

MEMORIA FINAL DEL PROYECTO

DATOS IDENTIFICATIVOS:

1. *Título del Proyecto*

Minería de Datos Educativos para la Detección de Recursos Clave

2. *Código del Proyecto*

2015-2-5005

3. *Coordinador/es del Proyecto*

| Nombre y Apellidos | Departamento | Código del Grupo Docente |
|-----------------------------|---------------------------------|--------------------------|
| Amelia Zafra Gómez | Informática y Análisis Numérico | 4 |
| Eva Lucrecia Gibaja Galindo | Informática y Análisis Numérico | 4 |

4. *Otros Participantes*

| Nombre y Apellidos | Departamento | Código del Grupo Docente | Tipo de Personal |
|------------------------|---------------------------------|--------------------------|--------------------------|
| María Luque Rodríguez | Informática y Análisis Numérico | 4 | PDI |
| Juan Luis Olmo Ortiz | Informática y Análisis Numérico | 4 | Becario de Investigación |
| Aurora Ramírez Quesada | Informática y Análisis Numérico | 4 | Becaria de Investigación |
| Antonio Arauzo Azofra | Ingeniería Rural | 67 | PDI |

5. *Memoria del Proyecto*

Título

Minería de Datos Educativos para la Detección de Recursos Clave

Resumen

Este proyecto de innovación educativa se ha centrado en diseñar y desarrollar un nuevo módulo de Moodle que permita obtener modelos predictivos basados en árboles de decisión a partir de los datos de uso almacenados en Moodle. Debido a las características de los modelos basados en árboles de decisión, es posible relacionar el trabajo personal realizado por los alumnos en Moodle con la calificación final obtenida, proporcionándonos así una descripción de los recursos y actividades que resultan clave para superar/suspender una asignatura. Con esta información, el profesor podría detectar alumnos con alta probabilidad de no superar la asignatura e intentar reconducir la situación. Este módulo podría ser aplicado en cualquier asignatura de la plataforma. En este proyecto, tras su realización, se ha aplicado en una de las asignaturas impartidas por los profesores implicados en el proyecto.

Palabras clave

Minería de datos educativos, recursos clave, clasificación, árboles de decisión

Introducción

Las plataformas *e-learning* han tenido muy buena aceptación y se han convertido en una herramienta esencial en la comunidad educativa. Así, constituyen uno de los pilares fundamentales en la construcción del conocimiento, permitiendo nuevos canales de comunicación tanto del aprendizaje síncrono como asíncrono, y dotando de un gran potencial tanto a los docentes como a los alumnos [Bon2007].

Con la llegada de estas plataformas, nos encontramos con una gran cantidad de información almacenada sobre la actividad del estudiante cuyo análisis resulta de gran interés para comprender mejor su proceso de aprendizaje. El problema es que identificar y encontrar conocimiento útil en toda la información disponible es una tarea difícil. Una solución muy prometedora es el uso de técnicas de extracción de conocimiento o minería de datos. Cuando estas técnicas son aplicadas sobre datos educativos, se habla de minería de datos educativos (*Educational Data Mining*, EDM). La *Minería de Datos Educativos* [Rom2010] se define como un proceso no trivial que nos permite transformar los datos almacenados en los sistemas de enseñanza en conocimiento útil que pueda utilizarse para tomar decisiones. Es una disciplina emergente que aplica minería de datos en el ámbito de la enseñanza con el objetivo de obtener una mejor comprensión del proceso de aprendizaje de los estudiantes, orientado a la mejora de la calidad y la rentabilidad del sistema educativo. Una de las aplicaciones más populares de la EDM es la predicción y entendimiento de los resultados académicos de un estudiante. En este contexto, en el proyecto presentado se pretende predecir el rendimiento académico y los factores que influyen en él (*recursos clave*) en función de la actividad realizada por los alumnos en una plataforma de *e-learning*. Para llevar a cabo esta tarea partimos de la información sobre la actividad de los estudiantes que está almacenada en los registros *log* de la plataforma. Estos alumnos son evaluados y etiquetados como aprobados o suspensos. A continuación, se construye un modelo predictivo utilizando un algoritmo de clasificación con los datos de la etapa anterior. El modelo nos permitirá predecir si un nuevo estudiante aprobará o suspenderá a partir de su información almacenada en la plataforma. Trabajaremos con modelos de clasificación interpretables que, además de dar una predicción, proporcionen una explicación con las evidencias en las que se han basado para llevar a cabo la clasificación. En este caso, estas evidencias serán las actividades y recursos (A&Rs) realizados en la plataforma y que han resultado clave para poder superar o no una asignatura. Con esta información, sería posible detectar alumnos con un alto riesgo de no superar la asignatura, para así tratar de tomar acciones preventivas que reconduzcan su proceso de aprendizaje.

