

Business Intelligence in E-learning, a case study of an Ecuadorian University

Malhena Sánchez Peralta
Universidad de Cuenca
Cuenca, Ecuador
malhena.sanchez@ucuenca.edu.ec

Abstract—E-learning platforms are widely used by universities, despite the advantages that these educational environments provide, there are problems in obtaining information about teacher's and student's behavior and level of use on these platforms. Consequently, it is necessary to promote tools for monitoring and analyzing the E-learning environment use. This article shows the application of *Business Intelligence* (BI) and tools for *Online Analytical Processing* (OLAP) in E-learning environments and presents a case study of an Ecuadorian University on how to apply these technologies in various sources of information together with the database of a *Learning Management System* (LMS). For this analysis, the Basic Computing courses have been selected which generates a great quantity of information. The study reveals that the majority of students belong to the Faculty of Economic Sciences and they are enrolled a second time. On the other hand, the results indicate that there is more teaching activity in the Moodle LMS platform in full-time teachers. The study also allowed classifying and predicting the approval of students in the subject of Basic Computing based on significant input attributes.

Keywords- *Business Intelligence, Data Warehouse, Data Mining, E-learning, LMS, OLAP.*

I. INTRODUCCIÓN

El campo de la analítica surge junto con el escenario de crecimiento exponencial de la información almacenada en el mundo digital, y como una solución tecnológica que permite recopilar y analizar datos provenientes de fuentes heterogéneas (servidores, computadoras, dispositivos móviles, etc.) para estimar acciones futuras que generen mejores decisiones y ayuden a adoptar estrategias más eficientes a organizaciones o individuos [1]. En la actualidad es importante que las universidades analicen los datos y la información existente en el Sistema de Administración del Aprendizaje (LMS) instalado en su infraestructura tecnológica, con el fin de convertirlos en conocimiento que sirva de apoyo en la toma de decisiones sobre el proceso de aprendizaje electrónico. Por lo que contar con un sistema de *Business Intelligent* (BI) se convierte en una solución bastante atractiva.

La Universidad de Cuenca, es una institución pública de educación superior del Ecuador, cuenta con varios *sistemas de información* (SI), entre ellos un LMS Moodle donde los docentes en su mayoría con formación en TIC lo emplean como una herramienta de apoyo a la docencia con modalidad presencial (b-learning). Así también, posee en internet un SI integrado conocido como eSIUC para estudiantes, docentes y empleados. En el análisis se ha seleccionado a estudiantes de la asignatura de Informática Básica de carácter obligatorio para

todos los estudiantes que cursan el primer semestre de estudios, lo que genera gran cantidad de información.

Este artículo está estructurado en 7 secciones, además de la introducción, la sección 2 enuncia la problemática que motivó el estudio; la sección 3 describe las actividades y resultados desarrolladas con la metodología planteada; las secciones 4 y 5 explican la implantación del Data Warehouse y del Data Mining respectivamente; la sección 6 presenta las conclusiones y la sección 7 plantea el trabajo futuro.

II. EL PROBLEMA

La Universidad de Cuenca tiene en producción varios SI para el desarrollo de sus gestiones, estos sistemas al ser heterogéneos hacen difícil la integración de información para el análisis y la evaluación de la situación actual que sustente la toma de decisiones. Enfocados en el ámbito de enseñanza aprendizaje mediados por TIC, es perentorio integrar la información almacenada en diferentes fuentes y formatos para analizarla y conocer cuál ha sido su impacto en docentes y estudiantes a fin de disponer acciones que apoyen la práctica.

III. METODOLOGÍA

Se utilizó la metodología Hefesto (ver Fig. 1) creada por Darío Bernabeu que parte de la recolección de requerimientos de información del usuario, seguido de los procesos de *extracción, transformación y carga de datos* (ETL) hasta definir un esquema lógico para la organización, sean estos Data Marts o *Data Warehouse* (DW) [2].



Fig. 1. Los cuatro pasos de la Metodología Hefesto (Bernabeu, 2010).

