

Hacia un sistema de detección temprana de estudiantes en riesgo en entornos de enseñanza-aprendizaje en línea

Ana Elena Guerrero-Roldán, M. Elena Rodríguez, David Bañeres, Cristina Pérez,
Javier Panadero, Abdulkadir Karadeniz

Estudios de Informática, Multimedia y Telecomunicaciones, eLearn Center
Universitat Oberta de Catalunya
Barcelona

{aguerrror, mrodriguezgo, dbaneres, cperezsola, jpanaderom, akaradeniz}@uoc.edu

Resumen

El uso de técnicas de la inteligencia artificial está teniendo un gran impacto en la educación. La disponibilidad de extensas colecciones de datos educativos ha permitido desarrollar modelos de predicción que permiten inferir conocimiento sobre el comportamiento de los estudiantes. Una de las líneas de investigación se centra en la identificación de estudiantes en riesgo, con el objetivo de reducir el riesgo de fracaso y de abandono. En este trabajo se presenta un sistema de detección temprana de estudiantes en riesgo en un entorno de enseñanza-aprendizaje en línea. Asimismo, se describe la prueba piloto del sistema realizada en una asignatura de bases de datos del Grado de Ingeniería Informática de la Universitat Oberta de Catalunya, y los resultados obtenidos de dicha experiencia.

Abstract

The use of artificial intelligence techniques is having a great impact on education. The availability of extensive educational datasets has allowed the development of predictive models that can be used to infer knowledge about students' behavior. One of these research lines focuses on the identification of at-risk students, with the aim of reducing the risk of failure and dropout. This paper presents a system for the early detection of at-risk students in an online setting. Moreover, a case study has been carried out in a database subject of the Computer Engineering Degree at the Universitat Oberta de Catalunya, and the results obtained from this experience are presented.

Palabras clave

Educación en línea, inteligencia artificial, modelos de predicción, sistema de detección temprana, estudiante en riesgo, mecanismo de intervención.

1. Motivación

Una de las mayores preocupaciones de las universidades es el índice de abandono de los estudiantes en las asignaturas que estos se matriculan, y cómo mitigar las consecuencias que de este hecho se derivan [4,5]. Una posible solución es intentar detectar tan pronto como sea posible cuándo los estudiantes están en riesgo en el contexto de cada una de las asignaturas donde están matriculados, con el objetivo de que se puedan activar los mecanismos de apoyo y acompañamiento pertinentes para incidir en la mejora de su proceso de aprendizaje y, por lo tanto, mejorar el rendimiento académico, y reducir el índice de abandono. Previamente, ya se había tratado en la JENUI la creación de perfiles de estudiantes con minería de datos [1][6] para analizar el abandono, pero los estudios se habían realizado a nivel teórico.

En el caso de las universidades en línea esta preocupación en torno al abandono es aún mayor, debido principalmente a dos factores. En primer lugar, el perfil del estudiante corresponde a una edad promedio superior al estudiante de la universidad tradicional, y con obligaciones profesionales y familiares. Compaginar estas obligaciones con el estudio de una carrera universitaria supone un reto continuo, y en general, los estudiantes tienden a concentrar sus esfuerzos en aquellas asignaturas en las que están obteniendo mejores resultados académicos, abandonando aquellas en donde no lo están consiguiendo. En segundo lugar, la mayor parte de la comunicación entre los estudiantes y profesores es asíncrona y se basa en comunicación escrita. Aunque se proporcionan mecanismos para favorecer dicha comunicación, y se proporciona retorno personalizado cualitativo y cuantitativo a los estudiantes sobre su progreso, establecer una comunicación efectiva y a tiempo requiere un esfuerzo por ambas partes.

Precisamente, para proporcionar soluciones a esta problemática, surge el proyecto LIS. Se trata de un proyecto de 3 años, financiado por la Universitat

Oberta de Catalunya (UOC) a través del eLearn Center, y en el que participan profesores de los Estudios de Informática, Multimedia y Telecomunicación. El objetivo del proyecto es desarrollar un sistema de aprendizaje inteligente adaptativo basado en técnicas de inteligencia artificial, que pueda ser aplicable en todas las titulaciones de la universidad, y que ayude a los estudiantes en su proceso de aprendizaje a través de la detección de estudiantes en riesgo, propuesta de caminos de aprendizaje, envío automatizado de retorno personalizado, y técnicas de gamificación para mejorar la retención y reducir el abandono.

Una de las funcionalidades de dicho sistema, aprovechando la gran cantidad de datos generados en un entorno de aprendizaje en línea, es la de proveer un sistema de detección temprana de estudiantes en riesgo. El sistema debe ser capaz de dar información comprensible y precisa a los estudiantes acerca de sus posibilidades de superar las asignaturas, y de brindarles apoyo y motivación (especialmente a aquellos en riesgo), proporcionándoles retorno preciso y valioso para reconducir la situación, y reducir así el índice de abandono. Asimismo, el sistema debe suministrar a los profesores, en el contexto de sus asignaturas, información convenientemente agregada para que puedan detectar de forma eficiente a dichos estudiantes, y puedan actuar de forma más eficaz.

