

# **Análisis de comportamiento de cursos a distancia de UPAE en Moodle utilizando técnicas de Business Intelligence**

**Esther Hochsztain (\*) , Andrómaca Tasistro (\*\*)**

esther@ccee.edu.uy, andromaca.tasistro@agesic.gub.uy

(\*) Departamento de Métodos Cuantitativos.

Facultad de Ciencias Económicas y de Administración. UDELAR

(\*\*) AGESIC. Agencia de Gobierno Electrónico y Sociedad de la Información

## **Resumen**

En este artículo se analiza el comportamiento de los participantes en cursos a distancia Unidad de Perfeccionamiento y Actualización de Egresados (UPAE) y en la plataforma Moodle en base al Framework MoDaWeEd, que integra Moodle, Data Warehouse y Data Mining en el ámbito de la Educación.

En la plataforma Moodle pueden realizarse algunos análisis a partir de la observación de logs de accesos y de los informes disponibles, pero que ello implica una explotación muy reducida de las potencialidades que brindan los datos disponibles.

El preprocesamiento de los logs de acceso permite identificar, genéricamente, intereses de los alumnos, patrones de actividad, horarios y frecuencia, y vincularlo con los períodos de revisiones, exámenes, vacaciones, feriados, entre otros. La aplicación de técnicas de Business Intelligence permite identificar patrones de comportamiento en estudiantes, profesores y administradores. Por último el conocimiento adquirido permite que los profesores adopten decisiones tendentes a mejorar sus cursos, los materiales disponibles y la interacción con los alumnos.

En este artículo en base a los server logs de la plataforma Moodle se aplican técnicas de Data Warehouse y Data Mining.

En base al Data Warehouse se pueden hacer consultas con herramientas OLAP (On Line Analytical Processing). Su principal ventaja es que son rápidas e intuitivas, permiten a los usuarios cambiar su perspectiva de análisis de los datos en forma interactiva recibiendo respuesta en forma inmediata. Las consultas OLAP constituyen un marco estándar y predecible que permite realizar análisis de datos con buen desempeño (performance) y que facilita la presentación de los

mismos. Es muy aplicable al contexto de análisis y toma de decisiones en el ámbito educativo, porque permite identificar con facilidad tanto los comportamientos más frecuentes como los cambios de conducta inesperados.

La aplicación de técnicas de Data Mining es útil para predecir el comportamiento de los participantes de cursos a distancia y segmentarlos, entre otras potencialidades. Permite que los profesores se basen en los análisis realizados para mejorar el vínculo con sus alumnos, anticipar el desempeño de los mismos y de esta forma, si fuera necesario, lo antes posible tomar acciones correctivas tanto para evitar las bajas calificaciones como la deserción.

Debe tenerse en cuenta que mucho se ha avanzado en Uruguay en la aplicación de Moodle en cursos a distancia, pero que falta mucho aún por hacer. Resulta importante lograr la integración de Moodle con los sistemas de bebedías, tanto para facilitar la labor del alumno como para poder contar con un panorama más global del desempeño del alumno.

*Palabras clave:* TIC's, Moodle, educación a distancia, Data Mining, Data Warehouse

## 1. Introducción

Como se ya dijo hace varios años (Minian 1999) “ Hoy, la computadora pasó de ser una sofisticada máquina de calcular veloz, a una máquina para comunicarse. Este es un cambio de paradigma muy importante, ...nos permite transmitir información a través de textos, y ya el proceso de transmisión de información esta en el ámbito del entorno multimedia, en donde el sonido, la voz, el texto y la capacidad de trabajar conjuntamente a distancia es una realidad. Tanto para la educación como para el desarrollo de los pueblos este hecho adquiere relevancia. Lo que antes era imposible ahora es posible. ” ..... “Pensar informáticamente supone operaciones mentales distintas y por lo tanto una propuesta pedagógica específica. No se puede pensar que el poder de la tecnología por sí sólo va a conseguir que los viejos procesos funcionen mejor. Su uso debe servir para que las organizaciones sean capaces de romper los viejos moldes y creen nuevas formas de trabajo y funcionamiento. El planteo debe ser cómo usar las tecnologías para hacer las cosas que todavía no podemos hacer y no sólo cómo poder usarlas para mejorar aquéllas que ya hacemos.”