A continuación se describe cada una de las actividades principales desarrolladas con el método y los resultados más significativos que se obtuvieron.

A. Análisis de Requerimientos

El caso de estudio abordó los siguientes requerimientos:

- Conocer cuántos estudiantes por carrera o por facultad se han matriculado en el curso virtual de Informática Básica agruparlos por docente en un semestre determinado, indicando el número de matrícula realizada en la asignatura (primera, segunda o tercera).
- Conocer los docentes que con mayor frecuencia crean cursos virtuales y el tipo de relación laboral que mantienen (dedicación a tiempo completo, medio tiempo, o tiempo parcial, si es un docente principal, agregado, contratado, etc.).
- Clasificar a los alumnos de Informática Básica según su rendimiento y carrera a la que pertenecen y predecir la aprobación de la materia con anticipación en función del análisis de variables de aprovechamiento inicial y del entorno.

Para la construcción del DW, en base en los requerimientos anteriores se hallaron los siguientes indicadores y perspectivas:

- Total de estudiantes matriculados por carrera, por facultad, por docente, por número de matrícula en un semestre determinado. *Indicadores:* Total Matriculados. *Perspectivas:* estudiante, carrera, facultad, docente, número de matrícula, tiempo.
- Docentes por carrera con cursos virtuales según su relación laboral. *Indicadores:* Total Actividad Docente. *Perspectivas:* docente, curso virtual en Moodle, carrera, facultad, dedicación docente, tipo de docente, tiempo.

Luego se construyeron los modelos conceptuales correspondientes.

B. Análisis de los OLTP

Para el análisis del *On Line Transaction Processing* o *Proceso de Transacción en Línea* (OLTP) se conformaron los siguientes indicadores:

- Total matriculados. *Hechos:* Total matriculados. *Función de sumarización:* SUM. *El indicador representa:* la sumatoria de estudiantes matriculados de una carrera o facultad por docente, por número de matrícula en un semestre determinado.
- Actividad Docente. *Hechos:* Total Actividad Docente. *Función de sumarización:* SUM. *El indicador representa:* la sumatoria de docentes que tienen cursos virtuales en la plataforma Moodle por carrera, por tipo de docente y según su tiempo de dedicación en un semestre dado.

Se realizó el análisis de las fuentes OLTP para establecer correspondencias, en primer lugar se recopilamos los datos provenientes de diferentes fuentes a citar: tabla de usuarios de la base de datos MySQL del LMS Moodle, datos de docentes y

estudiantes de archivos en Excel exportados desde el eSIUC, resultados de consultas sql a *Base de Datos* (BD) y datos de archivos de texto con información complementaria.

En consecuencia se obtuvo un volumen considerable de información para el análisis y evaluación. Cabe indicar que en el LMS Moodle solo residía información del último semestre de la asignatura de Informática Básica, ya que por políticas administrativas la información previa es eliminada. Sin embargo del sistema eSIUC se obtuvo el historial de registros de calificaciones desde el año 2014 que sirvieron para el requerimiento de clasificación y predicción.

Uno de los resultados más significativos de esta actividad fue la selección de tablas de la base de datos MySQL del LMS Moodle luego de analizar su diagrama entidad relación.

Se definió un nivel de granularidad para cada perspectiva, varios campos de las fuentes no fueron considerados por no ser relevantes para el análisis. En base en lo anterior, se obtuvieron Modelos Conceptuales Ampliados, como es el caso para el total de Matriculados (ver Fig. 2).

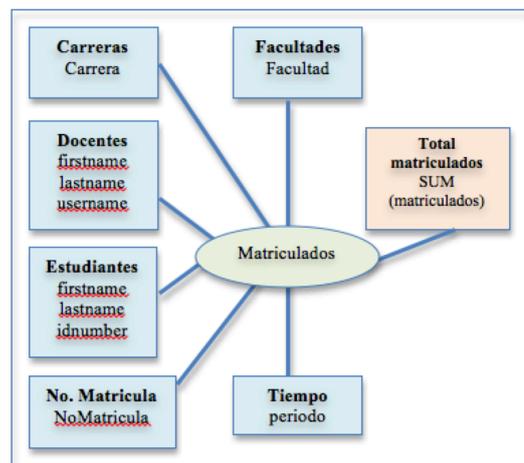


Fig. 2. Modelo Conceptual Ampliado Total Matriculados.

