

Análisis del desempeño del profesorado universitario en el uso de MOODLE a través de técnicas de minería de datos: propuestas de necesidades formativas

Analysis of university faculty performance in the use of MOODLE through data mining techniques: training needs proposals

Pedro Ernesto Camacho Chacón

Universidad Autónoma de Yucatán. Mérida, México
pedrocamacho@live.com.mx

Alfredo Zapata González

Universidad Autónoma de Yucatán. Mérida, México
zgonzal@correo.uady.mx

Víctor Hugo Menéndez Domínguez

Universidad Autónoma de Yucatán. Mérida, México
mdoming@correo.uady.mx

Pedro José Canto Herrera

Universidad Autónoma de Yucatán. Mérida, México
pcanto@correo.uady.mx

Resumen

Los Sistemas de Gestión de Aprendizaje son empleados por numerosas instituciones educativas de nivel superior para organizar y distribuir cursos en línea, convirtiéndose en fuente valiosa de conocimiento de la actividad de estudiantes y profesores. Este trabajo está orientado al análisis del desempeño de 484 profesores de la Universidad Autónoma de Yucatán (México) durante el período enero - julio del año 2016, que utilizaron la plataforma MOODLE como apoyo tecnológico a su actividad docente, con el objetivo de identificar sus patrones de comportamiento en el uso de las actividades y recursos que contiene esta herramienta tecnológica. La actividad de los profesores en la plataforma "UADY Virtual" fue analizada a través del Método de Extracción de Conocimiento en Bases de Datos (KDD), generando grupos mutuamente excluyentes, reglas de asociación y clasificación que establecen perfiles de profesores y herramientas que deben ser promovidas en esquemas de capacitación. Adicionalmente, se realizó un análisis estadístico del uso de la plataforma. Se determinó que no existen diferencias significativas en el uso de la plataforma MOODLE y el área de conocimiento del profesor, lo que significa que la manera en la que interactúan los profesores con "UADY Virtual" es igual para todas las áreas de conocimiento.

Palabras clave

Procesamiento de datos, Interpretación de datos, Usos de la tecnología en la educación.

Abstract

The Learning Management Systems are used by numerous higher education institutions to organize and distribute online courses, becoming a valuable knowledge source of the students and teachers activity. This work is oriented to the performance analysis of 484 educators at the Autonomous University of Yucatan (Mexico) during the period January - July 2016, who used the MOODLE platform as a technological support to their teaching activity, in order to identify their behavior patterns in the use of the activities and resources that this technological tool contains. The activity of the teachers in the "UADY Virtual" platform was analyzed through the Knowledge Discovery in Databases Method (KDD), generating mutually exclusive groups, association and classification rules that establish educators profiles and tools that should be promoted in schemes training. Additionally, a statistical analysis of the use of the platform was made. It was determined that there are no

significant differences in the use of the MOODLE platform and the teacher's knowledge area, which means that the way in which teachers interact with "UADY Virtual" is the same for all areas of knowledge.

Key words

Data Processing, Data Interpretation, Technology Uses in Education.

1. Introducción

En las últimas décadas, el tema de la comunicación en la sociedad se ha transformado aceleradamente debido a los avances tecnológicos y al acceso, cada vez más fácil, que tienen las personas para usar tanto la Internet como los dispositivos digitales. El investigador Drucker (1992) afirma que el concepto de sociedad del conocimiento surge a consecuencia de la implementación de las Tecnologías de Información y Comunicación (TIC) en la vida cotidiana, y afecta las relaciones sociales, culturales y económicas, en donde se eliminan las barreras de espacio y tiempo, y se facilita una comunicación ubicua y asíncrona.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) expone que las TIC contribuyen a la mejora de la educación en el mundo, tanto al facilitar el acceso universal de la misma y el desarrollo de los profesores, como en la eficiencia en los aspectos administrativos y de gestión escolar. Asimismo, este organismo promueve que la educación en línea es pieza clave para lograr una sociedad basada en el conocimiento (UNESCO, 2002).

Las instituciones educativas de diferentes países se han ajustado desde principios de los años noventa a dicha influencia, impulsada por la expansión de las TIC; asimismo, la globalización tiende y promueve una nueva cultura mundial, en donde se rompen las barreras geográficas y económicas (García, 2013). En este período, el mayor crecimiento se experimentó en los Estados Unidos de América (EUA), donde se comenzaron a utilizar recursos económicos para incrementar el acceso a Internet, y a finales de la década, más del 90% de las escuelas contaban con ordenadores y conexiones a ésta (Anderson & Becker, 2001). En cuanto a América Latina, en ese lapso de tiempo se abrió el paso a las políticas públicas que vinculan las TIC con la educación. En los últimos años, esta región ha sido muy dinámica en la integración de las tecnologías a su sector educativo, por lo que han surgido políticas que se establecen principalmente en tres experiencias (Lugo, López & Toranzo, 2014):

- a. Reducción de la brecha digital,
- b. Democratización del acceso a las TIC y
- c. La consideración de la situación en la que se encuentre cada país al momento de implementar las iniciativas.

Asimismo, el Banco Mundial (2013) comienza a apoyar la introducción de las TIC en el aula para los países de América Latina, con el objetivo de expandir el acceso a la educación, para en consecuencia, abatir la pobreza al lograr incrementar las oportunidades de empleo de los ciudadanos, al prepararlos con los conocimientos de alcance mundial más recientes.

En México, se pueden apreciar políticas encaminadas a incorporar las TIC en el desarrollo del sector educativo, tales como el Programa Especial de Ciencia, Tecnología e Innovación (2014-2018). Este programa, respaldado por el Plan Nacional de Desarrollo 2013-2018, tiene como punto de partida el objetivo de “hacer del desarrollo científico, tecnológico y la innovación, pilares para el progreso económico y social sostenible” (p. 5), con lo que se persigue una educación de calidad. Para ello el gobierno federal estableció cinco estrategias, entre ellas “contribuir al fortalecimiento de la infraestructura científica y tecnológica del país (p. 41). De lo anterior radica la importancia que México le está dando a la incorporación de las TIC a la educación del país (CONACYT 2014).

La educación basada en TIC generalmente se imparte a través de sistemas de gestión del aprendizaje, los cuales resguardan grandes cantidades de información relacionada con los profesores y estudiantes, por ejemplo, horarios de conexión, cantidad de páginas visitadas, duración de estadía en cada página, tipos de herramientas empleadas, número de tareas entregadas, mensajes compartidos en foros, entre otros. De acuerdo con Trandafil, Allkoçi, Kajo y Xhuvani (2012), las instituciones educativas se encuentran abrumadas por la gran cantidad de información almacenada en sus entornos virtuales de aprendizaje. Lo anterior ha originado una complejidad en el proceso de análisis de información. Transformar en conocimiento la actividad que se almacena cada día dentro del entorno de estos sistemas representa un gran desafío científico y tecnológico.

Desafortunadamente, en muchas ocasiones las instituciones educativas no aprovechan del todo esta información. De acuerdo a lo anterior, es necesario realizar investigaciones que permitan extraer conocimiento del uso de estos entornos virtuales de aprendizaje.

Esta propuesta se enfoca en el análisis del desempeño del profesorado universitario en un Sistema de Gestión del Aprendizaje a través del modelo denominado Descubrimiento del Conocimiento en Bases de Datos. El objetivo general de esta propuesta es reconocer los patrones de comportamiento del profesorado universitario en el uso de las actividades y recursos que contiene el entorno de esta herramienta tecnológica. Este objetivo general se detalla a través de los siguientes objetivos específicos:

- Determinar si existen diferencias significativas en el uso de las actividades y recursos del sistema por parte de los profesores universitarios en función de su área de conocimiento.
- Establecer si existen necesidades de formación en el uso de las actividades y recursos del sistema de gestión de aprendizaje.

La parte experimental se desarrolló en el entorno virtual de una universidad pública ubicada en el sureste de México. Para el análisis de la información se implementaron diversas técnicas de minería de datos. Adicionalmente, se desarrollaron tanto un análisis descriptivo como inferencial. El artículo continúa con una explicación breve de la fundamentación teórica, el entorno de la investigación, una descripción de la metodología seguida, el análisis de los resultados y las principales conclusiones.

2. Fundamentación teórica

En esta sección se describen conceptos teóricos que se relacionan con el trabajo de desarrollado. Se inicia con el aprendizaje electrónico y su evolución. Posteriormente, se realiza una descripción de las características de los Sistemas de Gestión del Aprendizaje, así como algunos ejemplos de dichos sistemas; enseguida, se explica el proceso para la extracción de conocimiento en bases de datos, para el cual se profundizará en la técnica de minería de datos. Finalmente, se describen algunas investigaciones relacionadas con esta propuesta.

2.1. El aprendizaje electrónico y su evolución

De acuerdo con Cooke (2004), el aprendizaje electrónico (en inglés, *e-learning*) surge como necesidad por parte de las empresas para capacitar a sus trabajadores. Alrededor de los años noventa, diversas compañías comenzaron a usar videos para sus cursos de capacitación, sin embargo, presentaban varias limitaciones: incapacidad por ajustar el contenido a las necesidades de los trabajadores, costo de mantención elevado y el contenido no podía ser actualizado fácilmente. Adicionalmente se requería del equipo adecuado para ello, y el seguimiento y evaluación de los trabajadores resultaba difícil (Hoffmaster, 2006).

Con el continuo avance de las TIC y con la introducción de estrategias pedagógicas a los cursos de capacitación, cuyo contenido se distribuía a través de la Internet, el aprendizaje electrónico evoluciona, y se denomina como aprendizaje en línea (en inglés, *online learning*) (Clarke, 2002).

En la literatura se encuentran diversas definiciones sobre aprendizaje en línea; ante esta diversidad se han seleccionado algunas de ellas:

- Es el uso de nuevas tecnologías de multimedia e Internet para el mejoramiento de la calidad del aprendizaje al facilitar el acceso a recursos y servicios, así como el intercambio y la colaboración (Alonso, López, Manrique, & Viñes, 2005).
- Es el proceso educacional que utiliza a las TIC para mediar el aprendizaje a través de actividades síncronas y asíncronas (Jereb & Šmitek, 2006).
- Es definido como las TIC usadas para el soporte del estudiante, con el propósito de mejorar su aprendizaje (Ellis, Ginns & Piggot, 2009).
- De acuerdo con Sangra, Vlachopoulos y Cabrera (2012), la definición del aprendizaje en línea posee cuatro dimensiones para su interpretación, a saber: enfoque tecnológico, de sistemas, de comunicación y educativo.

En la actualidad existen diversos términos relacionados con el aprendizaje en línea, al mismo tiempo que diversos autores emplean de múltiples maneras dichos conceptos, lo que puede causar confusión respecto a la terminología; en consecuencia, a continuación se describen los principales términos relacionados con el concepto de aprendizaje en línea y se adopta una postura en cuanto al empleo del vocabulario para la presente investigación.

Naidu (2006), separa el concepto de aprendizaje electrónico de aprendizaje en línea, ya que asocia al primero con únicamente tecnologías informáticas, mientras que al segundo

con el medio para su distribución (Internet). Al respecto, para la presente investigación se considerarán sinónimos los conceptos aprendizaje electrónico y aprendizaje en línea, debido al uso de la Internet como medio indispensable para el desarrollo del aprendizaje digital.

Por otro lado el aprendizaje en línea se relaciona, de acuerdo con Dillenbourg, Schneider y Synteta (2000), al aprendizaje virtual (en inglés, *virtual learning*), el cual a su vez se relaciona con ambientes virtuales de aprendizaje (en inglés, *virtual learning environments*), y se refiere a los espacios diseñados para el consumo de información virtual, los cuales deben ser sociales y motivadores para el aprendizaje, cuyos estudiantes sean partícipes de su propio conocimiento. Cabe destacar que dicho concepto no está limitado a la educación en línea, ya que se asocia a las clases presenciales con apoyo de medios informáticos. Asimismo, en concordancia con (Lynch & Lynch, 2003), el aprendizaje en línea se relaciona con la educación basada de Internet (en inglés, *web based education*); misma que es totalmente en línea, ya que no existe interacción física entre el estudiante y profesor.

De igual manera, surgen otros conceptos relacionados en función a otras dimensiones, por ejemplo, de acuerdo al dispositivo empleado. En éste sentido, el aprendizaje en línea se relaciona con el aprendizaje móvil (en inglés, *mobile learning, m-learning*), el cual se refiere a la educación a través del uso de dispositivos móviles, tales como teléfonos inteligentes y tabletas.

De acuerdo a su impacto en la modalidad de la clase, el aprendizaje en línea se relaciona con los siguientes conceptos:

- Aprendizaje combinado (en inglés, *blended learning, b-learning*): Combina las interacciones cara a cara entre profesor y estudiante, y el uso de interacciones en línea para el mismo curso (Shibley, 2011).
- Clases invertidas (en inglés, *flipped learning environment*): Modelo educativo en donde el estudiante desarrolla las actividades de la casa en la escuela y viceversa (Honeycutt & Garrett, 2013); es decir, las tareas que comúnmente se desarrollarían en la aula, se efectúan en el hogar. Lo anterior se logra con el apoyo de las TIC.

