

Utilización de técnicas de minería de datos para la identificación de rasgos de comportamiento en procesos de aprendizaje colaborativo en modelos de e-learning y b-learning

Using data mining techniques to identify behavioral traits in collaborative learning processes in models of e-learning and b-learning

ÁNGEL COBO ORTEGA¹, PILAR NUÑEZ BLAS², ESMELIN NIQUIN ALAYO² Y
FRANKLIN PUELLES GONZALES²

RESUMEN

Actualmente, las modalidades e-learning y b-learning son cada vez más utilizadas y se acumula una gran cantidad de información que es muy valiosa para analizar el comportamiento de los estudiantes. En estos sistemas de aprendizaje colaborativo se ofrece una gran variedad de canales y espacios de trabajo para facilitar el intercambio de información y la comunicación entre los participantes, pero que resultan muy difíciles de analizar; aun con el uso de las herramientas estadísticas. Se utilizó técnicas de Minería de Datos y la metodología de Proceso de Análisis Jerárquico (AHP) para identificar los rasgos de comportamiento en el aprendizaje colaborativo, el seguimiento de las actividades, la participación en los foros y demás actividades a partir de información extraída de los logs que quedan en los servidores y de los informes de actividad que generan las plataformas de gestión de cursos virtuales.

Se estudió, igualmente, la interactividad en estos sistemas con el Proceso de Análisis Jerárquico. Es por ello que utilizaremos en esta oportunidad Expert Choice, Weka, softwares de AHP, Minería de Datos que nos ayudará a cumplir nuestro propósito.

Palabras clave: e-learning; b-learning; aprendizaje colaborativo; minería de datos; AHP.

¹ Universidad de «Cantabria». Cantabria, España.

² Universidad Nacional «Santiago Antúnez de Mayolo». Huaraz, Perú.

ABSTRACT

Currently e-learning modalities and b-learning are becoming increasingly popular and accumulate a large amount of information that is valuable to analyze the behavior of students. In these systems, collaborative learning provides a variety of channels and workspaces to facilitate information exchange and communication among participants, but this is very difficult to analyze, even with the use of statistical tools. Within this mass of data is very important hidden information that could lead to discovery with data mining techniques. It is used data mining techniques and methodology Hierarchical Analysis of criteria to identify behavior traits in collaborative learning , monitoring activities , participation in forums and other activities from information extracted of the logs that are on servers and activity reports generated by the management platforms online courses. Interactivity will also be studied in these systems with the Hierarchical Analysis Process. That is why this time we will use Expert Choice, Weka, AHP software, data mining that will help us fulfill our purpose.

Keywords: e-learning; b-learning; collaborative learning; data mining; AHP

ICHIKLLACHAW

Kanan qipamanqa e-learningwan b-learningqa pi maypis wanayan, kaychawqa imayka willakuypis quchuqan, kayqa allaapa allim yachakuqkunapa kayanqanta riqinapaq. Kay yachakuykunachawqa atska laayam witsikan hinaman uryana hatunyanmi, tsaywan atinalla willakuykunata nunapura trukanapaq, tsaynaw kaykaq sasam analisanapaqqa, tsaypaqqa istadistikataran wanantsik. Kay willakuychawqa allaapa alli pakakashqa willakuyran kan, kaytaqa Miniriyawanran tarishwan. Tsaypaq Miniriya willakuy tiknika nishqantaran wanakashqa hinaman analisis hirarkiku kanqan mituduluhiyataran (AHP) wanakashqa tsaywan yachakuychaw imanaw kayanqanta riqinapaq, qatiq ruraykunata, hurukunachaw kayanqanta, niykur huk ruraykunatapis logs nishqanpiq hurqakashqawan sirwidurkunachaw haqishqawan, hinaman karupiq yachakuy ruraykunapiq willakuywan.

Hinaman prusisu analisis hirarkikupiq ruraykuna kanqanpiq. Tsayraykurmi kaychawqa Expert Choise, Weka, AHPpiq sofwares, Miniriya willakuy nishqankunata wanashun, tsayran munashqantsik alli kanqa.

