Universitat d'Alacant
Universidad de Alicante

Propuesta de un conjunto de herramientas de minería de datos para evaluar el desempeño de los estudiantes y los procesos de enseñanza-aprendizaje en el ámbito de la educación en ingeniería

Diego Patricio Buenaño Fernández

Tesis Doctorales
UNIVERSIDAD de ALICANTE

Unitat de Digitalització UA
Unidad de Digitalización UA

Universitat d'Alacant
Universidad de Alicante

Instituto Universitario de Investigación
Informática

Escuela Politécnica Superior

Propuesta de un conjunto de herramientas de minería de datos para evaluar el desempeño de los estudiantes y los procesos de enseñanza - aprendizaje en el ámbito de la educación en ingeniería

Diego Patricio Buenaño Fernández

Tesis presentada para aspirar al grado de

Doctor por la Universidad de Alicante

Doctorado en Informática

Dirigida por:

Dr. Sergio Luján Mora
Dr. David Gil Méndez

Mayo 2020

Universitat d'Alacant
Universidad de Alicante

TESIS DOCTORAL EN FORMA DE COMPENDIO DE PUBLICACIONES

Propuesta de un conjunto de herramientas de minería de datos para evaluar el desempeño de los estudiantes y los procesos de enseñanza - aprendizaje en el ámbito de la educación en ingeniería

El presente documento contiene una síntesis del trabajo realizado por Diego Buenaño Fernández, bajo la dirección del Dr. Sergio Luján Mora y la Subdirección del Dr. David Gil, para optar por el grado de Doctor en Informática. Se presenta en la Universidad de Alicante y se estructura según la normativa establecida para la presentación de tesis doctorales en forma de compendio de publicaciones: una primera parte con una síntesis, una segunda parte que reproduce las publicaciones científicas realizadas y una tercera parte con las conclusiones.

Mayo 2020

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Dedicatoria

Esta tesis está dedicada a mi amada familia, a mi esposa Verónica y a mis hijos Esteban Adrián y Diego Andrés, ellos han transitado junto a mí en este arduo camino, se han alegrado con cada notificación de aprobación de artículo y me han impulsado cuando los ánimos decaían; a mis padres María Carmen y Julio, a mi hermana y a mi sobrina por su apoyo y preocupación en el avance de este proyecto.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Agradecimientos

Mi agradecimiento eterno al Arquitecto de la creación por el privilegio de la vida.

A Sergio Luján, mi director quien ha dedicado horas y horas a retroalimentar mis aprendizajes en el área de investigación, sin duda he aprendido mucho. Con su dedicación y ejemplo no solo me ha permitido alcanzar esta meta, sino que ha modelado de forma excelente la conducta de un verdadero maestro. Sergio, muchas gracias por tu valioso tiempo y sobre todo por la amistad brindada en este proceso.

A David Gil, mi subdirector quien me ha apoyado de forma significativa para poder cumplir con la culminación de este proyecto, gracias David por compartir tus consejos y directrices en este proyecto.

A la Universidad de Las Américas, y en particular, a la Dirección General de Investigación por apoyar mis proyectos y por el respaldo recibido para la publicación de varios artículos.

A William, Ángel, Santiago, Patricia, Luis, Tania, y Oswaldo, los amigos y compañeros del grupo de investigación que trabajamos bajo la dirección de Sergio Luján.

Mi agradecimiento de corazón a todos ustedes.

Alicante, Mayo de 2020
Diego Buenaño Fernández

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Resumen

Diariamente al rededor del mundo se genera una cantidad inmensa de datos producto de nuestra interacción permanente y creciente con la tecnología, ya sea para actividades laborales, académicas, personales o de ocio, entre algunas actividades puntuales tenemos los negocios digitales, el envío y recepción de correos electrónicos, la interacción con plataformas financieras, la interacción con redes sociales, la interacción con plataformas educativas, el uso de mapas virtuales, etc. son solo algunos ejemplos de las acciones que ejecutamos a diario y que producen una cantidad gigante y variada de datos susceptibles de ser analizados. En los próximos años esta tendencia se acelerará debido al incremento de dispositivos y sensores conectados a internet. Es importante y conveniente mencionar que en la situación actual que vive el planeta debido a la pandemia de la covid-19 el teletrabajo ha permitido mantener una situación lo más similar a la de normalidad, haciendo posible que muchos de los sectores no quebrasen, teniendo como soporte principal el uso de plataformas tecnológicas. Además, la mayoría de los servicios de internet se han visto completamente testeados y en la mayoría de casos han pasado las pruebas con éxito.

En el campo educativo, el incremento en el uso de sistemas de aprendizaje en línea, tales como entornos personales de aprendizaje, sistemas inteligentes de tutoría, sistemas de gestión de aprendizaje, así como también el aumento de la interacción estudiante - docente a través de blogs, wikis, redes sociales entre otros, genera una variada y extensa cantidad de información. Esta información, almacenada en las bases de datos institucionales, está siendo infrautilizada por estudiantes, docentes y administradores educativos, que utilizan las plataformas digitales simplemente como repositorios de información.

En los últimos años, se ha evidenciado en las bases de datos científicas un número significativo de investigaciones tanto teóricas como aplicaciones prácticas, que se enfocan en el ámbito de la minería de datos en entornos educativos y específicamente en el ámbito de la educación superior. La organización y análisis de este volumen gigante de datos tiene al menos dos posibilidades de enfoque, la minería de datos educativos y la analítica de aprendizaje. La primera desarrolla y adapta métodos estadísticos, de minería de datos y de aprendizaje automático, para analizar los datos generados por

Universidad de Alicante

estudiantes y docentes. Por otro lado, la analítica de aprendizaje se define como el proceso de medición, recopilación, análisis y presentación de datos relacionados con la interacción de estudiantes con las plataformas digitales. La analítica de aprendizaje tiene como objetivo entregar información que permita optimizar el logro de resultados de aprendizaje en el entorno en el que este se produce.

Los algoritmos tradicionales de minería de datos en entornos educativos no pueden aplicarse sin un análisis previo de las estrategias institucionales en las que se va a aplicar, ya que las instituciones de educación superior presentan diferentes comportamientos. Por ejemplo, un modelo educativo en una institución puede estar centrado en la enseñanza basada en la práctica e innovación mientras que otro modelo puede hacer énfasis en la investigación acción. Bajo esta premisa es importante tener una visión clara de los siguientes tres elementos para la aplicación de técnicas de minería de datos y analítica de aprendizaje: a) Estrategias institucionales en las que se aplican métodos de minería de datos educativos y analítica de aprendizaje, b) Métodos de minería de datos aplicados en entornos educativos y c) Herramientas para la implementación de minería de datos en entornos educativos. La presente tesis presenta un conjunto de herramientas de minería de datos con el objetivo de reforzar la evaluación de procesos de enseñanza - aprendizaje en el ámbito de la educación en ingeniería. Esta propuesta se sustenta en los tres elementos mencionados anteriormente y sobre los cuales giran los objetivos y artículos científicos incluidos en el compendio.

En el momento que redacté este resumen tenía relativamente clara la importancia de la educación en línea y del análisis de datos que se generan en este campo. La situación actual de pandemia y confinamiento ha incrementado exponencialmente no sólo el uso de estos sistemas, sino que le ha conferido a la educación en línea la cualidad de imprescindible. En estos entornos se ha potenciado el uso de elementos tales como entornos personales de aprendizaje, sistemas inteligentes de tutoría, sistemas de gestión de aprendizaje, así como también el aumento de la interacción estudiante - docente a través de blogs, wikis, redes sociales entre otros, generando así una variada y extensa cantidad de información. La situación actual nos plantea el reto y oportunidad de aportar en el desarrollo de herramientas que permitan fortalecer el sistema de educación en línea. Esta es una responsabilidad de todos quienes estamos inmersos en el ámbito de la educación.

Universidad de Alicante

Abstract

Every day in the world an immense amount of data is generated as a result of our permanent and growing interaction with technology, whether for work, academic, personal or leisure activities. Digital businesses, sending and receiving emails, interaction with financial platforms, interaction with social networks, interaction with educational platforms, use of virtual maps, etc. These are just some examples of the actions that we execute on a daily basis and that generate a gigantic and varied amount of data that can be analyzed. In the coming years, this trend will accelerate due to the increase in devices and sensors connected to the internet.

In the educational field, the increase in the use of online learning systems, such as personal learning environments, intelligent tutoring systems, learning management systems (LMS). As well as the increase in student - teacher interaction through blogs, wikis, social networks, among others, generates a varied and extensive amount of information. This information, stored in institutional databases, is being underused by students, teachers and educational administrators, who use the digital platforms simply as repositories of information.

In recent years, a significant number of investigations, both theoretical and practical, have been evidenced in scientific databases, focusing on the field of data mining in educational settings and specifically in the field of higher education. The organization and analysis of this giant volume of data has two focus possibilities, educational data mining and learning analytics. The first develops and adapts statistical, data mining, and machine learning methods to analyze the data generated by students and teachers. On the other hand, learning analytics is defined as the process of measuring, collecting, analyzing and presenting data related to the interaction of students with digital platforms. Learning analytics aims to provide information that optimizes the achievement of learning results in the environment in which it occurs.

Traditional algorithms for data mining in educational settings cannot be applied

Universitat d'Alacant
Universidad de Alicante

without prior analysis of the institutional strategies in which it will be applied since higher education institutions exhibit different behaviors. For example, one educational model in an institution may be focused on practice-based teaching and innovation, while another model may emphasize action research. Under this premise it is important to have a clear vision of the following three elements for the application of data mining techniques and learning analytics a) Institutional strategies in which educational data mining methods and learning analytics are applied, b) Methods of data mining applied in educational settings and c) Tools for the implementation of data mining in educational settings. This thesis presents the proposal of an architecture based on data mining tools to evaluate teaching-learning processes in the field of engineering education. This proposal is based on the three elements mentioned above and on which the objectives and scientific articles included in the compendium revolve.

At the time of writing this summary, I was relatively clear about the importance of online education and the analysis of data generated in this field. The current situation of pandemic and confinement has exponentially increased not only the use of these systems but has also given online education the quality of being essential. In these environments, the use of tools such as personal learning environments, intelligent tutoring systems, learning management systems has been promoted, as well as the increase in student-teacher interaction through blogs, wikis, social networks, among others, thus generating a varied and extensive amount of information. The current situation presents us with the challenge and opportunity to contribute to the development of new tools that make it possible to strengthen the online education system. This is a responsibility of all of us who are immersed in the field of education.

Universitat d'Alacant
Universidad de Alicante

Glosario

Algoritmo Es un conjunto ordenado de instrucciones diseñadas para realizar una tarea específica orientada a hallar una solución a un problema. El proceso puede ser simple, como multiplicar dos números, o una operación compleja, cuando se refiere a instrucciones de inteligencia artificial.

Big Data Es un término que describe el gran volumen de datos que inundan a las organizaciones día a día. Sin embargo, el concepto realmente importante que subyace tras el término Big Data es el análisis que se pueda dar a esta cantidad y variedad de datos existentes. Esto con el objetivo de identificar información valiosa que conduzca a las organizaciones a tomar mejores decisiones estratégicas.

Calidad académica Se concibe como todas aquellas acciones que potencian el desarrollo de las capacidades académicas y sociales de la comunidad educativa. La calidad académica en una institución educativa promueve el desarrollo profesional de los docentes e influye directamente con su oferta educativa.

DM (Data Mining) Se define como el proceso de descubrir y extraer patrones ocultos en grandes volúmenes de datos, estos hallazgos pueden utilizarse para predecir comportamientos futuros en diferentes entornos. Para cumplir con este propósito se utilizan herramientas estadísticas y computacionales como redes neuronales y aprendizaje automático.

Deserción universitaria Se entiende como el abandono de estudiantes del sistema educativo, esta deserción se puede dar por diferentes factores que se generan tanto al interior del sistema educativo como en otros contextos a los que naturalmente pertenece el estudiante.

EDM (Educational Data Mining) Es una disciplina emergente en el campo de la minería de datos. Se orienta a la exploración de información en entornos educativos, para posteriormente, con la ayuda de técnicas, modelos y algoritmos, identificar patrones descriptivos e incluso realizar predicciones, con el objetivo final

Universitat d'Alacant
Universidad de Alicante

de aportar con información relevante para mejorar los procesos de enseñanza - aprendizaje.

EPM (Educational Process Mining) Es un campo emergente en la minería de datos educativos su objetivo es evidenciar el conocimiento no expresado de los procesos educativos llevados a cabo a través de plataformas tecnológicas. La EPM utiliza específicamente los datos de los registros recopilados específicamente de entornos educativos virtuales para descubrir, analizar y proporcionar una representación visual del proceso educativo completo.

IES (Instituciones de Educación Superior) Acrónimo utilizado ampliamente en contextos de educación superior para describir de forma general a todas las instituciones que forman parte del ecosistema de educación superior.

LA (Learning Analytics) Es una disciplina emergente encaminada a la medición, recopilación, análisis y presentación de datos sobre los estudiantes y sus contextos, con el fin de comprender y optimizar su aprendizaje y los entornos en los que ocurre.

LMS (Learning Management System) Son plataformas informáticas que se utilizan para administrar, distribuir y controlar las actividades académicas de instituciones educativas que tienen una modalidad de estudios en línea o híbrida. A través de este tipo de plataformas se fomenta una relación asincrónica entre estudiantes y docentes.

Minería de textos Es una forma específica de minería de datos. Su objetivo es extraer información útil a partir de diferentes tipos de formatos de documentos, tales como blogs, redes sociales, correos electrónicos, artículos de revistas, encuestas, etc. Esto se logra a través de la identificación de patrones dentro de los documentos, tales como tendencias en el uso de palabras, estructura sintáctica, estructura semántica,etc.

Universitat d'Alacant
Universidad de Alicante

Modelo educativo Es el conjunto de premisas, metodologías y conceptos que definen la forma en que se realiza el proceso de enseñanza - aprendizaje en una institución educativa determinada. El modelo educativo apunta a la obtención de una mejora continua en el aprendizaje de los educandos, y de esta manera impactar de forma positiva en la sociedad.

MOOC (Massive Open Online Course) Este tipo de cursos representan una opción de aprendizaje flexible donde los estudiante pueden acceder desde cualquier lugar y momento, además se caracterizan por ofrecer un sistema de aprendizaje acorde al ritmo de estudio de los participantes.

Sistemas de aprendizaje en línea Son sistemas de aprendizaje mediados por el uso de diversas herramientas tecnológicas. Entre estas herramientas se pueden destacar los sistemas de gestión de aprendizajes, los sistemas de gestión académica, las herramientas para evaluación del aprendizaje, los entornos personales de aprendizaje, los sistemas inteligentes de tutoría, entre otros.

TIC (Tecnologías de la información y la comunicación) Son un conjunto de recursos y herramientas de tipo tecnológico y comunicacional que han potenciado la gestión y distribución de la información. Se sustentan en la informática, la microelectrónica y las telecomunicaciones.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Índice general

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Índice de figuras

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Índice de tablas

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Parte I

SÍNTESIS

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

1 Introducción

1.1. Motivación

En las últimas décadas, el potencial de la minería de datos y la analítica consideradas como técnicas y metodologías que extraen información valiosa y procesable de grandes conjuntos de datos, ha transformado la gestión de la información y ha amplificado la investigación en estos campos (Baker y Inventado, 2014). El creciente uso de las tecnologías de la información y la comunicación (TIC) en entornos educativos y en particular en las instituciones de educación superior (IES) ha potenciado la generación de proyectos de innovación educativa (Johnson y otros, 2016) provocando que en la actualidad los estudiantes generen datos a una velocidad mucho mayor que hace unos pocos años (Romero, Ventura, y García, 2008). El registro histórico de calificaciones de los estudiantes, los resultados de pruebas y exámenes, los datos de interacción permanente con sistemas de gestión del aprendizaje (LMS, acrónimo en inglés de *Learning Management System*)[1], la información de las coordenadas de ubicación geográfica registrada en la matrícula estudiantil, el registro de asistencia, etc. son solo algunos ejemplos del volumen y variedad de información que dejan los estudiantes en su día a día (Beck, Chang, Mostow, y Corbett, 2008). La mayoría de estos datos se almacenan únicamente con fines administrativos, muchas veces para no volverlos a utilizar en ninguna otra ocasión. Sin embargo, estos datos pueden mejorar significativamente el proceso de toma de decisiones en las IES (Maldonado-Mahauad y otros, 2018) esto con el objetivo final de potenciar la calidad académica en los programas de estudio.

Los datos generados en las IES son almacenados en diferentes formatos y en variados tipos de repositorios tales como registros de LMS, blogs, bases de datos, documentos digitales, redes sociales, imágenes, videos, audios, metadatos, hipervínculos, etc. La cantidad de datos disponibles en estos repositorios es cada vez más grande y variada, lo cual implica que su procesamiento a través de técnicas basadas en estadística tradicional resulte insuficiente (Sin y Muthu, 2015). En muchas ocasiones, al no disponer

[1] Los acrónimos en inglés utilizarán el siguiente formato: minería de datos educativos (EDM, *educational data mining.*)

de técnicas y herramientas adecuadas para procesar este volumen de datos, se corre el riesgo de que toda la información almacenada se desperdicie o sub utilice. Es decir, se pierde la oportunidad de tomar decisiones estratégicas a partir de indicadores objetivos. Por lo tanto, estos datos requieren la aplicación de métodos o técnicas apropiadas para procesarlos y extraer de ellos conocimiento. En el campo educativo, estas técnicas se clasifican en lo que se conoce como minería de datos educativos (EDM, *Educational Data Mining*), analítica de aprendizaje (LA, *Learning Analytics*) y el descubrimiento de conocimiento en bases de datos (KDD, *Knowledge Discovery in Databases*) (Buenaño-Fernández y Luján-Mora, 2016). La International Educational Data Mining Society[2] define EDM como: *"an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students, and the settings which they learn in"* (Es una disciplina emergente, preocupada por desarrollar métodos para explorar los tipos únicos de datos que provienen de entornos educativos y usar esos métodos para comprender de mejor manera a estudiantes y sus entornos de aprendizaje). Por otro lado, se han identificado varias definiciones de LA en las que se evidencia que no todos los autores están de acuerdo con ellas (Siemens, 2012). Un enfoque ampliamente aceptado es el propuesto por la Society for Learning Analytics Research[3] que define a LA como *"the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs"* (La medición, recopilación, análisis y presentación de datos sobre los estudiantes y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que ocurre). El propósito final de LA es contribuir a la mejora del aprendizaje a través de la interpretación y contextualización de los datos educativos (Ferguson, 2012). La función principal de estas técnicas es la aplicación de varios métodos, herramientas y algoritmos que permitan al usuario descubrir y extraer patrones en los datos almacenados (Sin y Muthu, 2015).

La aplicación de la minería de datos (DM, *Data Mining*) en contextos educativos debe tomar en cuenta algunos elementos característicos que la hacen diferente en relación a su aplicación en otras áreas (Baker y Inventado, 2014), estos elementos son: los objetivos, los datos y las técnicas (Siemens, 2012). El objetivo de la DM en cada área de aplicación es diferente. Por ejemplo, en el área de los negocios, el objetivo principal de la DM es aumentar las ganancias de una empresa, lo que es tangible y puede medirse en términos monetarios y está relacionado directamente con en el número o con la eficiencia de los modelos de ventas. Sin embargo, en la EDM se presentan objetivos tales como mejorar y guiar el proceso de aprendizaje de los estudiantes, lograr una comprensión más profunda de los procesos educativos e identificar los modelos de enseñanza - aprendizaje más eficientes de acuerdo al perfil del estudiante. En algunas ocasiones estos objetivos son difíciles de cuantificar y requieren su propio conjunto de técnicas de medición que se adapten al modelo educativo que maneja cada institución. En lo referente a los datos, en los entornos educativos existe, además de un gran volumen una muy amplia variedad de datos que se encuentran disponibles para llevar a cabo el proceso de DM. Estos datos son específicos del área educativa y, por lo tanto,

[2] http://www.educationaldatamining.org
[3] https://www.solaresearch.org

tienen información semántica intrínseca, relaciones con otros datos y múltiples niveles de jerarquía significativa (Baker y Inventado, 2014). Además, también es necesario tener en cuenta aspectos pedagógicos implícitos en los sistemas educativos. En relación a las técnicas, los datos educativos tienen características especiales que requieren un procesamiento diferente. Aunque la mayoría de las técnicas tradicionales de la DM se pueden aplicar directamente, otras no pueden y deben adaptarse al problema educativo específico en cuestión (Buenaño-Fernández, Villegas-Ch, y Luján-Mora, 2019). La aplicación y usos de la DM en entornos educativos tiene una connotación particular de acuerdo a los objetivos que persiguen los diferentes agentes involucrados en el proceso educativo, tal como se puede ver en la Tabla 1.1.