C. Modelo Lógico del DW

Para construir el modelo lógico del DW se seleccionó el esquema estrella que mejor se adapta a los requerimientos y necesidades del usuario [3], de esta manera se definieron Tablas de Dimensiones, Tablas de Dimensiones Conformadas y Tablas de Hechos para cada perspectiva.

Integración de Datos

En esta fase de la metodología se realizó el proceso de *Extracción, Transformación y Carga de datos* (ETL) desde las fuentes detectadas para el análisis, para ello se usó el software Pentaho que posee un conjunto de herramientas de productividad para el análisis, minería y reporte de datos:

- Carga inicial:** Con ayuda de *Pentaho Data Integration* (PDI) o *Kettle-Spoon* se llevó a cabo el proceso ETL que permitió conectar los diversos orígenes de datos y transformarlos para cargarlos dentro de la estructura del DW. Entre las tareas de transformación que con mayor frecuencia se realizaron fueron: cambio de minúsculas a mayúsculas, concatenación de cadenas de texto, mapeo

de datos, selección de filas y ordenamientos de los flujos de datos, previo a la unión entre fuentes y carga en el DW.

- **Actualización:** La estructura del DW corresponde a una base de datos MySQL creada a través de la herramienta MySQLWorkbench.

IV. IMPLANTACIÓN DEL DATA WAREHOUSE

El Data Warehouse está estructurado por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema. A continuación se describen las actividades realizadas durante su implementación.

A. OLTP y Gestión de Carga

Los elementos del OLTP y la gestión de carga se completaron en la fase anterior con Hefesto, donde se realizó el proceso ETL y se creó la base de datos *ElearningBI*.

B. Gestión del DW

Un DW está basado en un modelo multidimensional para mejorar el análisis y el resumen de los datos. Este modelo de datos es un espacio n-dimensional llamado cubo de datos o un hipercubo [4]. En la Gestión del DW se combinó el sistema de gestión de BD SGBD de MySQL que almacena los datos de forma multidimensional a través de tablas de hechos y tablas de dimensiones y la creación de cubos multidimensionales que agilitan la recuperación de datos en los procesos de consulta.

La creación de cubos OnLine Analytical Processing para el procesamiento analítico en línea (OLAP) se realizó con la herramienta Schema-Workbench de Pentaho.

C. Gestión del Consulta

El componente de gestión de consulta del DW, Query Manager, realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales y de análisis de datos. Recibe las consultas de los usuarios y las aplica a la estructura de datos correspondiente como un cubo multidimensional y devuelve los resultados obtenidos. Para el caso de estudio se realizó operaciones sobre los cubos a través de la herramienta Saiku Analytics de Pentaho BI-Server.

Sobre el cubo “Matriculados” se realizó la operación Número Total de Estudiantes por Número de Matrícula agrupados por Facultad. Los valores indican que mayoritariamente los estudiantes están cursando su segunda matrícula en la asignatura de Informática Básica y son pocos los que cursan su tercera matrícula. Por otro lado la mayor cantidad de estudiantes se concentra en la Facultad de Ciencias Económicas y Administrativas.

Sobre el cubo “Actividad Docente” se realizó la operación Total Actividad Docente por Tipo y Dedicación del docente, los resultados obtenidos indican que existe mayor actividad docente en la plataforma LMS Moodle en docentes de tipo agregado con dedicación a tiempo completo, mientras que la menor actividad se registra en docentes de tipo ocasional y docentes con dedicación a tiempo parcial. Cabe indicar que los docentes de tipo contratado y con dedicación a tiempo completo superan ampliamente la actividad frente a docentes principales bajo cualquier tipo de dedicación.

D. Herramientas de Consulta y Análisis

Las herramientas permiten explorar los datos, son el nexo entre el DW y los usuarios, para ello utilizan la metadata de los cubos multidimensionales a fin de realizar consultas SQL.