El objetivo de este trabajo es, por un lado, presentar dicho sistema de detección y el modelo de predicción en el que se basa en su estado actual. Por otro lado, se describe la prueba piloto del sistema realizada en una asignatura obligatoria del área de bases de datos del Grado de Ingeniería Informática, y los resultados obtenidos en dicha experiencia.

El trabajo se estructura de la siguiente manera: la Sección 2 describe trabajo relacionado, mientras que la Sección 3 introduce el sistema de detección de estudiantes en riesgo. La Sección 4 presenta la prueba piloto y la discusión se realiza en la Sección 5. Finalmente, las conclusiones y líneas de trabajo futuro se incluyen en la Sección 6.

2. Trabajo relacionado

La proliferación de la educación en línea favorece la creación de nuevas herramientas para apoyar el proceso de enseñanza-aprendizaje. Algunas se centran en mejorar el proceso de distribución de contenidos, otras en la mejora del acompañamiento, pero también en el análisis y mejora de los procesos de retorno personalizado.

En el caso de la educación en línea, se generan gran cantidad de datos. La interacción de cada estudiante dentro de cualquier campus virtual o sistema de gestión de aprendizaje genera un rastro digital compuesto de datos, como serían los datos de navegación o datos que muestran los patrones de consulta

de los recursos de aprendizaje [9]. Además, los sistemas de información disponibles en las universidades guardan los datos académicos de sus estudiantes. Todo ello ha propiciado el desarrollo de áreas de conocimiento dedicadas a la analítica de datos. Este sería el caso de la Minería de Datos Educativos (o *Educational Data Mining*, EDM) y la analítica de aprendizaje (o *Learning Analytics*, LA). Mientras que la EDM busca nuevos patrones [12], y se concentra más en el descubrimiento automatizado de información a partir de los datos, LA permite presentar datos resumidos a través de cuadros de mando. Por lo tanto, LA posibilita que el profesorado pueda emitir juicios y tomar decisiones a través de la presentación de información significativa y resumida, extraída de grandes conjuntos de datos [13]. Esto permite conocer mejor a los estudiantes. Por ejemplo, saber los conceptos que aprenden más fácilmente; o, por el contrario, detectar los conceptos en los que encuentran más dificultades, qué recursos de aprendizaje son más apropiados para cada perfil de estudiante, e incluso facilita saber en qué punto crítico podemos como docentes ayudar a los estudiantes a tener éxito en su proceso de aprendizaje. Por otro lado, también hace posible que se puedan identificar estudiantes en riesgo y por lo tanto, diseñar intervenciones para motivarlos, para reducir la tasa de abandono en la asignatura e incluso mejorar la tasa de éxito [7,8] [10].

3. Descripción del sistema

En esta sección, describimos el sistema de detección temprana, incluyendo sus características técnicas y funcionalidades. Este sistema utiliza técnicas de inteligencia artificial para generar modelos de predicción en base a las notas de las actividades que se proponen a lo largo del semestre en la asignatura. Esto es así porque el modelo de evaluación de la universidad se basa en evaluación continua, que se puede usar tanto con fines de evaluación formativa como sumativa. La evaluación continua en cada asignatura se concreta en la propuesta de una serie de actividades de evaluación a lo largo del semestre, también denominadas Pruebas de Evaluación Continua (PEC). Además, algunas asignaturas incluyen actividades de evaluación que implican la creación de un producto, y que reciben el nombre de prácticas (PR). Para simplificar, denominaremos las PEC y PR simplemente actividades. Cada actividad se califica con la siguiente escala de notas: A (muy alta), B (alta), C+ (suficiente), C- (baja), y D (muy baja). Las calificaciones de C- y D significan que el estudiante suspende la actividad. Además, se utiliza otra calificación (N, no presentado) cuando un estudiante no presenta la actividad. La evaluación continua se complementa, generalmente, con un examen presen-

cial al final del semestre. La calificación final de la asignatura se calcula en base a una fórmula predefinida donde cada actividad, y el propio examen tienen un peso determinado dependiendo de su importancia.

El sistema de detección temprana también tiene asociado un sistema semiautomático de envío de retorno personalizado en base al nivel de riesgo del estudiante. Estos mensajes proponen recomendaciones a los estudiantes para evitar o superar la situación de riesgo según su nivel de alerta.

3.1. Modelo de predicción

El modelo de predicción, denominado modelo *Gradual At-Risk* (GAR), se basa en las notas de las actividades (PEC y PR) para intentar predecir si, a partir de las notas que ha obtenido un estudiante en las actividades, puede superar la asignatura. Existe un modelo por asignatura y, realmente, un modelo GAR está formado por un conjunto de submodelos, uno para cada actividad propuesta en la asignatura. Un submodelo toma como datos de entrada las notas de las actividades propuestas desde el inicio del semestre en la asignatura hasta la nota de la actividad a la que está asociado, y proporciona como salida la predicción de si un estudiante no superará la asignatura. Es decir, la información de salida es binaria (superará, suspenderá) la asignatura.