Esto último resulta de fundamental importancia, y está en la base de la propuesta de este artículo: crear una arquitectura que permita la aplicación de diversas técnicas de análisis, desde consultas SQL que si bien son bien conocidas, no siempre se contaba con datos para poder aplicarlas masivamente en pos de la mejora de la calidad de la enseñanza, hasta consultas OLAP, Data Mining y Web Usage Mining.

En ese sentido, (Martín-Blas and Serrano-Fernández 2009) muestran una aplicación

para complementar los cursos presenciales. Plantean como objetivo la creación de una comunidad de aprendizaje on-line que contribuya a que tanto profesores como estudiantes cuenten con un espacio virtual para compartir conocimiento a través de diferentes actividades supervisadas, chats y foros.

De acuerdo a (Romero, Ventura, and García 2008) Educational Data Mining es una disciplina, obviamente derivada de Data Mining, que se refiere al desarrollo de métodos para explorar los tipos de datos únicos provenientes del contexto educacional. En particular, presentan la instanciación de las etapas del proceso de Data Mining al caso específico de e-learning. Presentan el desarrollo de las siguientes fases: Recolectar datos, Preprocesar datos, Aplicar Data Mining, Interpretar, evaluar y aplicar los resultados. Como puede apreciarse, los nombres de las fases coinciden con las ya conocidas en el área de Data Mining (SEMMA, CRISP-DM), pero lo relevante es que presentan actividades específicas para el caso de la educación.

En particular, el preprocesamiento de los logs de acceso permite identificar intereses de los alumnos, patrones de actividad, horarios y frecuencia, y vincularlo con los períodos de revisiones, exámenes, vacaciones, entre otros. La aplicación de Data Mining permite identificar patrones de comportamiento en estudiantes, profesores y administradores. Por último el conocimiento adquirido permite que los profesores adopten decisiones tendientes a mejorar sus cursos, los materiales disponibles y la interacción con los alumnos.

Los logs de Moodle pueden estudiarse de varias formas: 1) a través de la inspección visual de los mismos, que permite identificar intereses y patrones de acceso, entre otros; 2) mediante los informes y estadísticas provistas por moodle, que resumen las características más importantes de los accesos, tales como por ejemplo el patrón temporal, 3) mediante el análisis de los logs disponibles en una base de datos relacional (MySQL y PostgreSQL son los DBMS más recomendables).

La aplicación de técnicas de Web Usage Mining para predecir las calificaciones de alumnos en el examen final de un curso a partir de datos de Moodle presentada por (Romero et al. 2010), muestra un camino de mejora continua en la explotación de datos de Moodle, dado que permite que los profesores utilicen este tipo de análisis para mejorar el vínculo con sus alumnos, así como para clasificar alumnos. La aplicación de estas técnicas permite anticipar el desempeño de los alumnos, y de esta forma estar en condiciones de tomar acciones correctivas, si fuera necesario, lo antes posible.

Debe tenerse asimismo en cuenta, esta arquitectura, como todo proyecto informático tiene tres etapas, y en cada una de ellas se deben sortear diferentes obstáculos: Inicial o sponsorización del proyecto, Desarrollo e implementación, y Operación y mejora continua.