Desde el punto de vista tecnológico, el aprendizaje en línea se relaciona con los siguientes conceptos:

- Sistemas de Gestión del Contenido (SGC, en inglés, *Content Management System, CMS*): Consiste en un sistema que colecta, almacena y publica contenido; asimismo ofrece diversas funcionalidades para las diferentes categorías de usuarios (Boiko, 2004).
- Repositorios de objetos de aprendizaje (ROA): son cualquier colección de recursos que son accesibles a través de una red sin el conocimiento previo de la estructura de la colección (IMS Global Learning Consortium, 2013).
- Sistemas de Gestión de Contenidos de Aprendizajes (SGCA, en inglés, *Learning Content Management System, LCMS*): "Son sistemas independientes o integrados con un entorno virtual de aprendizaje, que gestionan y administran

los contenidos de aprendizaje. Una vez que los contenidos están en este sistema, pueden ser combinados, asignados a distintos cursos o descargados, entre muchas otras opciones" (Zapata, 2013, p. 31).

- *Adaptive Intelligent Learning System (AILS)*: Ya que se basan del aprendizaje en línea, proveen una experiencia personal de aprendizaje para cada estudiante de acuerdo a sus hábitos, los AILS pueden comunicarse entre sí, están integrados a los entornos virtuales de aprendizaje y asociados con Objetos de Aprendizaje (Cemile, 2008).

Debido a que en la investigación se trabajó con un entorno educativo digital, basado en un Sistema de Gestión del Aprendizaje, en el siguiente punto se describirán sus características y algunos ejemplos estos.

2.2. Sistemas de Gestión del Aprendizaje

Actualmente, una de las tecnologías más empleadas por numerosas instituciones educativas de nivel superior para organizar y distribuir cursos en línea son los Sistemas de Gestión del Aprendizaje (en inglés, *Learning Management Systems, LMS*) (Kaap, 2003). Generalmente estos sistemas son fáciles de usar, se consideran flexibles en términos pedagógicos y eficientes con relación a los costos. De acuerdo con Abdullateef, Elías, Mohamed, Zaidan y Zaidan (2015), señalan que un LMS consiste en un sistema de comunicación que permite el desarrollo de un aprendizaje en línea. Su función principal es el almacenamiento y distribución de material educativo, que soporta la administración y comunicación de la enseñanza y el aprendizaje.

Existen varias maneras de alojar un LMS y que este funcione, siendo la principal el uso de un sistema complejo elaborado dentro de una institución, el cual estaría diseñado para llevar a cabo las funciones de gestionar cursos y el aprendizaje de los alumnos. De acuerdo con Sánchez (2009), los LMS contienen una gran variedad de herramientas, tales como:

- Distribución de contenido: es un espacio en donde el profesor comparte contenidos, por ejemplo: documentos, imágenes, videos, enlaces a sitios de interés, entre otros.
- Comunicación y colaboración síncrona y asíncrona: en las herramientas asíncronas el estudiante controla el momento para el consumo del contenido, mientras que en las síncronas se requiere conexión a la plataforma en un momento determinado, y la utilización de los contenidos es un por un espacio de tiempo preestablecido. El objetivo de las anteriores permite que los participantes puedan comunicarse y trabajar colaborativamente. Algunos ejemplos son: foros, salas de chat, mensajería interna, wikis, diarios, etc.
- Seguimiento y evaluación: en este rubro destacan las pruebas generadas por el profesor y autoevaluaciones del alumno.
- Administración y asignación de permisos: posibilitan la inscripción, y por consiguiente, el registro al curso.
- Complementarias: contiene diversas ayudas tales como el portafolio, el bloc de notas, los sistemas de búsqueda de contenidos, entre otras.

Adicionalmente a las características anteriores, Cabrero (2007) añade que son espacios de documentación compartida, debido a que posibilitan la administración de recursos por parte de profesores y estudiantes, por ejemplo, almacenamiento de documentos, y creación de carpetas para la gestión del contenido, entre otros. Para complementar la información anterior, las características claves de los SGA son (Itmazi, Gea, Paderewski y Gutiérrez, 2005):

- Administración de varios cursos o programas.
- Llevar registro de los participantes.
- Organizar y administrar el aprendizaje de los participantes.
- Capacidad de proporcionar informes.
- Programar y organizar los tiempos para el aprendizaje.

Los autores Menéndez-Domínguez y Castellanos-Bolaños (2014) describen que el uso de estos sistemas permite que:

- Los profesores coloquen a disposición de los alumnos los objetivos del curso, su contenido y su reglamentación.
- Los tutores y coordinadores supervisen el desarrollo del curso y el avance de cada alumno.
- Los alumnos accedan a los contenidos, realicen la ejercitación prevista, se comuniquen entre sí y con el tutor para resolver dudas y realizar trabajos en grupo.
- Los administradores obtengan información inmediata del progreso del curso y de las acciones administrativas relacionadas, tales como la inscripción de alumnos, el historial de cursos, etc.

Los LMS de acuerdo con su comercialización se clasifican en software propietario o de distribución libre. Los primeros se caracterizan por que han sido desarrollados por una empresa particular. Su modelo de negocio se basa a través de una licencia de uso que se paga bajo un esquema mensual o anual (el coste depende del número de usuarios). Algunos ejemplos son: Blackboard, Dokeos, E-ducativa, Sumtotal, eCollege, entre otros.

Algunas de las ventajas son:

- Suelen ser más estables y con funcionalidades que pueden adaptarse de acuerdo con las necesidades y el presupuesto.
- Incluyen en muchos casos el alojamiento (servidores y ancho de banda) desligando al cliente de lidiar con estos requerimientos.
- La capacitación de los profesores corre a cargo de la empresa contratada.

Algunas de sus desventajas son:

- Suelen ser muy costosas: hay empresas que no solo cobran un mantenimiento mensual por el alojamiento en sus servidores sino también un costo mínimo de implementación.

- No se tiene acceso al código fuente ya que las empresas no suelen vender el software, sino su implementación y posterior administración y mantenimiento, con el fin de asegurarse un abono mensual y/o anual.

Los LMS de distribución libre pueden instalarse tanto en ordenadores personales como en servidores web. El soporte técnico suele estar soportado por una comunidad mundial de desarrolladores. Algunos ejemplos son: ILIAS, ATutor, Sakai, MOODLE, entre otros.

Algunas de sus ventajas son:

- Permite realizar modificaciones sobre el código del sistema, ya sea para personalizarlo o agregarle nuevos componentes.
- Son compatibles con la mayoría de los formatos estándares.
- Su actualización es constante.
- Ausencia de malware al momento de instalarlo.

En contraste, algunas de sus desventajas son:

- Carece de soporte técnico, en caso de que se presente algún problema, se puede consultar en foros especializados en la web.
- Requiere implementar una infraestructura por parte de la institución que lo adopta.
- La capacitación de los profesores corre a cargo de la institución que adopta el sistema.

La fase experimental de este trabajo de investigación se desarrolló en el entorno de un LMS basado en el software de distribución libre MOODLE (*Modular object oriented distance learning environment*) (Dougiamas & Taylor, 2003), cuyo diseño está basado en un enfoque pedagógico constructivista social, que promueve la vinculación a una comunidad de aprendizaje como un potente estimulante para aprender, promoviendo una reflexión profunda y un replanteamiento de las propias opiniones y puntos de vista. Adicionalmente, se destaca que posee una arquitectura modular, lo que permite incorporar una gran diversidad de componentes y funcionalidades. Implementa una interfaz de navegador de tecnología sencilla, ligera, eficiente y compatible.

2.3. Extracción de conocimiento de bases de datos

El descubrimiento de conocimiento implica diversos conceptos con los que está relacionado, tales como: los datos, la información, la gestión de la información y la gestión del conocimiento. Los datos consisten en hechos, imágenes, o sonidos. Cuando se combinan con la interpretación y el significado, se convierte en información. Por lo tanto, la información consiste en datos que han sido formateados, filtrados, y resumidos (Chen, 2001).

Una vez establecidos los conceptos datos e información, se puede definir el término de gestión de la información como “el proceso que utiliza herramientas y técnicas para capturar, recoger, analizar, organizar, manipular y recuperar la información. De la

misma manera, controla y explota los recursos de la información de una organización” (Itmazi, 2005, p. 38). Por otro lado, la gestión del conocimiento, se define como una disciplina emergente cuyo objetivo es generar, compartir y utilizar el conocimiento tácito y explícito existente en un determinado espacio para dar respuesta a las necesidades de los individuos y de las comunidades en su desarrollo (Alavi y Leidne, 2001).

Finalmente, el Descubrimiento del Conocimiento en Bases de Datos (en inglés, *Knowledge Discovery in Databases*, KDD), se define como “el proceso no trivial de identificar patrones válidos, novedosos y potencialmente útiles y en última instancia, comprensible a partir de los datos” (Fayyad., Piatetsky-shapiro & Smyth, 1996, p. 40).

El KDD es un proceso iterativo e interactivo y de acuerdo con (Hernández, Ramírez & Ferri, 2004) su taxonomía se puede organizar en cinco fases:

- a) Integración y recopilación. En esta actividad se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas. Posteriormente, se transforman los datos a un formato común. Dado que los datos provienen de diferentes fuentes, pueden contener valores erróneos o faltantes.
- b) Selección, limpieza y transformación. Se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Además, se proyectan los datos para conseguir únicamente aquellas variables o atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de la misma sean más útiles.
- c) Minería de datos. Se decide cuál es la tarea a realizar (clasificar, agrupar, entre otras) y se elige el método a utilizar.
- d) Evaluación e interpretación. Los expertos evalúan y se analizan los patrones, y si es necesario se vuelve a las fases anteriores para una nueva iteración. Esto incluye resolver posibles conflictos con el conocimiento que se disponía anteriormente.
- e) Difusión y uso. Se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios. Este conocimiento se suele utilizar en procesos de toma de decisiones.

Cabe destacar, que la experimentación con las técnicas de minería de datos es una de las partes fundamentales para la extracción de conocimiento en esta investigación.

2.4. Minería de datos

La minería de datos trabaja con grandes volúmenes de información, procedentes en su mayoría de sistemas de información, con los problemas que ello conlleva (ruido, datos ausentes, intratabilidad, volatilidad de los datos, entre otros) y aplica técnicas adecuadas para analizar estos datos y extraer conocimiento novedoso y útil (Hernández et al., 2004). Otra definición importante en la literatura es la de Chen (2006), el cual señala que se asocia a bases de datos, en donde a través de diversas técnicas, se extrae información para generar conocimiento, el cual puede ser expresado a través de conceptos, reglas, leyes, patrones, entre otros.

La minería de datos se relaciona con diversas disciplinas emergentes, tales como: aprendizaje automático, aprendizaje profundo, reconocimiento de patrones, recuperación de información, aprendizaje estadístico, inteligencia artificial, entre otras. De acuerdo con (Hernández et al., 2004), los modelos de aprendizaje empleados en la minería de datos pueden dividirse en dos categorías principales: predictivos y descriptivos. A continuación se describen:

- Modelos predictivos o aprendizaje supervisado. Predicen el valor del atributo de un conjunto de datos a partir de información previamente conocida. El objetivo de este modelo es proporcionar características que describan las relaciones entre los datos. Los algoritmos aplicados en los modelos predictivos se catalogan en:
 - Clasificación. Son utilizados para clasificar datos, con la finalidad de predecir clases de objetos cuyas categorizaciones no se han definido.
 - Regresión. Predicen una o más variables continuas, como por ejemplo las pérdidas o los beneficios, basándose en otros atributos del conjunto de datos.
 - Predicción. Son aquellos que pueden ser usados para predecir tendencias de los valores en los datos o de las clases.
- Modelos descriptivos o aprendizaje no supervisado. Descubren patrones y tendencias en los datos. Sirven para llevar a cabo acciones y obtener un beneficio o conocimiento de ellas. Estos algoritmos se clasifican en:
 - Agrupamiento. Son aquellos que no requieren una clasificación predefinida para particionar los datos obteniendo el conocimiento de acuerdo a las características de los mismos.
 - Asociación. Se basan en el descubrimiento de reglas de asociación que muestran condiciones en los valores de los atributos que ocurren simultáneamente de forma frecuente en un determinado conjunto de datos.
 - Correlación y dependencias. Establecen un patrón en el que uno o más atributos determinan el valor de otro. Su función principal es predecir diferentes valores de los datos, como ejemplo: ganancias, ventas, tasas, entre otros.

En la minería de datos existen multitud de técnicas que pueden resolver una gran diversidad de problemas. A continuación se presenta un resumen de las técnicas más populares:

- Algebraicas y estadísticas. Se basan, generalmente, en expresar modelos y patrones mediante fórmulas algebraicas, funciones lineales, funciones no lineales, distribuciones o valores agregados estadísticos tales como medias, varianzas, correlaciones, entre otras. Algunos de los algoritmos más conocidos dentro de este grupo de técnicas son la regresión lineal (Edwards, 1976) y la regresión logística (Hosmer & Stanley, 2000).
- Redes bayesianas. Se basan en estimar la probabilidad de pertenencia (a una clase o grupo), mediante la estimación de las probabilidades condicionales inversas o a priori, utilizando el teorema de Bayes. Poseen varias ventajas, una de ellas es que pueden tratar muchos atributos y son muy robustas al ruido que pueden provocar los datos. Por el contrario, la expresividad es limitada y depende de la discretización. Algunos algoritmos populares son EM

(*Expectation-Maximization*) (Dempster, Laird & Rubin, 1977) y Naive Bayes (Zhang, 2004).