Ejemplo 1. Supongamos una asignatura con 5 actividades. El modelo GAR estará compuesto de 5 submodelos:

$$\begin{aligned} Pr_{A1}(Susp?) &= (Nota_{A1}) \\ Pr_{A2}(Susp?) &= (Nota_{A1}, Nota_{A2}) \\ Pr_{A3}(Susp?) &= (Nota_{A1}, Nota_{A2}, Nota_{A3}) \\ Pr_{A4}(Susp?) &= (Nota_{A1}, Nota_{A2}, Nota_{A3}, Nota_{A4}) \\ Pr_{A5}(Susp?) &= (Nota_{A1}, Nota_{A2}, Nota_{A3}, Nota_{A4}, Nota_{A5}) \end{aligned}$$

donde $Pr_{An}(Susp?)$ indica el nombre del submodelo para predecir si un estudiante suspenderá la asignatura después de la actividad A_n . Para hacer la predicción, cada submodelo utiliza las notas ($Nota_{A1}, Nota_{A2}, \dots, Nota_{An}$), es decir, las notas desde la primera actividad hasta la actividad A_n .

Los submodelos tienen una precisión determinada. Existen diferentes métricas para evaluar dicha precisión. En nuestro caso nos centramos en dos:

$$TNR = \frac{TN}{TN+FP} \qquad TPR = \frac{TP}{TP+FN}$$

donde TP indica el número de estudiantes correctamente identificados en riesgo, TN el número de estudiantes correctamente identificados que no están en riesgo, FP el número de estudiantes en riesgo no correctamente identificados, y FN el número de estudiantes que no están en riesgo y que no han sido correctamente identificados. Las métricas evalúan la precisión en la detección de estudiantes en riesgo (TPR) y la precisión en la detección de estudiantes

que no están en riesgo (TNR). Existen otras métricas que se podrían aplicar para evaluar los submodelos de forma global, como la precisión o el valor-F (F-score), pero se han seleccionado el TPR y el TNR ya que cada uno indica la precisión específica de detectar estudiantes en riesgo y no en riesgo, respectivamente.

Este modelo de predicción es útil para predecir si un estudiante tiene posibilidades de suspender una asignatura, pero no es muy útil para dar información a corto plazo, es decir, para la siguiente actividad. Por ello se define el modelo denominado *Next Activity At-Risk* (NAAR) que da información sobre qué nota se debe obtener en la siguiente actividad para tener garantías de superar la asignatura. El modelo NAAR explota el modelo anterior, ejecutando el submodelo correspondiente a la actividad donde queremos predecir la nota mínima. El modelo NAAR utiliza las notas del estudiante para las actividades anteriores ya evaluadas y simula las diferentes notas posibles para la actividad para la que se quiere dar una predicción.

Ejemplo 2. Supongamos el submodelo $Pr_{A1}(Susp?)$ asociado a la primera actividad del Ejemplo 1. Para saber la nota mínima que se debe obtener en esta actividad para tener una alta probabilidad de superar la asignatura, se ejecuta el modelo de predicción para las 6 posibles notas que se puede obtener para la primera actividad. Mostramos a continuación la salida del modelo para la asignatura de Uso de Bases de Datos que analizaremos en la Sección 4.

$$\begin{aligned} Pr_{A1}(Susp?) = (N) & \rightarrow Susp? = Sí \\ Pr_{A1}(Susp?) = (D) & \rightarrow Susp? = Sí \\ Pr_{A1}(Susp?) = (C-) & \rightarrow Susp? = Sí \\ Pr_{A1}(Susp?) = (C+) & \rightarrow Susp? = No \\ Pr_{A1}(Susp?) = (B) & \rightarrow Susp? = No \\ Pr_{A1}(Susp?) = (A) & \rightarrow Susp? = No \end{aligned}$$

donde podemos observar que un estudiante con una nota mínima de C+ en la primera actividad tiene una alta probabilidad de superar la asignatura. Existen estudiantes que suspendiendo esta actividad también superarán la asignatura, pero es menos frecuente. Esta predicción se realiza durante la realización de la actividad por parte del estudiante. Por lo tanto, recibir esta información antes de enviar la actividad para su evaluación, es una información muy valiosa para un estudiante para adaptar su esfuerzo.