Concretamente, en este proyecto se ha desarrollado un módulo en Moodle (*Module Object-Oriented Dynamic Learning Environment*) [Ros2008] para obtener información de cuáles son las A&Rs que resultan más relevantes para superar una asignatura. La selección de la plataforma Moodle para el desarrollo del módulo, se ha debido a que es la plataforma educativa que se encuentra implantada en la Universidad de Córdoba, y por tanto donde se encuentran almacenados todos los registros de actividad de los alumnos de todas las asignaturas de Grados y Másteres que se imparten en esta Universidad. Participantes de este proyecto, ya han desarrollado previamente otros proyectos basados en Moodle [Luq2014], [Zaf2014].

Los algoritmos de clasificación que se han aplicado están basados en árboles de decisión y están disponibles en el software de Weka [Hal2009]:

- Algoritmo J48 [Qui2014]. Este algoritmo genera un árbol de decisión a partir de una tabla donde cada fila es un patrón descrito por un conjunto de atributos.
- Algoritmo *RandomForest* [Bre2001]. A diferencia de J48, este algoritmo genera varios árboles de decisión.

La selección de estos algoritmos se debió a que generan modelos predictivos muy interpretables. Así, además de proporcionar una predicción sobre el rendimiento académico, nos permiten estudiar las A&Rs que se están considerando más relevantes para determinar si un estudiante tiene una alta probabilidad de obtener una calificación satisfactoria en una determinada asignatura, que es el fin que se pretende lograr con este proyecto.

El módulo diseñado e implementado permite utilizar la información de la actividad llevada a cabo por el alumno en la plataforma educativa y las calificaciones almacenadas en Moodle para aplicar algoritmos de clasificación y obtener un modelo predictivo que nos permita detectar si existen recursos que son determinantes para superar la asignatura. El módulo desarrollado presenta además las siguientes características:

- Cada uno de los algoritmos incorporados puede ser configurado considerando los parámetros más relevantes.
- La selección de los estudiantes y de la actividad que se quiere considerar es flexible permitiendo la selección de:
 - Un conjunto de estudiantes definidos por el profesor.
 - Un conjunto de A&Rs definidos por el profesor.
 - Un rango de fechas definido por el profesor.
- Se puede incorporar la calificación de los alumnos (en caso de que el profesor no esté utilizando el libro de calificaciones de Moodle).
- Se puede mantener el anonimato de toda la información personal utilizada. Esto facilitará la publicación de los resultados obtenidos en medios de difusión científica o docente.
- Es posible guardar los datos de *log* utilizados para generar los modelos en otros formatos para su posterior uso en otras herramientas de minería de datos.
- Es posible utilizar diversos niveles de desempeño para las A&Rs (accedido/no accedido, número de accesos, calificación).
- Es posible utilizarlo en inglés o en español.
- El módulo se puede utilizar en cualquier asignatura de la plataforma Moodle.

Desde un punto de vista técnico:

- El desarrollo ha seguido el formato estándar de Moodle, de manera que su manejo es lo más intuitivo posible y personas no expertas en minería de datos pueden aplicar estos algoritmos y obtener información relevante de las relaciones entre el trabajo de los alumnos y su calificación.
- El módulo incluye iconos de ayuda, con la finalidad de orientar al usuario.
- El módulo es robusto y fácil de ampliar y mantener en el futuro.