Pushaq shimikuna: e-learning; b-learning nishqan; alli yachakuy; miniriya datu nishqan; AHP nishqan.

INTRODUCCIÓN

La aplicación de técnicas de descubrimiento de conocimiento y minería de datos en sistemas educacionales basados en web es un área de investigación muy activa y con indudables aplicaciones prácticas. En el proceso de enseñanza-aprendizaje desempeñan un papel importante las tecnologías de la información, y en particular los entornos

virtuales de enseñanza y las herramientas de la Web social. Diversos estudios concluyen que una mezcla en el aprendizaje, *face-to-face* y el aprendizaje e-learning resulta muy adecuada para implementar nuevos modelos de enseñanza.

Las técnicas de minería de datos, concebidas como metodologías para el análisis inteligente de los datos, ayudan a descubrir conocimiento a partir de la información de utilización de los sistemas de *e-learning* y *b-learning*. Permiten descubrir relaciones entre actividades y estudiantes, ayudan a obtener una mayor retroalimentación de la enseñanza, conocer cómo los estudiantes aprenden, evaluar a los estudiantes por sus patrones de navegación, clasificar a los estudiantes en grupos que responden a comportamientos similares, etc. (Barba-Romero *et al.*, 2005).

MATERIALES Y MÉTODOS

Para el desarrollo de este trabajo de investigación ha sido necesario contar con los siguientes medios y materiales:

SOFTWARE: Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) es una colección de algoritmos de aprendizaje para tareas de Minería de Datos. Se trata de una herramienta totalmente desarrollada en el lenguaje de programación JAVA, que contiene una interfaz gráfica para la aplicación de los algoritmos directamente al conjunto de datos de origen, así como la posibilidad de aplicar los algoritmos desde un programa JAVA independiente.

Expert Choice (www.expertchoice.com) es un software de apoyo para la toma de decisiones, está basado en el Proceso Jerárquico Analítico (AHP, Analytic Hierarchy Process); asiste a los decisores organizando la información relacionada a la complejidad del problema en un modelo jerárquico consistente de un objetivo, escenarios posibles, criterios y alternativas.

Las metodologías utilizadas en la investigación fueron: Proceso de análisis Jerárquico, AHP es una técnica de decisión multicriterio propuesta por Saaty (1980) para resolver problemas complejos de criterios múltiples. El proceso requiere que quien toma las decisiones proporcione evaluaciones subjetivas respecto a la importancia relativa de cada uno de los criterios y que después especifique su preferencia con respecto a cada uno de las alternativas de decisión para cada criterio. El resultado del AHP es una jerarquización con prioridades que muestran la preferencia global para cada una de las alternativas de decisión.

El AHP, mediante la construcción de un modelo jerárquico, permite de una manera eficiente y gráfica organizar la información respecto de un problema, descomponerla y analizarla por partes, visualizar los efectos de cambios en los niveles y sintetizar.

El AHP consta de cuatro etapas básicas:

- a) Modelización del problema.
- b) Evaluación y priorización de alternativas y criterios.
- c) Síntesis del problema.
- d) Análisis de sensibilidad

Minería de Datos: Dada la orientación de esta investigación, cuyo objetivo es coleccionar datos de los modelos e-learning y b-learning, haciendo uso de técnicas de minería de datos, los pasos que se proponen son los siguientes: preparación de los datos, aplicación de las técnicas seleccionadas, evaluación del modelo obtenido, y uso del modelo sobre un caso concreto. Para ello se comenzará con la etapa de pre-procesamiento que incluye la selección y limpieza de datos; luego se empleará los algoritmos K-MEANS y EM como instrumento para la determinación de patrones de comportamiento similares entre los estudiantes. Todo ello se hará sobre dos experiencias concretas, una que responde a un modelo b-learning y otra que responde a un modelo totalmente on-line.