- Algoritmos evolutivos. Son métodos sistemáticos para la resolución de problemas de búsqueda y optimización, a los cuales se aplican los mismos métodos de la evolución biológica, tales como: selección basada en la población, reproducción sexual y mutación (Eiben & Smith, 2003).
- Conteos de frecuencias y tablas de contingencia. Se basan en contar la frecuencia en la que dos o más sucesos se presenten conjuntamente. Ejemplos de estos algoritmos son Apriori (Agrawal, Imieliński & Swami, 1993) y Predictive A priori (Scheffer, 2004).
- Árboles de decisión. Está técnica se encuentra dentro de una metodología de aprendizaje supervisado. Su representación es en forma de árbol, en donde cada nodo es una decisión, los cuales a su vez generan reglas para la clasificación de un conjunto de datos. Entre sus ventajas se encuentran: que admiten atributos discretos y continuos, también se desempeñan de forma eficiente con los atributos no significativos. Entre sus desventajas, se encuentran la sobreadaptación y el ruido generado por los datos. Algunos ejemplos de estos algoritmos son ID3 (Quinlan, 2007), J48 (Ye, & Baldwin, 2006) y C4.5 (Quinlan, 1993).
- Relacionales, declarativas y estructurales. La característica principal de este conjunto de técnicas es que representan los modelos mediante lenguajes declarativos, como los lenguajes lógicos, funcionales o lógico-funcionales. Las técnicas de ILP (en español, Programación Lógica Inductiva) son las más representativas y las que han dado nombre a un conjunto de técnicas denominadas minería de datos relacional.
- Redes neuronales artificiales. Se utilizan de manera frecuente para detectar categorías comunes en los datos, debido a que detectan y aprenden complejos patrones de comportamiento. Sus principales ventajas son: cuando están bien ajustadas, obtienen precisiones muy altas y se puede aplicar tanto en modelos predictivos como descriptivos. Su principal desventaja son los datos incompletos. Algunos ejemplos son las redes de kohonen (Kohonen, 1988) y el modelo de Retro-Propagación (Rumelhart, Hinton & Williams, 1986).
- Basadas en núcleo y máquinas de soporte vectorial. Se trata de técnicas que intentan maximizar el margen entre los grupos o las clases formadas. Existen muchas variantes, dependiendo del núcleo utilizado y la manera de trabajar con el margen.
- Estocásticas y borrosas. Este conjunto de técnicas contienen características que se describen por separado. Por una parte, las técnicas estocásticas se basan en las teorías del análisis de probabilidades. Contienen datos con alguna función de densidad de probabilidad asociada. Por otra parte, las técnicas borrosas, aunque también se basan en la asunción de un conocimiento impreciso, son distintas de las estocásticas. En vez de tener una función de densidad de probabilidad asociada, se caracterizan por una función de pertenencia que expresa el grado en que un elemento pertenece a un conjunto. Estas técnicas junto con las redes neuronales y los algoritmos evolutivos, forman lo que se denomina computación flexible.
- Basadas en casos, en densidad o distancia. Son métodos que se basan en distancias al resto de elementos, ya sea directamente, como los vecinos más

próximos o mediante la estimación de funciones de densidad. Algunos algoritmos conocidos son el K-medias (Macqueen, 1967) y COBWEB (Fisher, 1987).

Cabe destacar, que además de todo lo anterior existen diversas combinaciones que dificultan aún más realizar una clasificación razonable que contemple a todos los grupos de técnicas.

2.5. Minería de datos educativos

La Minería de Datos Educativos (en inglés, *Educational Data Mining*, EDM), es un área de conocimiento emergente y multidisciplinaria. Tiene como objetivo el desarrollo de métodos para la exploración de datos en un contexto educativo de manera que orienta y mejora el proceso de aprendizaje de los estudiantes (Suhirman, Tutut, Haruna & Jasni, 2014).

Existen diferencias significativas entre la minería de datos y la minería de datos educativos, en primer lugar, el primero es general, mientras que el segundo se especializa en datos provenientes de entornos presenciales y virtuales de aprendizaje, por ejemplo los LMS.

La aplicación de técnicas de EDM se puede ver desde distintos puntos de vista dependiendo del tipo de usuario (Romero & Ventura, 2010) (Romero & Ventura, 2013):

- Del estudiante. El objetivo es ayudar o realizar recomendaciones a los alumnos durante su interacción con el sistema de e-learning para mejorar su aprendizaje. Sus principales aplicaciones son: sugerir buenas experiencias de aprendizaje a los estudiantes, adaptar el curso según el progreso del aprendiz, ayudar a los estudiantes dando sugerencias y atajos, recomendar caminos más cortos y personalizados, entre otros.
- Del profesor. El objetivo es apoyar a los profesores de los sistemas de e-learning para que puedan mejorar el funcionamiento o rendimiento de estos sistemas a partir de la información generada por los alumnos al utilizar los cursos. Sus principales aplicaciones son: obtener una mayor realimentación de la enseñanza, conocer más sobre cómo los estudiantes aprenden en la Web, evaluar a los estudiantes por sus patrones de navegación, reestructurar los contenidos en el sitio Web para personalizar los cursos, clasificar a los estudiantes en grupos, entre otros.
- De los investigadores. El objetivo es comparar técnicas de minería de datos que permitan recomendar la más útil para una tarea específica educativa o problema, para evaluar la efectividad del aprendizaje utilizando diferentes configuraciones y métodos, y así sucesivamente.
- De las instituciones educativas. El objetivo es obtener información útil que permita mejorar la eficiencia del sitio y adaptarlo a la conducta de los usuarios (tamaño óptimo del servidor, distribución del tráfico de red, entre otros). Otros objetivos son adquirir medidas que permitan organizar mejor los recursos institucionales (humanos y materiales) y determinar la eficiencia de los programas educativos del aprendizaje a distancia soportado por ordenador.

Como se ha señalado, la minería de datos provee beneficios a la educación, Wang (2014), manifiesta que es posible analizar la administración de la clase, a través de la información de los estudiantes, como por ejemplo, situación familiar, desempeño académico, número de ausencia del alumno en un mes, entre otros; asimismo se puede entrenar los talentos personales, analizar el impacto académico y la enseñanza.

Por otro lado Charlton, Mavrikis y Katsifli (2013), señalan que es posible predecir el desempeño académico del estudiante al estimar un valor desconocido de la variable que describe al estudiante, el cual puede ser numérico (regresión) o categórico (clasificación). El análisis de regresión permite relacionar las variables dependientes e independientes. La predicción es una de las aplicaciones más antiguas y populares en minería de datos aplicados a la educación.

De igual forma asegura que es posible predecir el riesgo de deserción al analizar el comportamiento del estudiante, además, que es posible una retroalimentación inteligente por parte de los LMS con el propósito de mejorar el desempeño académico, es posible recomendar cursos, estimar las destrezas del estudiante y detección del comportamiento a través de actividades basadas en comunidad y juegos.

Por otro lado, existe una variedad de paquetes software para implementar técnicas de minería de datos en los entornos educativos. Cada uno de ellos posee características apropiadas para realizar determinadas tareas o para analizar cierto tipo de datos. A continuación se mencionan algunos ejemplos:

SPSS (Statistical Product and Service Solutions), este software propietario permite realizar análisis estadístico y gestión de información, es capaz de trabajar con datos procedentes de diversos formatos, como *.XLSX (Excel Microsoft)*, genera desde gráficos y estadística descriptiva hasta estadísticos complejos. Es uno de los más utilizados actualmente en el área de educación. El software es compatible tanto con el sistema operativo *Windows* de *Microsoft* como *Macintosh* de la compañía *Apple* (IBM Corp., 2017).

KEEL (Knowledge Extraction based on Evolutionary Learning), fue desarrollado por 5 grupos de Universidades españolas, permite utilizar y construir diferentes modelos para minería de datos y tiene como característica importante la inclusión de una librería de algoritmos de aprendizaje evolutivo con código abierto en Java. Las características principales son: algoritmos de preprocesamiento, librería de algoritmos de extracción de conocimiento, librería de herramientas estadísticas para el análisis de los algoritmos y aplicación para su uso vía Web; además contiene cuatro entornos (Alcalá-Fernández, et al., 2009).

DBMiner: A System for Mining Knowledge in Large Relational Databases, es un software de distribución libre desarrollado por la Universidad de Simon Fraser, Canadá. Está concebido para la extracción del conocimiento en bases de datos relacionales, almacenes de datos y Web, su arquitectura de diseño incorpora OLAP (*online analytic processing*) y OLAM (*online analytic mining*) (Han et al., 1997).

RapidMiner: Data mining use cases and business analytics applications, este sistema antes era conocido como *YALE (Yet Another Learning Environment)*, está desarrollado en Java y es de código abierto. Cuenta con una interfaz gráfica fácil de utilizar y además

integra un mecanismo sencillo para desarrollar extensiones que hacen posible integrar nuevos operadores y con ello adaptar el software para los requerimientos personales (Arimond, Kofler & Shafait, 2010).

GISMO (*Graphical Interactive Student Monitoring*), es una herramienta de monitorización gráfica interactiva que proporciona una visualización útil de las actividades de los estudiantes en los cursos en línea a los instructores. Con esta herramienta los instructores pueden examinar diversos aspectos de los estudiantes a distancia, tales como la asistencia a los cursos, la lectura de los materiales, la presentación de las asignaciones, entre otras (Mazza & Milani, 2004).

En esta investigación se utilizó el software de distribución libre WEKA (*Waikato Enviroment for Knowledge Analysis*) (Witten, Frank & Hall, 2011), el cual es una herramienta visual de distribución libre desarrollada por los investigadores de la Universidad de Waikato en Nueva Zelanda.

WEKA consiste en un conjunto de algoritmos de aprendizaje automático, los cuales están basados en el lenguaje de programación JAVA; su propósito es aplicar técnicas de minería de datos y extraer conocimiento de las bases de datos. El sistema posee una licencia de uso GNU (*General Public License*), lo cual lo convierte en un sistema de código abierto, lo que significa que los usuarios poseen la libertad de modificar su código fuente.

Una de las principales características del software es su interfaz visual, lo que facilita el uso del sistema. Es posible adjuntar bases de datos en diversas extensiones, de las cuales destaca la extensión .ARFF. El sistema se compone de cuatro entornos de trabajo, de acuerdo con Hernández, Fernández y Baptista (2006) se definen como:

- a. *Explorer*: entorno visual para emplear los paquetes (algoritmos) del sistema.
- b. *Experimenter*: diseñado para experimentos de gran escala, permite automatizar diversas tareas.
- c. *Knowledge Flow*: basado en flujo de información, permite análisis de minería de datos.
- d. *Simple CLI*: entorno consola que permite emplear los paquetes del sistema directamente con código.

De los cuatro entornos de trabajo mencionados anteriormente, para el desarrollo de esta propuesta se utilizó la versión 3.8 del WEKA, en particular se utilizó el entorno de trabajo visual denominado *explorer*. A continuación, se describen sus principales componentes:

- *Preprocess*: se aplican diversos filtros tanto de algoritmos supervisados como no supervisados.
- *Classify*: se aplican algoritmos supervisados (empleado para la creación de modelos predictivos)
- *Cluster*: se aplican a los algoritmos no supervisados.
- *Associate*: se aplican algoritmos de asociación.

2.6. Investigaciones relacionadas

En la revisión de la literatura se han encontrado diversos trabajos académicos que se enfocan en el análisis del desempeño de los profesores en el uso de un LMS. En primera instancia, se describen algunas investigaciones donde se utilizó software de análisis de datos y métodos estadísticos para la extracción del conocimiento:

La propuesta de Andone, Ternauciuc y Vasiu (2017), se realizó en el entorno del campus virtual de la Universidad Politécnica de Timisoara (Rumania) en un período de seis años. Cabe destacar, que en este país el uso del sistema MOODLE es amplio y que la formación de los profesores para el uso de estas herramientas web 2.0 está a cargo del departamento correspondiente en cada universidad. El objetivo principal se enfocó en analizar las interacciones entre alumnos y profesores en el entorno de este sistema. Para ello, tomaron en cuenta el número de sesiones de chat y mensajes intercambiados; así como también el historial académico y las estructuras de los cursos. La consulta en las bases de datos de los archivos almacenados se realizó a través de un complemento del sistema MOODLE denominado Ad-hoc database queries. En el análisis de los resultados se detectaron que las actividades más populares utilizadas por los profesores fueron el blog, los foros y el cuestionario. En las conclusiones se observó que la formación impartida a los profesores influyó en el incremento exponencial del uso de las herramientas que promueven la interacción con los estudiantes, lo que puede influir de forma positiva en que el alumnado alcance el desempeño deseado. Como área de oportunidad se detectó que no todas las herramientas disponibles se han utilizado al máximo.

Los autores Cantabella, López-Ayuso, Muñoz y Caballero (2016), desarrollaron un trabajo que se centró en la interacción que tiene lugar entre los coordinadores de titulación y un Entorno Virtual de Aprendizaje (EVA) basado en el software Sakai en la Universidad Católica San Antonio de Murcia (España) con el objetivo principal de realizar labores de seguimiento del desempeño del profesorado, especialmente en la modalidad en línea. Para ello, se establecieron tres objetivos específicos que guiaron la metodología seguida: (1) analizar las limitaciones de monitorización del profesorado de los EVA actuales; (2) desarrollar una herramienta de monitorización para superar las limitaciones encontradas; y (3) evaluar empíricamente la herramienta propuesta. El análisis de los datos se realizó a través de la herramienta OnlineData. Se destaca el uso de los colores para la visualización de las alertas sobre el uso de las diversas actividades y recursos que contiene el sistema. Adicionalmente, para el análisis del uso de las herramientas se implementaron dos métodos empíricos: grupos de discusión y encuestas de valoración. Los resultados obtenidos mostraron que esta herramienta ayuda a los coordinadores a analizar de forma intuitiva y eficiente el estado de las asignaturas de su titulación.