Finalmente, es importante indicar que en la situación de excepcionalidad en la que vivimos producto de una pandemia que no conoce precedentes, el análisis que nos corresponde hacer de los datos contextuales es fundamental porque posiblemente nos encontremos ante uno de los mayores cambios en la historia de la educación. Este cambio supone un reto y la oportunidad para construir un entorno mejor en la educación en línea. Tenemos que ser capaces de analizar muchos factores para extraer todos los beneficios de la situación actual. Por ejemplo hay una evidencia en la mejora en la sostenibilidad (se ha visto como ha mejorado la ecología y el medio ambiente en las ciudades disminuyendo drásticamente la contaminación durante el confinamiento) y también en la productividad, ya que estamos siendo capaces de sacar tareas desde casa muchas veces en una cantidad menor de tiempo. En el ámbito educativo nos va a permitir realizar seguimientos no presenciales pero sí completamente dirigidos al estudiante con una cantidad ilimitada de recursos tecnológicos que enriquecerán los entornos de aprendizaje virtuales (Viner y otros, 2020).

1.2. Definición del problema

En las últimas décadas, la educación superior ha incrementado su oferta académica gracias a la implementación de modelos educativos en línea y modelos híbridos en los que predomina el uso de plataformas interactivas de aprendizaje. Estos nuevos esquemas están ocurriendo paralelamente a la llegada e incorporación de tecnologías innovadoras tales como *Facial recognition tech, SmartCity, Big Data, Internet of Things (IoT), Cloud y Open Data, etc.* Estas tecnologías están impactando en todas las facetas de nuestras vidas, desde cómo generamos energía hasta cómo recogemos, almacenamos y presentamos la información. Existen además muchos datos que apuntan a que posiblemente la situación actual de la covid-19 conduzca hacia situaciones en los que predominen los modelos educativos híbridos. Las instituciones de educación superior no pueden ni deben quedarse atrás en el uso y desarrollo de estas nuevas tecnologías. La inclusión de estas tecnologías en entornos universitarios ocasiona que día a día, estudiantes, docentes y administradores educativos generen un continuo e ingente volumen de información como resultado de su interacción permanente con estas tecnologías. Este enorme volumen de información proveniente de diversas fuentes (alta variedad) no puede ser procesada a través de métodos tradicionales. Esto significa que se está dejando pasar la oportunidad de analizar estos datos para tomar acciones de mejora sobre los procesos de enseñanza - aprendizaje.

Interesados / Usuarios / Involucrados	**Objetivos de la DM en entornos educativos**
Estudiantes	Personalizar entornos de aprendizaje.
	Recomendar actividades, recursos y tareas de aprendizaje.
	Sugerir experiencias de aprendizaje interesantes.
Docentes / Instructores / Tutores	Predecir el rendimiento estudiantil.
	Analizar modelos de enseñanza - aprendizaje.
	Analizar la conducta y comportamiento de los estudiantes y detectar quienes requieren apoyo.
	Identificar patrones regulares e irregulares de rendimiento estudiantil.
	Mejorar la adaptación y personalización de cursos, y materiales de enseñanza, etc.
Investigadores educativos / Desarrolladores de contenido	Evaluar y mantener el material didáctico.
	Evaluar la estructura y contenido del curso y su efectividad en el proceso de aprendizaje.
	Predecir el rendimiento del estudiante.
	Comparar diferentes técnicas de DM para poder recomendar las que eventualmente sean más útiles para cada tarea.
	Desarrollar herramientas específicas de DM con fines educativos; etc.
Universidades	Mejorar los procesos de decisión.
	Encontrar los mecanismos más óptimos para mejorar la retención estudiantil.
	Ayudar a admitir estudiantes que tendrán buenos resultados en la universidad.
	Mejorar la eficiencia en el proceso de toma de decisiones.

Tabla 1.1: Usuarios versus objetivos de la minería de datos en entornos universitarios

Universitat d'Alacant
Universidad de Alicante

La implementación de estrategias de mejora educativa basadas en técnicas y herramientas de LA y EDM, debe tomar en cuenta parámetros específicos del entorno educativo. Es decir, antes de implementar una solución desarrollada para un contexto específico, es necesario realizar un estudio preliminar para adaptar estos trabajos a la realidad de una región, país, universidad e incluso una determinada facultad. Este requerimiento hace que proyectos específicos como el propuesto en la presente tesis tomen relevancia, ya que se orientan a proponer soluciones particulares en entornos educativos puntuales, en este caso en el ámbito de la educación en ingeniería. Por otro lado, es importante apoyar iniciativas regionales como el proyecto *Learning Analytics Latin America* (LALA), que están trabajando sobre el tema de LA en instituciones de educación superior en Latinoamérica.

1.3. Objetivos

La presente tesis tiene como objetivo general proponer herramientas de DM orientadas a evaluar procesos de enseñanza - aprendizaje en el ámbito de la educación en ingeniería. Los objetivos específicos son:

OE. 1. **Evaluar el estado de la cuestión** en relación a la aplicación de la DM en el ámbito de la educación en ingeniería.

OE. 2. **Analizar enfoques y herramientas** para la aplicación de DM en entornos educativos en ingeniería.

OE. 3. **Evaluar el rendimiento de estudiantes** en el campo de la educación en ingeniería a través de la aplicación de diferentes técnicas y herramientas de DM.

OE. 4. **Evaluar procesos de enseñanza - aprendizaje** en el ámbito de la educación en ingeniería a través de la aplicación de técnicas de minería de textos.

OE. 5. **Proponer herramientas de DM** para evaluar el rendimiento de estudiantes y los procesos de enseñanza - aprendizaje en el ámbito de la educación en ingeniería.

Los objetivos expuestos han sido cubiertos en las diferentes publicaciones presentadas a lo largo del programa de doctorado tanto en revistas indexadas así como en congresos arbitrados por pares. En la Tabla 1.2 se puede observar una alineación entre las publicaciones realizadas y los objetivos planteados. Los artículos dentro de cada objetivo se han ordenado de acuerdo a su cronología de publicación, de la más antigua a la más nueva.

1.4. Método de trabajo

En la primera etapa de este proyecto se trabajó en la revisión de conceptos básicos que sustentan el desarrollo de este trabajo de investigación, además, se trabajó en una

Universitat d'Alacant
Universidad de Alicante

Objetivo	Publicación asociada
OE. 1	The use of tools of data mining to decision making in engineering education—A systematic mapping study
OE. 2	Exploring approaches to Educational Data Mining and Learning Analytics, to measure the level of acquisition of student's Learning Outcomes
	Comparison of applications for educational data mining in Engineering Education
OE. 3	Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses
	A Hybrid Machine Learning Approach for the Prediction of Grades in Computer Engineering Students
	Application of machine learning in predicting performance for computer engineering students: A case study
OE. 4	Improvement of massive open online courses by text mining of students' emails: A case study
	Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach
OE. 5	Improvement of massive open online courses by text mining of students' emails: A case study
	Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses
	A Hybrid Machine Learning Approach for the Prediction of Grades in Computer Engineering Students
	Application of machine learning in predicting performance for computer engineering students: A case study
	Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach

Tabla 1.2: Alineación de objetivos versus publicaciones

Universitat d'Alacant
Universidad de Alicante

revisión sistemática de la literatura (Buenaño-Fernández y otros, 2019). Esta revisión abarcó el período comprendido entre los años 2012 y 2018. La revisión sistemática de la literatura estuvo relacionada con el uso de herramientas de DM en la toma de decisiones en el ámbito de la educación en ingeniería. El trabajo contribuyó a la consecución del OE. 1 (Evaluar el estado de la cuestión). El estudio descrito en esta revisión proporciona a los investigadores una visión general del progreso realizado hasta la fecha en temas de DM en entornos educativos, e identifica áreas en las que falta profundizar la investigación. Para la búsqueda se utilizaron diferentes bases de datos científicas. Un resultado importante de esta investigación fue identificar las principales acciones o estrategias educativas que pueden ser potenciadas con la aplicación de técnicas y métodos de DM.

Tomando como base los resultados obtenidos en la revisión sistemática de literatura, en una segunda fase, se procedió a realizar un análisis de los principales enfoques y herramientas para la aplicación de DM en entornos educativos en ingeniería (OE. 2) (Analizar enfoques y herramientas). En ese sentido, en el artículo (Buenaño-Fernández y Luján-Mora, 2017) desarrollamos un caso de estudio práctico cuyo objetivo fue comparar las características técnicas de tres herramientas de código abierto (RapidMiner[4], Knime[5] y Weka[6]. Estas características fueron evaluadas sobre los registros académicos de tres programas de ingeniería en una universidad ecuatoriana. Las herramientas evaluadas han facilitado la implementación de algoritmos complejos para identificar patrones ocultos de información en bases de datos académicas. Un segundo trabajo que contribuyó a la consecución del OE. 2 (Analizar enfoques y herramientas) fue el artículo (Buenaño-Fernández y Luján-Mora, 2016). En este trabajo se identificaron las principales categorías de aplicación de la EDM y la LA, así como los principales métodos y técnicas usados en entornos educativos. En una tercera fase y teniendo como base la información recabada en el OE. 2 (Analizar enfoques y herramientas) se trabajó en la aplicación de diferentes técnicas y herramientas de MD, con el objetivo de evaluar el rendimiento de estudiantes en el campo de la educación en ingeniería. En el artículo presentado en el número especial "Big Data Research For Social Sciences and Social Impact" de la revista Sustainability (Buenaño-Fernández, Gil, y Luján-Mora, 2019), se aplicaron técnicas de aprendizaje supervisado con el objetivo de predecir las calificaciones finales de los estudiantes en función de su rendimiento histórico de calificaciones. Esta propuesta se aplicó sobre la información académica histórica de estudiantes matriculados en la carrera de ingeniería informática en una universidad ecuatoriana. Este artículo propone una metodología, tal como se puede ver en la Figura 1.1, en la cual inicialmente se ejecuta el proceso de recopilación y preprocesamiento de datos, luego en una segunda etapa, se lleva a cabo la agrupación de estudiantes con patrones similares de rendimiento académico. En la siguiente fase, en función de los patrones identificados, se seleccionó el algoritmo de aprendizaje supervisado más apropiado, y luego se llevó a cabo el proceso experimental. Los resultados mostraron la efectividad de las técnicas de aprendizaje automático en la predicción del rendimiento de los estudiantes.

En la Tabla 1.2, se puede observar que existen dos artículos adicionales que sustentan

[4] http://www.rapidminer.com
[5] http://www.knime.org
[6] http://www.cs.waikato.ac.nz

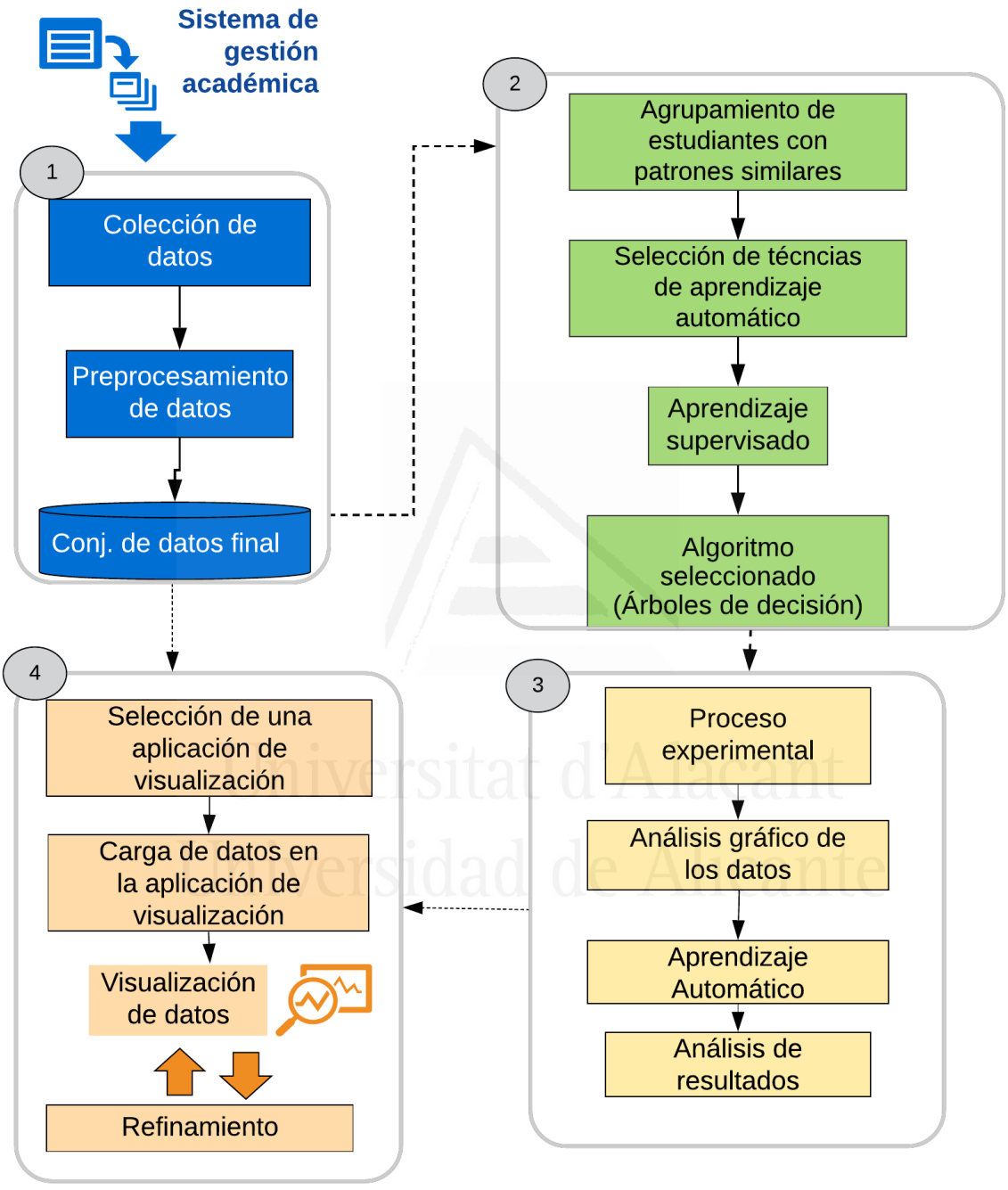


Figura 1.1: Metodología propuesta

el cumplimiento del OE. 3. (Evaluar el rendimiento de estudiantes). El trabajo presentado en (Buenaño-Fernández y Luján-Mora, 2019) describe el uso de herramientas de EDM y de minería de procesos para identificar las rutas de aprendizaje de estudiantes con discapacidad visual. Finalmente se tiene el artículo presentado en el congreso RII-FORUM (Buenaño-Fernández, Luján-Mora, y Gil, 2019). Este documento propone la aplicación de un enfoque híbrido de aprendizaje automático, con el objetivo de sentar las bases para una futura implementación de un sistema de recomendación que permita a los estudiantes tomar decisiones relacionadas con su proceso de aprendizaje. En este artículo se propuso un estudio de caso sobre la información académica de los estudiantes de ingeniería. Los resultados obtenidos en este artículo muestran la efectividad de aplicar un enfoque híbrido de aprendizaje automático. La metodología propuesta en este trabajo se compone, por un lado, de técnicas de aprendizaje supervisado con el objetivo de clasificar los datos en grupos, y por otro lado, teniendo esta clasificación inicial, técnicas de aprendizaje no supervisadas con el objetivo de llevar a cabo un análisis predictivo de los registros de calificaciones históricas de los estudiantes.

La cuarta fase del trabajo se enfoca en el cumplimiento del OE. 4 (Evaluar procesos de enseñanza - aprendizaje), el mismo que se ha plasmado a través de dos artículos tal como se describe en la Tabla 1.2. Como se puede observar los objetivos planteados en esta tesis han crecido en aplicación y alcance. Este objetivo se enfoca en la evaluación de procesos de enseñanza - aprendizaje en el ámbito de la educación en ingeniería a través de la aplicación de técnicas de minería de textos. En el artículo (Buenaño-Fernández, Luján-Mora, y Villegas-Ch, 2017) se explora el uso de técnicas de minería de textos con el fin de evaluar la opinión de mensajes de correo electrónico de cursos en línea masivos y abiertos (MOOC, *Massive Open Online Courses*). La técnica de minería de opinión aplicada sobre correos electrónicos es una tarea compleja debido a la disparidad temática de los correos electrónicos, su tamaño y la profundidad del análisis lingüístico requerido. El propósito de este estudio fue analizar las opiniones de los estudiantes sobre sus cursos, sus instructores y las principales herramientas utilizadas en el curso. La investigación se centró en el cálculo y análisis de la frecuencia de términos, el análisis de concordancias, agrupaciones y n-gramas. El estudio de caso utilizado en este documento fueron correos de un MOOC sobre la temática de desarrollo web con más de 40,000 estudiantes matriculados.

El otro artículo que favorece el logro de este objetivo es el presentado en un número especial "Advanced Data Mining Methods for Social Computing" de la revista IEEE Access (Buenaño-Fernández, Gonzalez, Gil, y Luján-Mora, 2020). Este artículo tiene como objetivo evaluar una metodología genérica basada en el modelado de temas y en el modelado de redes de texto, que permita a los investigadores recopilar información valiosa de encuestas que utilizan preguntas abiertas. Para lograr esto, la metodología propuesta fue validada a través de un estudio de caso en el que se analizaron las respuestas a una encuesta de autoevaluación de docentes en una universidad ecuatoriana. La principal contribución del artículo es la inclusión de algoritmos de agrupamiento para complementar los resultados obtenidos y luego ejecutar el modelado de temas.

Finalmente en la última fase del método de trabajo propuesto, que está relacionada con el OE. 5 (Propone herramientas de DM). Se observa en la Tabla 1.2 que seis de los ocho artículos propuestos en el compendio contribuyen a la consecución de este objetivo.

Universitat d'Alacant
Universidad de Alicante

Básicamente en este fase se resumen todos los aportes realizados en la presente tesis y que han sido descritos en las fases de la uno a la cuatro.

1.5. Trabajo desarrollado

El trabajo desarrollado en la presente tesis se enfoca en la identificación y aplicación de técnicas y herramientas de DM en el ámbito educativo, con el objetivo de reforzar las estrategias de gestión académica que se desarrollan en las IES. Cabe recalcar que el objetivo final de estas estrategias es fortalecer la calidad académica en las instituciones educativas. Los trabajos incluidos en el compendio se enfocan en cuatro estrategias de gestión académica que generalmente son implementadas a nivel institucional:

1. Predecir el rendimiento académico y la deserción estudiantil.
2. Mejorar el proceso de enseñanza - aprendizaje.
3. Evaluar las competencias o resultados de aprendizaje transversales en ingeniería.
4. Mejorar los planes de estudio y procesos educativos.

   Sobre estas estrategias académicas se han aplicado diferentes métodos y herramientas de DM que son comunmente usados en contextos de educación superior.

1.5.1. Estrategias académicas institucionales en las que se aplican métodos de minería de datos educativos y analítica de aprendizaje

En el OE. 1 de la presente tesis se propuso: "Evaluar el estado de la cuestión en relación a la aplicación de la DM en el ámbito de la educación en ingeniería". Este objetivo fue cubierto a través del artículo (Buenaño-Fernández y otros, 2019), en el cual, una de las preguntas de investigación de la revisión sistemática de literatura planteada en este artículo plantea la siguiente interrogante: *"In what areas or tasks of engineering education have decisions been made based on the results of the analysis generated by LA and EDM?"* (¿En qué áreas o tareas de la educación en ingeniería se han tomado decisiones basadas en los resultados del análisis generado por LA y EDM?). Esta pregunta se orienta a identificar las principales estrategias académicas sobre las cuales se han aplicado con mayor intensidad técnicas de LA y EDM. Como resultado de la investigación realizada en (Buenaño-Fernández y otros, 2019), se presentan las siguientes estrategias académicas:

1. Predecir el rendimiento académico y la deserción estudiantil.
2. Mejorar el proceso de enseñanza - aprendizaje.
3. Evaluar las competencias o resultados de aprendizaje transversales en ingeniería.
4. Mejorar los planes de estudio y los indicadores de gestión académica.

