

Establecimiento del estado del arte sobre la Minería de Datos Educativa en el Nivel Superior: Un Estudio de Mapeo Sistemático

Establishment of the state of the art on Educational Data Mining in the Higher Level: A Systematic Mapping Study

Marisa D. Panizzi¹

¹Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. Universidad de Morón. Argentina.

Manuscrito recibido: 5 de febrero de 2019; aceptado para publicación: 9 de abril de 2019

Contacto: Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. Universidad de Morón. Cabildo 134, (1708) Morón, Provincia de Buenos Aires. Argentina.

E-mail: mpanizzi@unimoron.edu.ar

Resumen

Los procesos de minería de datos permiten a las organizaciones descubrir conocimiento para la toma de sus decisiones. En los últimos años, se comenzó a aplicar la minería de datos en el dominio educativo, con el propósito de resolver diferentes tipos de problemas; como por ejemplo, la deserción y desgranamiento, el rendimiento académico de los alumnos, entre otros. Se presenta la aplicación del método de investigación, mapeo sistemático de la literatura. Este permite cotejar y sistematizar la evidencia empírica respecto a la aplicación de la minería de datos en el contexto educativo de Nivel Universitario. Se logró determinar cuáles son las metodologías utilizadas para los proyectos de minería de datos, las herramientas de trabajo y los lenguajes de programación que se emplean, como así también los algoritmos para la construcción de los modelos.

Palabras clave: Minería de Datos Educativa, Educación Superior, Estudio de Mapeo Sistemático

Abstract

Data mining processes allow organizations to discover knowledge for decision making. In recent years, data mining began to be applied in the educational domain, with the purpose of solving problems of different types of problems; as for example, the desertion, the academic performance of the students, among others. The application of the research method is presented, systematic mapping of the literature. This allows collating and systematizing the empirical evidence regarding the application of data mining in the educational context of the University Level. It was possible to determine which are the methodologies used for the data mining projects, the working tools and the programming languages that are used, as well as the algorithms for the construction of the models.

Keywords: Educational Data Mining, Higher Education, Systematic Mapping Study

Introducción

La Minería de Datos Educativos (*Educational Data Mining - EDM*) como exponen Romero *et al.* (Romero, C. *et al.*, 2010) es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizarla para mejorar dicho entorno. La información así obtenida se convierte en el insumo indispensable para la toma de decisiones.

En el contexto internacional, se han encontrado publicaciones previas, las cuales han demostrado que la minería de datos se puede utilizar para detectar alumnos en riesgo de deserción. En su trabajo, Luan aplicó exitosamente técnicas de minería de datos para predecir qué grupos de alumnos podrían abandonar los estudios (Luan, J., 2002). En un trabajo relacionado, Lin usó técnicas de minería de datos para optimizar los esfuerzos para retener estudiantes. (Lin, S., 2012). Investigadores de la Universidad Estatal de Bowie, Chacon, F. *et al.*, desarrollaron un sistema basado en minería de datos que permite a dicha institución, identificar y atender a estudiantes en riesgo de abandono de estudios (Chacon, F., 2012).

En esta comunicación, se presenta el desarrollo de un mapeo sistemático de la literatura (en inglés *Systematic Mapping Studies* o SMS). El SMS permite cotejar y sistematizar la evidencia empírica de la aplicación de la minería de datos en el contexto educacional de Nivel Superior. Para su desarrollo, se utiliza el proceso propuesto por Kitchenham *et al.* (Kitchenham, 2007) y Genero *et al.* (Genero *et al.*, 2014). El proceso propone tres actividades: la planificación de la revisión, su ejecución y por último el reporte de la revisión; las cuales serán presentadas en la sección Desarrollo.

Desarrollo

En esta sección se desarrolla el SMS, la actividad de planificación se presenta en la sección 1, la actividad de ejecución en la sección 2 y la actividad de reporte se sintetiza en la sección 3.