3.2. Sistema de detección temprana de estudiantes en riesgo

Mediante estos dos modelos se ha implementado el sistema de detección de estudiantes en riesgo. Este sistema ejecuta de forma automática las predicciones, ya que el sistema tiene acceso a un *data mart* [11] de la universidad que almacena los datos históricos del funcionamiento de todas las asignaturas y, además, se actualiza diariamente con los datos de utilización del campus virtual por parte de los estudiantes (acceso a



Figura 1: Cuadro de mando del estudiante.

aulas, y a los recursos de aprendizaje, interacciones con otros estudiantes etc.), así como otros datos de los estudiantes (por ejemplo, sus notas parciales de las asignaturas matriculadas). Actualmente el *data mart* de la universidad contiene datos verificados (es decir, correctos) desde el curso académico 2016/2017. Los diferentes submodelos del modelo GAR se entrenan con los datos históricos de la asignatura con diferentes algoritmos de clasificación. Concretamente, se utilizan Naïve Bayes (NB), CART Decision Tree (DT), K-Nearest Neighbors (KNN) y Support Vector Machine (SVM). Los algoritmos entrenados se testean con los datos del semestre anterior al actual, y el clasificador con la mejor suma entre el TPR y TNR se selecciona como el algoritmo que se utilizará para generar las predicciones. Maximizar esta suma tiende a dar el mejor algoritmo para detectar todo tipo de estudiantes, mientras que la maximización de una de las dos métricas tiende a penalizar la descartada y por lo tanto una de las dos posibles predicciones (superar, suspender).

Como resultado, se genera una visualización (o cuadro de mando) dirigida a los estudiantes que muestra la nota mínima que deben obtener en la siguiente actividad, cuando se evalúa la actividad previa. En concreto la Figura 1.a) muestra la visualización correspondiente al Ejemplo 2 de la asignatura de Uso de bases de datos una vez generada la predicción para la primera actividad mediante el modelo NAAR. El estudiante puede ver que con una nota mínima de C+ en la actividad tiene opciones de superar la asignatura, así como las probabilidades de cada opción. Estas probabilidades corresponden al TPR para el caso de suspender (estudiante en riesgo), y al TNR para el caso de superar la asignatura.

Es importante indicar en la visualización la precisión de la predicción, ya que de esta forma el estudiante se hace una idea de cómo debe interpretarla. En el caso de un valor bajo, quiere decir que el mode-

lo no puede asegurar la predicción. Sin embargo, en el caso de un valor alto, la predicción se vuelve más plausible. Como es muy posible que los estudiantes no entiendan correctamente la relación entre las métricas de predicción, y el nivel de riesgo asociado, el sistema indica un nivel de riesgo mediante un semáforo con colores rojo, amarillo (o ámbar) y verde, que significan, respectivamente, un nivel de riesgo alto, moderado y bajo de suspender la asignatura, similar al propuesto en [2]. Estos niveles se computan en base a la predicción y el valor de las métricas una vez se ha evaluado la actividad y el estudiante tiene la nota. El resultado se puede ver en la Figura 1.b) donde el estudiante se le asigna un nivel de riesgo mediante el semáforo de la parte superior y en ese momento se genera la predicción para la siguiente actividad. La visualización también muestra mediante triángulos la nota obtenida por el estudiante en la actividad (en este caso una A).

El nivel de riesgo se computa mediante el árbol de decisión mostrado en la Figura 2, donde se observa que los diferentes niveles de riesgo se fundamentan además en un umbral mínimo para el TPR y TNR. Este umbral se ha establecido al 75%, y todo submodelo con una precisión mayor o igual a este valor se considera un modelo de calidad. El color verde se asigna a los estudiantes cuando la nota de la actividad es superior a la predicción y el TNR es superior al 75%. Si la nota es inferior a C+ e inferior a la predicción, el nivel de riesgo es alto (rojo), y en el caso que el estudiante supere la actividad (nota superior o igual a C+) pero sea inferior a la predicción, el nivel de riesgo es amarillo. Finalmente, podemos observar que los niveles de riesgo verde y rojo se rebajan a amarillo, respectivamente, cuando la precisión del TNR y TPR son inferiores al 75%. Los profesores tienen un color adicional que es el negro. Este color identifica los estudiantes con posibilidades de abandonar la

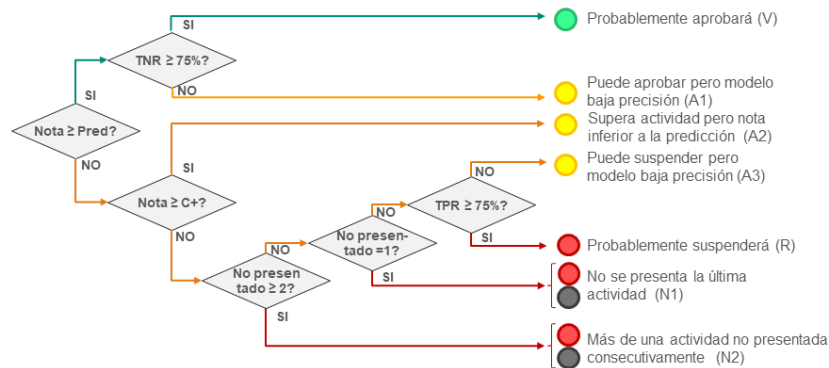


Figura 2: Árbol de decisión para computar el nivel de riesgo.

asignatura. Mientras que los estudiantes ven el color rojo de riesgo alto, los profesores ven el color negro.