A continuación se describen los métodos de LA y EDM aplicados en diferentes escenarios con el objetivo de contribuir con las estrategias académicas.

1.5.1.1. Predecir el rendimiento académico y la deserción estudiantil

Las IES realizan un sinnúmero de estrategias orientadas a predecir el rendimiento académico estudiantil y que puntualmente persiguen dos objetivos. Por un lado, fortalcer los procesos de enseñanza - aprendizaje de manera que se pueda garantizar una educación de calidad. Por otro lado, la crisis económica global ha ocasionado que se reduzcan los presupuestos en la educación superior en todo el mundo. Para solventar el vacío generado entre la disminución de ingresos y el aumento de los gastos, las IES experimentan una enorme presión por aumentar nuevos ingresos de estudiantes y, por otro lado, retener a los estudiantes existentes. Una de las medidas adoptadas por las instituciones es predecir el posible abandono de estudiantes en los primeros años. Las técnicas de EDM y LA pueden ayudar a predecir el posible abandono de los estudiantes a través de la identificación y el análisis de varios parámetros, como por ejemplo el diseño de sistemas de alerta temprana que muestran las calificaciones obtenidas en los cursos de requisito previo, las calificaciones obtenidas en pruebas, exámenes y tareas anteriores, la participación de los estudiantes en actividades, etc. Con el objetivo de predecir el abandono de los estudiantes en los primeros años de estudio se han realizado varias investigaciones y aplicaciones en los que se han combinado diferentes técnicas de aprendizaje automático.

En un estudio realizado en una universidad al norte de Taiwán (Lu y otros, 2018), se aplicaron los enfoques de LA y *big data* educativo con el objetivo de realizar la predicción temprana del rendimiento académico final de los estudiantes en un curso de la asignatura de cálculo. Este estudio aplicó el método de regresión del componente principal para predecir el rendimiento académico final de los estudiantes. En este trabajo, se incluyeron variables externas al curso, como los resultados de visualizar material multimedia fuera de clases, las prácticas fuera de horarios de clase, las calificaciones de tareas y exámenes, y las tutorías realizadas luego de los horarios de clase. En otra investigación (Polyzou y Karypis, 2016), los autores proponen el desarrollo de métodos que utilizan conjuntos de datos históricos de las calificaciones de los estudiantes por cursos, con el objetivo de estimar el rendimiento de los estudiantes. Su propuesta se basó en el uso de modelos lineales dispersos y factorizaciones matriciales de bajo rango. El trabajo evaluó el rendimiento de las técnicas propuestas en un conjunto de datos obtenidos de la Universidad de Minnesota que contenía calificaciones históricas de un período de 12 años. Este trabajo demostró que centrarse en los datos específicos del curso mejora la precisión de la predicción de calificaciones.

En el artículo (Buenaño-Fernández, Gil, y Luján-Mora, 2019), que ha sido incluido en este compendio en el capítulo 9, se trabajó en la aplicación de técnicas de aprendizaje automático para predecir las calificaciones finales de los estudiantes en función de su rendimiento histórico de calificaciones. La metodología usada se basó en la agrupación de estudiantes con patrones similares de rendimiento académico. Una tarea inicial del proceso de análisis de datos fue observar gráficamente, Figura 1.2, el comportamiento académico de los estudiantes en términos de sus notas, esto se lo realizó analizando el gráfico de los componentes de calificaciones parciales (PG1, PG2, PG3) versus el componente de calificación total (FG). En la Figura 1.2 se pueden observar tendencias generales de las calificaciones (picos resaltados en círculos rojos) por áreas de conocimiento. Estos picos evidencian que los estudiantes hacen mayor esfuerzo en sus

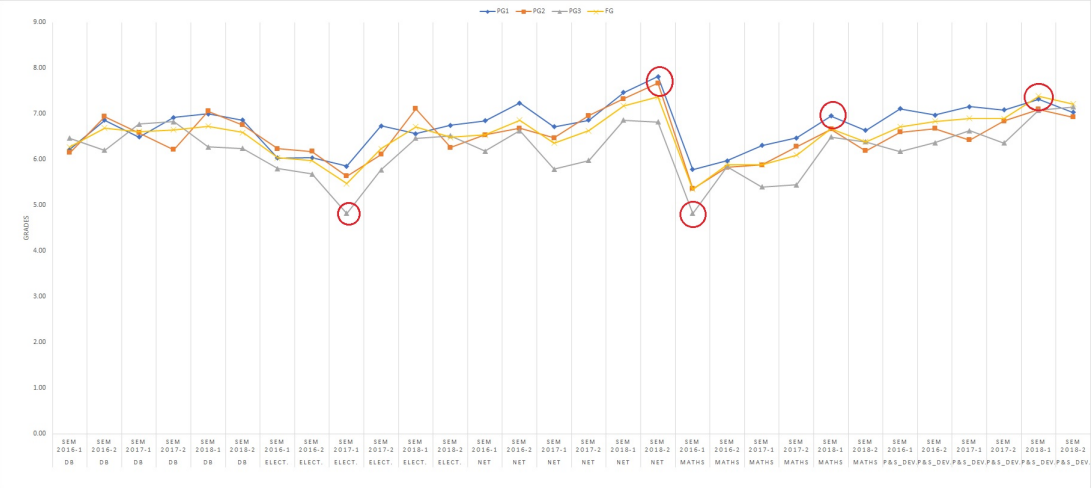


Figura 1.2: Tendencia de las calificaciones de los estudiantes en los períodos 2016 - 2017

primeras pruebas y luego sus calificaciones se deterioren a medida que avanza el curso. Este tipo de análisis inicial es importante ya que permite al investigador conocer de manera global el comportamiento estudiantil en periodos académicos específicos.

En el trabajo presentado por Buenaño-Fernández y Luján-Mora (2019) que está incluido en la presente tesis en el capítulo 7 y que participa específicamente en los OE. 3 (Evaluar el rendimiento de estudiantes) y OE. 5 (Propone herramientas de DM), se propuso la aplicación de mineria de procesos en el ámbito educativo (EPM, *Educational Process Mining*), con el objetivo de detectar puntos potencialmente peligrosos en las rutas de aprendizaje que siguen estudiantes con discapacidad visual. La EPM recibe como entrada los datos de los archivos de base de datos, archivos de registro o archivos de pistas de auditoría almacenados en plataformas educativas. Estos archivos contienen una secuencia de eventos producto de la interacción de estudiantes con las plataformas virtuales de aprendizaje. En la Figura 1.3 se puede observar el proceso a seguir para ejecutar una EPM. El objetivo de la EPM es descubrir, analizar y proporcionar una representación visual del proceso de aprendizaje completo de un estudiante (Buenaño-Fernández y Luján-Mora, 2019).

Por otro lado, a través de la aplicación de herramientas de LA es posible capturar, con diferentes niveles de granularidad, los rastros que dejan los estudiantes al interactuar con plataformas digitales de aprendizaje. Estos eventos van desde acciones de bajo nivel, como pulsaciones de teclas y clics del ratón, hasta eventos de mayor nivel, relacionados las actividades de aprendizaje de los estudiantes. La creciente investigación en analítica de aprendizaje ha generado un importante espacio para el desarrollo de herramientas de análisis de datos educativos. En este contexto toman fuerza las especificaciones técnicas y estándares para tecnologías de aprendizaje tipo xAPI o SCORM. Estas especificaciones permiten la creación de objetos pedagógicos estructurados, que cumple con los requerimientos de: adaptabilidad, accesibilidad, interoperabilidad y reusabilidad (Buenaño-Fernández y Luján-Mora, 2019). Con la implementación de las directrices propuestas en la especificación xAPI se podrá entender el significado de los datos almacenados en el repositorio de aprendizaje, conocido como repositorio

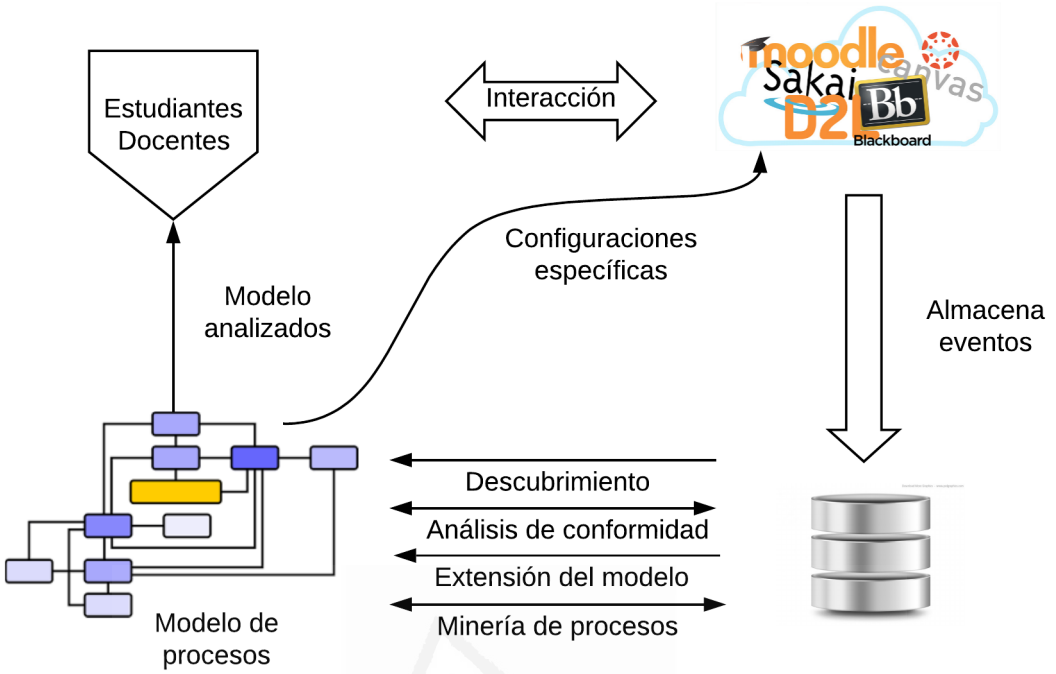


Figura 1.3: Esquema de trabajo para ejecutar minería de procesos educativos

de registros de aprendizaje (LRS, *Learning Record Store*). A través del análisis de los datos almacenados en el LRS de xAPI se pueden identificar las rutas de aprendizaje que siguen los estudiantes al navegar por un curso en línea. Además, se pueden identificar los posibles puntos críticos de aprendizaje que se generan en dichas rutas.

1.5.1.2. Mejorar el proceso de enseñanza - aprendizaje

Las estrategias de enseñanza - aprendizaje son acciones planificadas que se deben tomar sobre un curso, o sobre el plan de estudios completo de un programa de estudio. La definición de las estrategias de enseñanza - aprendizaje que se planifiquen para un curso debe tomar en cuenta las variables clave del entorno educativo. Estas variables incluyen las características de los estudiantes, los objetivos de aprendizaje y las definiciones de instrucción y evaluación del profesor. Una vez que se han analizado estas variables, se pueden tomar decisiones informadas sobre el contenido del curso, la estructura, los métodos de evaluación y otros componentes clave (Jovanović, Gašević, Dawson, Pardo, y Mirriahi, 2017). Del lado de los estudiantes, los estudios han demostrado que las estrategias de aprendizaje adoptadas por un estudiante están relacionadas con su capacidad de procesamiento de información. Por ende, es fundamental para los docentes identificar las habilidades cognitivas de sus estudiantes. Las estrategias de aprendizaje que utilizan los estudiantes son consideradas como construcciones que no pueden identificarse a simple vista a partir de las evidencias que generan los estudiantes al interactuar con una plataforma virtual. Por lo tanto, necesitan analizarse utilizando métodos y técnicas de analíticas de datos adecuadas. Por ejemplo, Los métodos no

supervisados, como la agrupación o la minería de procesos, han demostrado ser beneficiosos para extraer evidencias observables a partir del estudio de rutas de aprendizaje (Jovanović y otros, 2017). En la presente tesis se trabajó en un estudio de rutas de aprendizaje en el artículo (Buenaño-Fernández y Luján-Mora, 2019) que fue analizado en el capítulo anterior.

El diseño e implementación de nuevos modelos, estrategias y metodologías para mejorar el proceso de enseñanza - aprendizaje ha sido el foco de atención de muchos investigadores en el área educativa. Gracias al esencial aporte de las TIC se han implementado diferentes herramientas en la búsqueda de mejorar los procesos de enseñanza - aprendizaje. Así, por ejemplo, se ha utilizado la realidad virtual y aumentada, la simulación de juegos en procesos de enseñanza, laboratorios virtuales remotos, sólo por citar algunas de ellas. El uso de las TIC en el aula se ha identificado que es un componente que apoya las habilidades didácticas del docente, modificando así la estructura tradicional de enseñanza y contribuyendo a una educación de calidad.

Un aspecto importante del proceso de enseñanza - aprendizaje es la evaluación docente. Esta evaluación es un proceso formal y sistemático que permite medir el rendimiento docente. El establecimiento de estándares de enseñanza en las IES requiere que los docentes se desempeñen efectivamente para cumplir con dichos estándares (Akram y Zepeda, 2015). Por lo tanto, evaluar a los docentes en términos de identificar sus virtudes o defectos es un proceso de vital importancia, y que en cualquier caso debe de entenderse como un elemento de retroalimentación en la mejora de la enseñanza. Se espera que los docentes demuestren altos niveles de habilidades de enseñanza para cumplir con los estándares de responsabilidad requeridos, además de preocuparse profundamente por los estudiantes y su éxito.

En el trabajo presentado por Buenaño-Fernández y otros (2020) que está incluido en la presente tesis en el capítulo 11 y que participa específicamente en los OE. 4 (Evaluar procesos de enseñanza - aprendizaje) y OE. 5 (Propone herramientas de DM), se presentó un estudio de caso en el que se analizaron las respuestas a una pregunta abierta contenida en una encuesta de autoevaluación de docentes en una universidad ecuatoriana. El sistema de evaluación docente implementado en la universidad considerada consta de una serie de componentes: heteroevaluación, coevaluación y autoevaluación. En esta universidad, el proceso de autoevaluación de los docentes se lleva a cabo a través de una encuesta en línea que contiene 12 preguntas abiertas. Las preguntas abiertas proporcionan datos significativos sobre los tipos de encuestas y cuestionarios que se utilizan en los procesos de autoevaluación de los docentes. Tales datos pueden proporcionar a los investigadores información sobre las actitudes y opiniones de los encuestados, que no puede obtenerse fácilmente a partir de datos de preguntas cerradas. Sin embargo, el uso de preguntas abiertas tiene un conjunto asociado de problemas analíticos, particularmente en términos de identificación de temas coherentes que son compatibles con las preguntas planteadas (Ross y Bruce, 2007).

En el documento (Buenaño-Fernández y otros, 2020) se analizaron aproximadamente 900 respuestas de docentes a la pregunta: “Señale cuales estrategias ha adoptado para mejorar la retención de los estudiantes en sus clases sin afectar la calidad académica. Incluya ejemplos específicos respecto a sus estrategias”. La metodología propuesta se ejecutó en cuatro fases:

1. Recopilar datos y validar la base de datos generada.
2. Aplicación de técnicas de minería de textos y modelado de tópicos.
3. Aplicación de técnicas de modelado de redes.
4. Determinar la relevancia de los tópicos identificados.

En esta sección nos interesa analizar la aplicación de esta metodología con el objetivo de apoyar las estrategias académicas que apuntan a mejorar los procesos de enseñanza aprendizaje. En este sentido, luego de aplicar los métodos de modelado de tópicos y modelado de redes, se obtuvo un listado de estrategias aplicadas por los docentes las que se se pueden observar en la Tabla 1.3. De acuerdo a la información que se muestra en la tabla, los tópicos identificados son:

- Tópico 1. Investigación, análisis crítico y lectura.
- Tópico 2. Sin definición.
- Tópico 3. Tutorías.
- Tópico 4. Uso de tecnología.
- Tópico 5. Sin definición.
- Tópico 6. Aprendizaje práctico.
- Tópico 7. Aprendizaje práctico.
- Tópico 8. Estrategias de retención estudiantil.
- Tópico 9. Entorno de enseñanza - aprendizaje.
- Tópico 10. Estrategias de aprendizaje basado en la experiencia.
- Tópico 11. Mecanismos de evaluación.
- Tópico 12. Trabajo en equipo.

En el listado se puede observar que los tópicos 2 y 5, aparecen “sin definición”, esto se debe a que al realizar el análisis de las palabras que se les asignó a cada uno, no se encontró forma de identificarlos ya que las palabras eran muy dispersas. En cambio las palabras asignadas a los tópicos 6 y 7 aportaban al concepto de aprendizaje práctico. La identificación de tópicos basados exclusivamente en herramientas estadísticas y procesos computacionales puede ocasionar que el análisis semántico de un documento se vea afectado. Por lo tanto, los temas identificados por el algoritmo de modelado de tópicos pueden generar errores al analizar contenido semántico de las respuestas de los docentes. Como complemento al análisis automático de tópicos, en la fase inicial del proyecto, se llevó a cabo la lectura de aproximadamente un 10 % de las respuestas de los docentes. Este proceso se llevó a cabo con la participación de un experto en investigación educativa que trabajó en una primera identificación de los principales tópicos que los docentes mencionan al completar la encuesta.

Universitat d'Alacant
Universidad de Alicante

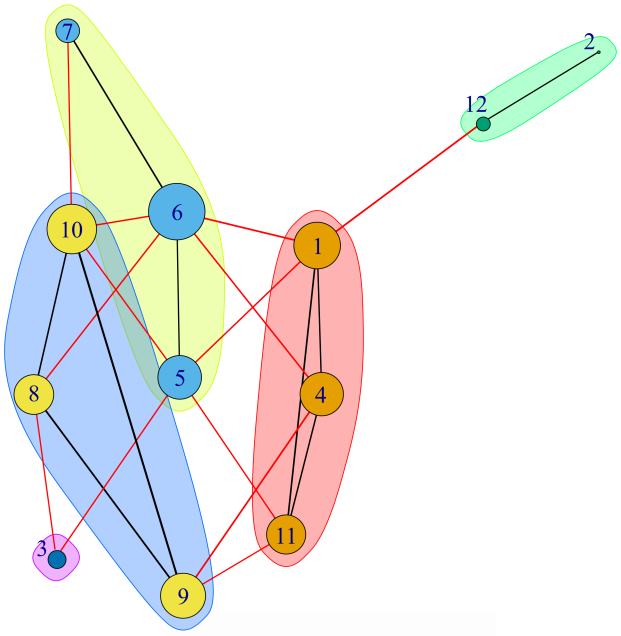


Figura 1.4: Estructura de tópicos utilizando el algoritmo de detección de comunidad de borde intermedio

Para un mejor análisis de los doce tópicos enumerados anteriormente, se los ha agrupado de acuerdo con los clusters observados en el gráfico de la Figura 1.4. La clasificación es la siguiente: cluster (amarillo) **Aprendizaje práctico**, incluye los tópicos 6 y 7; cluster (azul) **Iniciativa del docente**, incluye los tópicos 8, 9 y 10; cluster (rojo) **Uso de herramientas tecnológicas y estrategias de enseñanza tradicionales**, incluye los tópicos 1, 4 y 11, y el cluster (verde) **Estrategia de trabajo en equipo**, incluye únicamente tópico 12.

Los tópicos 6 y 7 se refieren al uso de estrategias que promueven el aprendizaje práctico como el mecanismo más utilizado por los docentes para apoyar la retención de los estudiantes. El aprendizaje práctico se evidencia a través del desarrollo de talleres, estudios de casos, uso de simuladores, ejercicios vinculados a un tema central, gamificación y dinámica para el trabajo interactivo.

