



SELECCIÓN DE ATRIBUTOS PREDICTIVOS DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES EN UN MODELO DE B-LEARNING

PREDICTIVE FEATURE SELECTION OF ACADEMIC EFFICIENCY IN A B-LEARNING MODEL

Angel Cobo Ortega ; acobo@unican.es

Rocio Rocha Blanco ; rochar@unican.es

Universidad de Cantabria (España)

Yurlenis Alvarez Díaz ; yurlead@gmail.com

Universidad de Holguín "Oscar Lucero Moya" (Cuba)

RESUMEN

La participación de estudiantes en iniciativas de *b-learning* genera una gran cantidad de datos e indicadores que no siempre son adecuadamente analizados por los docentes. Las plataformas de formación virtual permiten gestionar de manera óptima dichos indicadores. En este trabajo se aplican técnicas de minería de datos para identificar aquellos indicadores que puedan tener mayor valor predictivo, a la hora de medir el rendimiento de los estudiantes, en el contexto de una asignatura de grado que combina actividades docentes presenciales con actividades soportadas en aplicaciones de teleformación.

PALABRAS CLAVE: Aprendizaje a distancia, aprendizaje mixto, innovación educativa, selección de rasgos, entornos sociales de aprendizaje, minería de datos.

ABSTRACT

Student participation in b-learning initiatives generates a large amount of data and indicators that are not always properly used by teachers. Virtual learning platforms enable an optimal management of these indicators. In this paper we apply data mining techniques to identify indicators that may be used in assessing student performance in the context of a b-learning course that combines classroom teaching activities with virtual activities.

KEYWORDS: E-learning, blended learning, educational innovation, social learning environments, feature selection, data mining.



1. INTRODUCCIÓN

Los actuales modelos formativos proponen cambios en el paradigma educativo, donde el estudiante se sitúa en el centro del proceso del aprendizaje y se manifiesta la necesidad de metodologías orientadas al aprendizaje activo. El profesor deja de ser el centro de atención, pasando a ser el estudiante el verdadero protagonista del proceso de enseñanza-aprendizaje. Los entornos virtuales de enseñanza favorecen la implementación efectiva de estos nuevos modelos de enseñanza-aprendizaje, la interacción entre los actores participantes en el proceso, facilitan la difusión de materiales por medios electrónicos, y permiten que tanto el profesor como el estudiante se encuentren en diferentes entornos físicos. La utilización de las redes de comunicaciones y las comunidades virtuales generadas en el contexto de entornos virtuales de aprendizaje, se manifiesta como un elemento fundamental para la promoción de la participación activa, y condición básica para la creación de conocimiento. Para lograr estos objetivos las herramientas de comunicación, tanto síncrona como asíncrona, y muy especialmente estas últimas, deben ser elementos esenciales y cuya utilización en experiencias de *e-learning* debe ser cuidadosamente planificada. En especial, el profesor debe pasar a desempeñar un papel de moderador y guía en el proceso de construcción de conocimiento (Salmon, 2000).

Durante muchos años se defendió la tendencia de ir caminando hacia la utilización de las nuevas tecnologías para desarrollar modelos de enseñanza no presencial o virtual, los conocidos como modelos de *e-learning*. Esta modalidad educativa es totalmente a distancia y se desarrolla a través de ambientes virtuales flexibles, abiertos e interactivos. Estos ambientes ofrecen todos los materiales necesarios para el aprendizaje del estudiante e integran elementos interactivos de comunicación tales como foros, chat, videoconferencias, multimedia, presentaciones y documentos que guían el desarrollo cognitivo. Bartolomé y Aiello argumentan que los modelos de enseñanza no presencial se han mostrado eficaces para ciertas situaciones pero su aplicación conlleva ciertos límites a una gran masa de población que no posee las características adecuadas para llevar a cabo un aprendizaje basado en dichos modelos (Bartolomé, 2004; Bartolomé & Aiello, 2006). Los estudiantes se desorientan por la falta de comunicación u orientación directa (March *et al.*, 2004) y tienden a estar menos satisfechos con los cursos online cuando los comparan con las clases tradicionales, encuentran dificultades en la vinculación entre la información presentada en el modo *online* y la instrucción presencial debido a la falta de familiaridad en el modo de presentación. Diferentes análisis ponen de manifiesto unas mayores tasas de fracaso y abandono en los modelos cuando se aplican las nuevas tecnologías casi con exclusividad. Por todo ello, diversos estudios concluyen que una mezcla en el aprendizaje, *face-to-face* y el aprendizaje *online*, resulta más efectiva (Murphy, 2002). Ahora cada vez más se tiende a defender los modelos flexibles, semipresenciales o de enseñanza mixta que se agrupan bajo el término de *blended learning* (*b-learning*). Estos procedimientos mixtos buscan adaptar las nuevas tecnologías al aula tradicional y no plantear un proceso de enseñanza-aprendizaje a distancia, totalmente *on-line*. El *b-learning* es una modalidad que permite minimizar las limitaciones de espacio y tiempo que exige la enseñanza convencional, flexibilizar los procesos de aprendizaje y aprovechar al máximo los recursos de las tecnologías digitales (Pons & Moreno, 2005). Este modelo proporciona más responsabilidad a los estudiantes en su estudio individual, y mejora la calidad de las clases mediante el uso de herramientas tecnológicas. La emergencia del *b-learning*, en efecto, va respondiendo a un nuevo contexto social que demanda una renovada organización pedagógica que relacione el proceso



tecnológico y social de cambio con la innovación educativa (Aiello & Cilia, 2004). Este tipo de soluciones se requerirán cada vez más en entornos abiertos y a distancia que incorporen componentes de actividades presenciales; y por otro lado, también se requerirán en entornos presenciales que aprovechen las ventajas de la combinación de recursos y métodos de ambos mundos (Peñalosa *et al.*, 2010).