3.3. Sistema de retorno

Es interesante observar que el árbol de decisión utilizado para determinar el nivel de riesgo de los estudiantes diferencia entre tres niveles de amarillo (A1, A2 y A3) y dos niveles de posible abandono (B1 y B2). La razón de esta diferenciación es para la segunda característica del sistema de detección: el sistema semiautomático de envío de retorno personalizado. El sistema envía un mensaje diferente para cada posible nivel de riesgo al correo personal del estudiante en nombre del profesor de la asignatura. Es semiautomático porque, si bien el envío se realiza de forma automática cuando se computa el nivel de riesgo del estudiante tras la corrección de una actividad, el diseño del texto de los diferentes mensajes es una tarea manual. Teniendo en cuenta que el profesor de la asignatura es quien tiene el conocimiento y experiencia sobre cómo se comportan los estudiantes y qué necesitan a nivel de recursos de aprendizaje, conocimientos y competencias, el profesor diseña y redacta un mensaje para cada nivel de riesgo y actividad. Además, en el caso de estudiantes en niveles de riesgo alto, el profesor insta los estudiantes a ponerse urgentemente en contacto con él, para hacer un asesoramiento y un seguimiento más individualizado.

4. Prueba piloto

En esta sección presentamos el contexto donde se ha desarrollado la prueba piloto y los resultados obtenidos (precisión del modelo GAR, precisión en la correcta clasificación de los estudiantes en los niveles de riesgo y la opinión de los estudiantes).

4.1. Contexto

El piloto se ha llevado a cabo en la asignatura de Uso de Bases de Datos (UBD) del Grado de Ingeniería Informática. Es una asignatura obligatoria de 6 créditos ECTS que se cursa en el cuarto semestre. El número de matriculados se sitúa entre 200 y 250 estudiantes. La interacción entre estudiantes y profes-

sores se lleva totalmente en línea en un aula del campus virtual que contiene diferentes espacios de comunicación, el plan docente, el registro de evaluación continua y el conjunto de recursos de aprendizaje de la asignatura. En la asignatura los estudiantes aprenden qué son las bases de datos (y el *software* especializado que las gestiona), los fundamentos del modelo relacional, cómo crear y manipular bases de datos relacionales utilizando el lenguaje SQL (interactivo y embebido), gestión de transacciones y otros modelos de bases de datos (NoSQL).

Durante el transcurso de la asignatura, los estudiantes deben realizar 2 PEC (modelo relacional y álgebra relacional, y gestión de transacciones), un total de 3 PR en PostgreSQL (sentencias básicas de SQL, disparadores y procedimientos almacenados, y JDBC, respectivamente), y un examen final presencial al final del semestre (EX). Para el cálculo de la nota final (N_F), se aplica la siguiente fórmula:

$$N_F = \text{MAX}(32,5\% \text{ Nota}_{EX} + 32,5\% \text{ Nota}_P + 21\% \text{ Nota}_{PEC1} + 14\% \text{ Nota}_{PEC2}, 50\% \text{ Nota}_{EX} + 50\% \text{ Nota}_P)$$

Mientras que las PEC son opcionales, el estudiante debe presentar por lo menos dos de las PR para poder superar la asignatura. La nota final de las PR (Nota_P) se calcula como la media entre las dos mejores notas. Para aplicar la fórmula se exige una calificación mínima de C- en Nota_P y Nota_{EX} . Aunque hay estudiantes que aprueban la asignatura sin hacer las PEC, los profesores saben por experiencia que la mayoría de los estudiantes que aprueban las resuelven, ya que se trabajan conceptos útiles para la realización posterior de las PR y el examen final. El sistema ha realizado una predicción para cada estudiante y actividad (PEC/PR) con suficiente antelación a la fecha límite de entrega. Además, una vez evaluada la actividad, se ha actualizado el nivel de riesgo y se ha enviado un mensaje de recomendación al estudiante en base a su nivel de riesgo. El sistema ha realizado un total de 355 predicciones. Este volumen creemos que es suficiente para poder analizar la precisión del sistema.

Los profesores de la asignatura indican que, con un sistema de detección temprana, se podría incidir mejor en el retorno que se proporciona al estudiante,

	TP	FP	TN	FN	TNR	TPR	Alg.
P _{PEC1}	30	9	77	14	89.5	68.2	KNN
P _{PR1}	33	5	81	11	94.2	75.0	KNN
P _{PR2}	38	4	82	6	95.3	86.4	SVM
P _{PEC2}	37	3	83	7	96.5	84.1	NB
P _{PR3}	39	2	84	5	97.7	88.6	KNN

Cuadro 1. Precisión del modelo GAR para UBD

y acompañarlo mejor en su proceso de aprendizaje. El escenario, pues, ha sido realmente adecuado para testear el sistema y poder analizar los resultados. El piloto se ha desarrollado durante el semestre de otoño del curso académico 2019/2020, y han participado un total de 71 de 203 estudiantes matriculados (para participar en el piloto se firma un consentimiento informado, ya que se utilizan datos personales).