Materiales y métodos

Para el desarrollo del proyecto, se ha empleado un servidor del grupo de investigación KDIS de la Universidad de Córdoba. El bloque se ha implementado en PHP con XAMPP [Xam2016], que convierte al ordenador en un servidor que ofrece: Apache, MySQL, PHP y phpMyAdmin.

Para alcanzar los objetivos previamente comentados, el proyecto se ha desarrollado en las siguientes fases:

1. *Estudio previo de la plataforma Moodle.*
 - Se ha llevado a cabo un estudio de la base de datos del sistema Moodle para poder obtener la información necesaria de los registros *log*.
 - Se ha llevado a cabo un estudio de la API de desarrollo de Moodle, para poder aprovechar toda la funcionalidad que ya está incorporada en la plataforma educativa.
 - Se ha llevado a cabo un estudio de los módulos con los que interactuará el bloque a desarrollar.
2. *Diseño y desarrollo de la base de datos que se ha incorporado a Moodle.* Se ha definido la información que se debe almacenar o re-estructurar para poder aplicar los modelos predictivos.
3. *Diseño de un bloque de Moodle que permite aplicar las técnicas predictivas de clasificación.* Se ha llevado a cabo el diseño e implementación del módulo, de tal forma que permite: hacer anónima la información de los

alumnos (para facilitar la posterior publicación de resultados), seleccionar la información que se quiere utilizar para generar el modelo predictivo (A&Rs a considerar, estudiantes, fechas), introducir calificaciones de los alumnos (en caso de no estar utilizando el libro de calificaciones de Moodle), y seleccionar/configurar el algoritmo de clasificación que se desea aplicar.

4. *Instalación del software.* El software se ha instalado en un servidor propio.
5. *Validación del sistema.* Se ha llevado a cabo un proceso de validación del sistema para certificar y garantizar que el software está libre de errores y que lleva a cabo la tarea a desempeñar de manera correcta y eficiente.
6. *Análisis de la utilidad del bloque.* Tras desarrollar la nueva funcionalidad, se ha comprobado su utilidad en la evaluación y seguimiento de alumnos y A&Rs en la asignatura *Introducción a la Programación* del Grado en Ingeniería Informática (asignatura impartida por varios profesores que intervienen en este proyecto). Para ello, una vez finalizado un curso académico y teniendo las calificaciones finales de los alumnos, se han aplicado los dos algoritmos de árboles considerados, J48 y *RandomForest* y se evaluaron sus resultados. El análisis demuestra que los árboles de decisión obtenidos por los modelos, permiten determinar los factores más relevantes. Los resultados obtenidos, serán comentado en la siguiente sección, donde se muestra un caso de estudio.

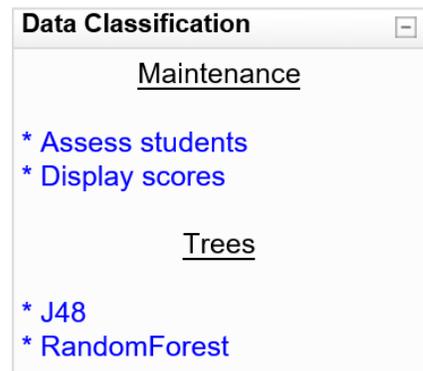
Resultados y conclusiones (concretar y discutir los resultados obtenidos con aportación de evidencias).

En este apartado describimos las características más relevantes del módulo desarrollado en este proyecto ilustradas con un caso de estudio en la asignatura *Introducción a la Programación* del Grado en Ingeniería Informática. Las imágenes están en inglés, uno de los requisitos del módulo era que la interfaz estuviese tanto en inglés, como en español.