El objetivo de este trabajo es mostrar cómo a través de las herramientas de gestión de entornos virtuales de enseñanza puede obtenerse valiosa información que, combinada con la obtenida directamente por el profesor en las actividades presenciales, puede ser útil para predecir el rendimiento de los estudiantes o identificar patrones de comportamiento. Esta información puede servir incluso de base para establecer estrategias de enseñanza-aprendizaje personalizadas. En concreto, en este trabajo, se tratará de aplicar técnicas de minería de datos para la selección de los rasgos más determinantes a la hora de realizar clasificaciones de estudiantes de acuerdo a su actividad y rendimiento académico.

2. DESCUBRIMIENTO DE CONOCIMIENTO CON MINERÍA DE DATOS EN ENTORNOS VIRTUALES DE APRENDIZAJE

En un proyecto de *b-learning*, el profesor puede obtener directamente en el contacto presencial con el estudiante información sobre la evolución en el proceso de adquisición de competencias por parte de éste. Pero además, los sistemas virtuales de aprendizaje aportan información igualmente valiosa, como por ejemplo, número y frecuencia de accesos, distribución temporal y física de dichos accesos, registro de actividades realizadas, de páginas visitadas, además de indicadores de rendimiento en tareas y actividades de evaluación soportadas por las herramientas de e-learning. Todos estos datos pueden ser analizados con técnicas de minería de datos para extraer nuevo conocimiento.

Todo proyecto de minería de datos debe pasar por cuatro etapas básicas (Ferri *et al.*, 2004; Witten & Frank, 2005):

1. Determinar los objetivos que se quieren alcanzar y obtener los datos a procesar.
2. Pre-procesamiento, que consistirá en limpiar, transformar y reducir los datos. Las operaciones que transforman los datos pueden ser la reducción o aumento de dimensionalidad, discretización (conversión de un valor numérico a un valor nominal ordenado) y normalización (transformación de los valores de las variables cuando tienden a tener rangos distintos). Esta etapa es la que consume más de la mitad del tiempo del esfuerzo de desarrollo.
3. Aplicación de técnicas de explotación de datos, determinando el modelo a utilizar, realizando análisis estadísticos y visualizando gráficamente los datos para tener una primera aproximación.
4. Análisis y evaluación de resultados obtenidos, verificando si los resultados son coherentes y cotejándolos con lo analizado estadística y gráficamente



Tomando como referencia estas 4 etapas básica, a continuación se describen las actividades realizadas en cada una de ellas en el contexto de este trabajo.

2.1. Determinación de los objetivos del trabajo y recopilación de datos

El objetivo principal ha sido tratar de identificar, de entre toda la información que puede ser recopilada, los rasgos o indicadores de los estudiantes que mejor pueden servir para predecir su rendimiento académico en el contexto de una iniciativa de b-learning. Los rasgos provienen de una cuantificación de la actividad del estudiante tanto en acciones presenciales como no presenciales; en este sentido se busca comprobar que tipo de indicadores (de actividad presencial o no presencial) tienen mayor poder discriminante a la hora de analizar el rendimiento.

Para realizar este estudio se optó por recopilar datos generados en el contexto de una asignatura de introducción a técnicas de optimización en el primer curso de dos titulaciones del ámbito de la economía y la administración de empresas en la Universidad de Cantabria. Dicha asignatura participa en un programa de innovación docente en el que parte de la docencia presencial es sustituida por actividades que se apoyan en la plataforma de enseñanza virtual *Moodle*. Igualmente se utiliza la plataforma para plantear diferentes acciones de evaluación a distancia durante el desarrollo del curso.

La asignatura está dividida en 4 grandes bloques temáticos en cada uno de los cuales se planifican un conjunto de actividades presenciales y no presenciales, estas últimas desarrolladas a través de la plataforma virtual y que buscan fomentar el trabajo colaborativo de los estudiantes a través de debates en los foros, planteamiento de trabajos y estudio de casos prácticos, realización de actividades de evaluación con soporte virtual y autoaprendizaje en la utilización de software de optimización. Algunas de las actividades no presenciales tienen asociadas diferentes acciones de evaluación presencial para comprobar la efectiva adquisición de conocimientos y competencias por parte de los estudiantes. Así por ejemplo, se realizan pruebas no programadas de contraste en las sesiones presenciales tras finalizar el periodo de realización de las pruebas de evaluación a distancia en el curso virtual. También se realizan pruebas de verificación de destrezas sobre las herramientas informáticas utilizadas.