Los términos inmersos en el tópico 10 se refieren a estrategias de aprendizaje experiencial, es decir, la inclusión de experiencias del campo profesional para motivar el aprendizaje de los estudiantes. En la Figura 1.4 se puede observar una relación particular del tópico 10 con los tópicos 5, 6 y 7. Esto tiene sentido ya que existe una relación muy estrecha entre el aprendizaje práctico y el aprendizaje experimental. En este mismo grupo se encuentran los tópicos 8 (retención académica del estudiante) y 9 (ambiente de enseñanza - aprendizaje). El tópico 9 (entorno de enseñanza - aprendizaje) agrupa los términos relacionados con la relación empática que debe existir entre docentes y estudiantes, una relación que genera confianza y motivación. Es interesante observar que los tópicos 8, 9 y 10 se relacionan con las estrategias grupales, que no son necesariamente parte de un plan de estudios, sino que son iniciativas particulares de algunos docentes que contribuyen a la relación estudiante-docente y, por lo tanto, mejorar el proceso de enseñanza - aprendizaje.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
1	realizar	aprendizaje	seguimiento	uso	trabajo	caso	conocimiento	estrategia	siempre	profesional	semestre	trabajo
2	preguntas	proyecto	tutorías	virtual	ejemplo	ejercicio	práctica	estudiante	tener	vida	final	actividad
3	taller	proceso	rendimiento	herramienta	estudiante	práctico	aplicación	retención	trato	estrategia	progreso	tutorías
4	lectura	evaluación	problemas	videos	grupos	análisis	metodología	interés	importancia	carrera	tutorías	grupo
5	presentación	docente	académico	participación	diferentes	reales	realiza	mejorar	estudiante	problema	exámenes	grupal
6	contenido	carrera	tutorías	actividad	mismo	conceptos	reforzar	ejemplo	docente	estudiante	evaluación	individuales
7	lecturas	parte	dificultades	tecnología	tiempo	ejemplos	realizar	nivel	actividad	real	nota	práctica
8	debate	enseñanza	tareas	a través	desarrollo	clínicos	parte	calidad	caso	personal	mejorar	personalizadas
9	discusión	retroalimentación	evaluación	material	primer	taller	permitido	académica	confianza	dar	notas	estudiante
10	base	evaluaciones	personales	taller	varios	juegos	mayor	principal	pueden	caso	examen	siguiente
11	participación	resultados	bajo	información	habilidad	estudio	realizan	puede	semestre	ejemplos	pruebas	cuenta
12	investigación	elaboración	horas	activa	técnica	realización	teórico	veces	ambiente	teoría	recuperación	dudas
13	relacionado	necesidades	constante	comunicación	semestre	dinámicas	pacientes	dentro	apoyo	conocer	presentan	equipos
14	a través	ejemplo	entrega	dentro	método	campo	simulación	pregunta	motivación	experiencias	forma	extras
15	ensayos	comprensión	permanente	entrega	investigación	explicación	nuevo	generar	importante	doy	semana	avance
16	exposiciones	virtual	estudiantil	ejercicios	pensamiento	proyecto	aplicar	lograr	atención	mundo	tareas	resultados
17	controles	diseño	tutoría	equipo	dos	estudios	estudiante	presenciales	ayuda	charlas	compañeros	permite
18	luego	pares	mejorar	programacion	fomentar	entender	haciendo	fin	vez	reales	adicionales	presentación
19	discusiones	autónomo	identificar	apoyo	utilización	propias	teóricas	muchas	inicio	adicionalmente	resumen	evaluar
20	información	basado	personalizada	técnica	docente	avances	teóricos	concepto	mantener	luege	tarea	errores

Tabla 1.3: Clasificación de palabras por tópico

Los tópicos 1 (lectura, análisis e investigación), 4 (uso de tecnología) y 11 (mecanismos de evaluación) son parte de un nuevo grupo. El tópico 4 está en el centro del grupo, lo que nos hace pensar que el uso de la tecnología en el aula debería ser un elemento central a la hora de captar la atención de los estudiantes. Por otro lado, los tópicos 1 y 11 deben verse como herramientas útiles para fortalecer las actividades programadas en entornos virtuales. Un especial análisis se debe hacer sobre el tópico 3 ya que en él se reflejan todos los términos relacionados con proceso de tutoría personalizada que realizan los docentes hacia los estudiantes.

Los resultados obtenidos en este estudio nos permiten ver claramente cómo la aplicación de técnicas de DM apoyan sobre manera a mejorar los procesos de enseñanza aprendizaje, en este caso desde el lado del docente.

1.5.1.3. Evaluar las competencias o resultados de aprendizaje del perfil de egreso

El modelo de aprendizaje basado en la adquisición de competencias exige al docente la aplicación de nuevos métodos de enseñanza que se adapten a este nuevo enfoque. Por otro lado, los estudiantes exigen criterios más específicos que les permita conocer cómo serán evaluados en su aprendizaje. En el trabajo propuesto por Menchaca, Guenaga, y Solabarrieta (2016), se definió el concepto de competencia en el área educativa como una combinación de conocimientos, habilidades y destrezas interrelacionadas, que permiten a las personas desempeñarse con éxito en contextos profesionales, educativos y de otro tipo. El concepto de competencia está muy relacionado con el concepto de resultados de aprendizaje, entendiéndose éstos como una declaración muy específica que describe exactamente lo que un estudiante podrá hacer de una manera medible.

El Comité de acreditación de ingeniería y tecnología de los EE.UU. (ABET, *Accreditation Board for Engineering And Technology*) ha definido 11 competencias que un ingeniero debe desarrollar a lo largo de su formación académica. Dichas competencias son, entre otras, la aplicación de conocimientos de matemáticas, ciencias e ingeniería, la identificación y resolución de problemas de ingeniería, la aplicación de herramientas de ingeniería avanzadas, la comunicación efectiva y la capacidad de llevar a cabo el trabajo en equipo. Un reto clave en el campo educativo es evaluar el aprendizaje de los estudiantes a través de medir la adquisición de habilidades prácticas o resultados de

Universitat d'Alacant
Universidad de Alicante

aprendizaje. Lamentablemente no es fácil evaluar la adquisición de estas competencias a través del uso de metodologías tradicionales de enseñanza. En este sentido se deben aplicar métodos de enseñanza fundamentados en la tecnología que fomenten el aprendizaje activo de manera que se puedan generar métricas de evaluación que permitan realizar el seguimiento del proceso de aprendizaje de un estudiante.

En el trabajo desarrollado por Conde, Colomo-Palacios, García-Peñalvo, y Larrucea (2018) se presenta una propuesta detallada que va mas allá del reconocimiento y evaluación de la competencia de trabajo en equipo. Los autores plantean una metodología para la definición de una herramienta que permita además de evaluar la adquisición de la competencia de trabajo en equipo, el desarrollo de funcionalidades para alimentar una ontología de competencias. La investigación se centra en la definición de una herramienta de LA que analice las evidencias de los estudiantes relacionadas con el desarrollo de la competencia.

El artículo presentado por Buenaño-Fernández y Luján-Mora (2016) que se incluye en el capítulo 5 del compendio apunta a la consecución del OE. 2 (Analizar enfoques y herramientas). Se presenta un resumen de los enfoques más relevantes del uso de técnicas y herramientas de EDM y LA, en sistemas de administración de aprendizaje, para medir el nivel de adquisición de resultados de aprendizaje en estudiantes universitarios.

1.5.1.4. Mejorar los planes de estudio y los indicadores de gestión académica

Los comités ejecutivos en las IES se enfrentan a la toma de decisiones complejas que tienen un impacto en el nivel estratégico, así como en los estudiantes, docentes, graduados, empleadores y toda la comunidad académica. Durante el proceso de toma de decisiones, las autoridades confían en herramientas informáticas para apoyar esta tarea. La mayoría de estas herramientas en la actualidad se basan en DM, estadísticas y análisis de redes sociales. Sin embargo, el uso de LA y EDM para apoyar la toma de decisiones en función de indicadores de gestión académica es aún un desafío dentro de la DM. A continuación se describen algunos indicadores de gestión de calidad académica en educación superior:

- Tasas de nueva inscripción estudiantil.
- Retención y deserción estudiantil.
- Tiempo y tasas de titulación estudiantil.
- Empleo e ingresos de los graduados.
- Evaluación de satisfacción estudiantil.

En la presente tesis se han incluido dos artículos (Buenaño-Fernández, Gil, y Luján-Mora, 2019; Buenaño-Fernández, Luján-Mora, y Gil, 2019) en los que se aplicaron técnicas de aprendizaje automático para predecir las calificaciones finales de los estudiantes en función de su rendimiento histórico de calificaciones. En estos trabajos se propuso una metodología en la cual se llevó a cabo inicialmente el proceso de recopilación y preprocesamiento de datos, y luego en una segunda etapa se realizó la agrupación

Universitat d'Alacant
Universidad de Alicante

de estudiantes con patrones similares de rendimiento académico. En la siguiente fase, con base a los patrones identificados, se seleccionó el algoritmo de aprendizaje supervisado más apropiado, y luego se llevó a cabo el proceso experimental. Finalmente, los resultados fueron presentados y analizados. Estos resultados, mostraron la efectividad de las técnicas de aprendizaje automático para predecir el rendimiento académico de los estudiantes.

Por otro lado, en el artículo presentado por Buenaño-Fernández y otros (2020) y que se encuentra incluido en la presente tesis, se aplicó técnicas de minería de textos con el objetivo de identificar las diferentes acciones que los docentes han realizado con el propósito de disminuir la tasa de deserción estudiantil. El listado de acciones identificadas permitirá a los administradores educativos canalizar jornadas de capacitación con el fin de proveer a los docentes todas las herramientas para cumplir con la estrategia académica de disminuir la deserción estudiantil.

1.5.2. Métodos de minería de datos aplicados en entornos educativos

Para cumplir con los objetivos de la DM en entornos educativos es necesario identificar los principales métodos, metodologías, técnicas y herramientas que se están utilizando para este propósito. En la actualidad existe una amplia variedad métodos que apoyan el proceso de DM en diferentes áreas. En las siguientes subsecciones de este apartado se pueden identificar los principales métodos que se aplican en EDM y LA de acuerdo a la información recabada en el estudio de mapeo sistemático de la literatura llevado a cabo durante esta investigación (Buenaño-Fernández y otros, 2019).

1.5.2.1. Clasificación y predicción

Las métodos de clasificación, como árboles de decisión, redes bayesianas, redes neuronales, máquinas de vectores de soporte, etc., se pueden aplicar a los datos educativos con el fin de predecir el rendimiento de los estudiantes, lo cual constituye un hito importante en entornos educativos. El rendimiento académico de un estudiante no es el resultado de un solo factor, depende en gran medida de diversos factores como variables personales, socioeconómicas, psicológicas e inclusive ambientales. Los métodos de clasificación se pueden implementar a través de la aplicación de técnicas de aprendizaje supervisado, en las que se usan observaciones previas para construir un modelo que permita predecir resultados. Un paso previo a la aplicación de un método de clasificación es la división del conjunto total de datos en dos conjuntos más pequeños de datos que se utilizarán con los siguientes objetivos: entrenamiento y prueba. El subconjunto de datos de entrenamiento es utilizado para estimar los parámetros del modelo, mientras que el subconjunto de datos de prueba es empleado para validar el comportamiento del modelo estimado. Cada registro del conjunto de datos debe ubicarse en uno de los dos subconjuntos. Para ello se ha utilizado la técnica de *cross-validation* (Cawley y Talbot, 2003) que permite una generalización de los conjuntos de entrenamiento y prueba. De esta forma se garantiza que son independientes de la partición entre datos de entrenamiento y prueba. El método completo consiste en repetir y calcular los valores de

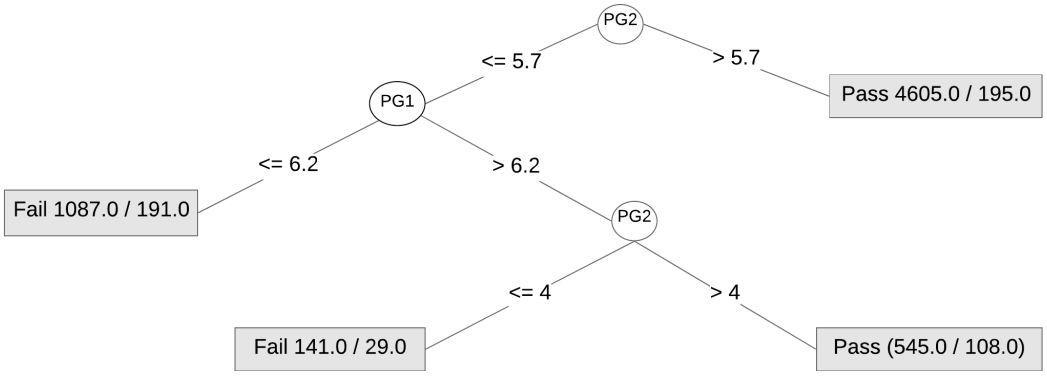


Figura 1.5: Árbol de decisión sin el componente PG3

precisión obtenidos sobre las diferentes particiones que normalmente son igual a 10, lo que lleva a realizar un total de 10 pruebas.

En el ámbito educativo las técnicas de clasificación y predicción se han utilizado para llevar a cabo, entre otras, las siguientes estrategias académicas: identificar grupos de estudiantes con características y reacciones similares frente a una estrategia pedagógica propuesta, predecir el rendimiento de los estudiantes y su calificación final, detectar estudiantes que hacen mal uso de recursos tecnológicos dentro de las IES, identificar a los estudiantes con baja motivación e identificar acciones para reducir las tasas de abandono escolar.

En el trabajo presentado por Buenaño-Fernández, Gil, y Luján-Mora (2019), que es parte del compendio presentado en esta tesis y que ha sido descrito en secciones anteriores, se aplicaron técnicas de aprendizaje supervisado para determinar un modelo predictivo que sentaría las bases para el desarrollo futuro de un sistema de recomendación para los estudiantes. En el proyecto se aplicó el algoritmo de árboles de decisión debido a su popularidad y sobre todo porque produce reglas de clasificación que son fáciles de interpretar en relación con otros métodos de clasificación; esto debido a su representación gráfica que resume un modelo de reglas de decisión implícitas. En la fase experimental del proyecto, se hicieron algunas pruebas eliminando alguno de los componentes de evaluación (PG1, PG2, PG3). Por ejemplo, en una primera prueba se eliminaron las calificaciones del componente PG3 (Figura 1.5) para hacer una predicción de las notas finales (FG). Con esta prueba se esperaba identificar el número de estudiantes que aprobaban las asignaturas sin este componente. Luego se intentó una predicción eliminando el componente PG2 (Figura 1.6). Lo más importante de estas pruebas fue sopesar la importancia de cada PG (PG1, PG2, PG3) en la calificación final (FG). Los resultados que se pudieron observar en estas pruebas mostraron árboles de decisión claros y coherentes.

La matriz de confusión que se presenta en la Figura 1.7 muestra información resumida de los árboles de decisión ejecutados sin los componentes PG3 y PG2 respectivamente. Los valores para la tasa de aprobación y reprobación que se muestran en esta figura evidencian que el modelo está identificando con éxito a los estudiantes que probablemente reprobarán una asignatura.

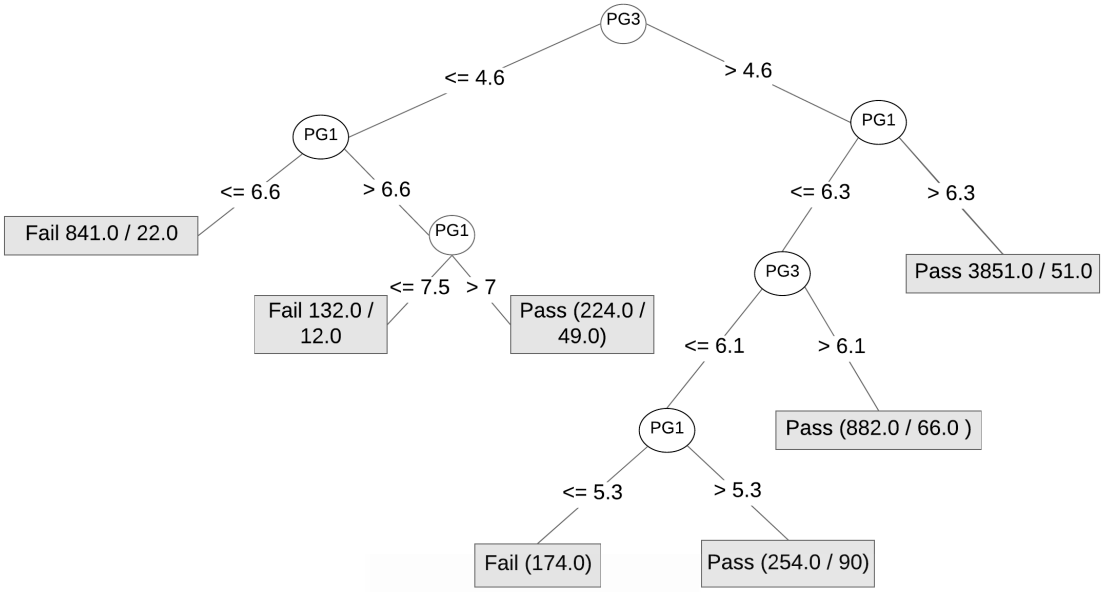


Figura 1.6: Árbol de decisión sin el componente PG2

Correctly Classified Instances	5815	91.5 %
Incorrectly Classified Instances	543	8.5 %
	= = = Confusion Matrix = = =	
a	b	<- classified as
961	330	a = Fail
213	4854	b = Pass

Correctly Classified Instances	5915	93 %
Incorrectly Classified Instances	443	7 %
	= = = Confusion Matrix = = =	
a	b	<- classified as
937	354	a = Fail
89	4978	b = Pass

Figura 1.7: Matriz de confusión para los árboles de decisión (sin PG3 y PG2, respectivamente)

Universitat d'Alacant
Universidad de Alicante

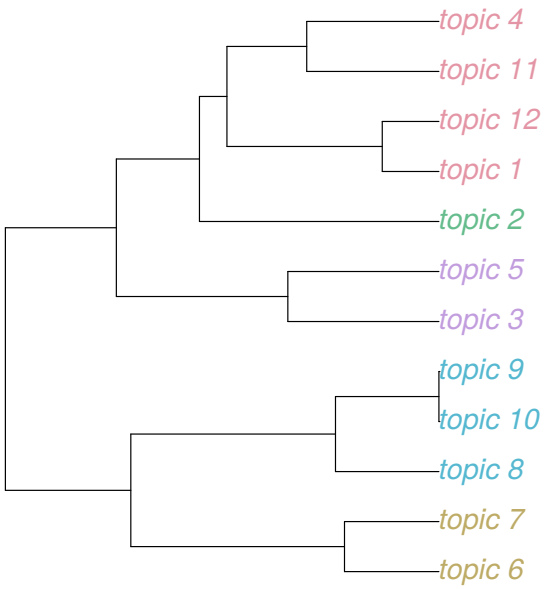


Figura 1.8: Agrupación de red de tópicos a través de un dendograma con 5 grupos

1.5.2.2. Agrupamiento

Las técnicas de agrupamiento hacen referencia al proceso de identificación de objetos que muestren similitud en algunas de sus características, de esta manera un conjunto de datos grande se divide en grupos de datos pequeños que tienen características similares. Los algoritmos de agrupamiento realizan una clasificación o partición no supervisada de patrones (observaciones, vectores de características, etc.) en grupos o subconjuntos (*clusters*), en función de su ubicación y conectividad dentro de un espacio n-dimensional (Romero y otros, 2008). En el campo educativo, las técnicas de agrupamiento se han utilizado para: encontrar grupos de estudiantes con características de aprendizaje similares, para promover el aprendizaje colaborativo grupal, para agrupar materiales similares de un curso o para agrupar a los estudiantes en función de sus patrones de aprendizaje e interacción. La naturaleza variada y voluminosa de los datos generados en entornos de aprendizaje en línea o híbridos ofrece una buena oportunidad para utilizar técnicas de agrupación (Sin y Muthu, 2015). En el trabajo presentado por Buenaño-Fernández y otros (2020) se utilizaron algoritmos de agrupamiento con el objetivo de complementar los resultados obtenidos en el proceso de modelado de tópicos. En la Figura 1.4 se pueden observar cinco grupos marcados con diferentes colores en los que se han agrupado, de acuerdo a sus relación, los diferentes temas identificados en el proceso de minería de tópicos. En la sección 1.5.1.1 se ha detallado la composición de estos grupos. En el artículo también se utilizó la teoría de grafos a través de una red bipartita para descubrir la relación entre los tópicos identificados. La Figura 1.8 muestra una proyección que relaciona los tópicos que son comunes en los documentos, proporcionando así una medida de cuán fuerte es su relación.

1.5.2.3. Detección de valores atípicos

Los métodos de detección de valores atípicos se refieren al proceso de identificación de puntos de datos que son significativamente diferentes al resto de los datos. El método de detección de valores atípicos se puede utilizar para identificar caídas anormales en el rendimiento de los estudiantes o docentes y para identificar a los estudiantes en los extremos del espectro de su rendimiento (Ray y Saeed, 2018).