1. Presentación del Módulo

El módulo desarrollado consiste en un bloque de Moodle que da las opciones de:

1. **Mantenimiento.** Para obtener un modelo de clasificación de rendimiento académico es necesario disponer de una calificación para los alumnos considerados. En caso de que el profesor no haya utilizado el libro de calificaciones de Moodle (algo que suele ser habitual), el módulo da la opción de introducir manualmente una calificación (*assess students*) y de visualizar las calificaciones (*display scores*). Estas calificaciones son totalmente independientes del libro de calificaciones de Moodle.
2. **Trees.** Obtención de modelos predictivos mediante árboles de clasificación utilizando los algoritmos:
 - Algoritmo *J48*.
 - Algoritmo *RandomForest*.



2. Mantenimiento

2.1. Evaluar alumnos

Como se ha comentado, esta funcionalidad permite evaluar como aprobado (*pass*) o suspenso (*fail*) a los alumnos que no tienen calificación en el libro de calificaciones.

Se muestran todos los alumnos pendientes de evaluación y se despliega un combo que permite calificar al alumno.

En caso de error, también es posible recalificar a alumnos ya evaluados utilizando la parte inferior del formulario (*assessed students*).

Finalmente, la opción *reset* permite resetear a la vez las calificaciones de todos los alumnos del grupo asignándoles el valor *no score*.

Assess students

You are logged in as

IP_MP ► GII-1-IPG2_C1 ► Assess students

Important

Scores can be introduced here are completely independent of Moodle gradebook.

Reset

Reset scores ▼

Pending students to be assessed: 9

| | |
|--------------|---|
| Juan Luis | <input type="button" value="No score"/> ▼ |
| Daniel | <input type="button" value="No score"/> ▼ |
| Jose | <input type="button" value="No score"/> ▼ |
| Manuel | <input type="button" value="Pass"/> ▼ |
| Damian | <input type="button" value="Fail"/> ▼ |
| Inmaculada | <input type="button" value="No score"/> ▼ |
| Sergio | <input type="button" value="No score"/> ▼ |
| Pedro Javier | <input type="button" value="No score"/> ▼ |
| Javier | <input type="button" value="No score"/> ▼ |
| Ramon | <input type="button" value="No score"/> ▼ |

Students assessed: 69

| | |
|---------|---------------------------------------|
| Oscar | <input type="button" value="Fail"/> ▼ |
| Pilar | <input type="button" value="Fail"/> ▼ |
| Antonio | <input type="button" value="Fail"/> ▼ |
| Mariano | <input type="button" value="Pass"/> ▼ |
| David | <input type="button" value="Pass"/> ▼ |

2.2. Visualizar calificaciones

Esta funcionalidad permite ver un listado de todos los alumnos de la asignatura junto con su correspondiente evaluación (*fail* o *pass*).

Display scores

You are logged in as

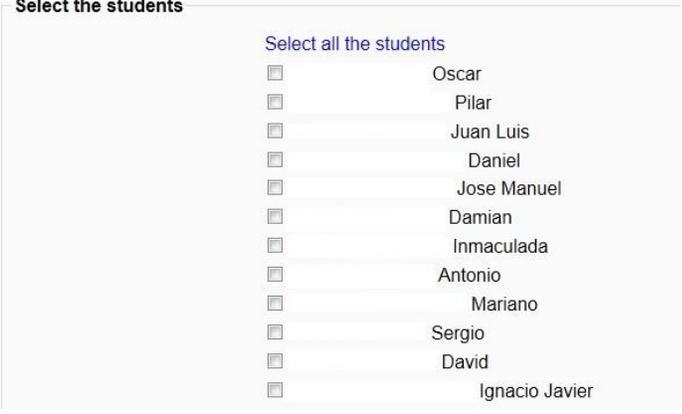
IP_MP ► GII-1-IPG2_C1 ► Display scores

Scores

| Student | Score |
|----------------|-------|
| Oscar | Fail |
| Pilar | Fail |
| Antonio | Fail |
| Mariano | Pass |
| David | Pass |
| Ignacio Javier | Fail |
| Barbara | Pass |