4.2. Resultados

Para poder valorar la correcta clasificación de los estudiantes a los diferentes niveles de riesgo, primero necesitamos evaluar la precisión del modelo GAR en UBD (véase el Cuadro 1), donde se muestran las métricas TNR y TPR para cada submodelo, junto con el clasificador seleccionado. Para entrenar los submodelos se han utilizado los datos desde el semestre de otoño 2016 hasta otoño 2018, y para testear los submodelos se han utilizado los datos del semestre de primavera de 2019. El modelo GAR tiene la mayoría de los submodelos con una precisión superior al umbral mínimo del 75%, excepto el TPR de la PEC1 (68,2%). Esto significa que el modelo de predicción no puede asegurar que un estudiante que no supere la PEC1 suspenderá la asignatura. Por el contrario, el hecho de superar la PEC1 es un buen indicador que el estudiante superará la asignatura (TNR 89,5%). Para el resto de las actividades, las métricas son superiores a 75%. Por lo tanto, los submodelos son de calidad.

También se ha comprobado si hay significación estadística en la distribución de la nota final comparando el semestre actual con respecto al semestre anterior, es decir, el semestre de primavera de 2019. El objetivo es ver si el sistema ha impactado en el

rendimiento del curso. Utilizamos la prueba de Wilcoxon-Mann-Whitney debido a la distribución no normal de la nota final. Aquí, asumimos como hipótesis nula que las notas son peores o iguales que en el semestre anterior. El valor p (p-value) es inferior a 0.02 y, por lo tanto, podemos rechazar la hipótesis nula. Además, la mediana de la nota final ha aumentado de 7 a 7.6, y el abandono ha disminuido del 26% al 17%. Aunque otras variables, como la dificultad de las actividades o los profesores que imparten docencia pueden influir en estos valores, es posible que la herramienta también haya ayudado.

Para saber si los estudiantes se han clasificado correctamente, se ha comparado el nivel de riesgo asignado a los estudiantes en cada actividad con el resultado obtenido por los estudiantes al finalizar la asignatura. El Cuadro 2 resume los resultados. Se puede observar, para cada nivel de riesgo (NR) y actividad (PEC/PR), el número (N.) de predicciones realizadas para el nivel de riesgo, y el porcentaje de estudiantes clasificados en el nivel que han suspendido (Susp.) y aprobado (Aprob.) al final del curso. Nótese que el modelo GAR es capaz de clasificar correctamente la mayoría de los casos. En un principio, en la PEC1, la mayoría de los estudiantes están clasificados en el nivel de riesgo bajo (V de verde) y el sistema acierta más del 78%. Los estudiantes en riesgo se asignan al nivel amarillo A3 ya que, como hemos indicado, el TPR es inferior al 75%. A pesar de ello, serían potencialmente estudiantes en riesgo, y vemos que el sistema identifica la mayoría de ellos. En el caso de los estudiantes que no presentan, observamos que la mayoría acabarán suspendiendo, probablemente porque ni empiezan el estudio de la asignatura. También podemos argumentar que el resto de los estudiantes, gracias a las recomendaciones realizadas, acaban realizando el número mínimo de PR y el examen, y superan la asignatura. Para el resto de las actividades, la tasa de acierto se incrementa, sobre todo en las PR, y para los diferentes niveles de riesgo.

Finalmente, se ha analizado la opinión de los estudiantes. El sistema, una vez producida una predicción, ofrecía al estudiante un breve cuestionario (responderlo era opcional) con tres preguntas: 1)

NR	PEC1			PR1			PR2			PEC2			PR3		
	N.	Susp.	Aprob.	N.	Susp.	Aprob.	N.	Susp.	Aprob.	N.	Susp.	Aprob.	N.	Susp.	Aprob.
V	55	21,82%	78,18%	49	10,20%	89,80%	47	6,38%	93,62%	46	6,52%	93,48%	54	12,96%	87,04%
A1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
A2	--	--	--	2	100,00%	--	2	50,00%	50,00%	4	50,00%	50,00%	--	--	--
A3	9	66,67%	33,33%	--	--	--	--	--	--	--	--	--	--	--	--
R	--	--	--	15	73,33%	26,67%	5	80,00%	20,00%	1	100,00%	--	1	--	100,00%
N	7	71,43%	28,57%	3	100,00%	--	13	84,62%	15,38%	5	60,00%	40,00%	--	--	--
N2	--	--	--	2	100,00%	--	4	100,00%	--	15	93,33%	6,67%	16	100,00%	--

Cuadro 2. Tasa de acierto del nivel de riesgo en comparación al resultado final del estudiante para UBD