1.5.2.4. Minería de textos

La minería de textos se refiere al conjunto de métodos utilizados para el análisis de textos no estructurados con el objetivo de derivar información de alta calidad a partir de datos textuales. Existen diversas técnicas de minería de textos que se aplican a problemas del mundo real. En el libro escrito por Weiss, Indurkhya, y Zhang (2010) se proporciona una detallada lista de aplicaciones, entre las que se pueden destacar las siguientes: web semántica, redes sociales, minería de opinión y análisis de sentimiento, métodos de síntesis y de organización, marketing y comercio electrónico, aprendizaje virtual (*e-learning*), aplicaciones de asistencia *help desk*, entre otros. En el área educativa, la minería de textos se ha utilizado para analizar el contenido de foros de discusión, blogs, chats, páginas web, documentos, respuestas de encuestas con preguntas abiertas, etc. (Buenaño-Fernández y otros, 2017).

La ausencia de una metodología genérica para la ejecución de análisis de textos se ha convertido en un gran desafío y una brecha de investigación en el campo de la minería de textos. Esto es un problema, ya que los modelos de minería de textos utilizados son diferentes para cada caso, debido a que cada área tiene un conjunto de palabras específicas con diferentes significados semánticos. Por ejemplo, el modelo de minería de textos utilizado para analizar mensajes en la red social Twitter es muy diferente del modelo de minería de textos utilizado para analizar las respuestas a preguntas abiertas en una encuesta. En el estudio presentado por Buenaño-Fernández y otros (2020) se proponen cuatro aspectos que deben ser tomados en consideración en cualquier metodología utilizada para la aplicación efectiva del modelado de tópicos:

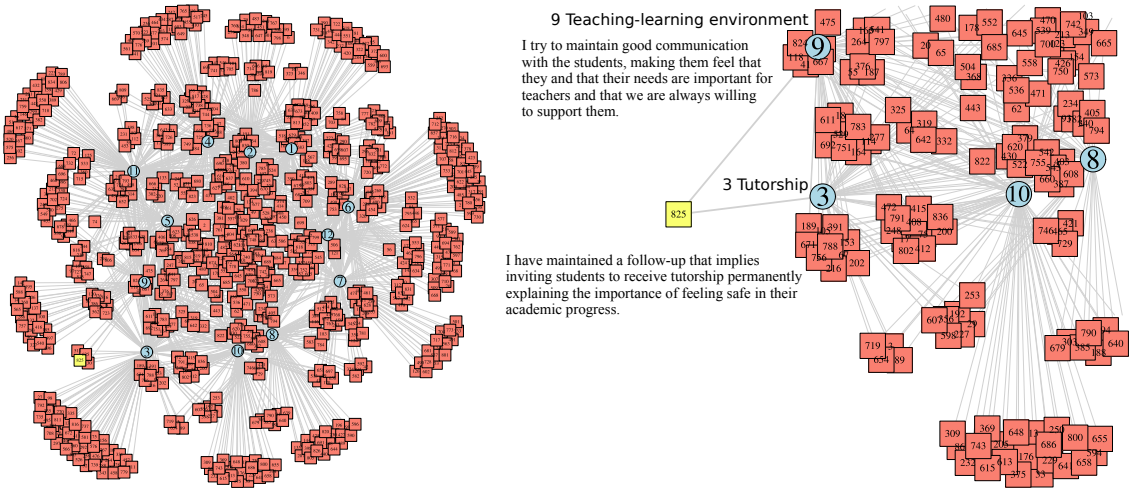
1. Una definición clara del proceso de recopilación y preprocesamiento de la base de datos de texto.

2. La selección correcta de los parámetros del modelado de tópicos.

3. Una evaluación de la fiabilidad del modelo.

4. Una interpretación exhaustiva de los tópicos identificados.

La técnica de modelado de tópicos permite analizar documentos que no están categorizados, por lo tanto, asume que cada documento es una mezcla aleatoria de tópicos. En el proyecto se aplicó el algoritmo *Latent Dirichlet Allocation* (LDA). Este es un algoritmo generativo probabilístico de temas, cuya idea básica es que un documento se compone de una mezcla aleatoria de temas latentes. Es decir, cada documento tiene una probabilidad de pertenecer a cada tema. Esto se puede interpretar como un gráfico



(a) (b)

Figura 1.9: (a) Representación de la red bipartita que conecta documentos (cuadrados rojos) con tópicos (círculos azules). (b) Parte ampliada de la red, incluida la respuesta del profesor (825) y la interrelación con los tópicos 3 y 9.

bipartito en el que cada documento está conectado a un número determinado de temas. Tales relaciones se presentan en la Figura 1.9.

Una importante contribución de la metodología propuesta en este artículo fue la inclusión de algoritmos de modelado de redes de texto que, junto con las contribuciones del algoritmo de modelado de tópicos LDA, proporcionaron resultados relevantes para el estudio de caso propuesto, que consistió en analizar las respuestas de una encuesta de autoevaluación docente que utiliza preguntas abiertas. El modelado de tópicos tiende a centrarse en la frecuencia de los términos, mientras que el análisis de la red de texto toma en cuenta tanto la estructura del texto como la secuencia de las palabras.

Universidad de Alicante

1.5.2.5. Minería de procesos

En el ámbito educativo, la minería de procesos tiene como meta principal extraer información oculta a partir de los registros de los eventos almacenados en entornos virtuales de aprendizaje. La minería de procesos proporciona una representación visual de dichos eventos. La EPM propone la construcción de modelos de procesos educativos completos y compactos a través de los cuales se pueda reproducir todo el comportamiento académico de un estudiante. A través de la EPM, se puede verificar si el comportamiento académico planificado coincide con el comportamiento observado (Buenaño-Fernández y Luján-Mora, 2019). La aplicación de la minería de procesos en el área educativa se ha extendido hacia diferentes ámbitos. En la investigación de Bogarín, Cerezo, y Romero (2017) se ha propuesto una clasificación de estos aportes en las siguientes áreas: aprendizaje colaborativo soportado por computadora, cursos de entrenamiento profesional, minería de currículum, registro de estudiantes, evaluaciones basadas en computador, entre otros. En el trabajo realizado por Buenaño-Fernández

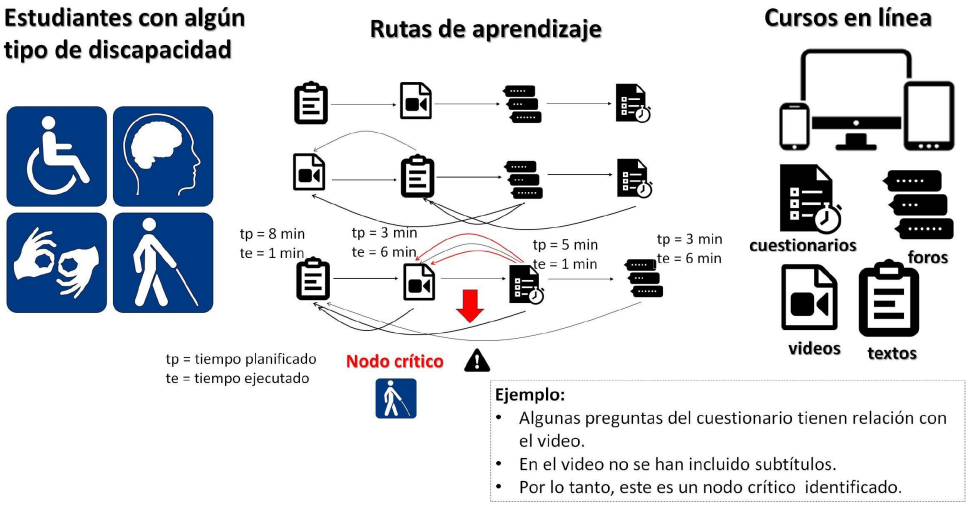


Figura 1.10: Esquema para la identificación de puntos críticos en rutas de aprendizaje de personas con discapacidad

y Luján-Mora (2019) se propuso la aplicación de EPM orientada a la identificación de puntos críticos en las rutas de aprendizaje seguidas por personas con discapacidad visual, en la Figura 1.10 se puede ver el esquema propuesto. En los cursos que se ejecutan en modalidad virtual, el oferente pone a disposición del estudiante un conjunto variado de recursos y actividades de aprendizaje (cuestionarios, foros, videos, textos, entre otros). Sin embargo, en muchas ocasiones estos recursos no contienen todos los criterios de accesibilidad que requieren las personas con algún tipo de discapacidad. En el proyecto se propuso la creación de dos cursos. El primer curso incluía en los recursos de aprendizaje los criterios de accesibilidad para personas con discapacidad auditiva, mientras que en el segundo no se incluían dichos criterios. El propósito de este esquema fue identificar cómo afecta la no inclusión de criterios de accesibilidad en el cumplimiento de las actividades de aprendizajes propuestas en el curso. Para capturar las trazas de navegación o actividades de aprendizaje del curso, se planteó utilizar herramientas que permitan recopilar las acciones de navegación web que realizan los estudiantes. Una vez almacenados los datos, se propuso la aplicación de técnicas de minería de procesos. El objetivo de aplicar estas técnicas es descubrir, analizar y proporcionar una representación visual del proceso educativo completo.

1.5.3. Herramientas para la implementación de minería de datos en entornos educativos

En el trabajo presentado por Buenaño-Fernández y otros (2019) se encontró una amplia gama de herramientas de DM desarrolladas específicamente para entornos educativos, a continuación se especifican algunas:

- **LOCO-Analyst**[7] Es una herramienta educativa que proporciona a los docentes retroalimentación sobre las actividades que realicen los estudiantes durante su proceso de aprendizaje en entornos basados en la web. Por lo tanto, les ayuda a mejorar el contenido y la estructura de sus cursos en línea. La herramienta tiene como objetivo proporcionar a los docentes comentarios relacionados con: todo tipo de actividades que sus alumnos realizaron durante el proceso de aprendizaje; el uso y la comprensión del contenido de aprendizaje que habían preparado e implementado; las interacciones sociales contextualizadas entre estudiantes (es decir, redes sociales) en el entorno de aprendizaje virtual.

- **GISMO**[8] Es una extensión (*plugin* para usarla en Moodle disponible en la base de datos de extensiones de la misma plataforma educativa. Es una herramienta interactiva de monitoreo y seguimiento de los estudiantes. Extrae datos de un curso en línea en Moodle, y genera reportes gráficos para el análisis de los profesores. Con GISMO, los instructores pueden examinar varios aspectos de los estudiantes a distancia, como la asistencia a cursos, la lectura de materiales y la presentación de tareas. Además, se puede filtrar por fechas para visualizar rangos temporales específicos.

- **OnTask**[9]. Es una herramienta desarrollada por un equipo de investigadores de la Universidad Edimburgo que ayuda al personal docente a proporcionar comentarios oportunos, personalizados y prácticos para los estudiantes.

- **NoteMyProgress**[10]. Es una herramienta de mentoría diseñada para fomentar el uso de las estrategias de autorregulación de los estudiantes en cursos en línea de forma automática y personalizada.

- **Early Warning Dropout System**[11]. El sistema de alerta temprana de riesgo de abandono es una herramienta desarrollada en el contexto del proyecto LALA con el objetivo de identificar a tiempo posibles casos de deserción estudiantil.

Muchas de estas herramientas han sido desarrolladas a la medida de las necesidades de las IES. Las instituciones educativas que trabajan con estas herramientas han iniciado procesos complejos y rigurosos de evaluación de aprendizajes, análisis de retención estudiantil, rediseño del modelo educativo, evaluación de su planta docente, por mencionar algunas de las medidas importantes que se han tomado (Buenaño-Fernández y Luján-Mora, 2017).

En la línea de minería de procesos existen diversas herramientas de software gratuitas y licenciadas que se pueden utilizar para llevar a cabo tareas relacionadas con este tema. En el trabajo presentado por Buenaño-Fernández y Luján-Mora (2019) se identificaron tres aplicaciones que se usan específicamente en el EPM. La herramienta ProM[12] es un

[7] http://jelenajovanovic.net/LOCO-Analyst/
[8] https://moodle.org/plugins/
[9] https://www.ontasklearning.org/
[10] http://tech4dlearn.com/
[11] https://git.cti.espol.edu.ec/LALA-Project-EN/ESPOL
[12] http://www.promtools.org

marco extensible que admite una amplia variedad de técnicas de minería de procesos en forma de complementos. Es independiente de la plataforma, ya que se implementa en Java y se puede descargar de forma gratuita, por lo tanto la más comúnmente usada en tareas de EPM. El software Disco[13] es una herramienta de propósito general aplicada en el área de minería de procesos, implementada mediante un software de pago que se utiliza con frecuencia en el campo educativo. Finalmente, la herramienta SoftLearn[14] es una aplicación de propósito específico utilizada en tareas relacionadas con minería de procesos en el ámbito educativo.

1.6. Esquema de la tesis

La estructura de la presente tesis consta de tres partes: la introducción y resumen del trabajo realizado en la Parte I; los trabajos publicados que han aportado a los objetivos planteados en la tesis y que son parte de este compendio se presentan en la Parte II; finalmente, en la Parte III se plantean las conclusiones identificadas en este trabajo y además se hace un planteamiento de propuestas de investigaciones futura.

La estructura capitular de la tesis es la siguiente:

El capítulo 1 Introducción, incluye motivación, definición del problema, objetivos, método de trabajo, trabajo desarrollado, esquema de la tesis y finalmente convenciones de escritura.

El capítulo 2 Publicaciones y visibilidad, incluye publicaciones en revistas, congresos y finalmente se expone la visibilidad del autor en redes científicas.

El capítulo 3 Compendio de publicaciones, recoge todos los artículos publicados en el lapso de tiempo de duración del programa de doctorado.

El capítulo 4 The use of tools of data mining to decision making in engineering education—A systematic mapping study, este artículo se publicó en la revista *Computer Applications in Engineering Education Journal*; se incluye referencia, contribución y texto completo.

El capítulo 5 Exploring approaches to educational data mining and learning analytics, to measure the level of acquisition of student's learning outcomes, es un artículo publicado en los *Proceedings of the International Conference on Education and New Learning Technologies* (EDULEARN); se incluye referencia, contribución y texto completo.

El capítulo 6 Comparison of applications for educational data mining in engineering education, es un artículo publicado en los *Proceedings of the IEEE World Engineering Education Conference* (EDUNINE); se incluye referencia, contribución y texto completo.

[13] https://fluxicon.com/disco/
[14] https://tec.citius.usc.es/SoftLearn/

El capítulo 7 Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses, este artículo se publicó en la revista *RISTI - Revista Ibérica de Sistemas y Tecnologías de la Información*; se incluye referencia, contribución y texto completo.

El capítulo 8 A hybrid machine learning approach for the prediction of grades in computer engineering students, es un artículo publicado en los *Proceedings of the International Research & Innovation Forum* (RIIFORUM); se incluye referencia, contribución y texto completo.

El capítulo 9 Application of machine learning in predicting performance for computer engineering students: A case study, este artículo se publicó en la revista *Sustainability*; se incluye referencia, contribución y texto completo.

El capítulo 10 Improvement of massive open online courses by text mining of students' emails: A case study, es un artículo publicado en los *Proceedings of the International Conference on Technological Ecosystems for Enhancing Multiculturality* (TEEM); se incluye referencia, contribución y texto completo.

El capítulo 11 Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach, este artículo se publicó en la revista *IEEE Access: The Multidisciplinary Open Access Journal*; se incluye referencia, contribución y texto completo.

El capítulo 12 Conclusiones, incluye las conclusiones del trabajo y las principales contribuciones al área de estudio.

El capítulo 13 Trabajos futuros, incluye los principales proyectos en los que se está trabajando actualmente y las ideas que el autor tiene para continuar con la linea de investigación desarrollada.

Universidad de Alicante

1.7. Convenciones de escritura

En el desarrollo de este trabajo de doctorado se utilizaron varios acrónimos para hacer referencia a diferentes métodos, técnicas y tecnologías comúnmente utilizadas en la aplicación de tecnología en el ámbito de la educación superior. El formato que se ha utilizado es escribir la primera vez aparece el término su definición y entre paréntesis el acrónimo. Por ejemplo, instituciones de educación superior (IES).

Se ha optado por usar algunos acrónimos en inglés ya que es así como comúnmente se los utiliza en la literatura científica. Por ejemplo, minería de datos educativos (EDM, acrónimo en inglés de *educational data mining*).

Las citas se reproducen en el idioma original de la referencia de donde provienen.

Las Figuras y Tablas presentadas en este documento son de elaboración propia del autor, a menos que se indique lo contrario en el título.

Algunas de las figuras que se incluyen provienen de las publicaciones que conforman el compendio. Por esta razón, varias de ellas están en inglés.

Las palabras en un idioma distinto al castellano se presentan en letra cursiva. Por ejemplo, *English, francais*.

Los textos que se han incluido textualmente de documentos en otro idioma se han escrito en su idioma original en formato de letra cursiva y a continuación se ha colocado entre paréntesis su respectiva traducción al español.

Debido al origen del autor, cuando existan sinónimos se prefieren los vocablos más utilizados en Sudamérica. Por ejemplo, computador por ordenador.

Las cifras numéricas de miles están separadas por coma y las cifras decimales están separadas por punto, siguiendo las normas internacionales.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

2 Publicaciones y visibilidad

2.1. Publicaciones

En este capítulo se presentan las publicaciones científicas realizadas progresivamente entre los años 2016 y 2020 y que han aportado al desarrollo de la presente tesis doctoral. Un total de 15 publicaciones se han presentado tanto en revistas con índice *Journal Citation Reports* (JCR) así como en memorias de congresos arbitrados por pares. De este total, 3 artículos corresponden a revistas con índice JCR y 1 artículo en una revista con índice *Scimago Journal & Country Rank* (SJR). Además, 11 artículos se presentaron en congresos indexados. Del total de artículos publicados en congresos, 4 tienen una relación directa con los objetivos planteados en la presente tesis doctoral y tienen más relación entre sí formando una unidad temática. Por otro, lado los 7 artículos restantes que se publicaron en este período abordan tópicos indirectamente relacionados con la unidad temática de esta investigación. El detalle de las publicaciones es el siguiente:

Universidad de Alicante

2.1.1. Revistas

Esta subsección incluye las publicaciones presentadas en revistas científicas con índice de impacto JCR o SJR. Estas publicaciones forman el eje central del presente trabajo de investigación ya que sus objetivos aportan de forma significativa con los objetivos de esta tesis doctoral. Los detalles generales de las publicaciones se pueden revisar en la Tabla 2.1. En la segunda columna de la tabla se presenta el nombre de la revista con su ISSN; en la tercera columna se muestra el factor de impacto del JCR; la cuarta columna describe el factor de impacto SJR; la última columna presenta la indexación de la revista, ya sea que el artículo haya sido indexado en Scopus (SCO), Web of Science (WOS) o en el Directory of Open Access Journals (DOAJ).

1. "The use of tools of data mining to decision making in engineering education—A systematic mapping study" (Buenaño-Fernández y otros, 2019). Las especifica-

ciones de este artículo se muestran en la Tabla 2.1, en la fila J1. En el capítulo 4 de este documento se detalla el contenido completo del artículo.

2. "Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses" (Buenaño-Fernández y Luján-Mora, 2019). Las especificaciones de este artículo se muestran en la Tabla 2.1 en la fila J2. En el capítulo 7 de este documento se detalla el contenido completo del artículo.

3. "Application of machine learning in predicting performance for computer engineering students: A case study" (Buenaño-Fernández, Gil, y Luján-Mora, 2019). Las especificaciones de este artículo se muestran en la Tabla 2.1, en la fila J3. En el capítulo 9 de este documento se detalla el contenido completo del artículo.

4. "Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach" (Buenaño-Fernández y otros, 2020). Las especificaciones de este artículo se muestran en la Tabla 2.1, en la fila J4. En el capítulo 11 de este documento se detalla el contenido completo del artículo.

Las especificaciones de los artículos enumerados anteriormente se muestran en la Tabla 2.1. El orden en el que aparecen las publicaciones está relacionado con su aporte a los objetivos de la presente tesis y se corresponde con el orden de aparición en el compendio en los siguientes capítulos.