Dashboard: Es una herramienta de consulta y análisis que permite al usuario realizar consultas al DW a través de una interfaz gráfica e intuitiva. Para el caso de estudio se empleó el editor de cuadro de mando *Community Dashboard Editor* (CDE) de Pentaho en la creación de un cuadro de mando para los modelos multidimensionales diseñados.

V. MINERÍA DE DATOS

Data Mining o *Minería de datos* (DM) es el proceso de análisis de datos desde diferentes perspectivas que resume información útil y permite encontrar correlaciones o patrones entre docenas de campos en grandes bases de datos y almacenes [5].

En las actividades realizadas sobre minería de datos, en el caso de estudio se utilizó la herramienta gráfica *Waikato Environment for Knowledge Analysis* (Weka) que posee una colección de algoritmos de aprendizaje automático para tareas de minería de datos como son: pre-procesamiento de datos, clasificación, regresión, agrupamiento o clustering, reglas de asociación, y la visualización de información [6]. Como fuentes de datos se trabajó con los resultados almacenados en las tablas de hechos que poseen valores numéricos, y utilizando Pentaho Kettle-Spoon se guardaron los resultados en Excel con valores nominales previo el almacenamiento en la BD del DW.

A continuación se describe únicamente la minería de datos del proceso de clasificación, el de predicción emplea los mismos que de clasificación con la diferencia de que los datos a predecir corresponden a valores numéricos continuos.

A. Clasificación

El problema de la clasificación es el más frecuente en la práctica. Una vez aplicados los algoritmos no supervisados de agrupamiento y asociación se aplica la clasificación como un refinamiento en el análisis [7]. Es un método aplicado en problemas predictivos con un aprendizaje supervisado, es decir recibe previamente del experto los posibles valores de clasificación en función de la experiencia. El método devuelve como resultado un número finito de valores de clasificación.

- **Definición del problema:** Clasificar los estudiantes de Informática Básica en aprobados o reprobados en función de atributos significativos de entrada.
- **Obtención de datos para el análisis:** Los datos se obtuvieron de un archivo separado por comas creado a partir de tres archivos en Excel exportados del sistema eSIUC, cada uno correspondía a un semestre distinto con el historial de registros de calificaciones de los estudiantes de Informática Básica desde el año 2014. Se seleccionaron los atributos nominales más significativos como: carrera, nota de aprovechamiento que representa el 50% del total requerido para aprobación, semestre de estudios, número de matrícula y el atributo estado (aprobado, reprobado o suspenso)

con valores ingresados en base a la experiencia de docentes de la asignatura; se discretizó el valor del aprovechamiento en la fase de pre-procesamiento de Weka.

- **Elección del algoritmo de análisis:** El algoritmo seleccionado fue J48 que es una implementación del algoritmo C4.5, uno de los algoritmos de minería de datos más utilizado. Se trata de un refinamiento del modelo generado con OneR, supone una mejora moderada en las prestaciones, y podría conseguir una probabilidad de acierto ligeramente superior [7]. El parámetro más importante a tener en cuenta fue el factor de confianza para la poda, que influye en el tamaño y capacidad de predicción del árbol construido. Para el estudio se utilizó el valor por defecto de 25%.
- **Aplicación del algoritmo:** El algoritmo J48 se utilizó con un modo de evaluación “Percentage Split” que divide los datos en dos grupos.
- **Selección del modelo:** Como resultado se construyó un modelo para predecir la categoría “estado” con valores de aprobado y reprobado en función de los atributos de entrada: carrera, nota de aprovechamiento, semestre y número de matrícula. El modelo se considera un buen modelo ya que el porcentaje de clasificación correcta de los datos es del 85,7% con un error del 14.3%.
- **Análisis de los resultados:** Los resultados indican a los atributos “total de aprovechamiento” y “nummatricula” como los atributos más significativos para realizar la clasificación: (1) Estudiantes con aprovechamiento mayor a 0 y menor a 30 se clasifican como reprobados, (2) Estudiantes con aprovechamiento mayor a 31 y menor a 45 se clasifican como aprobados y (3) Estudiantes en su primera matrícula se clasifican como reprobados, mientras que aquellos que realizan su segunda matrícula en la asignatura se clasifican como aprobados. De los 763 casos presentados, 513 son correctamente clasificados como aprobados y 141 son correctamente clasificados como reprobados.