RESULTADOS

Medición de niveles de interactividad utilizando AHP

Cada estudiante que ha participado en los procesos de e-learning y b-learning tiene asociados una serie de atributos o indicadores relativos a su actividad en el curso virtual. En definitiva, a la hora de medir el grado de interactividad del estudiante con el curso virtual hay diferentes criterios a tener en cuenta. En este caso una metodología de decisión multicriterio como es la AHP resulta útil para la obtención de un solo índice de interactividad a partir de los diferentes indicadores o criterios seleccionados.

En las próximas secciones, tomando como referencia esas cuatro etapas básicas, se irá presentando los pasos realizados hasta la obtención de un conjunto de pesos que permitirán ponderar los indicadores de actividad de los estudiantes en el curso virtual. Para ello, además de mostrar el proceso de construcción algebraica, se utilizará el software Expert Choice, que implementa esta metodología de decisión.

Se trabajó con una muestra de estudiantes matriculados en dos asignaturas diferentes de la Universidad de Cantabria (España). Dicha universidad es una universidad presencial pero que cuenta con sistemas de enseñanza virtual que se utilizan como apoyo para diferentes asignaturas presenciales pero también para asignaturas impartidas íntegramente on-line.

En este caso se consideró dos muestras de estudiantes:

- a) Los 343 estudiantes inscritos en una asignatura sobre «Teoría de la Optimización» del primer curso de dos titulaciones: Economía y Administración y Dirección de Empresas. Esta asignatura respondía a un modelo b-learning.

- b) Los 57 estudiantes de primer curso de diferentes carreras inscritos en una asignatura 100% virtual sobre «Herramientas de Trabajo Colaborativo». Esta asignatura respondía a un modelo e-learning.

Aplicación de técnicas de minería de datos sobre los indicadores del modelo b-learning

En esta sección se hace uso de un algoritmo de clustering para tratar de identificar automáticamente grupos de estudiantes que respondan a patrones similares en el uso de la plataforma de enseñanza virtual en la experiencia b-learning (asignatura de Teoría de la Optimización). Una vez identificados automáticamente los grupos, se utilizará los pesos generados con AHP para ordenar los centroides de los grupos de acuerdo al nivel de interactividad global que tengan asociados. Eso permitirá identificar posteriormente los rasgos característicos de cada grupo. Para hacer esto se necesita asociar cada uno de los atributos que se utilizarán para la clasificación con los criterios que se utilizaron en el modelo AHP. La siguiente tabla 5.1 muestra esta asociación. Obsérvese que en alguno de los criterios no existen atributos seleccionados para los estudiantes que permitan obtener un valor para el mismo. En el caso de criterios asociados con más de un atributo, a la hora de la ponderación se considerará la media de los correspondientes atributos.

Tabla1. Criterios asociados con los atributos

| | Teoría de la Optimización | Herramientas de apoyo al trabajo colaborativo |
|--|---|--|
| Modelo de enseñanza | b-learning | e-learning |
| Número de estudiantes | 343 | 57 |
| Ciclo | 1º | 1º |
| Duración de la asignatura | 4 meses | 2 meses |
| Titulaciones | Economía/Adm. Empresas | Todas las de la UC |
| Actividades realizadas presencialmente | Clases teóricas y prácticas. Participación en prácticas de Laboratorio. Evaluaciones presenciales | Ninguna |
| Actividades realizadas virtualmente | Participación en foros, acceso a materiales docentes, tests de autoevaluación, entrega de tareas | Participación en foros, acceso a materiales docentes, tests de autoevaluación, entrega de tareas |
| Número de foros disponibles | 4 foros temáticos más un foro general y otro de noticias | 4 foros temáticos más un foro general y otro de noticias |
| Número de tests on-line a realizar | 4 | 4 |
| Plataforma virtual | Moodle | Moodle |

Actividades con el Algoritmo EM en el sistema b-learning

En la figura 1 aparece el resultado del proceso de clustering realizado con el algoritmo EM de Weka. Como puede apreciarse, el algoritmo identificó de forma automática 5 grupos de estudiantes.

A la vista de los valores medios de los diferentes atributos en cada grupo, además tomando como referencia esos valores medios y la vinculación con los criterios del modelo AHP, se podría obtener un indicador único de nivel de actividad para cada grupo.