## 2. Mining Educational Framework

Proponemos un framework que denominamos MoDaWeEd (ver Figura 1), que surge de la integración de Moodle, Data Mining y Web Usage Mining en el ámbito de la Educación. En base a la clásica arquitectura de tres niveles, se consideran tres perspectivas:

- Perspectiva Conceptual
- Perspectiva Lógica
- Perspectiva Física

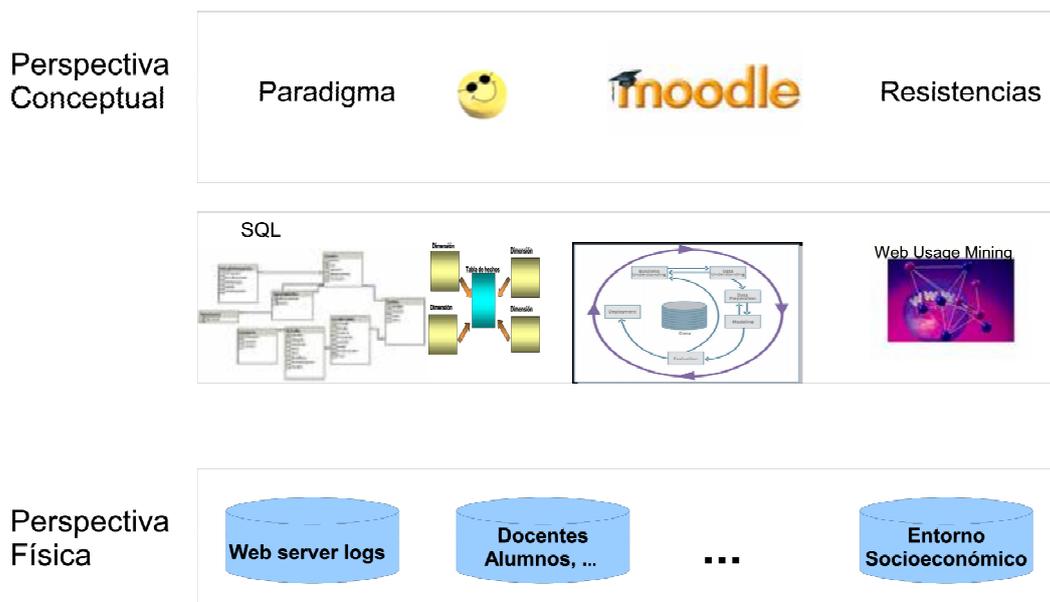


Figura 1: Framework MoDaWeEd

En la perspectiva conceptual se tienen en cuenta:

- Los paradigmas subyacentes
- Las necesidades y expectativas de los usuarios. Los usuarios pueden ser de diversos tipos: estudiantes, profesores, autoridades de la institución educativa, administradores y responsables del sistema informático, funcionarios de la institución
- Potencialidades de Moodle

- Resistencias por parte de los diversos actores involucrados

En la perspectiva lógica es necesario tener en cuenta la forma en que se utilizan las tecnologías disponibles en relación a los elementos planteados en la perspectiva conceptual. Aquí planteamos la existencia de herramientas de distintos niveles, tales como:

- SQL sobre bases de datos relacionales
- Data Warehouse
- Data Mining
- Web Usage Mining

En particular, en el Data Warehouse se pueden hacer consultas con herramientas OLAP (On Line Analytical Processing, Procesamiento Analítico en Línea). La principal ventaja del uso de estas herramientas es que su velocidad permite a los usuarios cambiar su perspectiva de análisis de los datos en forma interactiva, recibiendo respuesta en forma inmediata. Las consultas OLAP constituyen un marco estándar y predecible que permite realizar análisis de datos con buen desempeño (*performance*) y que facilita la presentación de los mismos. Es muy aplicable al contexto de toma de decisiones porque soporta cambios de conducta inesperados de los usuarios finales. Es extensible para ajustarse a nuevos datos (en algunos casos inesperados) y a nuevas decisiones de diseño (Hochsztain and Tasistro 2005).

En este caso se manejan las tablas de hechos (fact-tables) Accesos, Calificaciones On-Line y Calificaciones Off-OnLine. La tabla de hechos Accesos tiene las dimensiones Usuario, Objeto, Fecha y Hora, y como medida la duración. La tabla de hechos Calificaciones On-Line tiene como dimensiones Usuario, Objeto y Fecha, con las medidas: calificación y tiempo (que no siempre corresponde). La tabla de hechos Calificaciones Off-OnLine permite vincular las calificaciones obtenidas por el alumnos, tanto on-line como off-line. Tiene como dimensiones: estudiante, objeto, fecha, tipo de evaluación (on-off line), y como medidas la calificación y el tiempo (que no siempre corresponde).