1. Actividad de Planificación del SMS.

En la actividad "Planificación" del SMS, se realizan las siguientes tareas: a) identificar la necesidad del estudio, se continua por b) definir las preguntas de investigación que se busca contestar, luego la c) definición del protocolo de la revisión y por último, d) la validación.

a) *Identificación de la necesidad del estudio.*

Dado que en la Universidad de Morón se desarrolla un proyecto de investigación titulado: "Aplicación de tecnologías inteligentes de explotación de información para el análisis de perfiles de tesis de grado de carreras informáticas de la UM", en este contexto se requiere la construcción del estado del arte respecto a la minería de datos educacional en el nivel de Educación Superior, mediante un método que permite clasificar y analizar la literatura sobre un tema específico de la Ingeniería de software, como es el estudio de mapeo sistemático (Kitchenham *et al.*, 2016), se lo utiliza para identificar cómo se aplica la minería de datos en el entorno educacional de nivel superior.

b) *Definición de las preguntas de investigación.*

Se intenta identificar a los grupos, institutos o laboratorios de investigación que trabajan en minería de datos para resolver problemas en el contexto de la Educación Superior, en qué tipo de problemas se pone foco, las metodologías o procesos más empleados y las herramientas o lenguajes de programación que se utilizan.

A continuación se detallan las preguntas de investigación (PI) y su Motivación (MO):

PI1: ¿Qué se intenta resolver con EDM (*Educational Data Mining*)?

MO1: Descubrir qué se resuelve con minería de datos en el contexto educacional.

PI2: ¿Qué metodologías usan para aplicar minería de datos en instituciones académicas de nivel superior?

MO2: Identificar las metodologías más utilizadas en las instituciones académicas de nivel superior.

PI3: ¿Qué herramientas de trabajo y lenguajes de programación se utilizan para realizar minería de datos?

MO3: Descubrir las herramientas y lenguajes de programación más empleados.

PI4: ¿Qué algoritmos se aplican?

MO4: Determinar qué algoritmos de minería de datos son los más utilizados para resolver problemas en educación de nivel superior.

c) *Definición del Protocolo de la revisión.*

En esta tarea, se describen los elementos del protocolo para llevar a cabo el SMS, la definición de las estrategias

de búsqueda y la definición de los criterios de selección de estudios.

En la definición de las estrategias de búsqueda, se deciden los términos candidatos (T) para llevar a cabo la búsqueda como así también las cadenas de búsqueda a emplear. A continuación, se presentan los términos de búsqueda:

- T1: Minería de datos
- T2: *Data Mining*
- T3: Educación Superior
- T4: *Higher Education*
- T5: Minería de Datos Educacional
- T6: *Educational Data Mining*

Hay términos que se utilizarán como cadenas de búsqueda dado que con ellos se pueden encontrar estudios, estos casos son: Minería de Datos Educacional y *Educational Data Mining*.

Las cadenas de búsqueda a utilizar son:

C1: *Data Mining and Higher Education*

C2: *Educational and Data Mining*

C3: *Educational and Data Mining AND Higher Education*

C4: Minería de Datos y Educación Superior

Dentro de los criterios de selección de estudios se definen los criterios de inclusión y exclusión de los estudios primarios. El objetivo que se persigue es identificar cuáles de ellos muestran evidencia con respecto a las preguntas de

investigación propuestas en la sección b). Para este SMS, se definen los siguientes criterios de inclusión y exclusión:

- Criterios de inclusión:
 - Ámbito: académico universitario
 - Período de búsqueda: a partir del 2008
 - Idioma: Español e Inglés
 - Artículos primarios
- Criterios de exclusión:
 - Artículos que no estén accesibles para su lectura completa.
 - Artículos que correspondan al entorno educativo pero no universitario, por ejemplo, artículos que traten la educación primaria o secundaria.