3. Clasificación mediante árboles de decisión

A continuación se describe el módulo de clasificación desarrollado para la detección de recursos clave. Se describen primero los elementos comunes a los dos algoritmos considerados y después se hace una descripción detallada de cada uno de ellos. Cabe destacar que, al igual que hace Moodle, el usuario siempre tiene disponible información online sobre la utilización del módulo a través del icono de ayuda .

| | |
|--|--|
| <p>3.1.Ajustes comunes</p> <p><u>3.1.1. Selección personalizada de alumnos</u></p> <ul style="list-style-type: none"> • Se realiza mediante <i>check boxes</i>. • Se permite además seleccionar a la vez todos los alumnos mediante el enlace <i>select all the students</i>. |  <p>Select the students</p> <p>Select all the students</p> <ul style="list-style-type: none"> <input type="checkbox"/> Oscar <input type="checkbox"/> Pilar <input type="checkbox"/> Juan Luis <input type="checkbox"/> Daniel <input type="checkbox"/> Jose Manuel <input type="checkbox"/> Damian <input type="checkbox"/> Inmaculada <input type="checkbox"/> Antonio <input type="checkbox"/> Mariano <input type="checkbox"/> Sergio <input type="checkbox"/> David <input type="checkbox"/> Ignacio Javier |
| <p><u>3.1.2. Cifrado de los nombres de los alumnos</u></p> <p>Esta funcionalidad permite mantener la privacidad de los alumnos en caso de que se quieran publicar los resultados obtenidos. Se permiten dos tipos de cifrado: numérico y encriptado.</p> <p>En este primer ejemplo, vemos que la selección inicial de los alumnos a considerar da tres opciones. La primera consiste en mostrar el nombre y apellidos completos del estudiante (que hemos ocultado por motivos de privacidad).</p> | <p>Opciones de visualización</p> <p>Mostrar nombre y apellidos del estudiante Mostrar nombre y apellidos cifrados del estudiante Mostrar índice del estudiante</p> <hr/> <p>Seleccione los Alumnos</p> <p>Todos los Alumnos</p> <ul style="list-style-type: none"> <input type="checkbox"/> [redacted], Oscar <input type="checkbox"/> [redacted], Pilar <input type="checkbox"/> [redacted], Juan Luis <input type="checkbox"/> [redacted], Daniel <input type="checkbox"/> [redacted], Jose Manuel <input type="checkbox"/> [redacted], Damian <input type="checkbox"/> [redacted], Inmaculada |
| <p>La segunda opción, en el cifrado de los nombres de los alumnos, consiste en cifrar los nombres de los estudiantes con el algoritmo MD5.</p> | <p>Opciones de visualización</p> <p>Mostrar nombre y apellidos del estudiante Mostrar nombre y apellidos cifrados del estudiante Mostrar índice del estudiante</p> <hr/> <p>Seleccione los Alumnos</p> <p>Todos los Alumnos</p> <ul style="list-style-type: none"> <input type="checkbox"/> da2187a4c9277a885e40de738e467383 <input type="checkbox"/> 9a442410014d70101c087b2b2f8755df <input type="checkbox"/> 2ce2212703db37800070150d516a1d08 <input type="checkbox"/> dad9a20309712648f308e3a5a5950328 <input type="checkbox"/> 354b183b8036b83dfda18398d60438c8 <input type="checkbox"/> 7449ca905800df5cba42d07a37d7253 <input type="checkbox"/> 9d78886588c6d9dd2060ededf63bc88d |
| <p>La última opción consiste en asignar un índice a cada alumno y mostrarlos indicando su índice (<i>Estudiante1, Estudiante2, etc.</i>). Este es un método que genera nombres más intuitivos para el profesor que mediante MD5.</p> | <p>Opciones de visualización</p> <p>Mostrar nombre y apellidos del estudiante Mostrar nombre y apellidos cifrados del estudiante Mostrar índice del estudiante</p> <hr/> <p>Seleccione los Alumnos</p> <p>Todos los Alumnos</p> <ul style="list-style-type: none"> <input type="checkbox"/> Estudiante 1 <input type="checkbox"/> Estudiante 2 <input type="checkbox"/> Estudiante 3 <input type="checkbox"/> Estudiante 4 <input type="checkbox"/> Estudiante 5 <input type="checkbox"/> Estudiante 6 <input type="checkbox"/> Estudiante 7 |