Haciendo uso de esta fórmula sobre los centroides de cada grupo, la actividad virtual global de cada uno de ellos muestra el reparto de los estudiantes en los 5 grupos generados por el algoritmo EM y los 4 perfiles de estudiantes que se podrían identificar a partir de las características de cada clúster.

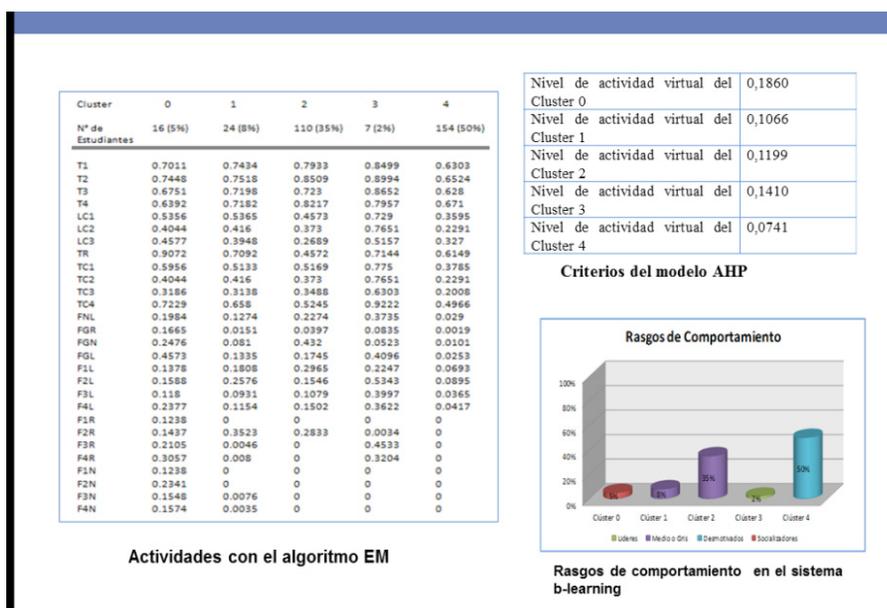


Figura 1. Actividades con el Algoritmo EM en el sistema b-learning

Análisis de la actividad de los estudiantes en los foros con el algoritmo SimpleK-Means en el sistema b-learning

Como segundo análisis se plantea la identificación de patrones de comportamiento de los estudiantes en los foros de debate habilitados en la asignatura.

En la figura 2 se muestra el resultado usando el algoritmo SimpleK-Means que divide a los estudiantes en 4 grupos. Igualmente, haciendo uso de la formula anterior sobre los centroides de dado grupo, en foro global de cada uno de ellos. Además, se muestra el reparto de los estudiantes en los 4 grupos generados por el algoritmo Simplek-Means y los 4 perfiles de estudiantes que se podría identificar a partir de las características de cada clúster.