Otra iniciativa es la de Hernández (2015), quién realizó un estudio en la institución de educación básica San Jorge de Inglaterra (Colombia) donde analizó la actividad realizada por los profesores del área de matemáticas a través del sistema MOODLE. La recolección de los datos se realizó a través de la aplicación de una encuesta virtual sobre el uso de las actividades y recursos que contiene este sistema. Adicionalmente, realizó reuniones con el personal docente. Con respecto a los resultados cuantitativos, se seleccionó la escala de medición aditiva tipo Likert para la formulación de las preguntas cerradas. Luego, se codificaron los resultados según las preguntas de investigación y se

transfirieron a una matriz de datos para analizarlos. En cuanto a los resultados cualitativos, se crearon categorías para agrupar los datos y utilizarlos posteriormente para la etapa del análisis. Entre los resultados se obtuvieron que los profesores desconocen varias de las actividades y recursos del entorno del sistema, como consecuencia utilizan con poca frecuencia las actividades interactivas disponibles en el aula virtual (chat, wiki, la lección, el taller, el cuestionario, entre otros.) por falta de capacitación tecnológica. De esta forma el aula virtual se convierte en un diario de clase en donde se consignan fechas y se suben contenidos, desaprovechando su carácter constructivista.

En segunda instancia, se describen algunas investigaciones donde se implementan técnicas de minería de datos para la extracción del conocimiento, las cuales han tenido un auge muy fuerte desde hace algunos años, tal como se ha documentado en las revisiones sistemáticas realizadas por (Romero & Ventura, 2007) y (Romero & Ventura, 2010). En ambos trabajos se observa que la gran mayoría de las publicaciones documentadas se han enfocado en el comportamiento de los estudiantes. Algunos ejemplos de estas investigaciones son:

La investigación de Romero, Ventura y García (2008), tuvo como propósito aplicar diversas técnicas de minería de datos en las actividades que realizaron alumnos de nivel superior dentro del entorno de un LMS basado en el sistema MOODLE. En cuanto a los métodos o técnicas de análisis de datos empleados, los autores recurrieron a la experimentación de técnicas de agrupamiento, reglas de asociación a través del software WEKA. También, experimentaron con los programas de distribución libre GISMO y KEEL. La aportación de esta propuesta fue el desarrollo de un sistema que permitía a los profesores modificar las actividades del curso, y detección a tiempo si un estudiante está en posibilidad de fracaso.

Otra investigación conducida de igual manera por Romero, Espejo, Zafra, Romero y Ventura (2010), tuvo como propósito aplicar la minería de uso de la web en sistemas de e-learning para predecir las calificaciones que los estudiantes universitarios obtendrán en el examen final de un curso. Para ello, desarrollaron una herramienta que funciona dentro del entorno del sistema MOODLE tanto para profesores experimentados como recién iniciados. Se comparó el rendimiento a través de diferentes técnicas de minería de datos para clasificar a los estudiantes, comenzando con los datos de uso en varios cursos en línea de ingeniería de la Universidad de Córdoba (España). Se implementaron varios métodos estadísticos, árboles de decisión, métodos de inducción de reglas difusas y redes neuronales. También se han utilizado técnicas de preprocesamiento de discretización y reequilibrio en los datos numéricos originales para probar de nuevo si se pueden obtener mejores modelos de clasificadores. Por último, se mostraron ejemplos de algunos de los modelos descubiertos y explican que un modelo clasificador apropiado para un entorno educativo debe ser preciso y comprensible para que los instructores y administradores del curso puedan utilizarlo en la toma de decisiones.

Algunas investigaciones más recientes son la de (Valsamidis, Kontogiannis, Kazanidis, Theodosiou y Karakos, 2012), quienes tuvieron como objetivo experimentar diversas técnicas de agrupamiento para analizar las actividades de los estudiantes de pregrado dentro del entorno del LMS MOODLE. Para ello, implementaron los algoritmos de cadenas de Markov y SimpleKMeans a través del software WEKA. Adicionalmente se

incluyó una visualización de los resultados a través del uso del software *3D graphs*. Como resultado se demostró que las métricas propuestas pueden ofrecer una clasificación preliminar del curso, que a su vez puede ser utilizada como entrada para los algoritmos de agrupamiento. Éstas sugieren acciones específicas a los instructores para que puedan mejorar el contenido de sus cursos, la usabilidad de los mismos y ayudarles para adaptarlos a las capacidades de los estudiantes. Adicionalmente, se mejoró la calidad de los cursos en línea mediante la reorganización del material educativo de manera uniforme, jerárquica y estructurada. También aumentaron la cantidad recursos mediante la incorporación de material educativo adicional. Sin embargo, algunos profesores se quejaron de que la organización de sus cursos no les ayudo a obtener las mejores puntuaciones finales de sus alumnos.

Una investigación de importancia, fue la realizada por Gobert, Sao Pedro, Raziuddin y Baker (2013), el objetivo del estudio fue el desarrollo de un método para evaluar las habilidades científicas de estudiantes de nivel superior. Para ello, implementaron técnicas de clasificación de minería de datos y utilizaron el software RapidMiner versión 4.6.

Magdin y Turcáni (2015), desarrollaron una investigación cuyo objetivo fue personalizar recomendaciones de cursos en línea a estudiantes. Para el estudio se aplicaron reglas de asociación. Los softwares empleados fueron *Module Interactive Element Stat* y *Excel de Microsoft*. Los principales resultados de esta investigación fueron la implementó un cuestionario que, de acuerdo a reglas de asociación, permitió la recomendación de materiales específicos para las preferencias de cada estudiante y a pesar de la recomendación de material, los estudiantes continuaron mostrando deficiencias en el análisis de las lecturas. El estudio concluyó que los estudiantes activos están más cómodos con actividades basadas en problemas y discusión en grupo, mientras que los estudiantes reflexivos prefieren más las lecturas, por consiguiente no es fácil determinar una estrategia de enseñanza estándar.

En contraste, las investigaciones orientadas a analizar el desempeño de los profesores son escasas. Es por ello, que se considera que es un área poco explorada y con un gran potencial de crecimiento. A continuación se detallan:

La propuesta de los autores Samaniego, Marqués y Gisbert (2015), se basa en el análisis de la actividad realizada en el sistema MOODLE de los profesores de la Facultad de Informática y Electrónica de la Escuela Superior Politécnica de Chimborazo (Ecuador). Para ello, se procesaron los accesos a los cursos en línea por un período de 8 semestres. Adicionalmente, se aplicaron 60 encuestas en línea y se realizaron 21 entrevistas semiestructuradas al profesorado. La población estuvo compuesta por un muestreo estratificado de 100 entornos virtuales de aprendizaje activos de esta Facultad y por los profesores que los crearon. Para el tratamiento y análisis *in situ* de los entornos virtuales de aprendizaje se diseñó un procedimiento basado en conceptos de las técnicas estadísticas de minería de datos. Los resultados evidenciaron que el crecimiento gradual de los niveles de interacción y uso de los entornos virtuales por parte de los profesores de esta Facultad, desde el semestre académico marzo 2009 / agosto 2009 hasta el semestre académico septiembre 2012 / febrero 2013. Los entornos virtuales de aprendizaje generalmente se emplean para distribuir materiales y realizar actividades no obligatorias, individuales y/o en pequeños grupos. De acuerdo a la tipología de recursos

utilizados los profesores priorizan la interacción comunicacional a la informacional con sus alumnos.

Otra investigación relacionada con este tema es la de Salcines, Romero, Ventura y de Castro (2008), quienes proponen un sistema de recomendación colaborativo denominado CIECoF (*Continuous improvement of e-learning courses framework*), el cual permite a profesores y expertos en educación intercambiar experiencias entre sí sobre cómo aprenden sus alumnos, de forma que este conocimiento les permita mejorar sus propios cursos en línea. Para la extracción del conocimiento se aplicaron diversas técnicas de minería de datos (asociación y clasificación de reglas). Las pruebas experimentales se realizaron teniendo en cuenta dos puntos de vista, el del profesor que realiza los cambios basándose en las recomendaciones que brinda el sistema y el del alumno que realiza el curso una vez modificado por el profesor. Entre los resultados se demostró que los problemas detectados se reducirían en consecutivas ejecuciones del curso y por otra, que las notas finales de los alumnos mejorarían en la medida que el profesor iba corrigiendo los problemas. Aunque los resultados de las notas demostraron una sensible mejoría, se observó que se requiere realizar un estudio más detallado con mayor cantidad de grupos de profesores para buscar diferencias más significativas.

De acuerdo con la revisión de los trabajos relacionados con el tema de investigación que se está planteando, se detectó que la implementación de técnicas de agrupamiento son las más recurrentes. Cabe señalar que algunas investigaciones utilizaron alguna técnica de minería de datos para realizar predicciones del desempeño académico de los estudiantes.

3. Entorno de la investigación

El entorno de esta investigación es la Universidad Autónoma de Yucatán (UADY), la cual es una importante institución de educación superior pública del sureste de México. Hoy en día, tiene una matrícula de alrededor de 26,000 alumnos y 1,800 profesores que imparten clases en los niveles de bachillerato, pregrado y posgrado.

La UADY ha implementado políticas para el uso de las TIC en sus procesos académicos y administrativos, las cuales se contemplan tanto en su Plan de Desarrollo Institucional (PDI) como en su Modelo Educativo para la Formación Integral (MEFI). En el PDI del 2010-2020 (UADY, 2010), la universidad plantea lo siguiente: "... la formulación de nuevos programas de licenciatura y posgrado utilizando modalidades no presenciales y semipresenciales, privilegiando la equidad y el uso de tecnologías de información y comunicación. La nueva oferta debe considerar su contribución a la operación de los programas presenciales, así como a la formación de los estudiantes inscritos en ellos" (p.76).

Por su parte, el MEFI (UADY, 2012) señala: "... la planeación deliberada y sistemática de nuevas propuestas para la solución de situaciones problemáticas y para la mejora continua de la práctica educativa que implica un cambio en el contexto y la práctica educativa misma, mediante la incorporación de recursos y medios educativos vanguardistas..." (p.39); respecto al profesor universitario: "14. Utilizar las tecnologías de información y comunicación como recurso didáctico en su ejercicio docente de manera pertinente." (p.49).

En respuesta a lo anterior, la UADY implementó una nueva modalidad para cursar los estudios de Educación Media Superior denominada Bachillerato en Línea (BeL), la cual se encuentra disponible en la siguiente dirección web: <http://www.bel.uady.mx/>.

En cuanto a los niveles de pregrado y posgrado, se ha implementado el LMS denominado UADY Virtual, el cual está basado en el software de distribución libre MOODLE. Este sistema se utiliza principalmente como una herramienta de apoyo tecnológico a las clases presenciales y sirve de escenario para algunas asignaturas que se imparten totalmente en línea.

El objetivo de UADY Virtual en conjunto con el MEFI es: “Impulsar la innovación académica de la institución en todos los niveles, a través de las diferentes modalidades educativas contempladas en el MEFI, especialmente con el apoyo de entornos virtuales de aprendizaje y la puesta en marcha de las modalidades mixta y no presencial (modalidades no convencionales).” (UADY, 2013, p.31).

El sistema UADY Virtual se encuentra en la siguiente dirección web: <http://es.uadyvirtual.uady.mx/>. Actualmente, tiene registradas aproximadamente 19,000 cuentas entre profesores y estudiantes. El entorno de este sistema cuenta con 21 herramientas, las cuales se agrupan en 15 actividades y 6 recursos que pueden ser usados por los profesores para el desarrollo de sus cursos (Tabla 1).

Tabla 1
Actividades y recursos de UADY Virtual

Actividades	
Asistencia	Permite al profesor pasar lista durante la clase y a los estudiantes les permite ver su propio registro de asistencias.
Base de datos	Permite a los participantes crear, mantener y buscar información en un banco de registros.
Chat	Permite que los participantes tengan una discusión sincrónica, basada en texto, en tiempo real en la web.
Consulta	Permite al profesor realizar una pregunta y especifica una serie de respuestas entre las cuales deben elegir los alumnos.
Cuestionario	Permite al profesor construir encuestas empleando una variedad de tipos de preguntas, para recolectar datos de sus usuarios.
Elección	Permite al profesor realizar una sola pregunta y ofrecer una selección de respuestas posibles. Los resultados obtenidos pueden publicarse después de que los estudiantes hayan contestado, o después de cierta fecha, o nunca. También pueden publicarse junto con los nombres de los estudiantes o de forma anónima.
Encuesta predefinida	Proporciona un número de instrumentos de encuesta verificada que han demostrado utilidad para evaluar y estimular el aprendizaje en el entorno en línea. El profesor puede utilizarlos para recopilar información que le ayude a conocer mejor su clase y reflexionar sobre su propia enseñanza.
Foro	Permite a los participantes tener discusiones asincrónicas (en diferentes tiempos), que suceden a lo largo de un largo período de tiempo.
Glosario	permite a los participantes crear y mantener una lista de definiciones, igual que un diccionario, o coleccionar y organizar recursos o información.
HotPot	permite a los profesores distribuir materiales de aprendizaje interactivos a sus