La asignatura se impartía a un total de 377 estudiantes, de dos titulaciones diferentes y repartidos en un total de 4 grupos docentes. Para determinadas actividades prácticas los grupos a su vez se desdoblaban. La asignatura se desarrollaba durante un cuatrimestre. Al inicio del curso se informó a todos los estudiantes de la posibilidad de participar voluntariamente en un proyecto de innovación docente en el que se plantearían diferentes actividades a lo largo del curso. Se advertía que tanto las actividades presenciales como las no presenciales, tendrían reflejo en la evaluación final de la asignatura. Aquellos estudiantes que en cualquier momento decidieran abandonar el proyecto tendrían la posibilidad de presentarse directamente al examen de evaluación final de la asignatura en el que podrían obtener también la calificación máxima. De los 377 estudiantes matriculados, 311 decidieron participar inicialmente en el proyecto, lo que supuso un 82,5% de los matriculados. A todos ellos se les dio de alta en un curso virtual sobre *Moodle* en el que podían encontrar los



materiales docentes de la asignatura, espacios de comunicación síncrona y asíncrona, espacios para la entrega de tareas, actividades de evaluación que se iban activando durante el desarrollo de la asignatura, calendarios de planificación de actividades, etc.

En particular, en lo relativo a los foros de debate, debe indicarse que la asignatura contaba con los siguientes foros de debate. En primer lugar un foro general de noticias, creado por el equipo docente para informar sobre aspectos importantes durante el desarrollo de la asignatura y en el que los estudiantes únicamente podían realizar operaciones de lectura. Se disponía también de un foro general abierto a todo tipo de participación; y cuatro foros temáticos, uno por cada bloque de la asignatura, abiertos igualmente a todo tipo de participación por los estudiantes, pero dirigidos a plantear dudas o comentarios sobre el correspondiente bloque temático. Como consecuencia de las actividades de los estudiantes en esos espacios virtuales de comunicación, y gracias a las funcionalidades del sistema *Moodle*, se obtuvieron para cada estudiante datos relativos a mensajes leídos, respuestas ofrecidas a otros mensajes y nuevos hilos de debate abiertos por el estudiante.

Al finalizar cada bloque temático se activaba una actividad de evaluación no presencial consistente en un test generado de manera aleatoria sobre una base de datos de preguntas. Cada estudiante debía realizar el test que le había correspondido de manera no presencial y con una franja temporal lo suficientemente amplia como para que su realización le permitiera igualmente preparar mejor el tema recién finalizado. *Moodle* almacenaba toda la información relativa a la realización de estos tests, en particular, los resultados globales, las respuestas a cada pregunta y los tiempos invertidos en la realización. También el profesor obtenía informes sobre los resultados globales de los estudiantes en las diferentes preguntas de la base de datos.

También se planteó a los estudiantes a través del curso virtual la realización de un trabajo práctico a distancia en el que debían demostrar la adquisición de destrezas informáticas relacionadas con la toma de decisiones. Este trabajo se realizaba de manera autónoma tras una serie de prácticas de laboratorio presenciales, y era entregado a través de la plataforma virtual para su corrección por el profesor.

En definitiva, la actividad no presencial de los estudiantes quedó reflejada en una serie de atributos cuantitativos, tal como se refleja en la Tabla 1.

Algunos datos llamativos pueden ser la participación media de los estudiantes en los foros, si bien como promedio los estudiantes leyeron una media de 78,3 mensajes, la participación activa en los foros fue sensiblemente inferior, con una media de 3,2 mensajes nuevos o respuestas dadas por estudiante. Las desviaciones típicas de los datos anteriores muestran una clara participación desigual, siendo la desviación típica de los mensajes leídos 87,6 y de los mensajes creados de 10,4. Como dato curioso se puede indicar que el número de mensajes leídos por el estudiante más participativo fue de 587, y el número máximo de mensajes creados por un mismo estudiante fue de 142.



Nombre	Tipo actividad	Descripción
FNL	Participación en foros	Número de mensajes leídos por el estudiante en el foro de noticias del curso virtual.
FGL	Participación en foros	Número de mensajes leídos en el foro general del curso.
FGR	Participación en foros	Número de respuestas dadas por el estudiante a mensajes del foro general del curso.
FGN	Participación en foros	Número de nuevos hilos de debate abiertos en el foro general del curso.
F1L, F2L, F3L, F4L	Participación en foros	Número de mensajes leídos en cada uno de los 4 foros temáticos del curso.
F1R, F2R, F3R, F4R	Participación en foros	Número de respuestas dadas a mensajes en cada uno de los 4 foros temáticos del curso.
F1N, F2N, F3N, F4N	Participación en foros	Número de nuevos hilos de debate abiertos en cada uno de los 4 foros temáticos del curso.
T1, T2, T3, T4	Evaluación no presencial	Resultados de los tests de evaluación on-line de cada uno de los bloques temáticos.
TR	Trabajo individual no presencial	Evaluación del trabajo sobre destrezas informáticas.

Tabla 1. Atributos que reflejan la actividad no presencial de los estudiantes.