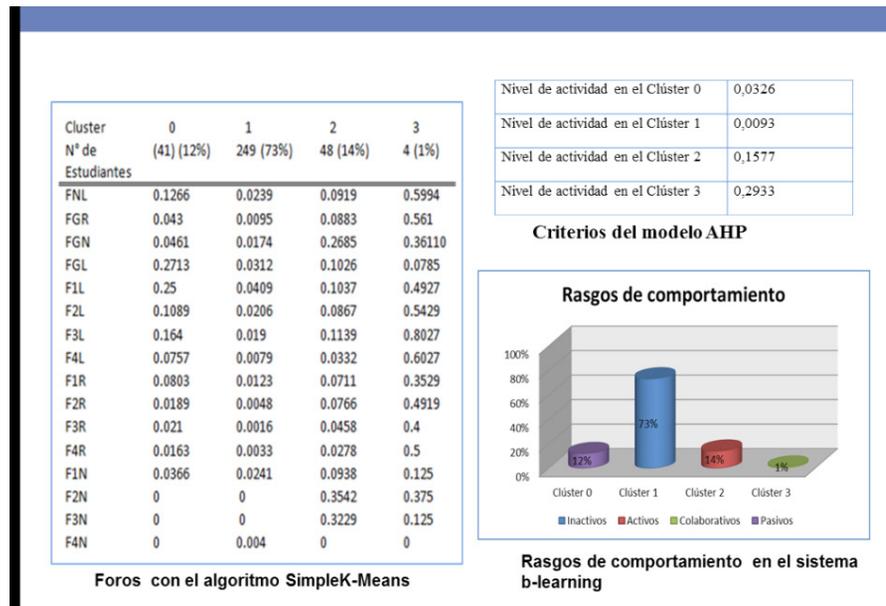


Figura 2. Foros con el algoritmo SimpleK-Means en el sistema b-learning

Aplicación de técnicas de minería de datos sobre los indicadores del modelo e-learning

Con el objeto de analizar si hay diferencias significativas entre los comportamientos de los estudiantes en los modelos b-learning y e-learning, también se ha estructurado en 4 bloques temáticos, cada uno con su foro de debate y su test on-line y un espacio para entrega de tareas. En este caso no hay indicadores de actividad en sesiones presenciales.

Los algoritmos que se utilizarán para realizar agrupamientos de estudiantes son los algoritmos EM y SimpleK-Means.

A. Análisis de las actividades virtuales con el algoritmo EM

Utilizando el algoritmo EM fueron identificados 5 grupos de estudiantes, al igual que en el análisis del curso impartido en modalidad b-learning.

En la Figura 3 aparecen los 5 grupos creados con los valores medios de los atributos de actividades.

Además, tomando como referencia esos valores medios y la vinculación con los criterios del modelo AHP, se podría obtener un indicador único de nivel de actividad para cada grupo.

Haciendo uso de esta fórmula sobre los centroides de cada grupo, la actividad virtual global de cada uno de ellos a la vista de todos los valores obtenidos se podría decir que las características de los diferentes grupos obtenidos serían las siguientes:

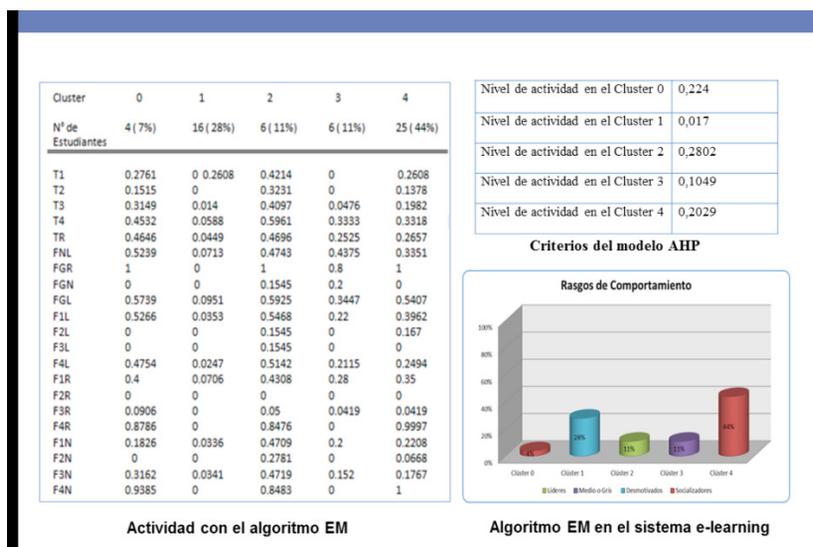


Figura 3. Actividades con el algoritmo EM en el sistema e-learning

Foros con el algoritmo SimpleK-Means en el sistema e-learning

Aplicando los mismos análisis en este modelo o sistema e-learning se plantea la identificación de patrones de comportamiento de los estudiantes en los foros de debate habilitados en la asignatura.