Se decide la búsqueda en las bibliotecas digitales y repositorios que se presentan en la **Tabla I**. Dado que el proyecto de investigación se desarrolla en la República Argentina, el grupo de investigación decidió considerar las publicaciones del Congreso Argentino de Ciencias de la Computación (CACIC), Congreso Argentino de Tecnología en Educación y Educación en Tecnología (TE&ET), Workshop de Investigadores en Ciencias de la Computación (WICC) y la Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología (Revista TR&ET); estas actividades científicas son organizados por la Red de Universidades Nacionales con carreras en Informática, RedUNCI¹.

Tabla I. Fuentes de búsqueda.

Bibliotecas y Repositorios digitales	Opciones
IEEE Explore	Publicaciones de congresos, revistas
Springer	Publicaciones de congresos, revistas
Dialnet	Publicaciones de revistas
Scielo	Publicaciones de Revistas
European Union Digital Library	Publicaciones de Revistas
ScienceDirect	Publicaciones de congresos, revistas
GoogleAcademic	Publicaciones de congresos, revistas
SEDICI ²	Libro de Actas: Congreso Argentino de Ciencias de la Computación (CACIC), Congreso Argentino de Tecnología en Educación y Educación en Tecnología (TE&ET), Workshop de Investigadores en Ciencias de la Computación (WICC) y Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología.

Validación del protocolo de revisión.

Esta actividad se llevó a cabo de manera conjunta con el resto de los investigadores del proyecto.

2.

Actividad "Ejecución del SMS"

a) En esta actividad, se definen las tareas para la ejecución de la revisión de la literatura: a) selección de los estudios primarios, b) evaluación de la calidad de los artículos, c) extracción de datos relevantes y d) síntesis de los datos extraídos.

Selección de los estudios primarios.

- En esta tarea, se realiza la selección de los estudios considerando algunos criterios utilizados en el estudio realizado por Sierra *et al.* (Sierra *et al.*, 2017), estos consisten en:

- aplicar los procedimientos definidos para la selección (estrategias de búsqueda, criterios de inclusión/exclusión),
- los resultados obtenidos en las fuentes de búsqueda,

extracción de datos, clasificación de los diferentes documentos.

- La cantidad de artículos seleccionados en cada una de las bases de datos se presentan en la **Tabla II**. Los mismos son clasificados según la propuesta de Panizzi (Panizzi, 2019) en: artículos relevantes, artículos no considerados y artículos primarios. El significado de esta clasificación se explica a continuación:

- **Artículos relevantes:** son los encontrados y seleccionados de los resultados de búsqueda iniciales que contienen los términos de búsqueda en el resumen, la introducción, las palabras clave o título.

- **Artículos no considerados:** son los artículos resultantes de la búsqueda que respetan los criterios de inclusión/exclusión y luego de su lectura no se consideran adecuados para la investigación.

Artículos primarios: son los artículos que han sido leídos de manera completa y se consideran adecuados para la investigación porque cumplen con los criterios de inclusión.

Tabla II. Base de datos y trabajos seleccionados para cada etapa.

Fuentes de búsqueda	Artículos relevantes	Artículos no considerados	Artículos primarios
IEEE Explore	61	14	47
Springer	2	2	0
Dialnet	12	5	7
Scielo	3	2	1
European Union Digital Library	2	1	1
ScienceDirect	7	5	2
GoogleAcademic	65	17	48
SEDICI	8	2	6

La cantidad de artículos encontrados ha sido 160 (artículos relevantes), el total de artículos no considerados ha sido 48 y la cantidad de artículos primarios ha resultado un total de 112.

En el Apéndice A, se presenta la lista de estudios primarios

utilizados para el SMS.

b) *Evaluación de la calidad de los artículos*

Para determinar la calidad de los artículos se adhiere al criterio que los estudios primarios se encuentran publicados en revistas indexadas y en eventos científicos con revisión

por pares según la propuesta de Genero *et al.* (Genero *et al.*, 2014).

c) *Extracción de datos relevantes.*

Los datos extraídos de los estudios primarios se registraron en una planilla de cálculo. El esquema de extracción se compone de dos partes:

- La primera con los metadatos de cada estudio primario, denominada Dimensión - Datos Generales. Las categorías son:

- ID: identificador de registración del artículo, cadena de búsqueda, Año, Título, Autor(es), Fuente (Nombre del Congreso o Revista), Tipo de publicación (R=Revista, C=Congreso), Fuente de búsqueda, País, Palabras clave, Cita (APA), Cantidad de citas, Problema, Propuesta, Resultados.