En la parte presencial de la asignatura también se realizaron actividades que generaban nuevos datos sobre el rendimiento de los estudiantes. Así, finalizado el plazo de entrega de los tests virtuales, se planteaban en la clase ejercicios que el estudiante debía realizar *in situ* y entregar al profesor para su corrección. Los resultados de estas evaluaciones eran contrastados con los de los correspondientes tests. Además, como complemento a las prácticas de laboratorio realizadas durante el curso, se organizaron 3 sesiones de evaluación práctica individual en el que cada estudiante debía resolver con ayuda del ordenador un problema similar a los planteados en las prácticas, y todo ello después de haberse instruido en los fundamentos básicos de las herramientas utilizadas a través del curso virtual.

En definitiva, la actividad presencial de los estudiantes quedó reflejada en una nueva serie de atributos cuantitativos, recogidos en la Tabla 2.

Nombre	Tipo actividad	Descripción
TC1, TC2, TC3, TC4	Ejercicios de evaluación	Resultados de los ejercicios propuestos para la evaluación después de la entrega de los tests de los bloques temáticos.
LC1, LC2, LC3	Evaluación en laboratorio	Resultados de pruebas presenciales de contraste de destrezas adquiridas en las prácticas de laboratorio

Tabla 2. Atributos que reflejan la actividad presencial de los estudiantes.

Finalmente, se consideró un atributo booleano que indicaba si el estudiante superó o no satisfactoriamente la asignatura tras el examen final (APROBADO). Por tanto, para cada estudiante se obtuvieron un total de 29 atributos o indicadores que medían su actividad y rendimiento en las actividades presenciales y no presenciales.



2.2. Pre-procesamiento de la información

Toda la información recopilada sobre los valores de los indicadores para cada estudiante fue introducida en una base de datos para facilitar su gestión. Parte de la información provenía de los informes de actividad generados por *Moodle* y el resto de la información proporcionada directamente por los profesores de la asignatura. Los datos fueron posteriormente importados desde la herramienta open source *Weka* (*Waikato Environment for Knowledge Analysis*)¹ de la Universidad de Waikato. *Weka* es un entorno para experimentación de análisis de datos que contiene múltiples módulos que facilitan el pre-procesamiento, clasificación, agrupamiento y análisis de datos. La Figura 1 muestra la carga de todos los valores de los indicadores en *Weka*, visualizando además el histograma correspondiente al atributo APROBADO, que marca la tasa de éxito final en la asignatura entre los alumnos participantes en el proyecto, tasa que se situó en el 55,3%.