	estudiantes vía Moodle (crucigramas, rompecabezas, entre otros) y ver reportes sobre las respuestas y resultados de sus estudiantes.
Lección	Permite al profesor proporcionar contenidos y/o actividades en una forma interesante y flexible. Un profesor puede usar una lección para crear un conjunto lineal de páginas de contenidos o actividades instruccionales que ofrecen una variedad de caminos o de opciones para el estudiante
Paquete SCORM	Permite que los paquetes SCORM (es una colección de archivos que están empaquetados de acuerdo a el estándar acordado para objetos de aprendizaje) se suban como un archivo ZIP y se añadan al curso. El contenido generalmente se muestra sobre varias páginas, con navegación entre ellas.
Taller	Permite la colecta, revisión y evaluación por pares del trabajo de los estudiantes, quienes pueden enviar cualquier contenido digital (archivos), tales como documentos de procesador de texto o de hojas de cálculo y también pueden escribir el texto directamente en un campo empleando un editor de texto (dentro de Moodle).
Tareas	Permite al profesor comunicar tareas, recolectar los trabajos, proporcionar calificaciones y retroalimentación. Los estudiantes pueden enviar cualquier contenido digital (archivos), tal como documentos de procesador de textos, hojas de cálculo, imágenes y clips de audio o video. De forma alterna, o adicional, la tarea puede requerir que el estudiante escriba texto directamente en un campo empleando un editor de texto.
Wiki	Permite a los participantes añadir y editar una colección de páginas web. Un wiki puede ser colaborativo, donde todos pueden editarlo, o puede ser individual, donde cada quien tiene su propio wiki que solamente esa persona puede editar.
Recursos	
Archivo	Permite al profesor proporcionar un archivo como recurso para un curso. Donde sea posible, el archivo se mostrará dentro de la interfaz del curso; de otra forma, se les pedirá a los estudiantes que descarguen el archivo.
Carpeta	Permite al profesor mostrar un número de archivos relacionados que están dentro de una sola carpeta, reduciendo la navegación de la página del curso.
Etiqueta	Permite insertar texto y multimedia en una página de curso junto con enlaces a otros recursos y actividades. Las etiquetas son muy versátiles y pueden ayudar a mejorar la apariencia de un curso si se emplean inteligentemente.
Libro	Permite a un profesor crear un recurso multi-página en un formato similar a un libro con capítulos y subcapítulos.
Página	Permite al profesor crear una página web empleando el editor de texto. En una página se pueden mostrar texto, imágenes, sonido, video, enlaces de internet y código incrustado (como los mapas de Google).
URL	Permite al profesor proporcionar un enlace de Internet para un recurso del curso.

Esta propuesta se centra en el análisis del uso de las actividades y recursos por parte de los profesores en el entorno de UADY Virtual; de forma particular, intenta averiguar si existen diferencias de acuerdo al área de conocimiento, en el caso de esta institución se dividen en cinco campus:

- Arquitectura, Hábitat, Arte y Diseño (AHAD)
- Ciencias Biológicas y Agropecuarias (CBA)
- Ciencias Exactas e Ingenierías (CEI)

- Ciencias de la Salud (CS)
- Ciencias Sociales Económico-Administrativas y Humanidades (CSEAH)

Como se ha señalado, la UADY contempla la implementación de las TIC para sus procesos de enseñanza y aprendizaje, así como también plantea un desarrollo a corto y mediano plazo para ampliar su oferta educativa a través de los entornos virtuales de aprendizaje, recientemente se anunció la implementación de tres licenciaturas en modalidad virtual.

La UADY, consciente de la importancia del uso de las TIC y particularmente, de los LMS, ha desarrollado diversas acciones en función al reconocimiento de necesidades del personal académico, y en consecuencia la capacitación de estos para el buen uso de dichas tecnologías; a continuación, se mencionan aquellas acciones que han impactado en la formación de los docentes respecto al uso de las TIC y de las plataformas educativas (UADY, 2013).

En el año 2003, la Dirección General de Desarrollo Académico de la UADY impartió cursos para la habilitación de profesores para el uso del Sistema de Gestión del Aprendizaje, en ese entonces llamado Dokeos, "... así como en el diseño instruccional para la enseñanza en entornos virtuales." (p.24).

En el año 2007 se desarrolla e implementa el Módulo V, conocido como: "Innovación educativa: Uso de las TIC en educación" el cual "... ha formado a 145 profesores de los cinco campus y del Centro de Investigaciones Regionales Dr. Hideyo Noguchi de la UADY, para el uso de la plataforma Dokeos o MOODLE, bajo diferentes diseños instruccionales." (p.24).

Respecto a las tutorías en línea, entre los años 2009 y 2010 se capacitaron a 96 profesores a través del curso "e tutoring" impartido por el Consejo Británico; de igual manera, en ese mismo periodo se habilitaron a profesores de las Facultades de Economía, Odontología, Enfermería y Ciencias Antropológicas de la UADY para el uso de la plataforma MOODLE.

En el año 2011, la UADY estableció un convenio con Virtual Educa Cono Sur para especializar a un grupo de académicos y administrativos de TI en entornos virtuales de aprendizaje, dicha capacitación se realizó a través del Instituto Argentino de Formación Docente.

En relación con las necesidades de formación de profesores para el desarrollo de competencias en diseño instruccional, tutoría y tecnológicas para las modalidades no convencionales, es decir, no presencial y mixta, la UADY condujo una investigación para la cual se encuestó a 924 académicos de 15 Facultades, dos escuelas preparatorias y la unidad académica de bachillerato con interacción comunitaria.

Los principales resultados, de acuerdo con el Plan Desarrollo de UADY Virtual (UADY, 2013), fueron los siguientes:

- a. El 40% del personal académico consideró tener las competencias para poder desarrollar un curso en una modalidad no convencional.

- b. El 38% indicó poder desarrollarse en un nivel elemental como tutor en línea.
- c. El 80% manifestó poder hacer uso, al menos en un nivel elemental, de los recursos tecnológicos necesarios para la modalidad no convencional.
- d. El 80% aseguró estar interesados en participar en las modalidades no convencionales.
- e. El personal académico, en los comentarios, recomendó mayor capacitación, reorganización de las funciones y mejores servicios tecnológicos.

Sin embargo, a pesar de los esfuerzos antes descritos, existe un área de oportunidad que consiste en el análisis de la actividad realizada por los profesores en el desarrollo de sus cursos en línea, los cuales en su gran mayoría se utilizan como una herramienta tecnológica de apoyo de los cursos presenciales.

4. Metodología

De acuerdo al autor Kerlinger (1969), este estudio es considerado como no experimental debido a que no se manipularon las variables, ni se asignaron aleatoriamente a los sujetos estudiados, los cuales fueron todos los profesores que utilizaron UADY Virtual en al menos un curso; asimismo, es *ex post-facto* ya que los eventos, al momento de la investigación, ya habían ocurrido al igual que sus efectos.

En cuanto a la temporalidad, de acuerdo con León y Montero (1997), el proyecto fue transversal, ya que se desarrolló en un momento determinado y debido a que permitió establecer la relación entre las variables de interés. La metodología empleada en esta investigación está basada del KDD (Figura 1).

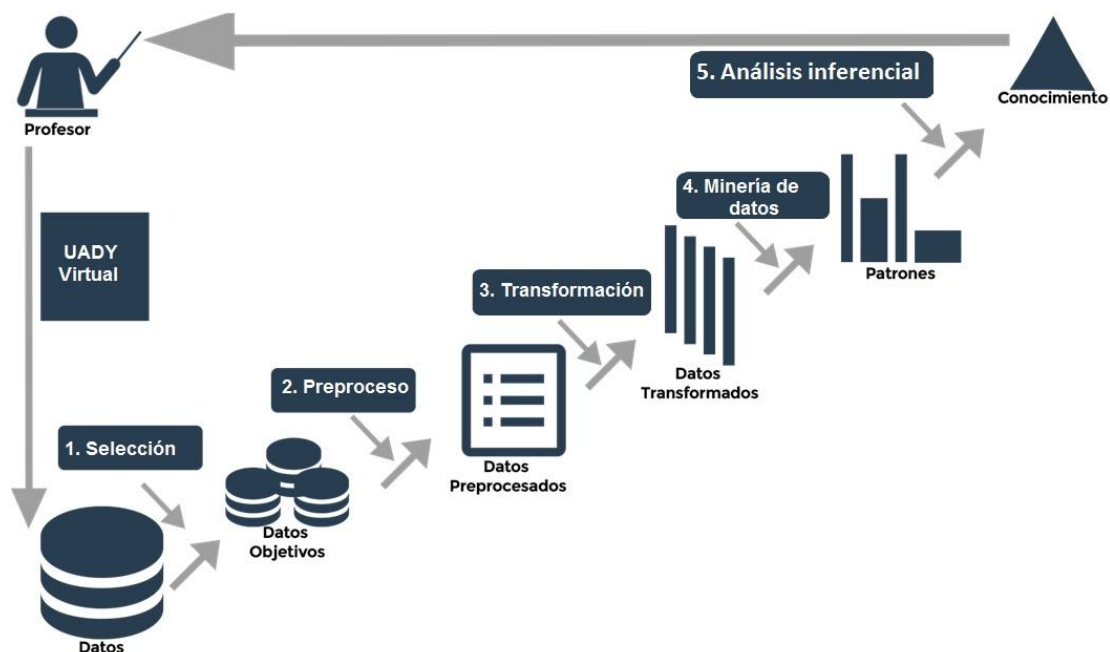


Figura 1. Método de Extracción del Conocimiento en Bases de Datos. Fuente: Fayyad et al. (1996).

De acuerdo con la Figura 1, las cinco fases del método KDD se describen a continuación:

1. Selección.

Los datos consistieron en archivos de acceso correspondientes a la actividad de 484 profesores de pregrado y posgrado que utilizaron UADY Virtual en al menos un curso, ya sea como una herramienta tecnológica de apoyo a sus clases presenciales o para impartir un curso totalmente en línea durante el periodo enero - julio del año 2016. Como se observa en la Figura 1, la distribución de la población del estudio es muy variable.

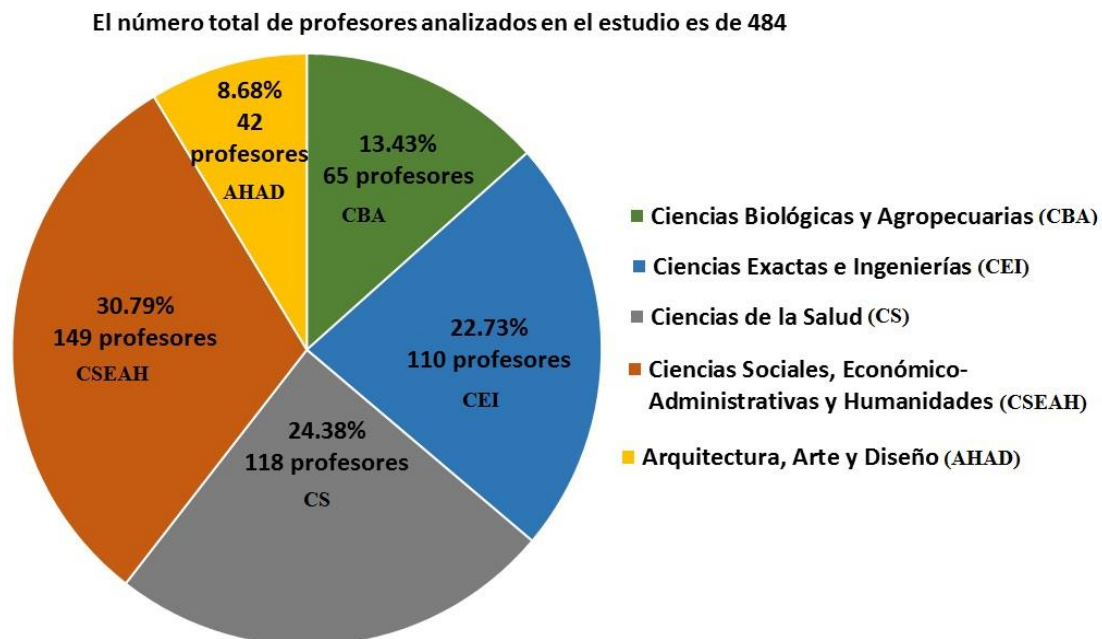


Figura 2. Profesores que impartieron al menos un curso en el período enero - julio 2016 con apoyo de UADY Virtual divididos por área de conocimiento.

La Figura 2, revela el porcentaje y número de profesores por área de conocimiento que usaron UADY Virtual. Se destaca que el campus de Ciencias Sociales, Económico Administrativas y Humanidades agrupa a la mayoría de los profesores que utiliza este sistema con 149 profesores (30.79%). En contraste, el campus de Arquitectura, Arte y Diseño con menor participación está representado por 42 profesores (8.68%). Esto se debe a la diferencia significativa que existe en las poblaciones del profesorado que están adscritos a estas áreas de conocimiento.

Por otra parte, se detectó que el número de las actividades y herramientas que utilizaron los profesores adscritos a las diversas áreas de conocimientos variable (ver Tabla 2).

Tabla 2

Actividades y recursos que utilizaron los profesores de UADY Virtual en las diversas áreas de conocimiento.

Campus por áreas del conocimiento de la UADY					
Actividad/ recurso	Ciencias biológicas y Agropecuarias	Ciencias Exactas e Ingeniería	Ciencias de la salud	Ciencias Sociales, Económico- Administrativas y Humanidades	Arquitectura, Arte y Diseño
Asistencia	Si	Si	Si	Si	Si
Base de datos	No	Si	Si	Si	Si
Chat	No	Si	Si	Si	Si
Consulta	Si	Si	Si	Si	Si
Cuestionario	Si	Si	Si	Si	Si
Encuesta	Si	Si	Si	Si	No
Encuesta predefinida	No	Si	Si	Si	No
Foro	Si	Si	Si	Si	Si
Glosario	Si	Si	Si	Si	Si
HotPot	No	No	Si	Si	No
Lecciones	No	Si	Si	Si	Si
SCORM	No	Si	No	Si	No
Taller	No	Si	Si	Si	No
Tareas	Si	Si	Si	Si	Si
Wiki	No	Si	Si	Si	Si
Archivo	Si	Si	Si	Si	Si
Carpeta	Si	Si	Si	Si	Si
Etiqueta	Si	Si	Si	Si	Si
Libro	Si	Si	Si	Si	Si
Página	Si	Si	Si	Si	Si
URL	Si	Si	Si	Si	Si
Total	13	20	20	21	16

Como se observa en la Tabla 2, los profesores que pertenecen al campus de Ciencias Sociales, Económico Administrativas y Humanidades son los que utilizaron más actividades y recursos de UADY Virtual. En contraste, los profesores del campus de Ciencias Biológicas y Agropecuarias fueron los que menos herramientas utilizaron en el desarrollo de sus cursos en línea.