En la figura 4 se muestra el resultado usando el algoritmo SimpleK-Means que divide a los estudiantes en 4 grupos. Haciendo uso de esta fórmula sobre los centroides de cada grupo, en foros globales de cada uno de ellos a la vista de todos los valores obtenidos se podría decir que las características de los diferentes grupos obtenidos serían las siguientes:

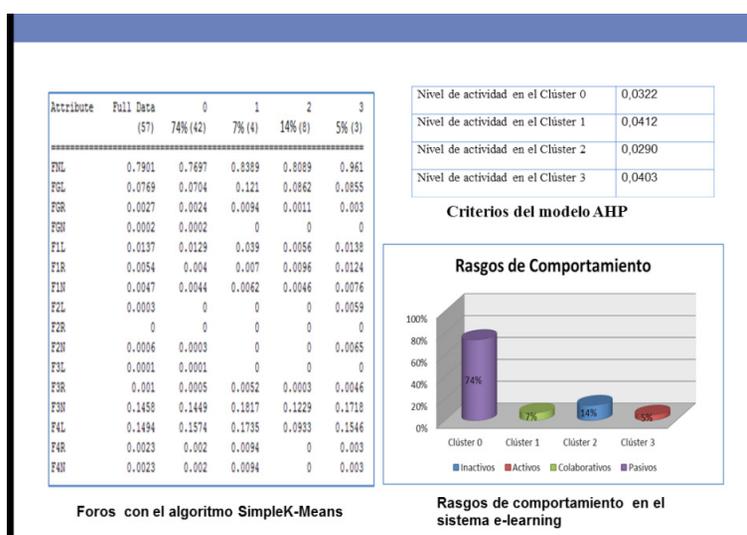


Figura 4. Foros con el algoritmo SimpleK-Means en el sistema e-learning

Comparación de los modelos b-learning y e-learning

En esta fase se compara tanto en los modelos e-learning, y b-learning, los rasgos de comportamiento evaluados con el proceso de análisis jerárquico y la minería de datos. Los resultados se muestra y en las figura 5 y 6.

Comparación de las actividades entre los modelos B-learning y E-learning

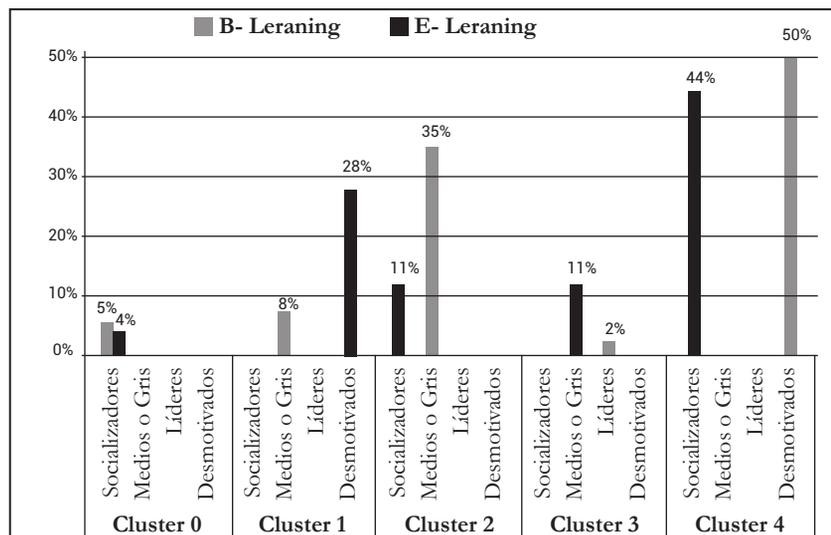


Figura 5. Comparación de las actividades entre los modelos b-learning y e-learning

En la figura 5, observamos la participación de las actividades realizadas en los modelos b-learning y e-learning de acuerdo a los grupos formados de estudiantes. Es así que en el Clúster 0 en el rasgo de comportamiento son similares los estudiantes Socializadores en estos modelos, pues en el Clúster 1 tiene mayor valor significativo el modelo e-learning con el rasgo de comportamiento desmotivados; mientras que en el modelo b-learning destaca el rasgo de comportamiento Medio o Gris. Continuando con el siguiente grupo Clúster 2, el valor significativo está representado por el modelo e-learning con el rasgo de comportamiento de estudiantes Medio o Gris, seguido del rasgo de comportamiento de los estudiantes Socializadores; pero en el modelo b-learning, siguiendo con los rasgos de comportamiento de los estudiantes, en el Clúster 3 observamos que el más destacado es el modelo e-learning con comportamiento Medio o Gris, mientras que en modelos b-learning destacan los estudiantes Líderes. Finalmente, en el Clúster 4 el más significativo es el modelo e-learning con el rasgo de comportamiento de estudiantes Socializadores, dado que en el modelo b-learning destacaron los estudiantes Desmotivadores.

