	88.6%
	+	16	52	Especificidad	76.5%
				MCC	92.1%
					0.68

Cuadro 3: Evaluación del modelo de clasificación.

La sensibilidad del modelo de clasificación es mejor. Es decir el modelo predice mejor el abandono, lo que lo hace más apropiado para el propósito de actuar de forma temprana.

Con los datos del rendimiento académico de la cohorte 2014-15 (197 registros) se ha probado que la variable *CrSup1S* tiene una alta correlación (0.888) con la variable *CrSup* y que las correlaciones de estas variables con la variable *Abandono* son muy parecidas (-0.312 y -0.362 respectivamente), por lo que el número de créditos aprobados en el primer semestre proporciona casi tanta información sobre el proceso de integración del estudiante como el número de créditos aprobados al cabo del año. Este hecho, junto con los modelos obtenidos por distintas técnicas, nos hace conjeturar que es factible obtener modelos de predicción del abandono, con una sensibilidad razonable, basándose en los créditos superados en el primer semestre.

Se ha analizado también la influencia en el proceso de integración de cada una de las asignaturas. Las cuatro asignaturas más importantes del primer semestre son: AM (Análisis matemático), FC (Fundamentos de computadores), P (Fundamentos de programación), y MD (Lógica y matemática discreta). Las calificaciones de todas ellas tienen correlación negativa con el abandono con p-valores menores que 0.05, luego son significativas a un nivel de confianza del 95 %. Tienen también altas correlaciones positivas entre ellas (lo que significa que pueden proporcionar información redundante). El número medio de asignaturas aprobadas entre estas 4 es 2.1 para la población que permanece y 0.9 entre los que abandonan. El número de aprobados en cada asignatura en las poblaciones de permanencia y abandono se ve en el cuadro 4.

Asignat.	Aprob.	Abandonan	Permanecen
AM	101	9	92
FC	45	4	41
P	101	8	93
MD	127	9	118

Cuadro 4: ENI (2014) que superan cada asignatura.

De los 34 alumnos que abandonan, 17 no han aprobado ninguna de las cuatro asignaturas. La tasa de abandono entre los que aprueban al menos tres es solamente del 2 %. Cabe destacar que cada una de las cuatro asignaturas tiene un porcentaje significativo de estudiantes no presentados. La tasa de abandono entre los ENI presentados a las cuatro es 8.1 %. Por este motivo se sugiere la implantación de metodologías que promuevan la participación y disminuyan el absentismo.

Finalmente, comentar que al disponer de las calificaciones de las asignaturas de primer semestre solo

para una cohorte de ENI, no hay datos suficientes para entrenar un buen modelo de predicción del abandono basado en las calificaciones del primer semestre, por lo que se sugiere mejorar los procesos de alimentación de ASIA^{1a} para poder incorporar de modo ágil esta información al final de cada semestre.

6. Conclusiones

Las universidades deben implantar protocolos sistemáticos de tratamiento automático de la información sobre factores relevantes en relación con el proceso de integración y el abandono de los estudiantes de nuevo ingreso. Analizando la evolución histórica, se puede constatar el impacto de las distintas actuaciones (programas de información previa, acogida, mentorización, cursos de nivelación, etc.).

El número de créditos superados el primer año, usado en combinación con la nota de acceso a la universidad, es un buen indicador del éxito/fracaso en la integración de los ENI. Pero el número de créditos superados en el primer semestre, del que, con un protocolo ágil de tratamiento automático de datos, se puede disponer antes, también proporciona información significativa sobre el proceso de integración y permite detectar antes a los estudiantes en riesgo de abandono, lo que posibilita habilitar medidas preventivas.

Las metodologías de aprendizaje activo, que suelen tener repercusión positiva en el número de presentados, favorecen la integración, y se constata que los ENI que se presentan a todas las asignaturas abandonan en mucho menor porcentaje. Cabe esperar que las acciones que mejoren la calidad de la enseñanza y conduzcan a mejores resultados en las asignaturas del primer semestre, tengan un reflejo inmediato en la disminución del abandono.