¿Encuentras útil la predicción obtenida?; 2) ¿Continuarás en la asignatura después de esta predicción?; 3) Indica tu estado de ánimo después de la predicción. Estas preguntas eran de tipo Likert con 5 posibles opciones (donde el valor 1 indica totalmente desfavorable y el valor 5 totalmente satisfactorio). Los cuestionarios, al estar dentro del sistema, no son anónimos. Esto permite relacionar la opinión con el estudiante y su nivel de riesgo para la actividad (PEC/PR). Los resultados se muestran en el Cuadro 3 donde, para cada nivel de riesgo y actividad, se resume el número de respuestas (N.) y la media para las diferentes preguntas (Útil, Cont. y Ánimo). Se han recibido 28, 31, 37, 30 y 32 respuestas de un total de 71 estudiantes. Esto significa que el margen de error es de aproximadamente un 11% con un intervalo de confianza del 95%. No es una buena ratio para extraer conclusiones, pero permite analizar las opiniones de los estudiantes. En el caso de los estudiantes con nivel de riesgo bajo (V), las respuestas son positivas en todas las actividades, es decir, los estudiantes encuentran útiles las predicciones, continuarán en la asignatura y su estado de ánimo es bueno con una ponderación superior al 70%. Además, las valoraciones se incrementan a medida que van superando actividades. El resto de los niveles de riesgo son difíciles de valorar por el bajo número de respuestas recibidas (por razones de espacio, solo se muestran los niveles de riesgo para los que se han recibido respuestas). A pesar de ello, en comparación, cuando hay más de una respuesta, la valoración tiende a ser inferior. Aunque tienden a valorar positivamente su utilidad, muchos estudiantes dudan de su capacidad para continuar la asignatura y esto se relaciona con su estado ánimo que, en general, es inferior a 50%.

5. Discusión

Los sistemas de detección de estudiantes en riesgo pueden ser de gran utilidad en cursos con un alto número de estudiantes. Permiten escalar de forma efectiva el envío de retorno a los estudiantes. En este trabajo, se ha combinado el sistema de detección con el mecanismo de intervención basado en recomendaciones. En el caso de entornos de enseñanza-aprendizaje en línea, es un sistema muy valorado por los estudiantes, puesto que permite optimizar esfuerzos, a la vez que promueve la comunicación entre profesores y estudiantes.

Estos sistemas no solo son útiles en instituciones en línea. Su desarrollo solo depende de una política de gestión de la información perfectamente definida por la institución y que los datos almacenados puedan ser explotados por estos sistemas. Por ejemplo, la universidad de Purdue [2] desarrolló hace más de 10 años un sistema de detección. Esta universidad presencial implementó otros mecanismos de

PEC1				
NR	N.	Útil	Cont.	Ánimo
V	27	3,59	3,67	3,74
N1	1	4	4	4
PR1				
NR	N.	Útil	Cont.	Ánimo
V	25	3,84	3,72	3,64
A2	1	4	4	4
R	4	3,5	3,25	3,25
N1	1	3	3	1
PR2				
NR	N.	Útil	Cont.	Ánimo
V	30	3,7	3,8	3,83
A2	1	5	5	3
R	1	3	3	3
N1	3	3,67	2,33	2,68
N2	2	2	3	3,5
PEC2				
NR	N.	Útil	Cont.	Ánimo
V	26	3,92	3,88	3,88
A2	1	3	3	3
N1	1	4	4	4
N2	2	3	2,5	1,5
PR3				
NR	N.	Útil	Cont.	Ánimo
V	29	3,86	3,86	4
R	1	4	2	2
N2	2	3,5	3,5	3

Cuadro 3. Opinión de los estudiantes de UBD clasificado por nivel de riesgo

intervención para ayudar a sus estudiantes. Además de ofrecer cuadros de mando y envío de recomendaciones por correo electrónico, se aplicaron otros mecanismos como un servicio de orientación telefónico y/o presencial con el profesor o tutor asignado, y cursos de iniciación, donde se explicaba técnicas de gestión del tiempo o formas efectivas de estudio a los estudiantes.

En nuestro caso, no hemos explorado estos mecanismos en profundidad, ya que en un entorno virtual son difícilmente escalables. Es complicado que profesores y estudiantes puedan coincidir presencialmente o por teleconferencia. De todas formas, en los casos de riesgo alto o de posible abandono, el mensaje enviado por el profesor urgía al estudiante a ponerse en contacto con él/ella para explicarle los problemas que hubiera podido tener. Es decir, el mensaje propuso iniciar una conversación con el profesor para poder solucionar cualquier problema que pudiera tener [3]. Esta opción fue utilizada por 7 estudiantes en la asignatura. Aunque sean pocos, los estudiantes valoraron muy positivamente la actitud proactiva del profesor.