Comparación de los foros entre los modelos b-learning y e-learning

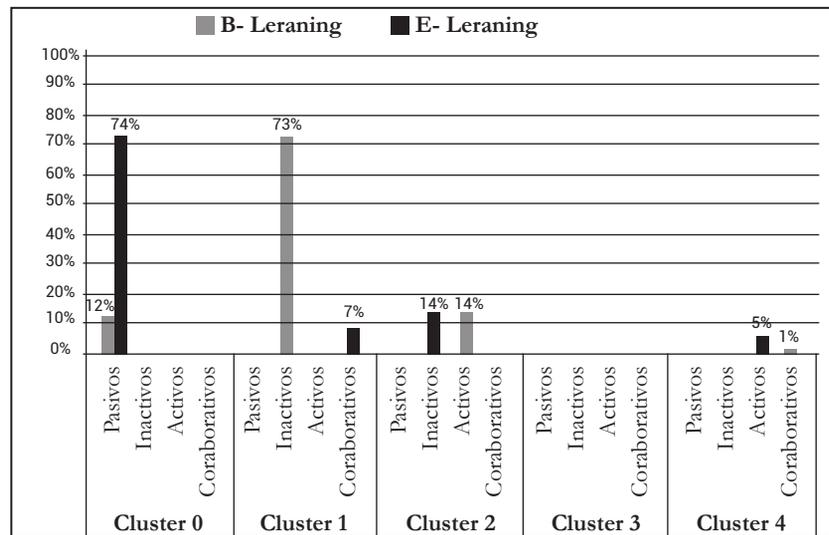


Figura 6. Comparación de los foros entre b-learning y e-learning

En la figura 6, observamos la participación de los foros realizados en los modelos b-learning y e-learning de acuerdo a los grupos formados de estudiantes. En el Clúster 0 ambos modelos son representados por estudiantes Pasivos, pero el más significativo es el modelo e-learning con respecto al modelo b-learning. En el clúster 1 el más significativo es el modelo b-learning, con rasgo de comportamiento de estudiantes Inactivos, mientras que en el modelo e-learning destacan los estudiantes Colaborativos. Siguiendo con la comparación de comportamiento de estos modelos, observamos en el Clúster 2 que ambos modelos son iguales en representación pero en rasgos de comportamiento en el modelo e-learning son estudiantes Inactivos, mientras que en el modelo b-learning son Activos. Finalmente, en el Clúster 3 el más significativo es el modelo e-learning con rasgo de comportamiento de los estudiantes Activos, mientras que en el modelo b-learning destaca el rasgo de comportamiento de estudiantes Colaborativo.

DISCUSIÓN

En este trabajo se ha utilizado técnicas de minería de datos para identificar grupos de estudiantes de acuerdo a su patrón de comportamiento y su rendimiento, así como el acceso a los foros. Se ha observado un predominio de estudiantes que adoptan posiciones más pasivas, pero que, sin embargo, no necesariamente obtienen bajos rendimientos académicos. Sí que es cierto que aquellos estudiantes que están más predispuestos a la colaboración por lo general obtienen mejores resultados.

Lograr la implicación de todo el estudiantado en las iniciativas de los modelos e-learning, b-learning, requiere implementar prácticas pedagógicas que favorezcan no solo

entornos de aprendizaje personalizados sino que se centren en las expectativas e intereses de los estudiantes, logrando una mayor interactividad e implicación de los estudiantes Inactivos y Pasivos en este proceso. Esto significa que, al igual que en el caso de los procesos de formación fuera de línea, no existe una relación de enseñanza y aprendizaje lineal y no se puede prever con precisión el desarrollo y los resultados de los procesos de b-learning y e-learning programados al mismo tiempo. Estas diferencias de rendimiento de los estudiantes revelan que a fin de comprender lo que sucede en las experiencias de e-learning es necesario para ver a los participantes con más interactividad.