| | |
|---|--|
| <p><u>3.1.3. Selección personalizada de fechas</u></p> <ul style="list-style-type: none"> Se realiza mediante desplegaables. | <p>Select dates (The course started: 14/09/2012)</p> <p>Start date: 14 September 2012</p> <p>End date: 10 February 2013</p> |
| <p><u>3.1.4. Selección personalizada de A&Rs</u></p> <ul style="list-style-type: none"> Se realiza mediante <i>check boxes</i>. Se permite además seleccionar a la vez todos los recursos de un tipo concreto. | <p>Select the assignments</p> <p>Select all the assignments</p> <ul style="list-style-type: none"> <input type="checkbox"/> R2. Condicionales <input type="checkbox"/> R2. Iterativos <input type="checkbox"/> R3. Funciones <input type="checkbox"/> R4. Cadenas <input type="checkbox"/> R4. Estructuras <input type="checkbox"/> R4. Matrices <input type="checkbox"/> R4. Vectores <p>Select the choices</p> <p>Select all the choices</p> <ul style="list-style-type: none"> <input type="checkbox"/> Estimación del número de presentados - Diciembre 2012 <input type="checkbox"/> Estimación del número de presentados - Febrero 2013 <input type="checkbox"/> Estimación del número de presentados - Septiembre |
| <p><u>3.1.5. Opciones de agrupamiento</u></p> <p>Se puede generar el modelo a partir del acceso y calificación del alumno respecto a:</p> <ul style="list-style-type: none"> A&Rs individuales (<i>grouping = No</i>). Tipos de A&Rs (<i>grouping = Yes</i>). | <p>Grouping options</p> <p>Group by activities/resources ? <input type="button" value="No"/> <input type="button" value="Yes"/></p> |
| <p><u>3.1.6. Tipos de datos utilizados para generar el modelo</u></p> <ul style="list-style-type: none"> <i>Hits</i>. Numero de accesos (<i>hits</i>) a cada A&R. <i>Hits binary</i>. Si el alumno ha accedido o no a una determinada A&R independientemente del número de accesos. <i>Score</i>. En lugar del acceso considera la puntuación. Según la configuración del agrupamiento y el tipo de A&R considera: <ul style="list-style-type: none"> <i>grouping = No</i> <ul style="list-style-type: none"> Cuestionarios: la calificación obtenida. Tareas: si se ha subido o no una tarea. Foros y wikis: si se ha participado o no. Consultas: si se ha realizado o no. <i>grouping = Yes</i> <ul style="list-style-type: none"> Grupo de cuestionarios: la nota media. Grupo de tareas: el número de tareas terminadas. Grupo de foros o wikis: el número de <i>posts</i>. Grupo de consultas: el número de consultas realizadas. <i>Hits & Scores</i>. Combina el método <i>Hits</i> y el método <i>Scores</i>. <i>Hits Binary & Scores</i>. Combina el método <i>Hits Binary</i> y el método <i>Scores</i>. | |
| <p>Data processing</p> <p>Data to use in the execution ?</p> <ul style="list-style-type: none"> Hits (Decimal system) Hits (Binary system) Scores for activities/resources Hits (Decimal system) & Scores for activities/resources Hits (Binary system) & Scores for activities/resources | |

3.2. Configuración del algoritmo Random Forest

Se han incluido todos los parámetros que proporciona weka para el algoritmo *RandomForest*:

- *MaxDepth*. Profundidad máxima de los árboles generados (0 para no poner límite a la profundidad).
- *NumFeatures*. Número de atributos del conjunto de datos original que serán seleccionados aleatoriamente para generar cada árbol.
- *NumTrees*. Número de árboles a generar.
- *Seed*. Semilla para el generador de números aleatorios.