Id	Revista	JCR IF	SJR	Indexado
J1	Computer Applications in Engineering Education. ISSN: 1099-0542. Estados Unidos	1.435	0.395	DOAJ, SCOPUS, WOS
J2	RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao. ISSN: 1646-9895. Portugal	S/N	0.217	DOAJ, SCOPUS
J3	Sustainability. ISSN: 2071-1050. Suiza	2.592	0.549	DOAJ, SCOPUS, WOS
J4	IEEE Access. ISSN: 2169-3536. Estados Unidos	4.098	0.61	DOAJ, SCOPUS, WOS

Tabla 2.1: Descripción las revistas en donde se han publicado los artículos de la presente tesis

2.1.2. Congresos

Los congresos científicos internacionales en los que se presentaron los artículos se detallan en la Tabla 2.2. En la tabla se muestra la información de acuerdo a la siguiente descripción: segunda columna el nombre del congreso; tercera columna la indexación en Scopus o WOS; cuarta columna el país y ciudad donde se realizó el congreso finalmente en la quinta columna, las fechas de realización del congreso. Todos los congresos en los

Id	Congreso	Indexado	País/Ciudad	Fecha
C1	EDULEARN, 8th International Conference on Education and New Learning Technologies	WOS	España, Barcelona	Julio 4-6, 2016
C2	EDUNINE, I IEEE World Engineering Education Conference	SCO, WOS	Brasil, Sao Paulo	Marzo 19-22, 2017
C3	RIIFORUM, The International Research and Innovation Forum	SCO	Italia, Roma	Abril 24-26, 2019
C4	TEEM, 5th International Conference on Technological Ecosystems for Enhancing Multiculturality	SCO	España, Cádiz	Octubre 18-20, 2017

Tabla 2.2: Descripción de las actas de congresos en donde se han publicado los artículos de la presente tesis

Id.	Perfil académico	URL
P1	ORCID	https://orcid.org/0000-0001-8123-2783
P2	Mendeley	https://www.mendeley.com/profiles/diego-buenao-fernandez/
P3	ResearchGate	https://www.researchgate.net/profile/Diego_Buenano-Fernandez
P4	Google Scholar	https://scholar.google.com/citations?user=idHTA9wAAAAJ&hl=es&oi=ao
P5	Scopus	https://www.scopus.com/authid/detail.uri?authorId=57194634841

Tabla 2.3: Perfiles en redes sociales académicas del autor de la tesis

que se han presentado artículos tienen procesos de revisión por pares. A continuación se presentan los artículos que han sido publicados en memorias de congresos, el orden en el que aparecen las publicaciones está relacionado con su aporte a los objetivos de la presente tesis y se corresponde con el orden de aparición en el compendio en los siguientes capítulos.

1. "Exploring approaches to Educational Data Mining and Learning Analytics, to measure the level of acquisition of student's learning outcomes" (Buenaño-Fernández y Luján-Mora, 2016). Este artículo fue publicado en el congreso C1 descrito en la Tabla 2.2. En el capítulo 5 se detalla el contenido del artículo.

2. "Comparison of applications for educational data mining in Engineering Education" (Buenaño-Fernández y Luján-Mora, 2017). Este artículo fue publicado en el congreso C2 descrito en la Tabla 2.2. En el capítulo 6 se detalla el contenido

del artículo.

3. "A Hybrid Machine Learning Approach for the Prediction of Grades in Computer Engineering Students" (Buenaño-Fernández, Luján-Mora, y Gil, 2019). Este artículo fue publicado en el congreso C3 descrito en la Tabla 2.2. En el capítulo 8 se detalla el contenido del artículo.

4. "Improvement of massive open online courses by text mining of students' emails: A case study" (Buenaño-Fernández y otros, 2017). Este artículo fue publicado en el congreso C4 descrito en la Tabla 2.2. En el capítulo 10 se detalla el contenido del artículo.

2.2. Visibilidad

La forma tradicional de comunicación científica se basa en la divulgación de resultados de investigación a través de medios tradicionales como revistas, congresos, informes académicos, entre otros. Sin embargo la consolidación de la web 2.0 ha dado lugar al aparecimiento de espacios interactivos para divulgar y visibilizar contenidos académicos tales como redes sociales, redes académicas, plataformas científicas, repositorios, blogs, wikis, gestores bibliográficos, entre otros. Estos espacios fomentan la transferencia de conocimientos desde la academia a la sociedad. Durante el período de estudios de doctorado, se crearon diferentes perfiles académicos con el objetivo de incrementar el impacto de la investigación realizada. En la Tabla 2.3 se presenta un resumen de los perfiles académicos del autor de esta tesis. Por otro lado, con el propósito de ganar mayor visibilidad en los productos de la investigación científica realizada, se tomó la decisión, en la medida de lo posible, de publicar en revistas de acceso abierto (*open access*). Este tipo de revistas dan la oportunidad para que la comunidad de académicos e investigadores tengan acceso libre y sin costo a los recursos publicados en estos medios. Por este motivo, tres de los cuatro principales artículos del compendio se publicaron como documentos de acceso abierto (Buenaño-Fernández, Gil, y Luján-Mora, 2019; Buenaño-Fernández y Luján-Mora, 2019; Buenaño-Fernández y otros, 2020).

El artículo (Buenaño-Fernández y otros, 2019) recibió un reconocimiento de Wiley Online Library por estar entre el primer 10 % de artículos más descargados en los 12 meses posteriores a la publicación en línea, entre enero de 2018 y diciembre de 2019. Esto significa que la investigación presentada generó un impacto inmediato y ayudó a aumentar la visibilidad de las aplicaciones informáticas en la educación en ingeniería.

A continuación se presenta un resumen del número de artículos publicados en algunas bases de datos científicas, entre ellas:

- Scopus presenta 17 artículos y 37 citaciones, tal como se muestra en la Figura 2.1.
- WOS presenta 18 artículos y 12 citaciones, tal como se muestra en la Figura 2.2.
- DBLP (Digital Bibliography & Library Project) presenta 4 artículos.
- GS (Google Scholar) presenta 23 artículos y 58 citaciones.

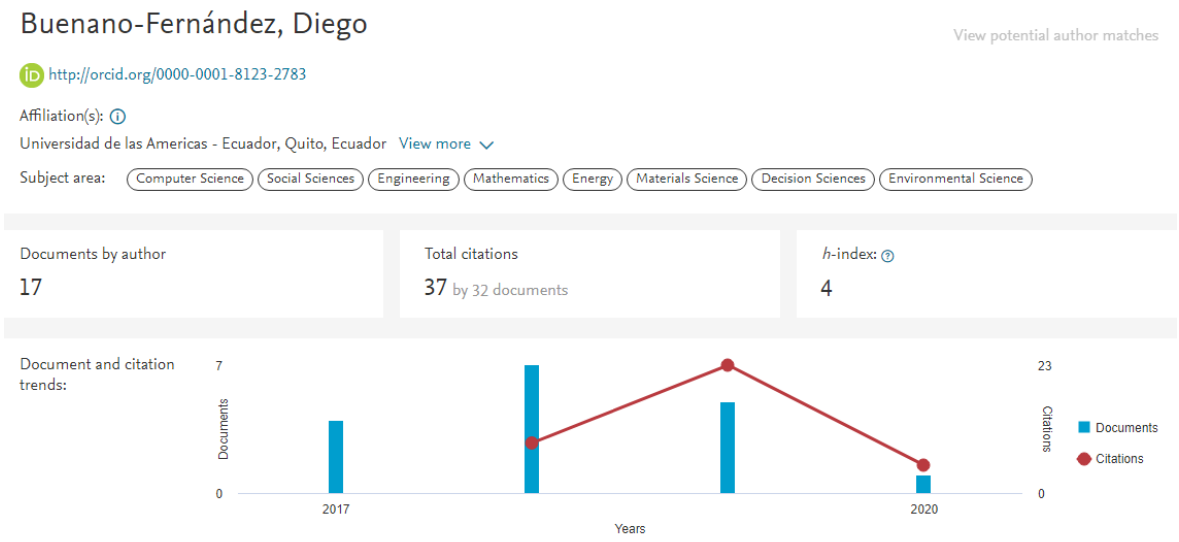


Figura 2.1: Perfil del autor en la base de datos Scopus

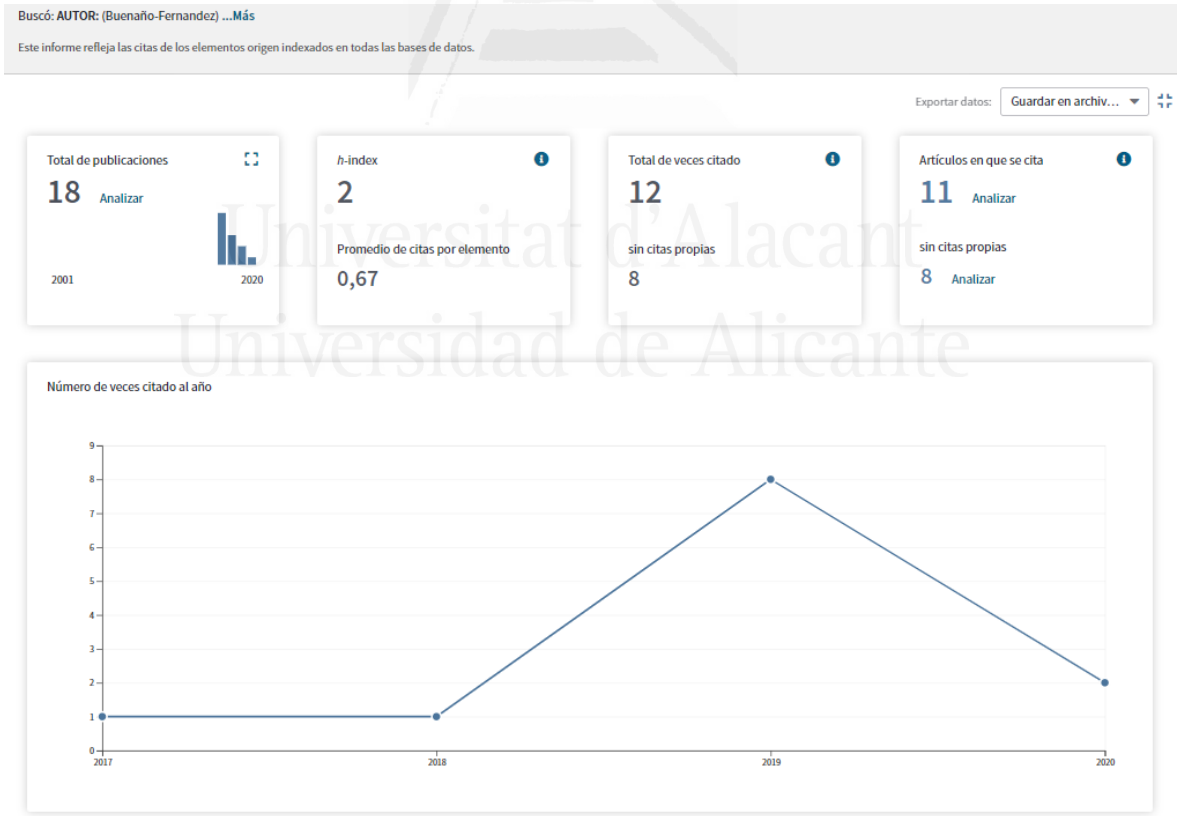


Figura 2.2: Perfil del autor en la base de datos Web of Science

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Parte II

COMPENDIO DE ARTÍCULOS

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

3 Compendio

En este capítulo se presentan las principales publicaciones producto de la investigación realizada en los cuatro años de estudios de doctorado, estos trabajos se han ordenado de acuerdo a los objetivos a los que apuntan y que han sido detallados en la Tabla 1.1. Cuatro de los trabajos incluidos en el compendio corresponden a artículos publicados en revistas científicas con un determinado factor de impacto. El primer artículo se publicó en una revista con factor de impacto 1.435 y clasificada en el segundo cuartil (Q2) de acuerdo al JCR de WOS (Buenaño-Fernández y otros, 2019). La siguiente publicación se realizó en una revista con factor de impacto 2.592 y clasificada en el segundo cuartil (Q2) de acuerdo al JCR de WOS (Buenaño-Fernández, Gil, y Luján-Mora, 2019). La tercera publicación se realizó en una revista con factor de impacto 4.098 y clasificada en el primer cuartil (Q1) de acuerdo al JCR de WOS (Buenaño-Fernández y otros, 2020). La cuarta publicación se realizó en una revista con factor de impacto 0.217 y clasificada en el tercer cuartil (Q3) de acuerdo al SJR de Scimago (Buenaño-Fernández y Luján-Mora, 2019).

Adicionalmente, se escogieron cuatro artículos que fueron presentados en congresos científicos arbitrados por pares académicos: (Buenaño-Fernández y Luján-Mora, 2016; Buenaño-Fernández y Luján-Mora, 2017; Buenaño-Fernández y otros, 2017; Buenaño-Fernández, Luján-Mora, y Gil, 2019).

Los detalles de las publicaciones se muestran en la Figura 3.1, en esta imagen se han considerado los ocho artículos que son parte del compendio. los códigos de cada publicación se corresponden con los definidos en las Tablas 2.1 y 2.2. Se ha considerado una ventana temporal que va desde el año 2016 hasta el 2020, considerando que en estos años se obtuvieron todos los trabajos para presentar el compendio de publicaciones. Cabe señalar que en la Figura 3.1 se incluyó una referencia al objetivo específico al que aporta cada trabajo y que guarda relación con la información registrada en la Tabla 1.2.

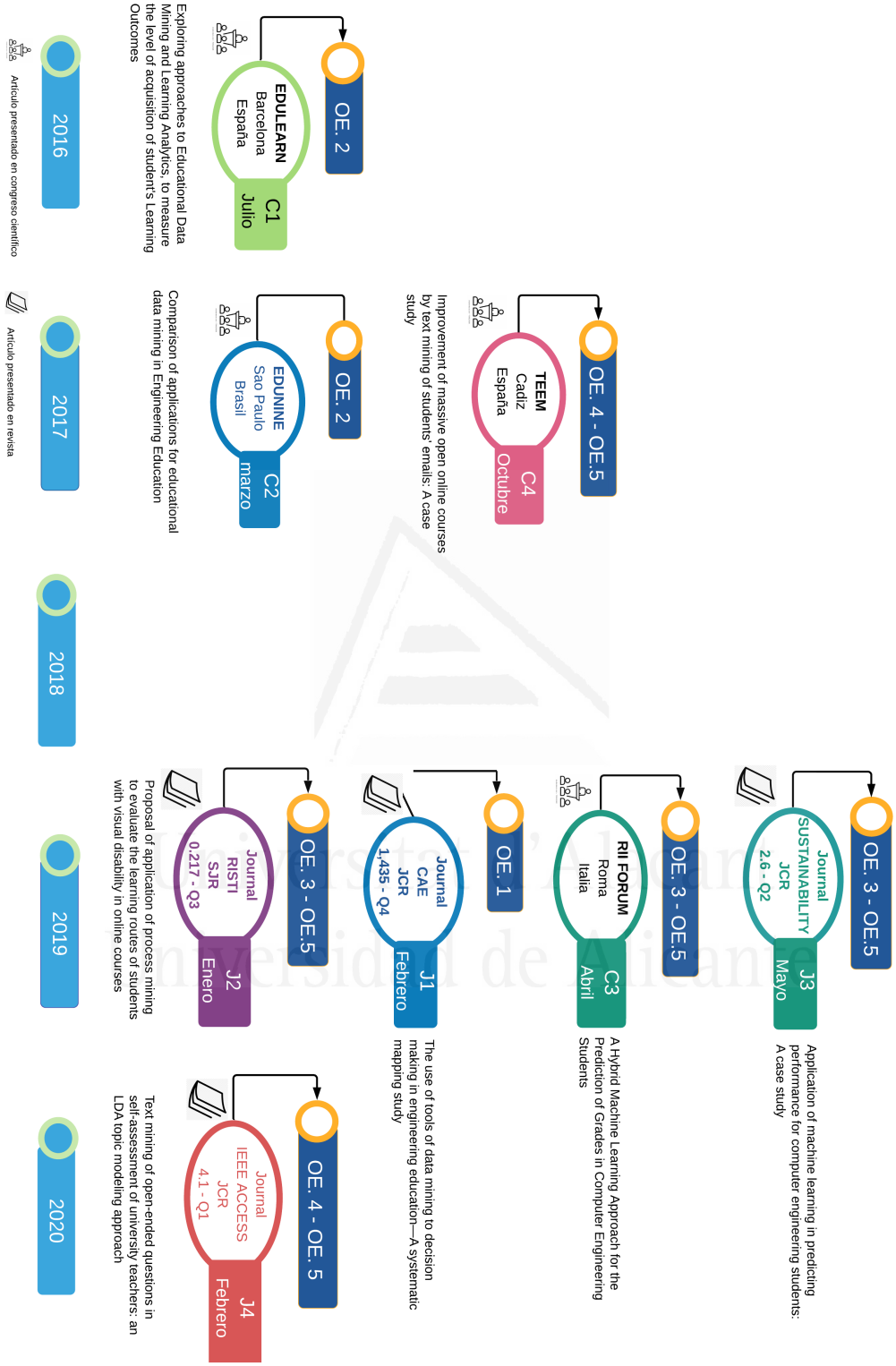


Figura 3.1: Línea de tiempo de publicaciones incluidas en el compendio

4 The use of tools of data mining to decision making in engineering education—A systematic mapping study

Referencia:

Buenaño-Fernandez, D., Villegas-CH, W., y Luján-Mora, S. (2019). The use of tools of data mining to decision making in engineering education—A systematic mapping study. Computer Applications in Engineering Education, 27(3), pp. 744-758. (Buenaño-Fernández y otros, 2019)

Disponible en:

- https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.22100
- https://doi.org/10.1002/cae.22100

Temas a los que aporta:

OE. 1 Evaluar el estado de la cuestión en relación a la aplicación de la minería de datos en el ámbito de la educativos en ingeniería.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Received: 13 March 2018 | Accepted: 16 December 2018
DOI: 10.1002/cae.22100

Check for updates

REVIEW

WILEY

The use of tools of data mining to decision making in engineering education—A systematic mapping study

Diego Buenaño-Fernandez[1] | **William Villegas-CH**[1] | **Sergio Luján-Mora**[2]

[1] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito, Ecuador

[2] Departmento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Alacant, Spain

Correspondence
Diego Buenaño-Fernandez, Facultad de Ingenierías y Ciencias Aplicadas, Universidad de las Américas, Quito, Ecuador.
Email: diego.buenano@udla.edu.ec

Abstract

In recent years, there has been an increasing amount of theoretical and applied research that has focused on educational data mining. The learning analytics is a discipline that uses techniques, methods, and algorithms that allow the user to discover and extract patterns in stored educational data, with the purpose of improving the teaching-learning process. However, there are many requirements related to the use of new technologies in teaching-learning processes that are practically unaddressed from the learning analytics. In an analysis of the literature, the existence of a systematic revision of the application of learning analytics in the field of engineering education is not evident. The study described in this article provides researchers with an overview of the progress made to date and identifies areas in which research is missing. To this end, a systematic mapping study has been carried out, oriented toward the classification of publications that focus on the type of research and the type of contribution. The results show a trend toward case study research that is mainly directed at software and computer science engineering. Furthermore, trends in the application of learning analytics are highlighted in the topics, such as student retention or dropout prediction, analysis of academic student data, student learning assessment and student behavior analysis. Although this systematic mapping study has focused on the application of learning analytics in engineering education, some of the results can also be applied to other educational areas.

KEYWORDS

decision making, educational data mining, engineering education, learning analytics, systematic mapping

Universidad de Alicante
Universidad de Alicante

1 | INTRODUCTION

Learning analytics (LA) and educational data mining (EDM) are disciplines that analyze educational data about learners and their contexts. This analysis is done through a variety of methods, techniques and statistical tools, including machine learning and data mining (DM). The objective of LA is to provide an analysis of the data that exist in educational repositories such as learning management systems (LMS), for the purpose of understanding and optimizing learning and the environments in which it occurs [64]. This analysis can significantly support decision making in terms of planning on the part of teachers and managers of education institutions [37]. This topic has been dealt with in several articles, taking into account the different requirements of the stakeholders involved including teachers, instructors, students, course developers, researchers, educational institutions, and private training companies. Depending on the stakeholder's

 © 2019 Wiley Periodicals, Inc.

Universitat d'Alacant
Universidad de Alicante

5 Exploring approaches to educational data mining and learning analytics, to measure the level of acquisition of student's learning outcomes

Referencia:

Buenaño-Fernández, D., y Luján-Mora, S. (2016). Exploring approaches to educational data mining and learning analytics to measure the level of acquisition of student's learning outcome. En actas del 8th Annual International Conference on Education and New Learning Technologies, pp. 1845-1850. (Buenaño-Fernández y Luján-Mora, 2016)

Disponible en:

- https://library.iated.org/view/BUENANOFERNANDEZ2016EXP
- https://doi.org/10.21125/edulearn.2016.1368

Temas a los que aporta:

OE. 2 Analizar enfoques y herramientas para la aplicación de minería de datos en entornos educativos en ingeniería.

Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

EXPLORING APPROACHES TO EDUCATIONAL DATA MINING AND LEARNING ANALYTICS, TO MEASURE THE LEVEL OF ACQUISITION OF STUDENT'S LEARNING OUTCOME

D. Buenaño Fernández[1], S. Luján-Mora[2]

[1] Universidad de las Américas Quito (ECUADOR)
[2] University of Alicante (SPAIN)

Abstract

The increase of interactive learning environments, such as learning management systems, personal learning environments and intelligent tutoring systems, generates large amounts of data. Nevertheless, in the majority of cases, the data stored in these environments, either in files or databases, are underused by teachers, students and institutions. The analysis of these data generates a space for the increase of research in topics related to evaluation of learning theories, early warning systems, and the development of future learning applications.

Several researches have shown the use of learning analytics, machine learning, data mining techniques and classifiers to predict on the issue of learning outcomes in students. Nevertheless, it is necessary to guide the study of methods and techniques of data mining and learning analytics classifying them according to the actors and use objectives. This paper presents a summary of the most relevant approaches of the use of techniques and tools of educational data mining and learning analytics, using Learning Management Systems platform's in higher education.

Keywords: Educational Data Mining, Learning Analytics, Learning Assessment, student modeling, Predicting Student's Performance.

1 INTRODUCTION

The process of virtual or hybrid learning is enriched by the interaction and discussion that occurs between teacher-student, student-student and student-learning materials. The teachers spend a lot of time preparing learning strategies that help maximize these interactions, the results of this interaction are stored in e-learning platforms. The advances in the data mining (DM) field make it possible to explore the educational data and the teachers can strengthen this process through the detailed interpretation of the results of these interactions. Nevertheless, in the majority of cases, the data stored in virtual platforms, either in files or databases, are underused by teachers, students and institutions.

The educational data mining community [2] defines educational data mining (EDM) as: "Educational data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in."

EDM is a field that exploits statistical, machine learning, and data mining (DM) algorithms over the different types of educational data. Its main objective is to analyze these types of data in order to resolve educational research issues [3].

As learning management systems (LMS) offer an ever wider range of online resources, the volume of data generated increases exponentially. These data, product of student interaction with learning resources, are stored in files or databases of virtual learning environments. The EDM provides very relevant information to higher education institutions. This information can be used in higher education institutions, such as student behaviours, course management systems, education student retention, decision support system in higher, and attrition. Giving rise to the creation of decision support systems are generated in the educational field [3].

In this paper, the main approaches discussed of EM and learning analytics (LA), as support tools to identify the effectiveness of the strategies proposed by teachers in the teaching-learning process [1].

Although the tools to analyse the log files stored on the LMS allow to measure, collect, evaluate and present all student data, these tools do not include artificial intelligence algorithms as a support

Universitat d'Alacant
Universidad de Alicante

6 Comparison of applications for educational data mining in engineering education

Referencia:

Buenaño-Fernández, D., y Luján-Mora, S. (2017). Comparison of applications for educational data mining in Engineering Education. En actas del IEEE World Engineering Education Conference (EDUNINE), pp. 81-85. (Buenaño-Fernández y Luján-Mora, 2017)

Disponible en:

- https://ieeexplore.ieee.org/document/7918187/
- https://10.1109/EDUNINE.2017.7918187

Temas a los que aporta:

OE. 2 Analizar enfoques y herramientas para la aplicación de minería de datos en entornos educativos en ingeniería.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

Comparison of applications for educational data mining in Engineering Education

Diego Buenaño Fernández
Facultad de Ingenierías y Ciencias Agropecuarias
Universidad de las Américas
Quito, Ecuador
diego.buenano@udla.edu.ec

Sergio Luján-Mora
Department of Software and Computing Systems
University of Alicante
Alicante, Spain
sergio.lujan@ua.es

***Abstract*— Currently there are many techniques based on information technology and communication aimed at assessing the performance of students. Data mining applied in the educational field (educational data mining) is one of the most popular techniques that are used to provide feedback with regard to the teaching-learning process. In recent years there have been a large number of open source applications in the area of educational data mining. These tools have facilitated the implementation of complex algorithms for identifying hidden patterns of information in academic databases. The main objective of this paper is to compare the technical features of three open source tools (RapidMiner, Knime and Weka) as used in educational data mining. These features have been compared in a practical case study on the academic records of three engineering programs in an Ecuadorian university. This comparison has allowed us to determine which tool is most effective in terms of predicting student performance.**

***Keywords*— Educational Data Mining; performance; Open Source; Software Tool, K-means.**

I. INTRODUCTION

The increasing use of information and communication technologies in the educational field entails the storage of large volumes of data in various formats. The applications used in educational environments save the information in many different repositories such as archives, blogs, documents, images, videos, audios, scientific data, meta data or hyperlinks, and many new data formats. The amount of data available in previous scenarios is so enormous that traditional processing techniques are insufficient when it comes to processing them [1]. This makes the information stored underutilized, and means that it is not taken into account in terms of strategic decision making. Therefore, these data require the application of appropriate methods or techniques to process them and to extract knowledge. In the educational field, these techniques are classified into what is known as Educational Data Mining (EDM), Learning Analytics (LA) and Knowledge Discovery in Databases (KDD) [2]. The main function of these techniques is the application of various methods and algorithms that allow the user to discover and extract patterns in the stored data [3]. In the educational field, the most commonly used algorithms are Regression, Nearest Neighbor, Clustering, Classification, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Nearest Neighbor method, etc. [4].

Nowadays, there are several open source tools that support the management of Data Mining (DM). Some of the principal tools are as follows: Hadoop, Orange, Weka, Knime, Rapid Miner, Keel, among other [5].

The large number of DM tools that are currently available in the market, and which are used in widely dispersed areas of data analysis, generate uncertainty for specialists who focus on the analysis of educational data [6]. This is especially in view of the need to take into account that this environment has very special features that makes it different from other environments. While it is true that specific applications have been developed for educational environments, especially in universities [7], these applications do not have all the functionalities compared with the most commonly used DM applications in the market.

The main objective of this work was to compare the technical characteristics of three DM tools (RapidMiner, Knime and Weka) in an educational environment. These characteristics were measured in terms of the application of clustering and segmentation techniques using a case study of the academic records of three engineering programs at an Ecuadorian university. The results of this comparison will be of value to people who wish to work in EDM and need to know how to choose the most appropriate tool to perform a particular study.

The content of this paper is organized as follows: Section II covers the description of the method used in this study, Section III includes a detailed description of the case study, and finally, Section IV presents the conclusions and future work.

II. METHOD

The method used to perform the comparison of the technical features of the tools will be through their use in each of the phases of the DM process. The different characteristics are compared in terms of aspects such as the results generated by each tool in the development of certain processes, the number of algorithms available for performing DM operations, and the working environment that each tool uses.

The DM process follows a sequence that generally includes the following elements [8] [9]:

978-1-5090-4886-1/17/$31.00 ©2017 IEEE

Universitat d'Alacant
Universidad de Alicante

7 Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses

Referencia:

Buenaño Fernández, D., y Luján-Mora, S. (2019). Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses, RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, pp. 1035-1047. (Buenaño-Fernández y Luján-Mora, 2019)

Disponible en:

- http://www.risti.xyz/issues/ristie17.pdf

Temas a los que aporta:

OE. 3 Evaluar el desempeño de estudiantes en el campo de la educación en ingeniería a través de la aplicación de diferentes técnicas y herramientas de minería de datos.

OE. 5 Proponer un conjunto de herramientas de minería de datos para evaluar el desempeño de estudiantes y los procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 04/09/2018
Aceitação/Acceptance: 09/11/2018

Propuesta de aplicación de minería de procesos para evaluar las rutas de aprendizaje de estudiantes con discapacidad visual en cursos en línea

Diego Buenaño-Fernández[1], Sergio Luján-Mora[2]

diego.buenano@udla.edu.ec, sergio.lujan@ua.es

[1]Universidad de Las Américas, José Queri, 170137, Quito, Ecuador.

[2] Universidad de Alicante, San Vicente del Raspeig, 03080, Alicante, España

Pages: 1035–1047

Resumen: La culminación exitosa de un curso en línea por parte de estudiantes con algún tipo de discapacidad está relacionada directamente con la inclusión o no de criterios de accesibilidad. Las aplicaciones web, y en particular los cursos en línea, están conformados por un sinnúmero de componentes cuyo objetivo es hacer más fácil la navegación de los usuarios. La exclusión de criterios de accesibilidad en el diseño de estos componentes puede llegar a convertirse en barreras de aprendizaje para estudiantes con algún tipo de discapacidad. En la literatura revisada, se han identificado varios trabajos relacionados con el análisis de problemas de accesibilidad en entornos educativos virtuales. Sin embargo, hay poca investigación vinculada con el análisis de rutas de aprendizaje seguidas por personas con algún tipo de discapacidad. En este trabajo se propone aplicar herramientas de minería de datos educativos y de minería de procesos para detectar puntos potencialmente peligrosos en las rutas de aprendizaje que siguen este grupo de estudiantes.

Palabras-clave: Analítica de aprendizaje; Discapacidad visual; Minería de procesos educativos; Web Content Accessibility Guidelines (WCAG) 2.1; World Wide Web Consortium (W3C).

Universitat d'Alacant
Universidad de Alicante

Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses

Abstract: The successful completion of an online course by students with some type of disability is directly related to the inclusion or not of accessibility criteria. Web applications, and in particular online courses, are made up of a number of components whose objective is to make it easier for users to navigate. The exclusion of accessibility criteria in the design of these components can become learning barriers for students with some type of disability. In the literature reviewed, several works related to the analysis of accessibility problems in virtual educational environments have been identified. However, there is little research linked to the analysis of learning routes followed by people with some type of disability. In this project, it is proposed to apply educational data mining and process mining tools

Universitat d'Alacant
Universidad de Alicante

8 A hybrid machine learning approach for the prediction of grades in computer engineering students

Referencia:

Buenaño-Fernandez, D., Luján-Mora, S., y Gil, D. (2019). A Hybrid Machine Learning Approach for the Prediction of Grades in Computer Engineering Students. En actas del The International Research & Innovation Forum, pp. 125-134. (Buenaño-Fernández, Luján-Mora, y Gil, 2019)

Disponible en:

- https://link.springer.com/chapter/10.1007/978-3-030-30809-4_13
- https://doi.org/10.1007/978-3-030-30809-4_13

Temas a los que aporta:

OE. 3 Evaluar el desempeño de estudiantes en el campo de la educación en ingeniería a través de la aplicación de diferentes técnicas y herramientas de minería de datos.

OE. 5 Proponer un conjunto de herramientas de minería de datos para evaluar el desempeño de estudiantes y los procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

A Hybrid Machine Learning Approach for the Prediction of Grades in Computer Engineering Students

Diego Buenaño-Fernandez, Sergio Luján-Mora and David Gil

Abstract The growing application of information and communication technologies (ICTs) in teaching and learning processes has generated an overload of valuable information for all those involved in education field. Historical information from students' academic records has become a valuable source of data that has been used for different purposes. Unfortunately, a high percentage of research has been developed from the perspective and the need of teachers and educational administrators. This perspective has left the student in the background. This paper proposes the application of a hybrid machine learning approach, with the aim of laying the groundwork for a future implementation of a recommendation system that allows students to make decisions related to their learning process. The work has been executed on the historical academic information of students of computer engineering degree. The results obtained in this article show the effectiveness of applying a hybrid machine learning approach. This architecture is composed of, on the one hand, techniques of supervised learning applied with the objective of classifying the data in clusters, and on the other hand, having this initial classification, unsupervised learning techniques applied with the objective of carrying out a predictive analysis of students' historical grade records.

Universitat d'Alacant
Universidad de Alicante

1 Introduction

Online educational platforms and academic management systems can capture, with different levels of granularity, all types of data generated from the interaction of

D. Buenaño-Fernandez ( )
Universidad de Las Américas, Quito 08544, Ecuador
e-mail: diego.buenano@udla.edu.ec

S. Luján-Mora · D. Gil
Universidad de Alicante, 03690 Alicante, Spain
e-mail: sergio.lujan@ua.es

D. Gil
e-mail: david.gil@ua.es

© Springer Nature Switzerland AG 2019

A. Visvizi and M. D. Lytras (eds.), *Research & Innovation Forum 2019*,
Springer Proceedings in Complexity,
https://doi.org/10.1007/978-3-030-30809-4_13

Universitat d'Alacant
Universidad de Alicante

9 Application of machine learning in predicting performance for computer engineering students: A case study

Referencia:

Buenaño-Fernández, D., Gil, D., y Luján-Mora, S. (2019). Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. Sustainability, 11(10), 2833.(Buenaño-Fernández, Gil, y Luján-Mora, 2019)

Disponible en:

- https://www.mdpi.com/2071-1050/11/10/2833
- https://doi.org/10.3390/su11102833

Universitat d'Alacant
Universidad de Alicante

Temas a los que aporta:

OE. 3 Evaluar el desempeño de estudiantes en el campo de la educación en ingeniería a través de la aplicación de diferentes técnicas y herramientas de minería de datos.

OE. 5 Proponer un conjunto de herramientas de minería de datos para evaluar el desempeño de estudiantes y los procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería.

Universitat d'Alacant
Universidad de Alicante

sustainability

MDPI

Article

Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study

Diego Buenaño-Fernández [1,*], David Gil [2] and Sergio Luján-Mora [3]

[1] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Av. de los Granados E12-41 y Colimes, Quito EC170125, Ecuador
[2] Departamento de Tecnología Informática y Computación, Universidad de Alicante, San Vicente del Raspeig, 03690 Alicante, Spain; david.gil@ua.es
[3] Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, San Vicente del Raspeig, 03690 Alicante, Spain; sergio.lujan@ua.es
* Correspondence: diego.buenano@udla.edu.ec; Tel.: +59-39-8449-8347

Received: 7 April 2019; Accepted: 14 May 2019; Published: 17 May 2019

Abstract: The present work proposes the application of machine learning techniques to predict the final grades (FGs) of students based on their historical performance of grades. The proposal was applied to the historical academic information available for students enrolled in the computer engineering degree at an Ecuadorian university. One of the aims of the university's strategic plan is the development of a quality education that is intimately linked with sustainable development goals (SDGs). The application of technology in teaching–learning processes (Technology-enhanced learning) must become a key element to achieve the objective of academic quality and, as a consequence, enhance or benefit the common good. Today, both virtual and face-to-face educational models promote the application of information and communication technologies (ICT) in both teaching–learning processes and academic management processes. This implementation has generated an overload of data that needs to be processed properly in order to transform it into valuable information useful for all those involved in the field of education. Predicting a student's performance from their historical grades is one of the most popular applications of educational data mining and, therefore, it has become a valuable source of information that has been used for different purposes. Nevertheless, several studies related to the prediction of academic grades have been developed exclusively for the benefit of teachers and educational administrators. Little or nothing has been done to show the results of the prediction of the grades to the students. Consequently, there is very little research related to solutions that help students make decisions based on their own historical grades. This paper proposes a methodology in which the process of data collection and pre-processing is initially carried out, and then in a second stage, the grouping of students with similar patterns of academic performance was carried out. In the next phase, based on the identified patterns, the most appropriate supervised learning algorithm was selected, and then the experimental process was carried out. Finally, the results were presented and analyzed. The results showed the effectiveness of machine learning techniques to predict the performance of students.

Keywords: educational data mining; learning analytics; machine learning; big data; prediction grades

1. Introduction

Quality education is one of the Sustainable Development Goals (SDGs) approved by the United Nations forum in 2015 [1] and is a fundamental challenge to support sustainable development worldwide. A key element that must be taken into account when talking about sustainable development

Universitat d'Alacant
Universidad de Alicante

10 Improvement of massive open online courses by text mining of students' emails: A case study

Referencia:

Buenaño-Fernández, D., Luján-Mora, S., y Villegas-Ch, W. (2017). Improvement of massive open online courses by text mining of students' emails: A case study. En actas del 5th International Conference on Technological Ecosystems for Enhancing Multiculturality. pp. 1-7. (Buenaño-Fernández y otros, 2017)

Disponible en:

- https://dl.acm.org/doi/10.1145/3144826.3145393
- https://doi.org/10.1145/3144826.3145393

Universitat d'Alacant
Universidad de Alicante

Temas a los que aporta:

OE. 4 Evaluar procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería a través de la aplicación de técnicas de minería de texto.

OE. 5 Proponer un conjunto de herramientas de minería de datos para evaluar el desempeño de estudiantes y los procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería.

Universitat d'Alacant
Universidad de Alicante

Improvement of massive open online courses by text mining of students' emails: a case study

Diego Buenaño-Fernández
Facultad de Ingeniería de Ciencias Agropecuarias
Universidad de Las Américas
Quito
170125
Ecuador
diego.buenano@udla.edu.ec

Sergio Luján-Mora
Departament of Software and Computing Systems
University of Alicante
Alicante
Spain
sergio.lujan@ua.es

W. Villegas-Ch
Facultad de Ingeniería de Ciencias Agropecuarias
Universidad de Las Américas
Quito
170125
Ecuador
william.villegas@udla.edu.ec

ABSTRACT

In recent years, the constant increase in the number of online courses has led to radical changes in the education sector. These new online learning environments present a series of challenges that are difficult to manage using traditional methods. The challenges relate to the level of commitment and motivation shown by students on this type of course. Several articles have been identified from the analysed literature related to the application of text or opinion mining techniques for the analysis of comments made in social networks. In the educational field, articles related to the topic that focus on the analysis of opinion have been identified based on entries included in discussion forums for online courses. Many publications are geared towards solutions in the English language, and the nature of linguistic analysis of this type of study makes it necessary to adapt them for languages other than English. In this paper, we explore the opinion mining through text mining in emails from Massive Open Online Courses (MOOC). The opinion mining expressed in emails is a complex task due to the thematic disparity of emails, their size and the depth of linguistic analysis required. The purpose of this study is to analyse students opinions about their courses, their instructors, and the main tools used on the course. The research focus on the calculation and analysis of the frequency of terms, the analysis of concordances, groupings and n-grams. The case study used in this paper is a MOOC on the topic of web development with more than 40,000 enrolled students.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

TEEM 2017, October 18–20, 2017, Cádiz, Spain
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5386-1/17/10…$15.00
https://doi.org/10.1145/3144826.3145393

CCS Concepts

• Computing methodologies~Information extraction • Computing methodologies~Supervised learning • Information systems~Sentiment analysis

Keywords

Opinion mining; Massive Open Online Course; MOOC; Supervised learning; Text mining.

ACM Reference format:

Diego Buenaño-Fernández, Sergio Luján-Mora and W. Villegas-Ch. 2017. Improvement of massive open online courses by text mining of students' emails: a case study. In *Proceedings of 5th International Conference on Technological Ecosystems for Enhancing Multiculturality, Cádiz, Spain, October 2017 (TEEM 2017), 7 pages.*
DOI: https://doi.org/10.1145/3144826.3145393

1: INTRODUCTION

Globalization and the proliferation of Massive Open Online Courses (MOOC) has radically altered the model of education. New technology in this field offers the opportunity to increase the availability of courses to a far greater audience than that provided in the traditional setting. However, the implementation has significant challenges that must be overcome to allow students to take full advantage of them [1]. The flexibility of the platforms in which MOOC operate, and the wealth of learning resources they provide, allows for the inclusion of large numbers of students across a greater geographical base. This interaction between students and systems produces large-scale learning behaviour data and leaves traces of the educational process on systems that are useful for analysis [2].

Given the large volume of emails that are usually generated in a MOOC, it is useful to develop methods that are oriented to the automatic processing of texts with an acceptable reliability [3], and which should become valuable tools to support the educational process.

Interactions between students and MOOC can be explored using text mining techniques, with the aim of improving

RIGHTSLINK

Universitat d'Alacant
Universidad de Alicante

11 Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach

Referencia:

Buenaño-Fernandez, D., González M., Gil, D., y Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach. IEEE Access, vol. 8, pp. 35318-35330. (Buenaño-Fernández y otros, 2020)

Disponible en:

- https://ieeexplore.ieee.org/document/9003400
- https://doi.org/10.1109/ACCESS.2020.2974983

Temas a los que aporta:

OE. 4 Evaluar procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería a través de la aplicación de técnicas de minería de texto.