VI. CONCLUSIONES

Este artículo muestra el uso de BI en el entorno de E-learning de la Universidad de Cuenca, revela cómo aplicar estas tecnologías en la integración de diversas fuentes de información junto con la base de datos del sistema de gestión de aprendizaje LMS Moodle que posee la institución.

El desarrollo de BI requiere de una adecuada y efectiva implantación de metodologías apoyadas de tecnología de punta, pero ante todo y lo más importante es poseer un conocimiento profundo de todos los factores que afectan los datos que se generan, almacenan y eliminan en la organización, tanto interna como externamente, es de suma importancia mantener históricos de información relevante ya que son el insumo necesario para iniciar con un sistema de BI.

A través del proceso de análisis de las necesidades de información, las etapas de construcción y despliegue del DW, y la minería de datos con la generación de dashboards para la interacción del usuario con la información, se demuestra la

complejidad pero también los beneficios del uso del BI en el sistema de E-learning de la universidad.

Los resultados más interesantes que el estudio revela sobre el comportamiento de los docentes y los estudiantes en la plataforma LMS Moodle indican que los estudiantes de Informática Básica en su mayoría pertenecen a la Facultad de Ciencias Económicas y están cursando su segunda matrícula; también exponen que existe mayor actividad docente en la plataforma LMS Moodle en docentes de tipo agregado con dedicación a tiempo completo, mientras que la menor actividad se registra en docentes de tipo ocasional y docentes con dedicación a tiempo parcial. Por otro lado el estudio permitió también clasificar y predecir la aprobación de los estudiantes en la asignatura de Informática Básica en función de atributos significativos de entrada.

VII. TRABAJO FUTURO

El BI implantado proporciona información útil, sin embargo quedan áreas pendientes de desarrollo.

La creación del sistema de inteligencia de negocios para la gestión de conocimiento del E-learning en la universidad debería ser una solución integrada a los sistemas de información ya existentes y a nuevas fuentes de datos, de tal forma que permita a las autoridades universitarias y en general a los usuarios finales el análisis y visualización sencilla de resultados consolidados que aporte significativamente en la toma de decisiones académicas y administrativas en el ámbito de la educación mediada por TIC.

A un futuro el estudio podría ser ampliado para ayudar a los docentes a optimizar los materiales de aprendizaje electrónico mediante el descubrimiento de brechas de competencias antes y después de los procesos de educación y mediante el reconocimiento de estilos de aprendizaje electrónico de los participantes, para estos será necesario la automatización de los procesos de ETL y el desarrollo de la capa de presentación para el usuario final.

REFERENCIAS

- [1] Lemos dos Santos, H., Cechinel, C., Carvalho Nunes, J., & Ochoa, X. (2017). An Initial Review of Learning Analytics in Latin America. *Twelfth Latin American Conference on Learning Technologies (LACLO)*.
- [2] Bernabeu, D. (2010). Data PRIX. Obtenido de Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/-metodologia-hefesto/52-descripcion>
- [3] Peralta, V. (2014). Diseño Lógico de Data Warehouses a partir de Esquemas Conceptuales Multidimensionales. Uruguay.
- [4] ZORAN, N. (2012). Data warehouse for an elearning platform. *LATEST TRENDS on COMPUTERS, II*.
- [5] Julija Lapuh, B., & Darko, B. (2015). Business Intelligence in e-Learning. *International Conference on Interactive Collaborative Learning (ICL)*
- [6] Documentación de las herramientas de Pentaho en el sitio oficial. (s.f.). Obtenido de www.pentaho.com
- [7] García Jiménez, M., & Álvarez Sierra, A. (2014). Análisis de Datos en WEKA – Pruebas de Selectividad.