Parameter settings

maxDepth ? 0
numFeatures ? 0
numTrees ? 10
seed ? 1

Clear form

Run

3.3. Configuración del algoritmo J48

Se han incluido todos los parámetros que proporciona weka para el algoritmo J48:

- *SaveInstanceData*. Indica si los datos de entrenamiento se salvan para su posterior visualización.
- *Binary splits*. Indica si al generar el árbol se utilizarán particiones binarias en los atributos nominales.
- *Confidence factor*. Factor de confianza para la poda (pequeños valores producen mayor poda).
- *MinNumObj*. Número mínimo de instancias por hoja.
- *NumFolds*. En *reducedErrorPruning* un *fold* se utilizará para la poda y el resto para generar el árbol.
- *Unpruned*. Indica si se realizará poda del árbol o no.
- *SubtreeRaising*. Indica si se realizará *subtree raising* al realizar la poda.
- *ReducedErrorPruning*. Indica si se aplicará *reduced-error pruning* en lugar de la poda básica de C4.5
- *Seed*. Semilla para el generador de números aleatorios del proceso *reducedErrorPruning*.
- *useLaplace*. Si se utilizará en el conteo en las hojas del árbol la corrección de Laplace.

Parameter settings

binarySplits ? False ▾
confidenceFactor ? 0.25
minNumObj ? 2
numFolds ? 3
reducedErrorPruning ? False ▾
saveInstanceData ? False ▾
seed ? 1
subtreeRaising ? True ▾
unpruned ? False ▾
useLaplace ? False ▾

Clear form

Run

3.4. Caso de estudio

Para mostrar la utilidad del módulo desarrollado se presenta un caso de estudio para la asignatura *Introducción a la Programación* de Grado en Ingeniería Informática. Esta asignatura incluye 6 cuestionarios que los alumnos deben ir realizando a lo largo del cuatrimestre, conforme se explica en clase la teoría necesaria para completarlos. Se enumeran en el mismo orden en que están secuenciados en la asignatura:

- Tipos de datos, operadores, expresiones E/S
- Estructuras de control
- Modularización
- Tipos de datos compuestos
- Vectores y Matrices
- Matrices y estructuras

El objetivo de este caso de estudio es determinar cuáles de estos cuestionarios son recursos clave de la asignatura. Para ello, se han seleccionado como datos de entrada todos alumnos (cifrándolos por índice) y los 6 cuestionarios. Después, se ha ejecutado el algoritmo J48 con sus parámetros por defecto obteniendo el siguiente árbol de decisión:

```
*Cuestionario modularización <= 5: Fail (22.0)
*Cuestionario modularización > 5
| *Cuestionario tipos de datos, operadores, expresiones E/S > 4.67: Pass (3.0)
| *Cuestionario tipos de datos, operadores, expresiones E/S <= 4.67
| | *Cuestionario modularización > 7.33: Pass (18.0/7.0)
| | *Cuestionario modularización <= 7.33: Fail (22.0/3.0)
```

De este árbol de decisión se puede deducir que:

- De los 6 cuestionarios considerados son dos los que pueden considerarse recursos clave: *modularización* y *tipos de datos, operadores y expresiones E/S*.
- El cuestionario que además parece más determinante es el de *modularización*, debido a que si se suspende el correspondiente a tipos de datos, se requiere de una calificación relativamente alta (7.33) para superar la asignatura.

Este modelo permite que el profesor obtenga conocimiento que puede ayudar a detectar alumnos en situación de riesgo de suspender la asignatura y poder así tomar medidas preventivas. Por ejemplo, todos los alumnos deberían tener muy claros los conceptos vistos en *modularización*.