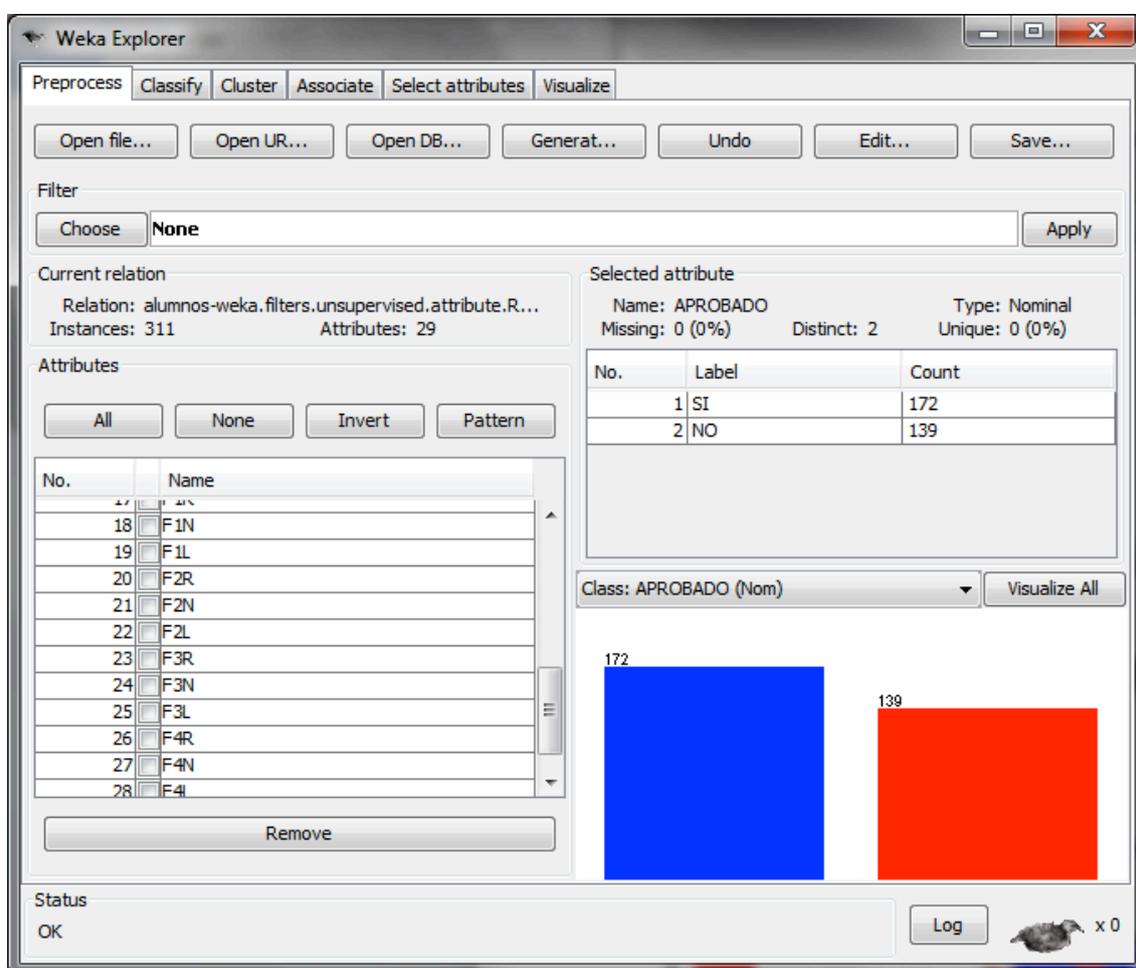


Figura 1. Preprocesamiento de los datos con la herramienta Weka.

Desde *Weka* la primera labor realizada fue la transformación de los datos, de manera que para facilitar su comparación todos los atributos numéricos fueron normalizados en el rango [0,1]. Las Figuras 2 y 3 muestran los histogramas de distribución de valores normalizados de los atributos correspondientes a actividades no presenciales y presenciales,

¹ <http://www.cs.waikato.ac.nz/ml/weka/>



respectivamente. En ambas figuras se diferencian los estudiantes según el éxito final o no en la asignatura.

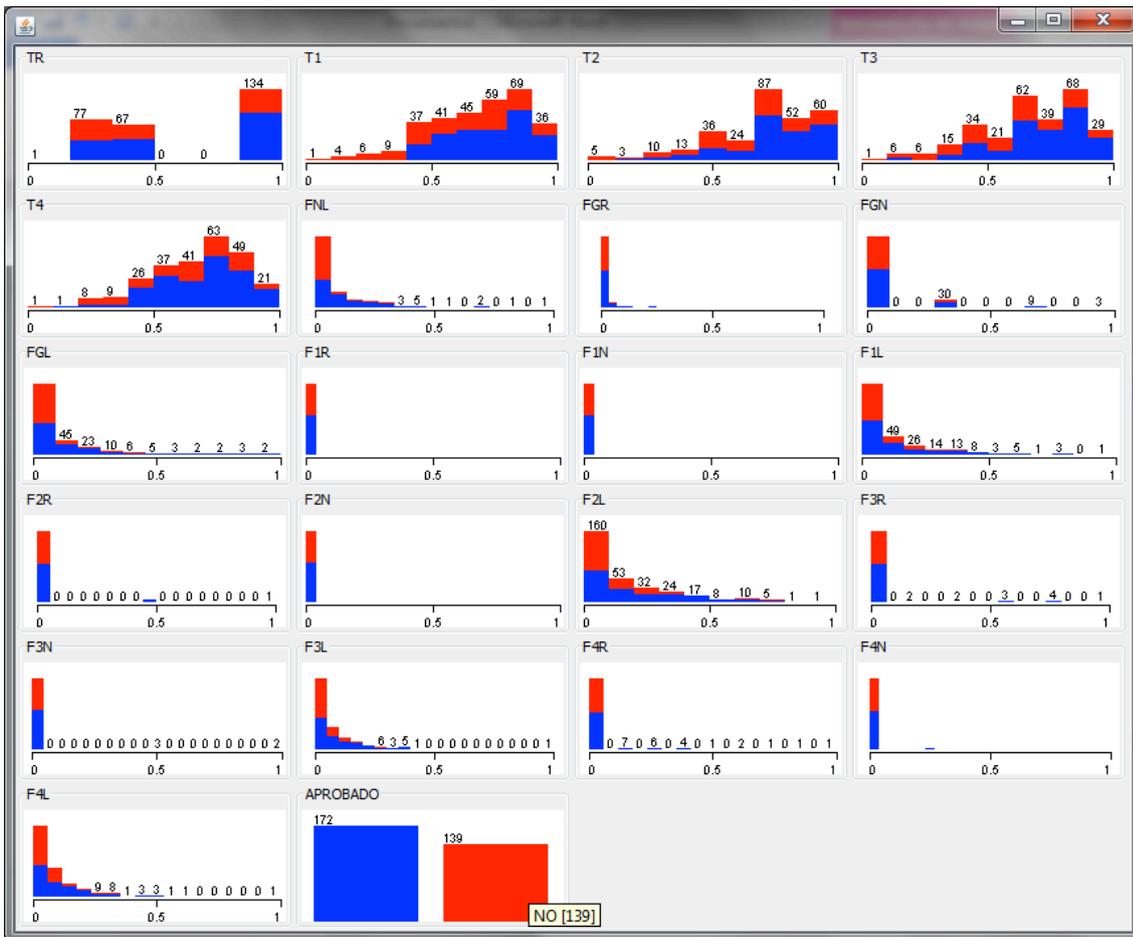


Figura 2. Histogramas generados por Weka para visualizar los valores normalizados de los 21 atributos correspondientes a actividades no presenciales.