- La segunda parte, compuesta por dimensiones y categorías específicas del esquema definido para clasificar los estudios primarios seleccionados. Las dimensiones consideradas son: Metodologías, Contribución, Propósito, Algoritmos y Herramientas y lenguajes de programación. La definición de estas dimensiones y sus categorías se define en función de las preguntas de investigación definidas en la Actividad "Planificación de la Revisión". A continuación se detallan las categorías de cada una de las dimensiones definidas:

- Dimensión - Metodologías: KDD³, CRISP-DM⁴, SEMMA⁵, No menciona

- Dimensión - Contribución:

- Rendimiento académico
- Comportamiento estudiantil
- Deserción estudiantil
- Análisis/Clasificación de perfiles
- Calidad Educativa
- Procrastination Estudiantil
- Diferencias de género en postulantes universitarios
- Empleabilidad estudiantil
- Educación cooperativa
- No menciona
- Dimensión - Propósito:
- Evaluar
- Validar
- Proponer una solución
- Informar una experiencia

- Comparar

- Opinar

- Dimensión - Algoritmos: en esta dimensión se consideran los algoritmos para minería de datos (J48, ID3, K-means, etc.). Es importante aclarar que la inclusión de una categoría nueva en la dimensión, se ha realizado en función de su aparición en los estudios relevantes.

- Dimensión - Herramientas y lenguajes de programación: en esta dimensión se consideran las herramientas y los lenguajes de programación utilizados para los procesos de minería de datos (WEKA, Rapid Miner, Knime, Orange, Python, etc.). Es importante aclarar que la inclusión de una categoría nueva en la dimensión, se ha realizado en función de su aparición en los estudios relevantes.

d) *Síntesis de los datos extraídos.*

Para resumir los datos extraídos de los artículos, se realiza una síntesis cuantitativa del SMS, mediante el uso de gráficos y luego se realiza un análisis de los artículos primarios obtenidos para dar respuesta a cada una de las preguntas de investigación.

A continuación se presenta, la síntesis cuantitativa; en la **Fig. 1**, se muestra la cantidad de artículos primarios por el año de publicación.

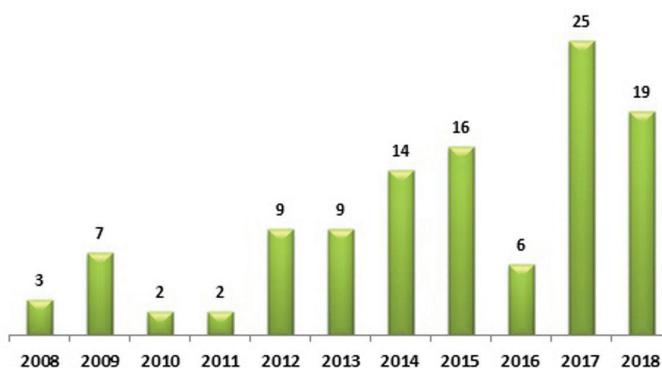


Figura 1. Cantidad de artículos primarios por año de publicación.

En la **Fig. 2**, se presenta que la mayoría de artículos encontrados han sido publicados en congresos, con un total de 75 artículos y los artículos publicados en revistas, con un total de 37 artículos.