En el nivel físico, tienen en cuenta esencialmente los datos que servirán para satisfacer las necesidades de los usuarios, tales como:

- Web server logs de moodle
- Información acerca de las funcionalidades disponibles en moodle

- Datos de alumnos
- Datos de profesores
- Datos de estructura de la universidad.
- Datos del entorno socioeconómico

Debe tenerse en cuenta que en Moodle pueden realizarse algunos análisis a partir de la observación de logs de accesos y de los informes disponibles, pero que ello implica una explotación muy reducida de las potencialidades que brindan los datos disponibles. El framework MoDaWeEd permite realizar cluster de alumnos, predecir calificaciones, clasificar alumnos y docentes, así como facilitar el estudio de las causas de la deserción estudiantil.

### 3. Estudio empírico

El uso de la plataforma varía considerablemente según los participantes, como se puede apreciar en las Figuras 2 y 3.

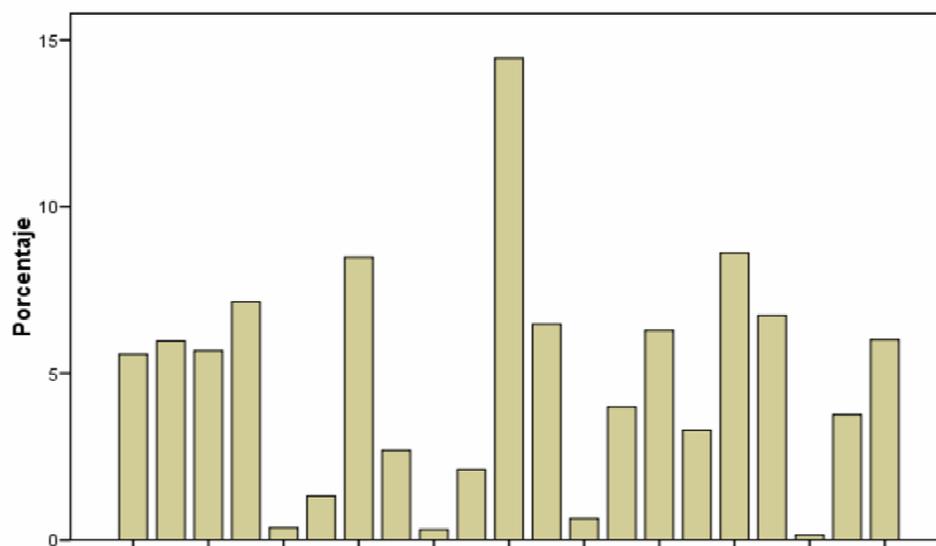


Figura 2: Porcentaje de accesos de los participantes

Por otra parte en base a la Figura 4 se pueden identificar los aspectos que más explican las calificaciones.