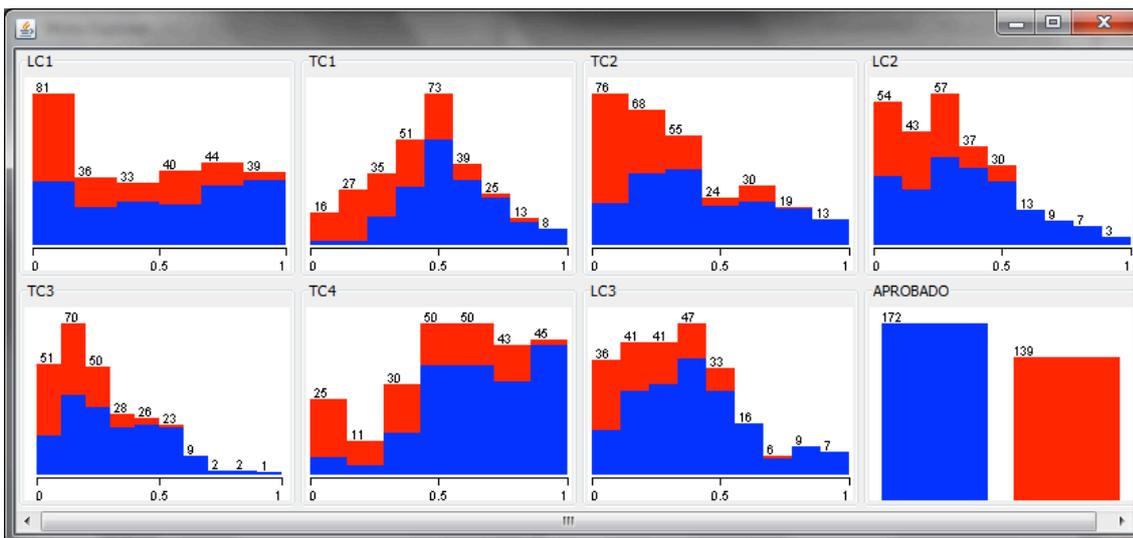


Figura 3. Histogramas generados por Weka para visualizar los valores normalizados de los 7 atributos correspondientes a actividades presenciales.



2.3. Aplicación de técnicas de selección de atributos

De los 28 atributos recogidos en las Tablas 1 y 2, no todos son igual de válidos a la hora de predecir el rendimiento final de un estudiante (valor del atributo APROBADO). En este trabajo se ha pretendido, por tanto, la utilización de estrategias de selección automática de rasgos para identificar los atributos con mayor poder discriminante. *Weka* incorpora una gran variedad de técnicas de selección de atributos que tratan de explorar qué subconjuntos de atributos son los que mejor pueden clasificar la clase de la instancia. Esta selección de atributos tiene dos componentes:

1. Un método de evaluación que determina la calidad del conjunto de atributos para discriminar la clase. Se pueden distinguir dos categorías de métodos de evaluación, en la primera se utiliza directamente un clasificador específico para medir la calidad del subconjunto de atributos a través de la tasa de error del clasificador. Estos métodos necesitan un proceso completo de entrenamiento y evaluación en cada caso de búsqueda, por eso resultan de un elevado coste computacional. La alternativa es la utilización de métodos que no utilizan un clasificador específico, por ejemplo el método *CfsSubsetEval* que se encuentra implementado en *Weka* y que se basa en calcular la correlación de la clase con cada atributo, y eliminar atributos que tienen una correlación muy alta como atributos redundantes. Según este método los subconjuntos preferidos son aquellos altamente correlacionados con el atributo que define las clases y con poca correlación entre ellos.

2. Un método de búsqueda que determina la forma de realizar la búsqueda de conjuntos. La evaluación exhaustiva de todos los posibles subconjuntos se convierte en un problema combinatorio inabordable cuando el número de atributos es elevado. Por tanto, se necesitan estrategias de búsqueda más eficientes. Una de las estrategias más efectiva, por su rapidez, es el *ForwardSelection*, que se basa en elegir primero el mejor atributo, y realizar un proceso iterativo de ir añadiendo atributos que aporten más información hasta llegar a la situación en la que añadir un nuevo atributo empeora la situación.

En la Tabla 3 se puede observar los subconjuntos de atributos obtenidos por *Weka* utilizando el método de evaluación *CfsSubsetEval* y diferentes métodos de búsqueda. En todos los casos el atributo que define las clases es el atributo APROBADO, utilizado para comprobar el éxito final del estudiante en la asignatura. Como puede apreciarse, los subconjuntos obtenidos no son iguales pero tienen una gran similitud, en concreto en la última fila de la tabla se incluyen los atributos seleccionados por todos los métodos de búsqueda.