Los registros de acceso de los profesores estaban distribuidos en 197 archivos (144 MB) en formato CSV (del inglés, *Comma Separated Value*). Cada base de datos estaba compuesta por siete atributos los cuales se describen en la Tabla 3.

Tabla 3
Atributos de las bases de datos.

#	Atributo	Descripción
1	Hora	Registra el momento exacto en el que se crea el archivo de registro.
2	Contexto del evento	Almacena la dirección web que describe la ruta del elemento seleccionado.
3	Componente	Identifica si la dirección web corresponde a una actividad o recurso del sistema.
4	Nombre del evento	Muestra la dirección web de la acción del elemento seleccionado.
5	Descripción	Identifica al usuario que efectuó alguna acción y describe lo que realizó.
6	Origen	Señala el origen del archivo de registro.
7	Dirección IP (<i>Internet Protocol</i>)	Identifica la dirección IP del usuario que accedió al sistema.

Posteriormente se agruparon los 197 archivos .CSV en un archivo .XLSX, lo cual permitió consolidar toda la información en un mismo documento, con el propósito de obtener todos los registros de acceso de los profesores, y particularmente, determinar las actividades y recursos de UADY Virtual.

2. Preproceso

Esta es la fase que requirió la mayor cantidad de tiempo en el proceso de extracción del conocimiento. Su principal objetivo consistió en ubicar a los profesores a través de un Número de Identificación Único (en inglés, *Identification*, ID) y determinar las herramientas (así como la frecuencia de uso de éstas) empleadas para cada uno de los profesores. Para ello, se utilizó el archivo .XLSX creado en la fase anterior, y el software *Excel de Microsoft* versión 2010.

Con el objetivo de reconocer las herramientas que utilizó cada profesor en el entorno de UADY virtual, se seleccionaron los siguientes atributos: Nombre del evento y Descripción. Posteriormente, se realizó la limpieza de datos para identificar la

frecuencia de uso para cada herramienta utilizada por el profesor. Para ello, se realizaron las siguientes acciones:

- a. Se filtraron los archivos de registro considerando únicamente las herramientas que pueden usar los profesores, las cuales se obtuvieron del atributo “Nombre del evento”.
- b. Se creó la primera versión de la base de datos que relacionaba el ID del profesor con la herramienta utilizada. La identificación se obtuvo con el atributo “Descripción”.
- c. Se contabilizaron cada una de las herramientas con la función Subtotal del programa *Excel de Microsoft*.
- d. Se creó la base de datos UADY Virtual V2.0 en la que se relacionaba el ID del profesor y los archivos de acceso correspondientes para cada herramienta usada por cada profesor.
- e. Se complementó la información de la base de datos UADY Virtual V2.0 con la Facultad a la que correspondía y el número de identificación del profesor.

3. Transformación

En esta fase se cuantificó para cada uno de los 484 profesores la frecuencia de uso de cada una de las actividades y recursos del sistema UADY Virtual. Para ello, se transformaron los valores de las herramientas de numérico a nominales: Muy bajo, Bajo, Medio, Alto y Muy alto.

Para determinar los nuevos valores nominales, se identificaron los valores máximos y mínimos de cada herramienta; posteriormente, se calculó el rango (valor máximo menos el valor mínimo, dividido entre cinco) para conformar la escala Likert. Finalmente se convirtió el tipo de archivo de las dos bases de datos de .XLSX a .ARFF (*del inglés, Attribute-Relation File Format*).

4. Minería de datos

En esta etapa se aplicaron diversas técnicas de minería de datos a través del software de distribución libre WEKA, en el que se experimentaron con diversos algoritmos de agrupamiento, asociación y clasificación. A continuación, se describen:

Técnicas de agrupamiento: el conjunto de datos se transformó al formato ARFF (del inglés, *Attribute-Relation File Format*), el cual es compatible con el software WEKA. Después de experimentar con los algoritmos EM (Dempster et al., 1977), FilteredClustered (Cartan, 1937) y SimpleKMeans (MacQueen, 1967), se obtuvieron resultados más eficientes con este último algoritmo (Tabla 2).

Tabla 4

Resultados obtenidos con el algoritmo de agrupamiento SimpleKmeans a través del software WEKA

Variable	Grupo 1 (2 profesores)	Grupo 2 (219 profesores)	Grupo 3 (263 profesores)
Actividades			
Asistencia	Muy bajo	Muy bajo	Muy bajo
Base de Datos	Muy bajo	Muy bajo	Muy bajo
Chat	Muy bajo	Muy bajo	Muy bajo
Consulta	Muy alto	Muy bajo	Bajo
Cuestionario	Muy bajo	Muy bajo	Muy bajo
Encuesta	Muy bajo	Muy bajo	Muy bajo
Encuesta predefinida	Muy bajo	Muy bajo	Muy bajo
Foro	Muy bajo	Muy bajo	Muy bajo
Glosario	Muy bajo	Muy bajo	Muy bajo
HotPotatoes	Muy bajo	Muy bajo	Muy bajo
Lecciones	Muy bajo	Muy bajo	Muy bajo
SCORM	Muy bajo	Muy bajo	Muy bajo
Taller	Muy bajo	Muy bajo	Muy bajo
Tareas	Muy alto	Bajo	Medio
Wiki	Muy bajo	Muy bajo	Muy bajo
Recursos			
Archivo	Muy bajo	Medio	Bajo
Carpeta	Medio	Bajo	Muy bajo
Etiqueta	Muy alto	Medio	Bajo
Libro	Muy alto	Muy bajo	Muy bajo
Página	Muy alto	Muy bajo	Muy bajo
URL	Muy bajo	Muy bajo	Muy bajo

En la Tabla 4, se observa que las actividades y recursos mutuamente excluyentes que definen a los tres grupos son: Consulta, Tareas, Archivo, Carpeta y Etiqueta. El resto de las herramientas con las que cuenta el sistema UADY Virtual se utilizan en un nivel muy bajo. En el análisis de los resultados obtenidos, se distingue que dos profesores han tenido un desempeño extraordinario. En contraste, 482 de los profesores que participaron en el estudio tuvieron un desempeño extraordinariamente bajo.

Técnicas de asociación: para la obtención de este tipo de reglas de conocimiento, se consideraron las mejores 10 (Tabla 5) con las siguientes condiciones: soporte de 0.95, y confianza de 0.9. Después de varios experimentos, se obtuvo mejores resultados con el algoritmo *A priori* (Agrawal et al., 1993).

çTabla 5

Reglas de asociación de las actividades y recursos de UADY Virtual obtenidas con el algoritmo A priori (Agrawal et al., 1993).a través del software WEKA.

#	Reglas de asociación	Interpretación en lenguaje natural
1.	Hot Potatoes = Muy poco ==> Lecciones = Muy poco	Si el profesor usa la herramienta Hot Potatoes muy poco, entonces usará muy poco la herramienta Lecciones.
2.	SCORM = Muy poco ==> Taller = Muy poco	Si el profesor usa la herramienta SCORM muy poco, entonces usará muy poco la herramienta Taller.
3.	SCORM = Muy poco ==> Libro = Muy poco	Si el profesor usa la herramienta SCORM muy poco, entonces usará muy poco la herramienta Libro.
4.	SCORM = Muy poco Libro = Muy poco ==> Taller = Muy poco	Si el profesor usa la herramienta SCORM muy poco y Libro muy poco, entonces usará muy poco la herramienta Taller.
5.	SCORM = Muy poco Taller = Muy poco ==> Libro = Muy poco	Si el profesor usa la herramienta SCORM muy poco y Taller muy poco, entonces usará muy poco la herramienta Libro.
6.	SCORM = Muy poco ==> Taller = Muy poco Libro=Muy poco	Si el profesor usa la herramienta SCORM muy poco, entonces usará muy poco las herramientas Taller y Libro.
7.	Hot Potatoes = Muy poco Taller = Muy poco ==> Lecciones = Muy poco	Si el profesor usa la herramienta Hot Potatoes muy poco y Taller muy poco, entonces usará muy poco la herramienta Lecciones.
8.	Hot Potatoes = Muy poco Libro = Muy poco ==> Lecciones = Muy poco	Si el profesor usa la herramienta Hot Potatoes muy poco y Libro muy poco, entonces usará muy poco la herramienta Lecciones.
9.	Lecciones = Muy poco SCORM = Muy poco ==> Taller = Muy poco	Si el profesor usa la herramienta Lecciones muy poco y SCORM muy poco, entonces usará muy poco la herramienta Taller.
10.	Lecciones = Muy poco SCORM = Muy poco ==> Libro = Muy poco	Si el profesor usa la herramienta Lecciones muy poco y SCORM muy poco, entonces usará muy poco la herramienta Libro.

Como se observa en la tabla 5, las reglas de asociación están conformadas por diversas actividades que se relacionan debido al muy poco uso entre el profesorado. Esto nos

permite detectar que es lo que menos se utiliza al momento de implementar un curso en línea,

Técnicas de clasificación: se experimentaron con diversos algoritmos que generan este tipo de reglas, obteniendo resultados óptimos con el algoritmo J48 (Ye, & Baldwin, 2006). Cabe destacar, que se seleccionó la actividad Tareas como elemento clasificador debido a que tiene una frecuencia muy alta de uso entre los profesores (representa el 80% de toda la actividad en UADY Virtual en el período evaluado). Como resultado se obtuvo el siguiente modelo predictivo (Figura 3) representado a través de un diagrama de árbol de decisión:

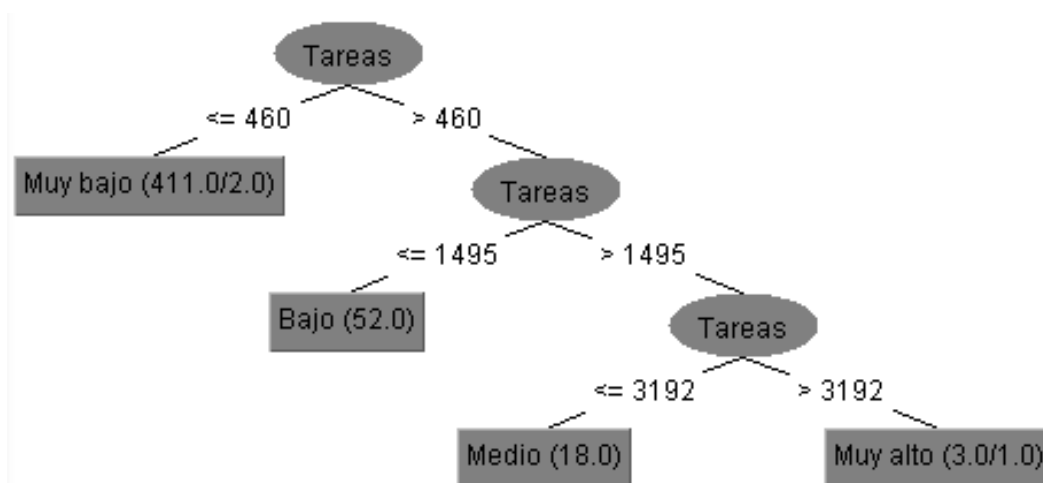


Figura 3. Reglas de clasificación con base a la actividad Tareas

En lenguaje natural, el diagrama anterior puede ser interpretado de la siguiente manera, Si la actividad tareas:

- Se usa por encima de las 3,192 veces, entonces, la calificación de profesor será muy alto.
- Es menor o igual a 3,192 y mayor a 1,495, entonces, la calificación del profesor será medio.
- Es menor o igual a 1,495 y mayor a 460, entonces, la calificación del profesor será bajo.
- Es menor o igual a 460, entonces, la calificación del profesor será muy bajo.

5. Análisis inferencial.

El propósito de la aplicación del análisis inferencial, fue determinar si existieron diferencias significativas en el uso de UADY Virtual de acuerdo al área de conocimiento, para ello, en primer lugar se realizó la prueba de normalidad de Kolmogorov – Smirnov (Lilliefors, 1967) ya que la población de los profesores de la UADY era mayor a 50 sujetos. El objetivo de la aplicación de la prueba de normalidad fue determinar si la distribución de cada una de las variables numéricas de cada una de las herramientas de UADY Virtual tenía una distribución normal o anormal (no paramétrica).

Para la prueba de normalidad se analizaron tanto cada una de las variables como una variable global (la cual consideraba todas las variables al mismo tiempo), en todos los análisis se consideró como hipótesis nula que, la distribución de la variable a analizar es igual a la distribución normal, mientras que la hipótesis alterna consistió en que la distribución de la variable a analizar era diferente de la distribución normal. Se consideró un nivel de significancia de 5 % (.05) para la región de rechazo. Los análisis se desarrollaron con el software estadístico SPSS (*Statistical Package for the Social Sciences*) versión 22.