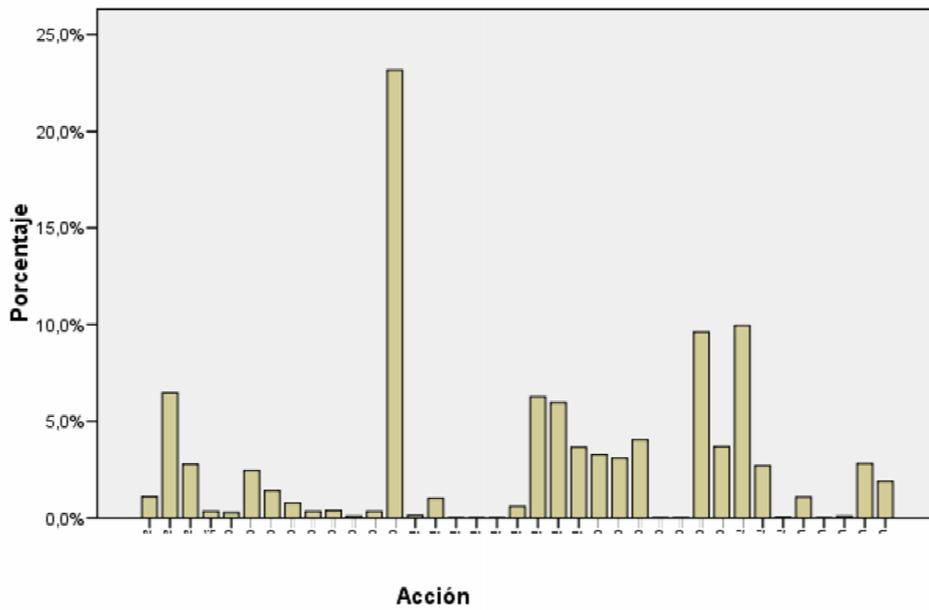
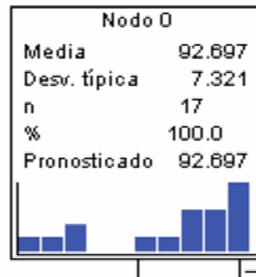


Figura 3: Acciones

La cantidad de accesos no se presenta correlacionada con la calificación obtenida, como lo muestra la Figura 5.

Total del curso

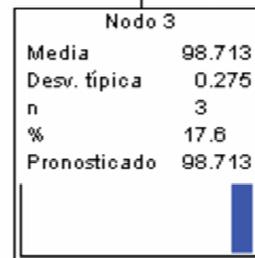
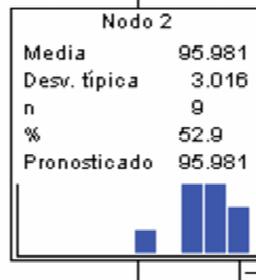
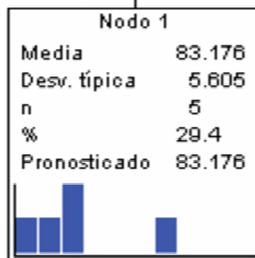


Tarea: Tarea 2.1. Gráficos univariados en R  
 Valor P corregido=0.002, F=23.228,  
 gl1=2, gl2=14

<= 80

(80, 95]; <perdido>

> 95

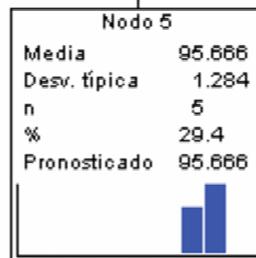


Cuestionario: Cuestionario 4.1.  
 Modelos Lineales. R  
 Valor P corregido=0.018, F=8.436,  
 gl1=2, gl2=6

<= 10

> 10

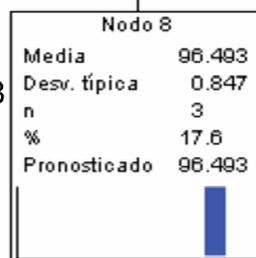
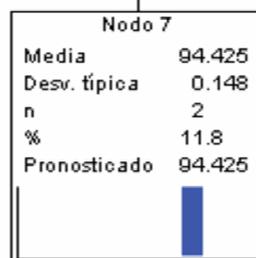
<perdido>



Cuestionario: Cuestionario 1.1  
 Valor P corregido=0.047, F=10.577,  
 gl1=1, gl2=3