Método Búsqueda	Nº atributos	Atributos
Best first	12	LC1, TC1, T1, TC2, TC3, T3, TC4, FNL, FGL, F3L, F4R, F4L
SubsetSizeForwardSelection	11	LC1, TC1, T1, TC, TC3, T3, TC4, FNL, FGL, F3L, F4L
GeneticSearch	10	LC1, TC1, T1, TC2, TC3, T3, TC4, FNL, FGL, F4L
GreedyStepwise	11	LC1, TC1, T1, TC2, TC3, T3, TC4, FNL, FGL, F3L, F4L
LinearForwardSelection	12	LC1, TC1, T1, TC2, TC3, T3, TC4, FNL, FGL, F3L, F4R, F4L
RankSearch	19	LC1, TC1, T1, TC2, T2, LC2, TC3, T3, TC4, T4, FNL, FGR, FGL, F2L, F3R, F3L, F4R, F4N, F4L
ScatterSearchV1	11	LC1, TC1, T1, TC2, TC3, T3, TC4, FNL, FGL, F3L, F4L
TabuSearch	11	LC1, TC1, T1, TC2, TC3, T3, TC4, FNL, FGL, F3L, F4L
Atributos presentes en todos los subconjuntos	8	LC1, TC1, T1, TC3, TC4, FNL, FGL, F4L

Tabla 3. Subconjuntos de atributos seleccionados con el evaluador de atributos CfsSubsetEval.

Orden	Valor	Atributo	Orden	Valor	Atributo
1	71,4263	TC1	15	16,5185	FGR
2	67,9752	FNL	16	14,7469	LC3
3	57,0588	TC2	17	13,6315	F4R
4	55,3716	FGL	18	10,7716	T4
5	39,8525	TC3	19	10,0869	F3R
6	37,4271	TC4	20	5,7872	F4N
7	37,1090	T1	21	0	TR
8	33,7923	F4L	22	0	F3N
9	25,9176	LC2	23	0	F1N
10	25,8653	F3L	24	0	FGN
11	21,4115	T3	25	0	F1R
12	20,3212	T2	26	0	F2N
13	19,9006	F2L	27	0	F1L
14	17,8280	LC1	28	0	F2R

Tabla 4. Ranking de atributos según su calidad para medir la tasa de éxito.

Además de los evaluadores de subconjuntos de atributos como el utilizado anteriormente, *Weka* dispone de prorratedotes de atributos (*AttributeEval*) que no seleccionan atributos sino que los ordenan por relevancia de acuerdo a un *ranker* establecido. Utilizando el prorratedor *ChiSquaredAttributeEval* que evalúa el valor de un atributo mediante el cálculo del estadístico chi-cuadrado con respecto a la clase, los 28 atributos de los datos son ordenados tal como se muestra en la Tabla 4. Como puede observarse, los 8 atributos que



son seleccionados por todos los métodos de selección de atributos se encuentran entre los 14 primeros del ranking anterior.

2.4. Análisis y evaluación de resultados

En esta sección se extraen algunas conclusiones a partir de los resultados obtenidos al aplicar las técnicas de selección de rasgos sobre la información procesada del conjunto de estudiantes.

En primer lugar, los 8 atributos que son seleccionados por todos los métodos de selección utilizados incluyen valoraciones o indicadores tanto de actividades presenciales como no presenciales; de hecho se reparten al 50%. Sin embargo, se observa que los indicadores de acciones presenciales tienen en general mayor poder predictivo a la hora de analizar el éxito del estudiante en la asignatura. De hecho, en el ranking de rasgos obtenido, los atributos de actividades presenciales tienen un valor medio de 37,8 frente a 16,1 de valor medio en los atributos de actividades no presenciales.

Los resultados de las pruebas de contraste presenciales realizadas con posterioridad a las evaluaciones on-line se muestran como elementos con mayor poder predictivo de éxito. Así 3 de las 4 pruebas (TC1, TC3 y TC4) se encuentran siempre en los subconjuntos de rasgos seleccionados automáticamente. Además el atributo TC1 es el que ocupa la primera posición en el ranking de atributos. En el caso de las pruebas de contraste de destrezas en el aula de informática parece que con los resultados de una sola de ellas (LC1) sería suficiente, y las otras dos, de estructura muy similar, parece que no aportan ninguna información adicional de interés.

Los tests on-line no parecen tener un gran valor predictivo, pero al menos uno de ellos (T1) es seleccionado dentro del subconjunto de atributos. El primer test es dentro de las pruebas de evaluación a distancia el que obtiene una mayor puntuación en el ranking de atributos. Considerar los resultados de un test y compararlo con el de la prueba de contraste correspondiente puede aportar información muy útil sobre la forma en la que el estudiante se enfrenta a esas pruebas a distancia. Debe recordarse que estas pruebas no contaban con supervisión, y la experiencia demuestra que la actitud con la que los estudiantes se enfrentaban a ellas era muy variable, observándose un sector cuyo único interés era obtener buenos resultados en ellas y no tanto en considerarlas como un instrumento para mejorar su capacitación.