En ninguno de los casos (análisis individual de las variables y análisis de los datos global) el nivel de significancia fue mayor a .05, es decir, en todos los análisis $P < .05$, por lo que se rechaza (en todos los casos) la hipótesis nula. Entonces, la distribución de la variable a analizar fue distinta de la distribución normal, comprobando que los datos tenían una distribución asimétrica, es decir no paramétrica o anormal.

A continuación, se muestra la Figura 4 que comprueba que la distribución de la variable global (total de herramientas) es distinta a la distribución normal.

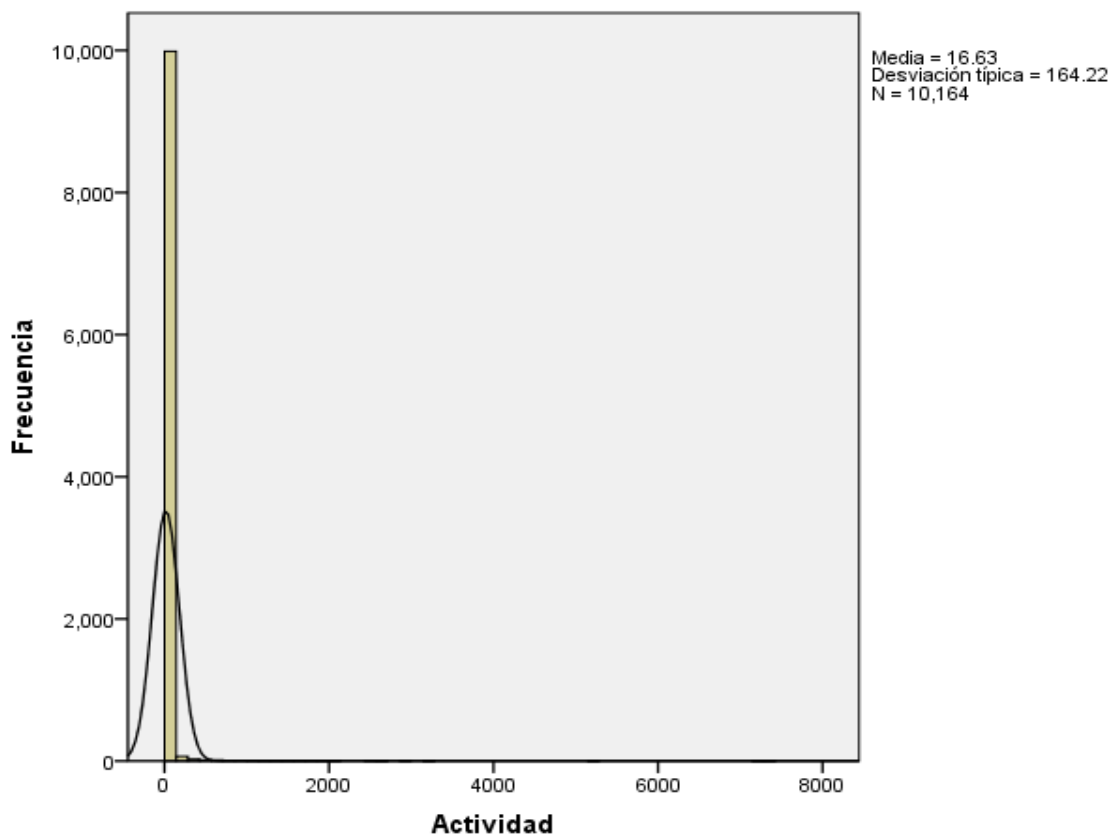


Figura 4. Prueba de normalidad para la variable global (total de herramientas usadas).

A continuación, se presenta la Tabla 6 que resume los resultados de normalidad aplicado a cada una de las herramientas de UADY Virtual.

Tabla 6

Prueba de Kolmogorov – Smirnov para cada una de las herramientas de UADY virtual.

Herramientas	Z de la prueba	Significancia
Actividades		
Asistencia	11.48	0.0
Base de datos	11.23	0.0
Chat	11.56	0.0
Consulta	10.92	0.0
Cuestionario	9.72	0.0
Encuesta	11.06	0.0
Encuesta predefinida	9.24	0.0
Foro	7.92	0.0
Glosario	10.77	0.0
Hot Potatoes	11.54	0.0
Lecciones	11.31	0.0
SCORM	11.37	0.0
Taller	11.16	0.0
Tareas	7.65	0.0
Wiki	11.46	0.0
Recursos		
Archivo	6.74	0.0
Carpeta	7.75	0.0
Etiqueta	5.54	0.0
Libro	10.45	0.0
Página	10.24	0.0
URL	8.48	0.0

Como se observa en la Tabla 6, debido a que ninguna de las variables (herramientas), tenían una distribución normal, se decidió aplicar una prueba de hipótesis para datos no paramétricos. Se utilizó la prueba de Kruskal-Wallis (Kruskal y Wallis, 1952) para realizar las siguientes comparativas:

- El uso de cada herramienta entre áreas de conocimiento, en caso de diferencias significativas se aplicó la comparación múltiple de Dunn (Daniel, 1990). Las pruebas estadísticas se consideraron significativas cuando $P < 0.05$ y se utilizaron los paquetes estadísticos STATGRAPHICS Centurion XVII v. 17.0.16 y nuevamente el SPSS versión 22. Cabe destacar que no se tomaron en cuenta el número de cursos que impartió cada profesor en el periodo evaluado ($n=484$), por lo que cada valor es una medida de frecuencia global para cada herramienta del sistema.
- El Total de Herramientas Usadas (THU) entre áreas de conocimiento (15 actividades y 6 recursos, siendo un total de 21), en caso de diferencias significativas se aplicó la comparación múltiple de Dunn. La variable THU puede describirse de la siguiente manera:

$$THU = \sum_{i=1}^{21} H_i$$

Donde:

- $H_i=1$ si el profesor usó la herramienta i -ésima
- $H_i=0$ si el profesor no usó la herramienta i -ésima. Así, para cada profesor se

tiene $0 \leq \sum_{i=1}^{21} H_i \leq 21$, es decir, $0 \leq THU \leq 21$

A continuación, se muestra el resultado (Tabla 7), donde se puede visualizar que el THU entre las áreas de conocimiento difirió significativamente ($H=42.4941$, $P<0.0001$, $gl=4$), resultando mayor en el campus de Ciencias Exactas e Ingeniería que en:

- Ciencias Sociales Económico-Administrativas y Humanidades
- Arquitectura, Hábitat, Arte y Diseño
- Ciencias Biológicas y Agropecuarias

Así como también, obteniendo un resultado mayor en el campus de Ciencias de la Salud que en:

- Ciencias Sociales Económico-Administrativas y Humanidades
- Arquitectura, Hábitat, Arte

Sin embargo, no difirió entre los campus de Ciencias Exactas e Ingeniería y Ciencias de la Salud.

Tabla 7

Total de herramientas usadas entre áreas de conocimiento.

Campus	n	Promedio	DE	Rango Promedio	Mediana	Mínimo	Máximo
CSEAH	149	3.86577	2.51927	202.19 a	4	1	12
AHAD	42	3.85714	2.32244	202.62 a	3	1	11
CBA	65	4.01538	1.90798	220.06 a b	4	1	10
CS	118	5.01695	2.94967	261.26 b c	5	1	15
CEI	110	5.88182	2.8308	305.46 c	5.5	1	13

CSEAH: Ciencias Sociales Económico-Administrativas y Humanidades

AHAD: Arquitectura, Hábitat, Arte

CBA: Ciencias Biológicas y Agropecuarias

CS: Ciencias de la Salud

CEI: Ciencias Exactas e Ingeniería

Nota: Los rangos promedio con distinta letra difieren ($P<0.05$), prueba de Dunn.

5. Conclusiones

En esta propuesta se analizó el desempeño de 484 profesores de los niveles de pregrado y posgrado que utilizaron al menos una vez el Sistema de Gestión del Aprendizaje denominado UADY Virtual en el período enero-junio del año 2016 como una herramienta de apoyo a sus clases presenciales o para impartir cursos totalmente en línea. El objetivo general que se planteó fue reconocer los patrones de comportamiento del profesorado universitario en el uso de las actividades y recursos que contiene el entorno de esta herramienta tecnológica.

En el estudio se experimentó con diversos algoritmos de agrupamiento, asociación y clasificación a través del software de distribución libre WEKA. En los resultados obtenidos se destaca lo siguiente:

- Se obtuvo la caracterización de tres grupos de profesores por la frecuencia de uso de las siguientes actividades y recursos: Consulta, Tareas, Archivo, Carpeta y Etiqueta. Cabe destacar, que las tres últimas, son herramientas cuyo propósito radica en la construcción de la interfaz del sitio web y para la navegación del mismo, por consiguiente, se podría afirmar que sólo las actividades Consulta, Tareas y Foro utilizan los profesores como apoyo a sus clases presenciales.
- Se generaron reglas de asociación que permitieron identificar las correlaciones más fuertes entre las siguientes actividades y recursos: *hotPotatoes*, lecciones, SCORM, taller y libro. Esto nos indica que dichas herramientas requieren mayor importancia para ser consideradas al momento de la capacitación del profesorado.
- Se obtuvieron reglas de clasificación con base a la actividad Tareas. Esto demuestra que sólo considerando esta herramienta se puede clasificar el desempeño de un profesor en UADY Virtual debido a su amplio uso.

Los resultados obtenidos fueron consistentes con la investigación efectuada por Romero et al. (2010), en cuanto a que en el manejo del tipo de dato (numérico o nominal) influyen considerablemente los resultados de la aplicación de los algoritmos, tanto de agrupamiento como de clasificación, asimismo, los algoritmos de clasificación mejoran sus porcentajes predictivos si se usa en totalidad la base de datos, es decir, sin filtro alguno. Respecto a las diferencias, de igual manera, considerando la investigación previamente mencionada, se obtuvo un porcentaje de efectividad del modelo superior al 90%, lo que significa que el modelo clasificó con mejor precisión las actividades de los profesores.

En cuanto al análisis inferencial, se efectuó un análisis de varianza de un solo factor para comparar el uso de UADY Virtual entre las cinco áreas de conocimiento con los que cuenta esta institución. Dicho estudio, determinó que no existieron diferencias significativas en el uso de este sistema de acuerdo con el área de conocimiento, lo que significa que la manera en la que interactúan los profesores con el entorno es igual para todas las áreas de conocimiento. Por lo tanto, se refuerza la conclusión que los 484 profesores que participaron en el estudio usan como principal herramienta la actividad tareas.

En cuanto a las necesidades de formación de los profesores en el uso de este sistema como apoyo a sus clases presenciales, se puede concluir que a pesar de que la población quedó agrupada con un desempeño muy bajo, al realizar los análisis por áreas de conocimiento se pueden apreciar diversos desempeños entre los profesores.

Los patrones de comportamiento de los profesores demuestran que utilizan muy poco las diversas actividades y recursos del sistema UADY Virtual. Aunque esto no significa que no saben utilizarlas, se reconoce como un área de oportunidad para el personal encargado de la formación de los profesores de esta institución educativa de nivel superior. Con base a lo anterior, se propone reforzar el programa de formación del profesorado en el uso de las actividades y recursos que contiene UADY Virtual. Algunas de las acciones concretas que se proponen son:

- Impartición de cursos personalizados tomando en cuenta el área de conocimiento y el desempeño de los profesores, donde se promuevan aquellas actividades y recursos que no reportan actividad.

- Creación de un canal de videos por parte de la institución donde se promuevan aquellas actividades y recursos que reporten un uso nulo en todas las áreas de conocimiento.
- Identificación de profesores con un desempeño por encima de la media para impartirles un curso avanzado, con el objetivo de que ellos se conviertan en instructores de esos cursos de formación personalizados.
- Desarrollo de tutores inteligentes que ayuden a los profesores en el diseño de sus cursos independientemente del nivel de desempeño que tengan.
- Implementación de herramientas de visuales para los administradores de UADY Virtual que permitan obtener indicadores en tiempo real sobre el uso de este sistema.
- Replicar cada período lectivo este experimento, a fin de medir la evolución del uso del sistema, de igual manera se recomienda un análisis más profundo de los datos, el análisis puede desarrollarse de acuerdo con los siguientes niveles: Facultad, programa educativo y asignatura.

Finalmente, se destaca que la investigación es de mucha utilidad para la UADY, debido a que planea implementar 3 licenciaturas en modalidad virtual en los próximos 2 años.

Presentación del artículo: 17 de febrero de 2018

Fecha de aprobación: 15 de marzo de 2018

Fecha de publicación: 31 de octubre de 2018

Camacho Chacón, P.E., Zapata González, A., Menéndez Domínguez, V.H., y Canto Herrera, P.J. (2018). Análisis del desempeño del profesorado universitario en el uso de MOODLE a través de técnicas de minería de datos: propuestas de necesidades formativas. *RED. Revista de Educación a Distancia*, 58. Consultado el (dd/mm/aaaa) en <http://www.um.es/ead/red/58/>

Financiación

Esta investigación no ha recibido ninguna subvención específica de los organismos de financiación en los sectores públicos, comerciales o sin fines de lucro.

Agradecimientos

Esta investigación ha sido posible gracias a la asignación de efectivos a través de la beca número 39 3709 del Consejo Nacional de Ciencia y Tecnología (CONACYT, México).

Un agradecimiento especial al Departamento de Investigación e Innovación Educativa de la UADY por las facilidades otorgadas para el análisis de las bases de datos de UADY Virtual.