OE. 5 Proponer un conjunto de herramientas de minería de datos para evaluar el desempeño de estudiantes y los procesos de enseñanza aprendizaje en el ámbito de la educación en ingeniería.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON ADVANCED DATA MINING METHODS FOR SOCIAL COMPUTING

Received January 30, 2020, accepted February 12, 2020, date of publication February 19, 2020, date of current version February 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974983

Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach

DIEGO BUENAÑO-FERNANDEZ[1], MARIO GONZÁLEZ[1], DAVID GIL[2], AND SERGIO LUJÁN-MORA[3]

[1]Faculty of Engineering and Applied Sciences, Universidad de Las Américas (UDLA), Quito 170504, Ecuador
[2]Department of Computer Technology and Computation, University of Alicante, 03690 Alicante, Spain
[3]Department of Software and Computing Systems, University of Alicante, 03690 Alicante, Spain

Corresponding author: Diego Buenaño-Fernandez (diego.buenano@udla.edu.ec)

This work was supported in part by the Universidad de Las Américas, Quito, Ecuador, under Project SIS.DBF.19.02 and Project SIS.MGR.20.02, in part by the Spanish Ministry of Science, Innovation and Universities through the Project ECLIPSE-UA under Grant RTI2018-094283-B-C32, and in part by the Lucentia AGI Grant.

ABSTRACT The large amount of text that is generated daily on the web through comments on social networks, blog posts and open-ended question surveys, among others, demonstrates that text data is used frequently, and therefore; its processing becomes a challenge for researchers. The topic modeling is one of the emerging techniques in text mining; it is based on the discovery of latent data and the search for relationships among text documents. In this paper, the objective of the research is to evaluate a generic methodology based on topic modeling and text network modeling, that allows researchers to gather valuable information from surveys that use open-ended questions. To achieve this, this methodology has been evaluated through the use of a case study in which the responses to a teacher self-assessment survey in an Ecuadorian university have been studied. The main contribution of the article is the inclusion of clustering algorithms in order to complement the results obtained when executing topic modeling. The proposed methodology is based on four phases: (a) Construction of a text database, (b) Text mining and topic modeling, (c) Topic network modeling and (d) The relevance of the identified topics. In previous works, it has been observed that the human interpretative contribution plays an important role in the process, especially in phases (a) and (d). For this reason, the visualization interfaces, such as graphs and dendograms, are of critical importance for researchers in order allow topic to efficiently analyze the results of the topic modeling. As a result of this case study, a compendium of the main strategies that teachers carry out in their classes with the aim of improving student retention is presented. In addition, the proposed methodology can be extended to the analysis of the unstructured textual information found in blogs, social networks, forums, etc.

INDEX TERMS Latent Dirichlet allocation, open-ended questions, teacher self-assessment, topic modeling, topic network.

I. INTRODUCTION

The absence of a generic methodology when it comes to performing text analysis has become a great challenge and a gap in research in the text mining field. This is a problem because the text mining models used are different for each case, since each area has a set of specific words with different semantic meanings. For example, the text mining model used to analyze messages on the social network Twitter is very different from the text mining model used to analyze the answers to open-ended questions in a given survey [1]. One of the most powerful and widely used techniques for text mining, the recovery of hidden information in texts and social network analysis is the topic modeling [2]. Based on these premises, it is appropriate to develop a model with generic criteria that guarantee the validity and reliability of the topic modeling that is applied to different areas. The present study emphasizes four aspects to be taken into consideration in any methodology aimed at the effective application of topic modeling: (a) a clear definition of the process of collection

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

Universitat d'Alacant
Universidad de Alicante

Parte III

CONCLUSIONES

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

12 Conclusiones

En la investigación realizada en la presente tesis se han identificado algunas estrategias académicas del campo educativo que pueden beneficiarse con la aplicación de técnicas de DM. Estas acciones son:

- Retroalimentación de la gestión docente.
- Retroalimentación a estudiantes.
- Predicción del desempeño estudiantil.
- Análisis de la interacción de la comunidad educativa.
- Análisis y visualización de datos para detectar y comprender tendencias.

En las publicaciones incluidas en el compendio se evidenció la aplicación de técnicas de DM a través de la realización de casos de estudio. Estos casos de estudio estuvieron orientados a trabajar sobre las estrategias académicas descritas en el párrafo anterior. Por ejemplo, en la publicación (Buenaño-Fernández, Gil, y Luján-Mora, 2019) se trabajó en una metodología para monitorear y predecir las calificaciones de estudiantes en el ámbito de la educación en ingeniería. Esta metodología fue evaluada sobre las calificaciones históricas de un grupo de estudiantes de ingeniería en una universidad ecuatoriana. La metodología propuso la agrupación de estudiantes que cumplían con ciertas condiciones comunes, por ejemplo, aquellos que tomaron las mismas materias y que aprobaron esas materias en el mismo período académico, afinidades por área de conocimiento, por desempeño académico por semestre. En el artículo (Buenaño-Fernández y otros, 2020) se presentó una propuesta metodológica basada en modelado de tópicos y modelado de redes para identificar las estrategias que utilizan los docentes en el ámbito de la educación en ingeniería para fortalecer la retención estudiantil. En el artículo (Buenaño-Fernández y Luján-Mora, 2019) se trabajó en una propuesta basada en minería de procesos con el objetivo de identificar inconvenientes en las rutas de aprendizaje de estudiantes con algún tipo de discapacidad.

Un tema importante para reflexionar tiene que ver con las dificultades existentes para obtener los datos con los cuales llevar a cabo los casos de estudio. Las IES manejan esta información con mucho sigilo, debido a que su mal uso puede generar consecuencias imprevistas que afecten a los involucrados (docentes y estudiantes). Por tal razón es fundamental que los resultados obtenidos en las investigaciones sean presentados a las IES de manera que puedan palpar los beneficios sobre los procesos de enseñanza - aprendizaje y por ende sobre la calidad académica. Un aspecto primordial en este tema es precautelar la información personal de los estudiantes y docentes, en ese sentido en esta investigación hemos anonimizado toda información personal que pudiese aparecer en los estudios de caso.

El análisis de los datos obtenidos en los procesos de EDM y LA debe poner especial atención a las variables externas que pudiesen afectar los resultados de los estudios de caso. Las variables externas hacen referencia a parámetros específicos del entorno educativo. Por ejemplo, al replicar una herramienta desarrollada a la medida para una institución en particular, se debe realizar un análisis comparativo del modelo educativo que rige en las dos instituciones. También se debe tomar en cuenta, por ejemplo, las características comportamentales de los estudiantes, no es lo mismo realizar el análisis de la información de estudiantes de una carrera de formación humanística que para estudiantes de carreras de ingeniería o medicina. Una característica particular identificada en los estudiantes de ingeniería es que generalmente tienen comportamientos académicos muy irregulares en la aprobación de las materias previstas en su plan de estudios. Esto está estrechamente ligado con el hecho de que, para los títulos de ingeniería, las tasas de repetición son altas, especialmente en materias relacionadas con las matemáticas o la ingeniería básica. Como resultado de la investigación realizada se han identificado iniciativas locales o regionales que están trabajando sobre el tema de EDM y LA en la educación superior. Es recomendable unirse a estas iniciativas y redes de investigación con el fin de beneficiarse de los proyectos que se ejecutan en el área del EDM.

La aplicación de técnicas de minería de textos orientadas a evaluar la interacción de estudiantes y docentes en procesos de enseñanza - aprendizaje ha demostrado que puede contribuir de forma significativa en el análisis de grandes volúmenes de información no estructurada. En la actualidad, en los entornos virtuales de aprendizaje se genera un gran volumen de datos textuales a través de comentarios en redes sociales, publicaciones de blog, encuestas con preguntas abiertas, entre otros. Esto demuestra que los datos de texto se usan con frecuencia y, por lo tanto, su procesamiento se convierte en un desafío para los investigadores. En la presente tesis se ha trabajado con técnicas de modelado de tópicos, técnicas de modelado de redes y técnicas de análisis de sentimientos para procesar datos provenientes de diferentes fuentes que implica, como se ha descrito a lo largo de este documento, uno de los mayores retos con el que nos encontramos en la actualidad del *Big Data*, y que es el de la variedad en las fuentes de datos. Un ejemplo lo tenemos en la información de correos electrónicos de cursos MOOC que se trabajó en el artículo (Buenaño-Fernández y otros, 2017) y contribuyó con el análisis de datos generados en la interacción de estudiantes a través de redes sociales. Finalmente, se trabajó en el análisis de datos de encuestas de autoevaluación docente con preguntas abiertas (Buenaño-Fernández y otros, 2020). El aporte de este último trabajo está

relacionado con la estrategia educativa de retroalimentación de la gestión docente.

El uso de herramientas de minería de datos para la toma de decisiones en la educación de ingeniería (Buenaño-Fernández y otros, 2019) permitió obtener información valiosa para alinear los casos de estudio con el plan de investigación propuesto. A partir de las preguntas de investigación que se plantearon en el mapeo sistemático se identificaron las principales técnicas y métodos de EDM y LA que son los mas usados en el campo de la educación en ingeniería. También se identificaron las principales estrategias académicas de la educación en ingeniería en las que se aplicaron técnicas de EDM y LA. A lo largo del programa de doctorado se profundizó en el estudio y desarrollo de casos de estudio enfocados en estas dos temáticas.

Universitat d'Alacant
Universidad de Alicante

Universitat d'Alacant
Universidad de Alicante

13 Trabajos futuros

Producto de la investigación desarrollada se han podido identificar líneas futuras de investigación que están relacionadas con la aplicación de DM en entornos educativos. Así, por ejemplo, en relación a las fuentes de datos de donde minar la información, el estudio (Buenaño-Fernández y otros, 2019) muestra que existe una brecha de investigación en el análisis de datos generados a través de estrategias de aprendizaje basado en juegos, bases de datos de portafolios educativos, bases de datos textuales, base de datos de multimedia, entre otros. Por otro lado, hay que poner especial énfasis en los nuevos entornos de aprendizaje que van apareciendo producto del incremento de la hiperconectividad. Así, por ejemplo, se deben analizar datos provenientes de plataformas de aprendizaje social, entornos de aprendizaje móviles o ubicuos, plataformas basadas en tutores cognitivos, MOOC, entre otros.

Una línea de investigación que queda abierta producto del presente trabajo gira en torno a la aplicación de técnicas de DM para evaluar el logro de resultados de aprendizaje en estudiantes de ingeniería. La evidencia de que un estudiante ha adquirido un resultado de aprendizaje proviene de diversas fuentes, por ejemplo, el registro multimedia de proyectos realizados en empresas externas, encuestas a empleadores, identificación de patrones de comportamiento, proyectos integradores de asignaturas, presentación de casos de estudio, entre otras. Esta amplia variedad de fuentes hace complejo el proceso de evaluación del logro de resultados de aprendizaje por lo que se plantea profundizar en temas como analítica de aprendizaje multimodal, minería de textos y comportamiento, aprendizaje automático y aprendizaje profundo, entre otros.

A partir de los resultados obtenidos en los trabajos sobre predicción del desempeño académico de estudiantes universitarios presentados en esta tesis, se ha generado una línea de investigación vinculada con el desarrollo de sistemas de aprendizaje autorregulado. Las IES desarrollan su planificación académica tomando en cuenta el desarrollo de habilidades de auto aprendizaje en los estudiantes. Estas habilidades están relacionadas con la capacidad de planificar, administrar y controlar su proceso de aprendizaje. Una línea de investigación futura es la identificación de estrategias de aprendizaje autorregulado efectivas que permitan hacer un seguimiento del comportamiento del estudiante.

Universitat d'Alacant
Universidad de Alicante

La investigación de las rutas de aprendizaje que siguen los estudiantes cuando navegan por plataformas educativas virtuales es un tema emergente y que genera desafíos en el campo de la investigación educativa y la minería de procesos. Existen trabajos que se han orientado a analizar dichas rutas de aprendizaje con varios objetivos, por ejemplo:

- Identificar patrones o estilos de aprendizaje.
- Realizar propuestas de diseño instruccional.
- Realizar propuestas de sistemas de aprendizaje autoregulado.
- Identificar razones de deserción estudiantil en entornos virtuales.

La identificación de puntos críticos en las rutas de aprendizaje constituye un desafío latente y abre una enorme oportunidad para generar aportes entorno a la calidad de la educación en entornos virtuales e híbridos. En la medida en la que se incremente la investigación orientada a identificar estos puntos críticos de aprendizaje, mayor será el aporte formativo de los cursos virtuales o híbridos. El estudio de técnicas de analítica web y minería de procesos aplicados al ámbito educativo permitirá abrir un abanico extenso de posibilidades de investigación, permitiendo así extrapolar dichas técnicas hacia contextos educativos particulares.

Una línea de investigación que se encuentra abierta es el análisis de comentarios en redes sociales, hemos iniciado el estudio de esta temática con el trabajo presentado por Buenaño-Fernández, Villegas-Ch, y Luján-Mora (2018). El alto número de estudiantes matriculados en entornos de aprendizaje mediados por tecnología y su interacción constante con las plataformas genera una gran cantidad de datos que es difícil de manejar con los métodos tradicionales de análisis de datos. Esta interacción se basa en entradas incluidas en foros de discusión, correos electrónicos o interacción en redes sociales. En este artículo, exploramos la interacción de los estudiantes a través de técnicas de minería de texto en diferentes entornos de interacción estudiantil en un curso MOOC. La investigación se centró en el cálculo y análisis de la frecuencia de los términos, el análisis de concordancias y agrupaciones en n-gramas.

Universidad de Alicante

Referencias

Akram, M., y Zepeda, S. (2015). Development and Validation of a Teacher Self-assessment Instrument. *Research and Reflections in Education*, *9*(2), 134–148. (citado en las páginas 16)

Baker, R. S., y Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. En *Learning analytics* (pp. 61–75). New York, NY: Springer New York. doi: 10.1007/978-1-4614-3305-7_4 (citado en las páginas 3, 4, 5)

Beck, J. E., Chang, K. M., Mostow, J., y Corbett, A. (2008). Does help help? Introducing the bayesian evaluation and assessment methodology. En *Proceedings of intelligent tutoring systems* (pp. 383–394). doi: 10.1007/978-3-540-69132-7-42 (citado en las páginas 3)

Bogarín, A., Cerezo, R., y Romero, C. (2017). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1), 1–17. doi: 10.1002/widm.1230 (citado en las páginas 26)

Buenaño-Fernández, D., y Luján-Mora, S. (2016). Exploring approaches to educational data mining and learning analytics to measure the level of acquisition of student's learning outcomes. En *Proceedings of 8th Conference on Education and New Learning Technologies (EDULEARN16)* (pp. 1845–1850). doi: 10.21125/edulearn.2016.1368 (citado en las páginas 4, 9, 20, 35, 41, 47)

Buenaño-Fernández, D., Gil, D., y Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability (Switzerland)*, *11*(10), 1–18. doi: 10.3390/su11102833 (citado en las páginas 9, 13, 20, 22, 34, 36, 41, 63, 77)

Buenaño-Fernández, D., Luján-Mora, S., y Gil, D. (2019). A hybrid machine learning approach for the prediction of grades in computer engineering students. En *The International Research & Innovation Forum (RIIFORUM)* (pp. 125–134). doi: 10.1007/978-3-030-30809-4_13 (citado en las páginas 11, 20, 36, 41, 59)

Buenaño-Fernández, D., Luján-Mora, S., y Villegas-Ch, W. (2017). Improvement of massive open online courses by text mining of students' emails: A case study. En *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM)* (pp. 1–7). doi: 10.1145/3144826.3145393

(citado en las páginas 11, 25, 36, 41, 67, 78)

Buenaño-Fernández, D., Villegas-Ch, W., y Luján-Mora, S. (2018). Using text mining to evaluate student interaction in virtual learning environments. En *Proceedings of the 2dn IEEE World Engineering Education Conference (EDUNINE)* (pp. 1–6). IEEE. doi: 10.1109/EDUNINE.2018.8450969 (citado en las páginas 82)

Buenaño-Fernández, D., y Luján-Mora, S. (2019). Proposal of application of process mining to evaluate the learning routes of students with visual disability in online courses. *Revista Ibérica de Sistemas y Tecnologias de Información (RISTI)*, *E17*(1), 1035–1047. doi: 10.1007/978-3-030-30809-4_13 (citado en las páginas 11, 14, 16, 26, 27, 28, 34, 36, 41, 55, 77)

Buenaño-Fernández, D., Villegas-Ch, W., y Luján-Mora, S. (2019). The use of tools of data mining to decision making in engineering education—A systematic mapping study. *Computer Applications in Engineering Education*, *27*(3), 744–758. doi: 10.1002/cae.22100 (citado en las páginas 5, 9, 12, 21, 27, 33, 36, 41, 43, 79, 81)

Buenaño-Fernández, D., Gonzalez, M., Gil, D., y Luján-Mora, S. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *IEEE Access*, *8*(1), 35318–35330. doi: 10.1109/ACCESS.2020.2974983 (citado en las páginas 11, 16, 21, 24, 25, 34, 36, 41, 71, 77, 78)

Buenaño-Fernández, D., y Luján-Mora, S. (2017). Comparison of applications for educational data mining in Engineering Education. En *Proceedings of the 1st IEEE World Engineering Education Conference (EDUNINE)* (pp. 81–85). doi: 10.1109/EDUNINE.2017.7918187 (citado en las páginas 9, 28, 35, 41, 51)

Cawley, G. C., y Talbot, N. L. (2003). Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognition*, *36*(11), 2585–2592. doi: 10.1016/S0031-3203(03)00136-5 (citado en las páginas 21)

Conde, M. A., Colomo-Palacios, R., García-Peñalvo, F. J., y Larrucea, X. (2018). Teamwork assessment in the educational web of data: A learning analytics approach towards ISO 10018. *Telematics and Informatics*, *35*(3), 551–563. doi: 10.1016/j.tele.2017.02.001 (citado en las páginas 20)

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304. doi: 10.1504/IJTEL.2012.051816 (citado en las páginas 4)

Johnson, L., Samantha, A., Michele, C., Victora, E., Alex, F., y Courtney, H. (2016). *Nmc horizon report: 2016 higher education*. The New Media Consortium. (citado en las páginas 3)

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., y Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*(1), 74–85. doi: 10.1016/j.iheduc.2017.02.001 (citado en las páginas 15, 16)

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., y Yang, S. J. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology and Society*, *21*(2), 220–232. (citado en las páginas 13)

Maldonado-Mahauad, J., Hilliger, I., Pérez-Sanagustín, M., Millecamp, M., Verbert, K., y Ochoa, X. (2018). The LALA Project: Building Capacity to Use Learning

Analytics to Improve Higher Education in Latin America. En *Proceedings of the 8th International Learning Analytics & Knowledge Conference (LAK)* (pp. 630–637). (citado en las páginas 3)

Menchaca, I., Guenaga, M., y Solabarrieta, J. (2016). Using learning analytics to assess project management skills on engineering degree courses. En *Proceedings of the 4th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM)* (pp. 369–376). doi: 10.1145/3012430.3012542 (citado en las páginas 19)

Polyzou, A., y Karypis, G. (2016). Grade prediction with models specific to students and courses. *International Journal of Data Science and Analytics*, *2*(3-4), 159–171. doi: 10.1007/s41060-016-0024-z (citado en las páginas 13)

Ray, S., y Saeed, M. (2018). Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education. , *5*(4), 135–160. doi: 10.21917/ijsc.2015.0145 (citado en las páginas 25)

Romero, C., Ventura, S., y García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. doi: 10.1016/j.compedu.2007.05.016 (citado en las páginas 3, 24)

Ross, J. A., y Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, *23*(2), 146–159. doi: 10.1016/j.tate.2006.04.035 (citado en las páginas 16)

Siemens, G. (2012). Learning analytics: envisioning a research discipline and a domain of practice. En *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK)* (pp. 4–8). doi: 10.1145/2330601.2330605 (citado en las páginas 4)

Sin, K., y Muthu, L. (2015). Application of Big Data in Education Data Mining and Learning Analytics a Literature Review. *Journal on Soft Computing: Special issue on Soft Computing models for Big Data*, *5*(4), 1035–1049. doi: 10.21917/ijsc.2015.0145 (citado en las páginas 3, 4, 24)

Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., ... Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *The Lancet Child & Adolescent Health*, *4*(5), 397–404. doi: 10.1016/S2352-4642(20)30095-X (citado en las páginas 5)

Weiss, S. M., Indurkhya, N., y Zhang, T. (2010). *Fundamentals of Predictive Text Mining*. London: Springer London. doi: 10.1007/978-1-84996-226-1 (citado en las páginas 25)