Referencias

- [1] Manuel Álvarez, Pilar Figuera y Mercedes Torrado. La problemática de la transición Bachillerato-Universidad en la Universidad de Barcelona. *REOP*, 22(1):15-27, 2011.
- [2] Francisco Araque, Concepción Roldán y Alberto Salguero. Factors influencing university drop-out rates. *Computers & Education*, 53:563-574, 2009.
- [3] Joseph B. Berger y John M. Braxton. Revising Tinto's Internationalist Theory of Student Departures through Theory Elaborations: Examining the Role of Organizational Attributes in the Persistence Process. *Research in Higher Education*, 39 (2):103-19, 1998.
- [4] Ana Bernardo, María Esteban, Alfonsa García y Luis J. Rodríguez. ASIA^{1a}, herramienta en línea para analizar el abandono universitario *Re-*

- vista d'Innovació Docent Universitària (RI-DU), 9: 148-159, 2017.
- [5] Shane Brown y Jacob Burnham. Engineering Student's Mathematics Self-Efficacy Development in a Freshmen Engineering Mathematics Course. *International Journal of Engineering Education*, 28(1):113-29, 2012.
- [6] Ana Casaravilla, José del Campo, Alfonso García y Rosario Torralba. Un análisis del pre-abandono en estudios de Ingeniería y Arquitectura. En *Actas de la II Conferencia Latinoamericana sobre el Abandono en Educación Superior (II CLABES)*, pp. 61-72. Porto Alegre (Brasil), Noviembre 2012.
- [7] Barbara Christe. Persistence Factors Associated with First-Year Engineering Technology Learners. *Journal of College Student Retention*, 17 (3): 319 -335, 2015
- [8] Philip Cifuo. Analysis of First-year Student Performance in an Engineering Program. *International Journal of Engineering Education*, 27 (5):1054-1060, 2011.
- [9] Enric Corominas Rovira. La transición a los estudios universitarios. Abandono o cambio en el primer año de Universidad. *Revista de Investigación Educativa*, 19 (1):127-51, 2001.
- [10] Christian J. Díaz. Factores de Deserción Estudiantil en Ingeniería: Una Aplicación de Modelos de Duración. *Información Tecnológica*, 20(5):129-145, 2009.
- [11] Alfonso García, Julio Blanco, Ana Casaravilla, Ángeles Castejón, Ángeles Mahillo y Misha Malinga. Protocolo de calidad para la tasa de permanencia a un año en la Universidad Politécnica de Madrid. En *Actas de la III Conferencia Latinoamericana sobre el Abandono en Educación Superior (III CLABES)*, pp. 985-995. Méjico, Noviembre 2013.
- [12] Alfonso García, Ana Lías, Ángeles Mahillo y Rosa M. Pinero. Abandono de primer año en Ingeniería Informática. En *Actas de las X Jornadas de Enseñanza Universitaria de Informática, Jenui 2014*, pp. 151-158. Oviedo, Julio 2014. Disponible (7 de febrero de 2017), en <https://jenui2014.uniovi.es/programa>
- [13] Alfonso García, Ana Lías y Mildred Martínez. Influencia de la formación previa en el proceso de la integración en la Universidad. Pendiente de publicación en *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 2017.
- [14] Külli Kori, Margus Pedaste, Eno Tonisson, Tauno Palts, Helio Altin, Ramon Rantsus, Raiivo Sell, Kristina Murtazin y Tiia Rütman. First-year dropout in ICT studies. En *Actas del 2015 IEEE Global Engineering Education Conference (EDUCON)*, 437-445, Marzo 2015.
- [15] Horacio Kuna, Ramón García y Francisco R. Villatoro. Identificación de Causales de Abandono de Estudios Universitarios. Uso de procesos de Explotación de la Información. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 5:39-44, 2010.
- [16] Francisco Michavila (Coord.) Análisis de las Políticas y Estrategias de Acogida e Integración de los Estudiantes de Nuevo Ingreso en las Universidades Españolas, 2012. Disponible en <http://www.catedraunesco.es/proyectosdestacados.html> (7 de Febrero de 2017)
- [17] Michael P. Ryan y Patricia A. Glenn. Increasing one-year retention rates by focusing on academic competence: an empirical odyssey. *Journal of College Student Retention*, 4(3):297-324, 2003.
- [18] Vincent Tinto. Dropout from higher education: a theoretical synthesis of recent research. *Rev. Educ. Res.* 45:89-125, 1975.
- [19] Vincent Tinto. Research and practice of student retention: What next? *Journal of College Student Retention*, 8 (1):1-19, 2007.
- [20] Lesley Willcoxson. Factors affecting intention to leave in the first, second and third year of university studies: a semester-by-semester investigation. *Higher Education Research & Development*, 29 (6): 623-39, 2010.