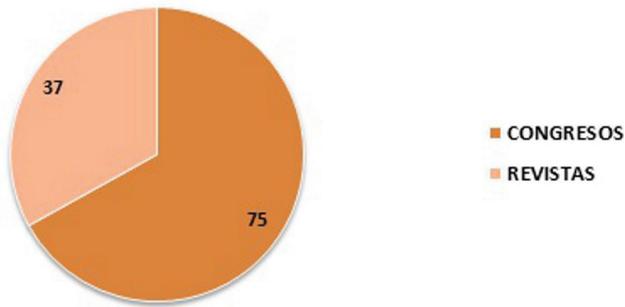


Figura 2. Artículos por tipo de publicación.

En la Fig. 3, se presenta una distribución de los congresos respecto a los continentes en los que han sido realizados. Esta estadística surge de los estudios encontrados por el grupo de investigación en las distintas fuentes de búsqueda, aunque la misma presenta una cantidad de artículos por continentes, se visualiza una tendencia incremental en los estudios realizados en esta línea de investigación. La mayor cantidad de los artículos corresponden a los congresos

realizados en el Continente Asiático (40 artículos), en un segundo lugar se encuentra en América del Norte (16 artículos) y en un tercer lugar el Continente Europeo (14 artículos).

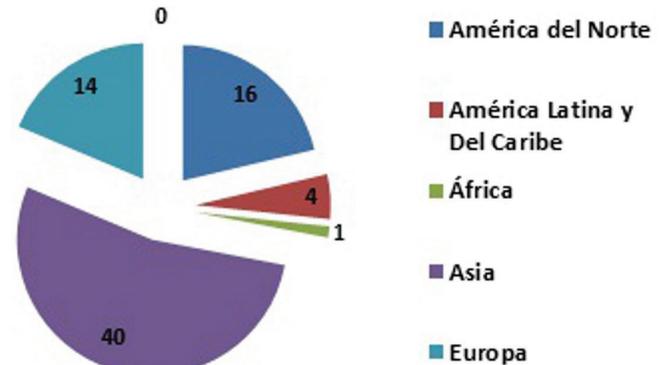


Figura 3. Distribución de los congresos respecto al continente de realización.

Se presenta en la Fig. 4, la cantidad de artículos primarios clasificados por las revistas en las que han sido publicados.