Por último, esta investigación nos ha permitido afirmar que los patrones de comportamiento del estudiante en los modelos e-learning y b-learning son muy dispares. Por lo tanto, los estudiantes pueden mostrar diferentes tipos de comportamiento que van desde la participación permanente en el curso virtual.

CONCLUSIONES

El Proceso de Análisis Jerárquico y las técnicas de minería de datos ha permitido obtener el rasgo de comportamiento de los estudiantes en las modalidades b-learning y e-learning, utilizando los pesos generados con AHP, y permitiendo identificar con mucha facilidad los rasgos característicos de cada grupo.

Dentro del campo de la educación, el área de aplicación de minería de datos en los modelos e-learning, b-learning resulta muy útil y de interés para los profesores, a fin de descubrir patrones de comportamiento de sus alumnos así como para tener información objetiva sobre el uso de sus cursos.

El modelo AHP nos ha permitido evaluar el nivel de interactividad de los alumnos, igualmente a identificar un índice cuantitativo de la interactividad del estudiante que puede variar en términos de la relevancia que cada profesor da a los diferentes indicadores utilizados para su construcción.

La herramienta utilizada (Expert Choice) y Weka para realizar el estudio de comportamiento de los estudiantes en las modalidades e-learning y el b-learning ha resultado ser eficiente y rápida.

REFERENCIAS BIBLIOGRÁFICAS

- Anaya, A. y Boticario, F. 2011. *Towards improvements on domain-independent measurements for collaborative assessment Proceedings*. 4th International Conference on Educational Data Mining EDM 2011. Eindhoven 317–318.
- Barba-Romero, Sergio. 2005. *Teoría de decisión multicriterio: conceptos, técnicas y aplicaciones*. Alianza D.L.

- Baker, R. 2011. *La minería de datos para la educación*. En: B. McGaw, E. Baker, y P. Peterson, eds. *Enciclopedia Internacional de educación*. 3rd ed. Vol. 7. Oxford: Elsevier, 112–118.
- Bodea, CN, Dascalu, MI, y Lytras, MD, 2012. *Un recomendador motor para la retroalimentación personalizada avanzada en e-learning entornos*. *Revista Internacional de Ingeniería Educación*. 28 (6), 1326–1333.
- Brusilovsky, P. 2003 «*Adaptive and Intelligent Web-based Educational Systems*», *International Journal of Artificial Intelligence in Education*, IO Press.
- Brusilovsky, P., Schwarz, E., & Weber, G., «*ELM-ART: An intelligent tutoring system on World Wide Web*», *Third International Conference on Intelligent Tutoring Systems, ITS-96* Berlín: Springer Verlag.
- Cobo, A; Rocha, R; Álvarez, 2012. *Análisis y Evaluación del Rendimiento Académico y Patrones de Comportamiento de los estudiantes en un entorno Interactivos B-Learning Utilizando Técnicas De Minería De Datos*. TBIT853838.
- De Wever, B., et al., 2010. *Roles as a structuring tool in online discussion groups: the differential impact of different roles on social knowledge construction*. *Computers in Human Behavior*, 26 (4), 516–523.
- García, A. and Suárez. C., 2011. *Interacción virtual y aprendizaje cooperativo*. Un estudio cualitativo. *Revista de Educación*, 354, 473–498.
- Itmazi, J.A.S., «*Sistema Flexible de gestión del e-learning para soportar el aprendizaje en las universidades tradicionales y abiertas*». PhD Thesis. University of Granada, Spain.
- Saaty, T. 1980. *The analytical Hierarchy Process*. Ed. Mc Graw Hill. New York.

Fecha de recepción: 9 de febrero 2016

Fecha de aceptación: 30 de mayo 2016

Correspondencia

Ángel Cobo Ortega

angel.cobo@unican.es