Respecto a la participación en los foros del curso virtual, puede llamar la atención el hecho de que se seleccionen únicamente atributos correspondientes a lecturas de foros (FNL, FGL y F4L) y no a los de generación de nuevos mensajes. Prácticamente todos los métodos de selección de atributos se decantan por este tipo de atributos de participación receptiva. En el ranking de atributos los seis primeros correspondientes a actividad en los foros se refieren a lecturas de mensajes. Los atributos seleccionados corresponden al foro general de noticias, el utilizado por el equipo docente para informar de los aspectos más importantes del desarrollo de la asignatura, el foro general y el foro del último bloque temático. Esto último puede significar que aquellos alumnos que llegan hasta el final del curso con interés por participar en los foros son los que se encuentran más motivados y animados por su rendimiento en la asignatura, por el camino otros alumnos han ido abandonando y desconectándose del desarrollo del curso virtual. El hecho de ser estudiantes con una



participación más activa en los foros, aportando puntos de vista en los debates abiertos o abriendo nuevos hilos, no presupone un mejor rendimiento final en la asignatura.

3. CONCLUSIONES

En los actuales modelos de enseñanza-aprendizaje centrados en el estudiante, los recursos tecnológicos proporcionados por las TICs se revelan como herramientas esenciales para propiciar espacios formativos (virtuales o presenciales) de participación y cooperación. En tal sentido la combinación de modelos presenciales con actividades apoyadas en plataformas virtuales de enseñanza favorece la implementación efectiva de estos modelos educativos.

Además del valor añadido para el estudiante, estos modelos de b-learning permiten al profesor obtener una mayor cantidad de información sobre la evolución del proceso de enseñanza. La información obtenida directamente en el contacto presencial en el aula con los estudiantes puede ser complementada por información que de manera automática generan las plataformas de enseñanza virtual. En ocasiones manejar tales volúmenes de información puede resultar una tarea compleja y las técnicas de minería de datos pueden permitir descubrir asociaciones, tendencias, y, en definitiva, generar nuevo conocimiento.

En este trabajo se ha pretendido mostrar una experiencia realizada en el contexto de la enseñanza de una asignatura de carácter matemático a estudiantes de primer curso de carreras universitarias. A pesar del elevado número de estudiantes implicados, la combinación de acciones presenciales y no presenciales permite obtener un elevado número de indicadores que pueden ser utilizados para predecir el rendimiento de los estudiantes. En este sentido, las técnicas de selección de atributos han permitido seleccionar aquellos indicadores que puedan tener un mayor poder predictivo. De los análisis realizados se desprende que las actividades de evaluación a distancia no son siempre adecuadas para predecir el nivel de adquisición de competencias por parte de los estudiantes. Parece que la evaluación debe tener componentes presenciales, que convenientemente combinadas con actividades no presenciales y un seguimiento de los niveles de participación de los estudiantes en los foros, puedan servir para que el profesor obtenga una imagen fiel de la evolución y éxito en sus objetivos docentes.

4. BIBLIOGRAFÍA

AIELLO, M.; CILIA, W. (2004). El blended learning como práctica transformadora. *Pixel-Bit. Revista de Medios y Educación*. Nº. 23, 21-26.

BARTOLOMÉ, A.; AIELLO, M. (2006). Nuevas tecnologías y necesidades formativas. Blended learning y los nuevos perfiles en comunicación audiovisual. *Revista TELOS. Nuevas tecnologías y necesidades formativas*. Nº 67.

BARTOLOMÉ, A. (2004). Blended learning. Conceptos básicos. *Revista Pixel-Bit Revista de Medios y Educación*. Nº 23, 7-20.



MARCH, G.; McFADDEN, A.; PRICE, B. (2004). *Blended instruction: Adapting conventional instruction for large classes*. [documento en línea]. Online Journal of Distance Learning Administration, Vol. 6, nº 4. Distance Education Center. University of Alabama, Institute for Interactive Technology. [Fecha de consulta: 20/4/2011]. <<http://www.westga.edu/~distance/ojdla/winter64/marsh64.htm>>

MURPHY, P. (2002). *The hybrid strategy: Blending Face-to-Face with virtual instruction to improve large lecture course*. [documento en línea]. University of California. Regents Teaching, Learning, and Technology Center. [Fecha de consulta: 20/4/2011]. <<http://www.ucop.edu/tltc/news/2002/12/feature.php>>

FERRI, C.; RAMÍREZ, M. J.; HERNÁNDEZ, J. (2004). *Introducción a la Minería de Datos*. Prentice Hall.

PEÑALOSA, E.; GARCÍA, H.; ROJAS, G. (2010). Modelo estratégico de comunicación educativa para entornos mixtos de aprendizaje: estudio piloto. *Pixel-Bit: Revista de medios y educación* Nº. 37, 43-55.

PONS, J.; MORENO, P. V. (2005). El Espacio Europeo de Educación Superior y las Tecnologías de la Información y la Comunicación. Percepciones y demandas del profesorado. *Revista de Educación*, Nº 337, 99-124

SALMON, G. (2000): *E-moderating: The key to teaching and learning online*. London, Kogan Page.

WITTEN, I.; FRANK, E. (2005). *Data Mining. Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers.

Para citar este artículo:

COBO, A.; ROCHA, R. & ÁLVAREZ, Y. (2011). Selección de atributos predictivos del rendimiento académico de estudiantes en un modelo de B-Learning. *EduTec-e, Revista Electrónica de Tecnología Educativa*, 37. Recuperado el dd/mm/aa de http://edutec.rediris.es/revelec2/revelec37/atributos_predictivos_rendimiento_academico_b-learning.html