No queremos finalizar esta sección sin apuntar las limitaciones de nuestro trabajo. En términos de validez interna, el sesgo de autoselección y la mortalidad pueden afectar el análisis cuantitativo y cualitativo efectuado. Los estudiantes dieron su consentimiento para ser incluidos en el piloto, porque así lo requiere nuestro comité ético de investigación. Son los estudiantes más comprometidos los que tienden a participar, y su rendimiento suele ser mejor que el promedio

de la asignatura. Por su parte, la mortalidad puede afectar el análisis cualitativo, ya que los estudiantes que abandonan dejan de interactuar con el sistema.

6. Conclusiones y trabajo futuro

En este trabajo hemos presentado un sistema de detección temprana de estudiantes en riesgo. Aunque el sistema se ha testeado en un entorno virtual, su uso es perfectamente viable en entornos presenciales. Además, estos sistemas permiten implementar mecanismos de intervención para ayudar a los estudiantes a superar la situación de riesgo. En este trabajo se ha aplicado el retorno personalizado, aunque existen otros mecanismos como tutorización u orientación.

Como trabajo futuro se estudiará la creación de nuevos modelos de predicción como sería la detección de caminos de aprendizaje óptimos y la recomendación automática de recursos y/o actividades adicionales para adquirir las competencias necesarias para superar una asignatura.

Agradecimientos

Este trabajo ha sido financiado por el proyecto New Goals 2018NG001 “LIS: Learning Intelligent System” del eLearn Center (UOC).

Referencias

- [1] Rosa Alcover, José Benlloch, Pedro Blesa, M. Ángeles. Calduch, Meri Celma, César Ferri, José Hernández-Orallo, Laura Iniesta, Jordi Más, M. José. Ramírez-Quintana, Antonio Robles, José. Miguel Valiente, M. José Vicent, y Lemus R. Zúnica, “Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos”. *En Actas de las XIII Jornadas de Enseñanza Universitaria de la Informática, Jenui 2007*, pp. 163-170, Te-ruel, 2007.
- [2] Kimberly E. Arnold y Matthew D. Pistilli “Course signals at Purdue: Using learning analytics to increase student success”. *Proceedings of the 2nd international conference on learning analytics and knowledge (LAK'12)*. pp. 267–270, 2012.
- [3] Alexander W. Astin. “Student involvement: A developmental theory for higher education”. *Journal of College Student Development*, vol. 40, no. 5, pp. 418–429, 1999.
- [4] Agustín Cernuda del Río, Sonia Hevia Vázquez, María del Carmen Suárez Torrente, y Daniel Gayo-Avello. “Un estudio sobre el absentismo y el abandono en asignaturas de programación”. *ReVisión*, vol. 6, no. 1, 2013
- [5] Alfonsa García, Ana Lías, Ángeles Mahillo, y Rosa M.ª Pinero, “Abandono de primer año en la Ingeniería Informática”, *En Actas de las XX Jornadas de Enseñanza Universitaria de la Informática, Jenui 2014*. pp. 151-158, Oviedo, 2014.
- [6] Carmen Lacave Rodero, Ana Isabel Molina Díaz, Miguel Ángel Redondo Duque, y Manuel Ortega Cantero. “Redes bayesianas para identificar perfiles de estudiante. Aplicación al estudio del abandono de las titulaciones de Informática en la Universidad de Castilla-La Mancha.” *En Actas de las XXIII Jornadas de Enseñanza Universitaria de la Informática, Jenui 2016*, pp. 85-92, Almería, 2016.
- [7] Ioanna. Lykourantzou, Ioannis Giannoukos, Vassilis. Nikolopoulos, George Mpardis, y Vassili Loumos, “Dropout prediction in e-learning courses through the combination of machine learning techniques”, *Computers & Education*, vol. 53, no. 3, pp. 950–965, 2009.
- [8] Leah P. Macfadyen, y Shane Dawson, “Mining LMS data to develop an early warning system for educators: A proof of concept”, *Computers & Education*, vol. 54, no. 2, pp. 588–599, 2010.
- [9] Leah P. Macfadyen, y Shane. Dawson, “Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan”, *Journal of Educational Technology & Society*, vol. 15, no. 3, pp. 149–163, 2012.
- [10] Carlos Márquez-Vera, Alberto Cano, Cristobal Romero, Amin Youseff M. Noaman, Habbid Mousa Fardoun, y Sebastian Ventura, “Early dropout prediction using data mining: a case study with high school students”, *Expert Systems*, vol. 33, no. 1, pp. 107–124, 2016.
- [11] Julià Minguillón, Jordi Conesa, M. Elena Rodríguez, y Francesc Santanach, “Learning analytics in practice: providing e-learning researchers and practitioners with activity data”, *Frontiers of Cyberlearning: Emerging Technologies for Teaching and Learning*, pp. 145–164, Lecture Notes in Educational Technology, Singapore: Springer, 2018.
- [12] Carlos Romero, y Sebastian Ventura, “Data mining in education”, *WIRES Data Mining Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.
- [13] George. Siemens, and Ryan S. J. d. Baker, “Learning analytics and educational data mining: towards communication and collaboration”, *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12)*, pp. 252–254. 2012.