6. Referencias bibliográficas

- Abdullateef, B. N., Elias, N., Mohamed, H., Zaidan, A., & Zaidan, B. (2015). Study on Open Source Learning Management Systems: a Survey, Profile, and Taxonomy. *Journal of Theoretical & Applied Information Technology*, 82(1), 93-105. Recuperado de <http://www.jatit.org/volumes/Vol82No1/10Vol82No1.pdf>
- Agrawal, R., Imieliński, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In *Acm sigmod record*, 22(2), 207-216.
- Alavi, M., Leidne, D. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS quarterly*, 25(1), 107-136.
- Alcalá-Fernández, J., Sánchez, L., Garcia, S., del Jesus, M. J., Ventura, S., Garrell, J. M., ... & Fernández, J. C. (2009). KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3), 307-318.
- Alonso, F., López, G., Manrique, D., & Viñes, J. M. (2005). An Instructional Model for Web-Based e-Learning Education with a Blended Learning Process Approach. *British Journal of Educational Technology*, 36(2), 217-235.
- Anderson, R., & Becker, H. (2001). *Report #8 School Investments in Instructional Technology*. Recuperado de http://www.crito.uci.edu/tlc/findings/report_8/REPORT_8.PDF
- Andone, D., Ternauciuc, A., & Vasiu, R. (2017, July). Using Open Education Tools for a Higher Education Virtual Campus. In 17th IEEE International Conference on Advanced Learning Technologies (*ICALT*)(pp. 26-30), Rumania.
- Arimond, A., Kofler, C., & Shafait, F. (2010). Distributed Pattern Recognition in RapidMiner. In *Proceedings of RapidMiner Community Meeting And Conference*.
- Banco Mundial (2013). About the World Bank Institute. Recuperado de: <http://wbi.worldbank.org/wbi/content/acerca-del-instituto-del-banco-mundial>.
- Boiko, B. (2004). *Content management bible*. John Wiley & Sons.
- Cabrero, J. (2007). *Nuevas Tecnologías Aplicadas a la Educación*. McGraw-Hill.
- Cartan, H. (1937). Theorie des filtres, filtres et ultrafiltres, C. R. Acad. Sci. Paris vol. 205 pp. 595-77.
- Cantabella, M., López-Ayuso, B., Muñoz, A., & Caballero, A. (2016). Una herramienta para el seguimiento del profesorado universitario en entornos virtuales de Aprendizaje. *Revista Española de Documentación Científica*, 39(4), 1-15. Recuperado de <http://redc.revistas.csic.es/index.php/redc/article/view/957/1437>

- Cemile, F. (2008). Intelligent learning system for online learning. *International Journal of Hybrid Intelligent Systems* , 129-141
- CONACYT (2014). Programa especial de ciencia, tecnología e innovación 2014-2018. Recuperado de: http://www.conacyt.mx/siicyt/images/PECiTI-2014_2018.pdf
- Cooke, M. (2004). Clomedia: The Evolution of e-Learning. Recuperado de: http://www.clomedia.com/content/templates/clo_webonly.asp?articleid=571&zoneid=78
- Charlton, P., Mavrikis, M & Katsifli, D. (2013). *The Potential of Learning Analytics and Big Data*. Recuperado de: <http://www.ariadne.ac.uk/issue71/charlton-et-al>
- Chen, H. (2001). *Knowledge Management Systems: A text Mining perspective*. University of Arizona, Tucson, Arizona. Knowledge Computing Corporation.
- Chen, A. (2006). *The Technology and Application of Data Mining*. Beijing: Science Press.
- Clarke, D. J. (2002). E-learning: Big bang or steady evolution. *Learning Technologies*, 7(1), 1-2.
- Daniel, W.W. (1990). *Applied Nonparametric Statistics. 2a Ed. Duxbury. Pacific Grove CA, USA*.
- Dempster, A.P, Laird, N.M & Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- Dillenbourg, P., Schneider, D., & Synteta, P. (2002). Virtual learning environments. In *3rd Hellenic Conference" Information & Communication Technologies in Education"* (pp. 3-18). Kastaniotis Editions, Greece.
- Dougiamas, M., Taylor, P.C. (2003). MOODLE : Using Learning Communities to Create an Open Source Course Management System. *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 171–178.).
- Drucker, P. (1992). *The Age of Discontinuity: Guidelines of Our Changing Society*. New York: Harper & Row.
- Edwards A. L. (1976). *Introduction to Linear Regression and Correlation*. San Francisco, EEUU: W. H. Freeman & Co. LTD,
- Eiben, A. E., Smith, J. . (2003). *Introduction to Evolutionary Computing* [versión Ebook]. Recuperado de http://csllt.riit.tsinghua.edu.cn/mediawiki/images/e/e8/Introduction_to_Evolutionary_Computing.pdf

- Ellis, R. A., Ginns, P., & Piggott, L. (2009). E-learning in higher education: some key aspects and their relationship to approaches to study. *Higher Education Research & Development*, 28(3), 303-318.
- Fayyad, U., Piatetsky-shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery. *AI magazine*, 17(3), 37-54.
- Fischer, S. (2001). Course and exercise sequencing using metadata in adaptive hypermedia learning systems. *Journal of Educational Resources in Computing*, 1(1). doi:10.1145/376697.376700
- García, A. (2013). De dónde venimos y hacia dónde vamos en Educación a Distancia. *Cátedra UNESCO de Educación a Distancia. UNED*. Recuperado de https://www.youtube.com/watch?v=J_A2wp7_5Z8
- Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, 22(4), 521-563.
- Han, J., Chiang, J. Y., Chee, S., Chen, J., Chen, Q., Cheng, S., ... & Lu, Y. (1997). DBMiner: A system for data mining in relational databases and data warehouses. In *Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research* (p. 8). IBM Press.
- Hernández, J., Ramírez, M.J., Ferri, C. (2004). *Introducción a la Minería de Datos*. Pearson Prentice Hall.
- Hernández, R., Fernández, C. y Baptista, P. (2006). *Metodología de la investigación* (4ª ed.). México: McGraw Hill.
- Hernández, G. (2015). Análisis del uso y manejo de la plataforma Moodle en docentes de matemáticas, para el desarrollo de competencias integrales en estudiantes de primaria. *Revista Q*, 10 (19). Recuperado de <http://dx.doi.org/10.18566/revistaq.v10n19.a01>
- Hoffmaster, D. (2006). The E-Learning Evolution. In *E-Learning Concepts and Techniques*. Pennsylvania.
- Honeycutt, B., & Garrett, J. (2013). *The flipped approach to a learner-centered class*. Magna Publications.
- Hosmer, D.W., Stanley, L. (2000). *Applied Logistic Regression*. New Jersey, EE.UU: John Wiley & Sons, Inc.
- IBM Corp, N. (2017). IBM SPSS statistics for windows. *Version*, 25.
- IMS Global Learning Consortium, I. (2003). IMS learning design information model. Recuperado de: https://www.imsglobal.org/learningdesign/ldv1p0/imsld_infov1p0.html

- Itmazi, J. (2005). *Sistema flexible de gestión del E-Learning para soportar el aprendizaje en las universidades tradicionales y abiertas* (Tesis doctoral, Universidad de Granada, España). Recuperada de <http://0-hera.ugr.es.adrastea.ugr.es/tesisugr/15508584.pdf>
- Itmazi, J. A., Gea, M. M., Paderewski, P., & Gutiérrez, F. L. (2005). A comparison and evaluation of open source learning management systems. *IADIS AC*, 1(11).
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318), 399-402.
- Jereb, E., & Šmitek, B. (2006). Applying Multimedia Instruction in e-Learning. *Innovations in Education & Teaching International*, 43(1), 15-27
- Kapp, K. (2003). Five Technological Considerations When Choosing an E-Learning Solution. *eLearn Magazine*, 7.
- Kerlinger, F. N. (1969). Research in Education. En R. Ebel, V. Noll, & R. Bauer (Eds.), *Encyclopedia of Education* (4th ed., pp. 1127-1134). New York, EE.UU: Macmillan.
- Kohonen, T. (1988). Self-organized formation of topologically correct feature maps. En E. Anderson, J.A. Rosenfeld (Eds.), *Neurocomputing: foundations of research* (pp. 509-521). Berlin, Alemania: Springer-Verlag.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621.
- León, O. y Montero, I. (1997). *Diseño de investigaciones: Introducción a la lógica de la investigación en psicología y educación* (No. 303.44). McGraw-Hill Interamericana.
- Lynch, T. D., & Lynch, C. E. (2003). Web-based education. *The Innovation Journal*, 8(4).
- Lugo, M. T., López, N., & Toranzo, L. (2014). *Informe sobre tendencias sociales y educativas en América Latina, 2014: políticas TIC en los sistemas educativos de América Latina*. Recuperado de: <https://goo.gl/ND86EQ>
- Mazza, R., & Milani, C. (2004). Gismo: a graphical interactive student monitoring tool for course management systems. In *International Conference on Technology Enhanced Learning, Milan* (pp. 1-8).
- Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. En *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.

- Magdin, M., & Turcáni, M. (2015). Personalization of Student in Course Management Systems on the Basis Using Method of Data Mining. *Turkish Online Journal of Educational Technology-TOJET*, 14(1), 58-67.
- Menéndez-Domínguez, V. H., & Castellanos-Bolaños, M. E. (2014). La Calidad en los Sistemas de Gestión del Aprendizaje. *Abstraction and Application Magazine*, 4.
- Naidu, S. (2006). *E-Learning, A Guidebook of Principles, Procedures and Practices*. Melbourne: Commonwealth Educational Media Center for Asia (CEMCA).
- Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. San Francisco, USA Morgan Kaufmann Publishers Inc.
- Quinlan, J R. (2007). Induction of Decision Trees. *Machine learning*, 1(1), 81–106.
- Romero, C., & Ventura, S. (Eds.). (2006). *Data Mining In E-Learning* (Vol. 4). Editorial Wit Press.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1), 135-146. Recuperado de <https://www.sciencedirect.com/science/article/pii/S0957417406001266>
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368–384.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618. Recuperado de <https://ieeexplore.ieee.org/abstract/document/5524021/>
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2010). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 21(1), 135-146.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. En J. L. Rumelhart, David E. and McClelland (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (pp. 318–362). MA, USA: MIT Press Cambridge
- Sánchez, J. (2009). Plataformas de Enseñanza Virtual para Entornos Educativos. *Pixel-Bit: Revista de Medios y Educación*, 217-233. Recuperado de <http://www.redalyc.org/pdf/368/36812036015.pdf>
- Sangra, A., Vlachopoulos, D., & Cabrera, N. (2012). Building an Inclusive Definition of E-Learning: An Approach to the Conceptual Framework. *International Review of Research in Open and Distance Learning*, 145-159.

- Salcines, E. G., Romero, C., Ventura, S., & de Castro, C. (2008). Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning. *IEEE-RITA*, 3(1), 19-30. Recuperado de <http://rita.det.uvigo.es/200805/uploads/IEEE-RITA.2008.V3.N1.A3.pdf>
- Samaniego, G., Marqués, L., & Gisbert, M. (2015). El profesorado universitario y el uso de Entornos virtuales de aprendizaje. *Campus Virtuales*, 4(2), 50-58. Recuperado de <http://uajournals.com/ojs/index.php/campusvirtuales/article/view/84>
- Suhirman, S., Tutut, H., Haruna, C., & Jasni, M. Z. (2014). Data Mining for Education Decision Support: A Review. *International Journal of Emerging Technologies in Learning (iJET)*, 9(6), 4-19.
- Scheffer, T. (2004). Finding Association Rules that Trade Support Optimally Against Confidence. *Intell. Data Anal.*, 9(4), 381-395.
- Shibley, I. (2011). Putting the learning in blended learning. *Faculty Focus*.
- Trandafilu, E., Allkoçi, A., Kajo, E., Xhuvani, A. (2012). Discovery and evaluation of student's profiles with machine learning. In *Proceedings of the Fifth Balkan Conference in Informatics* (pp. 174-179). ACM.
- UADY (2010). Plan de Desarrollo Institucional 2010-2020. *Universidad Autónoma de Yucatán*. Recuperado de <http://www.pdi.uady.mx/docs/pdi.pdf>
- UADY (2012). Modelo Educativo para la Formación Integral. *Universidad Autónoma de Yucatán*. Recuperado de http://www.dgda.uady.mx/media/docs/mefi_dgda.pdf
- UADY (2013). Plan de Desarrollo UADY Virtual 2013-2020. *Universidad Autónoma de Yucatán*. Recuperado de <http://es.uadyvirtual.uady.mx/>
- UNESCO (2002). *Information and Communication Technologies in Teacher Education. a Planning Guide*. UNESCO.
- Valsamidis, S., Kontogiannis, S., Kazanidis, I., Theodosiou, T., & Karakos, A. (2012). A Clustering Methodology of Web Log Data for Learning Management Systems. *Journal of Educational Technology & Society*, 15 (2), 154-167.
- Wang, J. K. (2014). A Research of Information Management System Based on Data Mining. *Applied Mechanics and Materials*, 616-619.
- Witten I.H., Frank E., Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- Ye, P., Baldwin, T. (2006). Semantic Role Labelling of Prepositional Phrases. *ACM Transactions on Asian Language Information Processing (TALIP)*, 5(3), 228-244.

Zapata, A. (2013). *Modelo Híbrido de Recomendación de Objetos de Aprendizaje*. Tesis doctoral de la Universidad de Castilla-La Mancha, España. Recuperada de <https://dialnet.unirioja.es/servlet/tesis?codigo=111073>

Zhang, H. (2004). The Optimality of Naive Bayes. In Z. Barr, V., Markov (Eds.), *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*.