<= 9

> 9



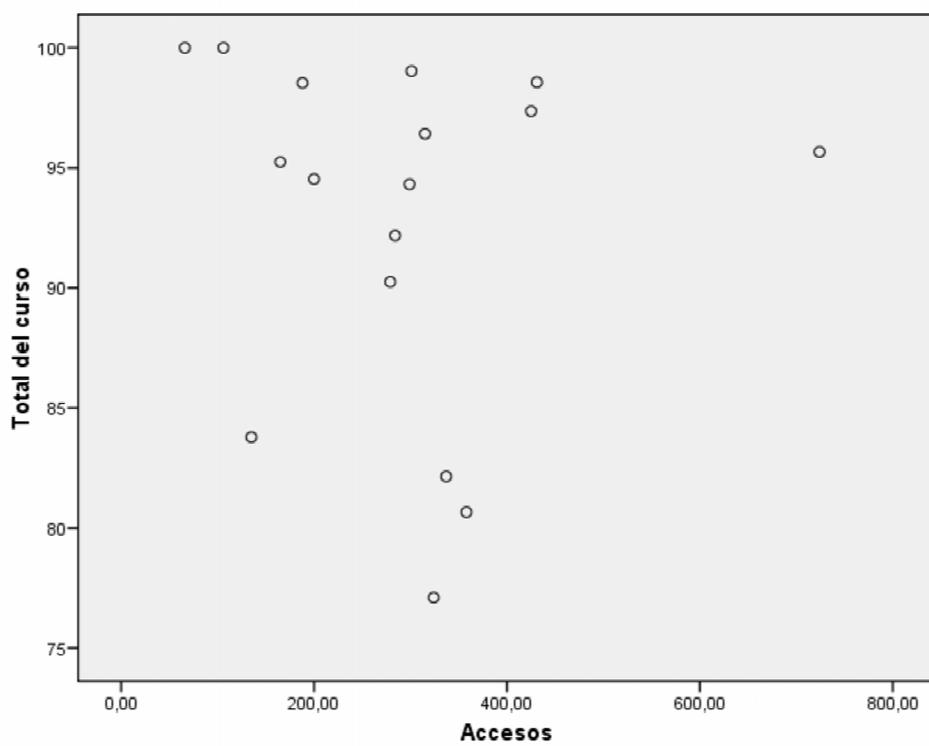


Figura 5: Diagrama de dispersión

## 4. Conclusiones

En este trabajo se propone un framework o marco conceptual que satisfaga las necesidades de los diversos actores involucrados en el entorno de Moodle. El mismo puede ser aplicado a los diversos niveles educativo, desde la educación inicial y primaria hasta la educación universitaria y de postgrados y educación permanente. Asimismo se presentan los resultados iniciales de un estudio de caso.

Debe tenerse en cuenta que mucho se ha avanzado en Uruguay en la aplicación de Moodle, pero que falta mucho aún por hacer. Resulta importante lograr la integración de Moodle con los sistemas de bebedías, tanto para facilitar la labor del alumno como para poder contar con un panorama más global del desempeño del alumno.

## Referencias

- Hochsztain, Esther, and Andrómaca Tasistro. 2005. *Diseño e Implementación de Bases de Datos y Data Warehouses*. Edited by Unidad de Perfeccionamiento y Actualización de Egresados (UPAE). Facultad de Ciencias Económicas y de Administración. Universidad de la República. Montevideo. Uruguay. Unidad de Perfeccionamiento y Actualización de Egresados (UPAE). Facultad de Ciencias Económicas y de Administración. Universidad de la República. Montevideo. Uruguay.
- Martín-Blas, Teresa, and Ana Serrano-Fernández. 2009. "The role of new technologies in the learning process: Moodle as a teaching tool in Physics." *Computers & Education* 52 (1): 35–44.
- Minian, Judit. 1999. "Aplicaciones del uso de la informática y las nuevas tecnologías de la información y comunicación en el ámbito educativo." *Revista electrónica Quaderns Digitals*. <http://www.quadernsdigitals.net> Última fecha consultada: 15/7/2011.
- Romero, Cristóbal, Sebastián Ventura, and Enrique García. 2008. "Data mining in course management systems: Moodle case study and tutorial." *Computers & Education* 51 (1): 368–384.
- Romero, Cristobal, Pedro G. Espejo, Amelia Zafra, Jose Raul Romero, and Sebastian Ventura. 2010. "Web usage mining for predicting final marks of students that use Moodle courses." *Computer Applications in Engineering Education*.