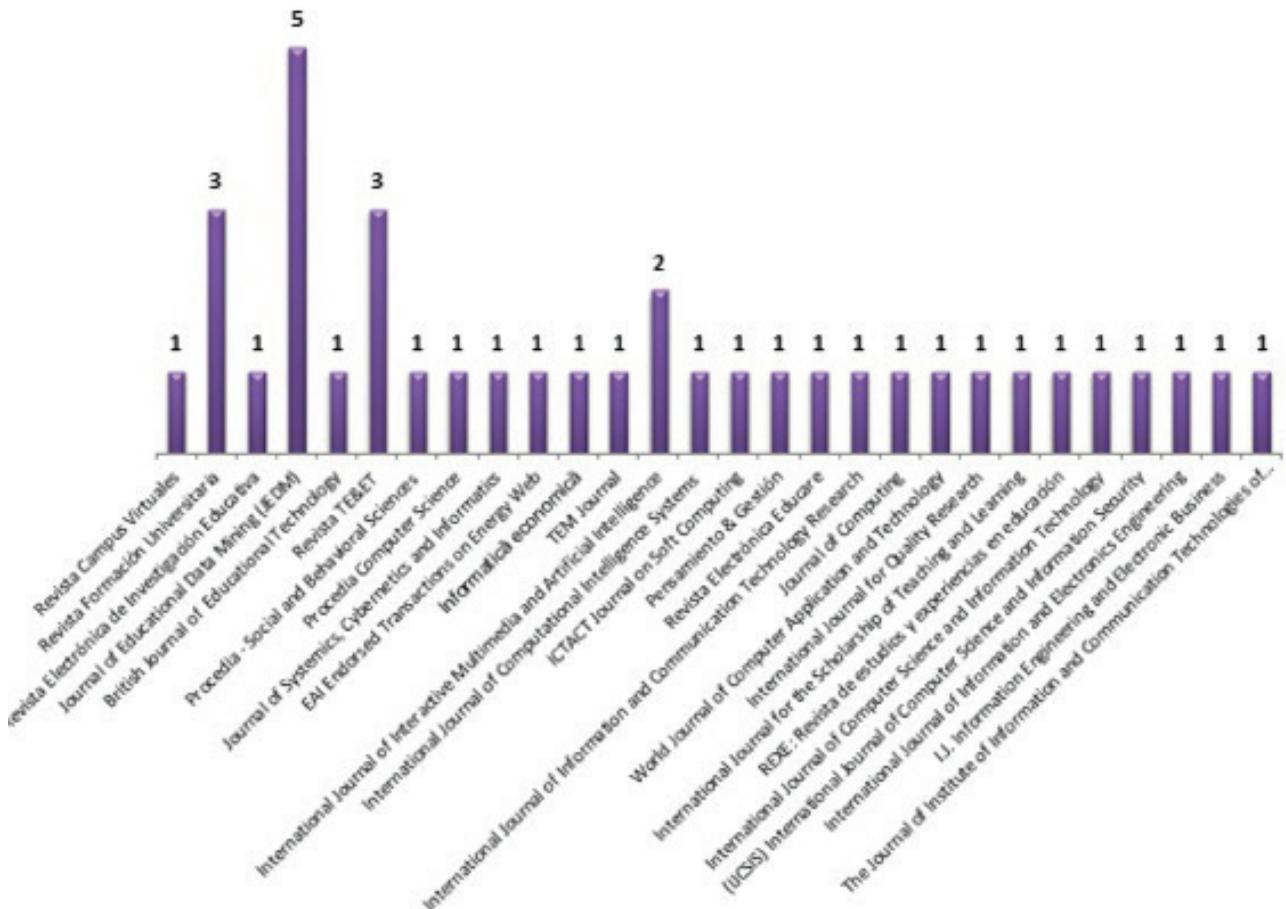


Figura 4. Revistas con las publicaciones de los estudios primarios.

A continuación, se da respuesta a cada una de las preguntas de investigación que motivaron la realización del SMS.

PI1: ¿Qué se intenta resolver con EDM (Educational Data Mining)?

En la Fig. 5, se visualizan los diferentes tipos de contribuciones de la minería de datos en la educación superior. La mayor cantidad de problemas que se intenta resolver con la minería de datos son de rendimiento académico, con una distribución porcentual del 38 % de los estudios primarios analizados. En segundo lugar, con una distribución porcentual del 20 %, se analizan y se clasifican los perfiles de los estudiantes. Un 17 % del total de los estudios utilizan los datos con el propósito de mejorar la calidad educativa y un 11 % de los estudios se focalizan en resolver problemas de deserción de los estudiantes. En la franja de porcentajes inferiores al 12 % del total de los estudios, se menciona la resolución del resto de tipos de problemas como así también, se ubican los estudios que no mencionan el tipo de problemas que se han resuelto.

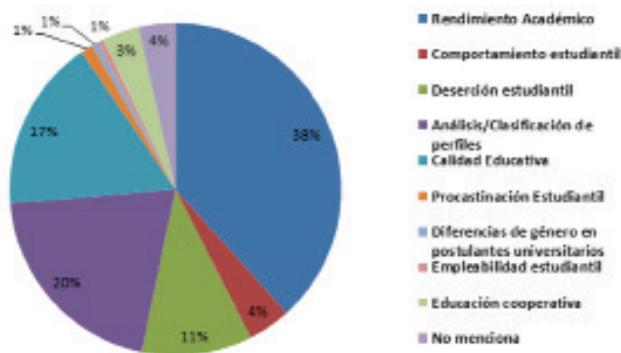


Figura 5. Contribuciones de la Minería de Datos.

PI2: ¿Qué metodologías se usan para aplicar minería de datos en instituciones académicas de nivel superior?