Cabe destacar que el módulo da la opción de guardar los datos utilizados en el formato *arff* de weka, por lo que se facilita que puedan ser posteriormente procesados directamente desde weka. A continuación, se muestra un extracto del fichero que se generaría. Debido a que los nombres de los alumnos fueron cifrados, en el fichero *arff* no aparecen sus nombres, sino su alias. No obstante, el profesor tiene accesible desde el módulo la correspondencia entre los nombres cifrados y los reales.

```
@relation J48
@attribute "*Cuestionario tipos de datos, operadores, expresiones E/S" numeric
@attribute "*Cuestionario estructuras de control" numeric
@attribute "*Cuestionario modularización" numeric
@attribute "*Cuestionario tipos de datos compuestos (homogéneos y heterogéneos)" numeric
@attribute "*Cuestionario vectores y matrices" numeric
@attribute "*Cuestionario matrices y estructuras" numeric
@attribute "Score" {"Pass","Fail"}

@data
7.33,8,7,0,8.17,6,"Fail" %Student 1
8,8,8,0,8,7,"Fail" % Student 2
7.33,4,9.33,0,0,8.5,"Fail" % Student 3
7.33,8,6.67,0,8,7.67,"Pass" % Student 4
8,6,7.33,0,7.5,7.33,"Pass" % Student 5
```

CORRESPONDENCIA

| Index | Name |
|-----------|--------------------|
| Student 1 | ██████████ Oscar |
| Student 2 | ██████████ Pilar |
| Student 3 | ██████████ Antonio |
| Student 4 | ██████████ Mariano |
| Student 5 | ██████████ David |

Conclusiones

En este proyecto se han cumplido todos los objetivos planteados. De este modo, es posible obtener información relevante (conocimiento), a partir de la información almacenada en la base de datos de Moodle y aplicando algoritmos de clasificación basados en árboles. Con esta información que hasta ahora estaba inaccesible a los profesores de forma directa, los docentes podremos responder a preguntas como “¿cuáles son los recursos más relevantes de nuestras asignaturas?”, que a su vez nos permitirá poder detectar situaciones que podrían conducir al alumno a suspender la asignatura, y poder actuar en consecuencia. Así, una vez tengamos suficiente actividad de los alumnos en la plataforma y una calificación, se pueden aplicar estas técnicas para obtener las A&Rs que podrían mejorar el rendimiento académico en los sucesivos cursos.

Bibliografía

- [Bon2007] Boneu, J. M. (2007). Plataformas abiertas de e-learning para el soporte de contenidos educativos abiertos. *RUSC. Universities and Knowledge Society Journal*, 4(1), 8.
- [Bre2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [Fra1992] Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI magazine*, 13(3), 57.
- [Hal2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- [Luq2014] Luque, M., Gibaja E., Zafra, A., Ventura, S. Proyecto de innovación docente para la autoevaluación y mejora de la metodología docente usando encuestas personalizadas en Moodle. *I Seminario Iberoamericano de Innovación Docente de la Universidad Pablo de Olavide*. Sevilla. 20 y 21 de noviembre 2014. ISBN: 84-697-2223-9.
- [Qui2014] Quinlan, J. R. (1993). *C4. 5: programs for machine learning*. Morgan Kaufmann Publishers, San Mateo, CA.
- [Rom2010] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), 601-618.
- [Ros2008] Ros Martínez de Lahidalga, I. (2008). Moodle, la plataforma para la enseñanza y organización escolar. *Ikastorratza, e-Revista de didáctica*, 2, 3-12.
- [Xam2016] XAMPP, 2015. Xampp for windows. <http://www.apachefriends.org/en/xampp.html> (Accedido, Septiembre 2016).
- [Zaf2014] Zafra, A., Gibaja E., Luque, M., Ventura, S. ExportQuiz: Módulo de Moodle para exportar cuestionarios a ficheros de texto. *I Seminario Iberoamericano de Innovación Docente de la Universidad Pablo de Olavide*. Sevilla. 20 y 21 de noviembre 2014. ISBN: 84-697-2223-9.