La mayor cantidad de los artículos primarios analizados (86 artículos) no presentan la metodología que emplearon para llevar a cabo el proceso de minería de datos. En los trabajos que hacen mención a las metodologías utilizadas, se visualiza la cantidad de 19 artículos en los cuales han

utilizado KDD, 7 artículos en los cuales han empleado CRISP-DM y en ningún trabajo se ha utilizado la metodología SEMMA. La clasificación metodológica se presenta en la Fig. 6.

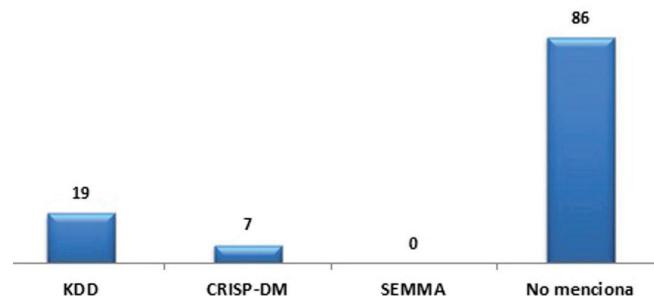


Figura 6. Metodologías utilizadas para proyectos de minería de datos.

PI3: ¿Qué herramientas de trabajo y lenguajes de programación se utilizan para realizar minería de datos?

En la Fig. 7, se visualiza que en los artículos primarios que mencionan la utilización de herramientas, la más utilizada es Weka, con un total de 34 artículos. Luego, siguen Rapid Miner con una aparición en 8 artículos, SPSS se menciona en 5 artículos, y el lenguaje de programación Python, con una aparición en 5 artículos.

El resto de las herramientas para minería de datos no denotan una aparición significativa ya que oscila entre 1 y 2 artículos. De los artículos analizados, en un total de 49 artículos no se mencionan las herramientas empleadas para los estudios realizados.

Figura 8. Algoritmos utilizados en los artículos analizados.

3. Actividad "Reporte del SMS".

El desarrollo del mapeo sistemático de la literatura permitió dar respuesta a las preguntas de investigación. Se lograron identificar los grupos, institutos o laboratorios de investigación que trabajan en minería de datos para resolver problemas en el contexto de la Educación Superior, en qué tipo de problemas se pone foco, las metodologías o procesos más empleados como así también las herramientas o lenguajes de programación que se utilizan.

Al analizar los datos extraídos de los artículos primarios, se comprueba que:

- Los problemas que se resuelven con la aplicación de procesos de minería de datos son de rendimiento académico en un primer lugar, con el objetivo de estudiar el desempeño de un estudiante universitario, ya sea durante en el transcurso o final de una carrera. En un segundo lugar, se aplica con el propósito de clasificar a los estudiantes y poder realizar un análisis sobre los distintos grupos por sus similitudes y diferencias entre ellos. En un tercer lugar, se utiliza para analizar políticas institucionales universitarias, a partir del conocimiento obtenido de la institución como herramienta para la toma de decisiones estratégicas.
- De las metodologías empleadas, se denota que el proceso KDD es el más utilizado y en segundo orden, CRISP-DM.
- De las herramientas y lenguajes de programación que se utilizan para minería de datos, la herramienta más utilizada es Weka. A continuación, Rapid Miner y SPSS. En el grupo de los lenguajes de programación, se destaca el uso de Python para este tipo de proyecto.
- Del uso de los algoritmos para la construcción de los modelos, se denota un uso significativo de las redes bayesianas, en particular el modelo Naive Bayes. De esto se deduce una fuerte tendencia hacia la construcción de modelos predictivos. En un segundo lugar, se observa el uso de algoritmos de *clustering* en general (sin especificar los algoritmos en particular). En un tercer lugar, se menciona el uso del algoritmo J4.8 y los árboles de decisión (sin especificar los algoritmos). El uso de algoritmos de

clustering y de árboles de decisión muestran la tendencia hacia la construcción de modelos descriptivos también.

Discusión y Conclusiones

Se presentó un proceso de revisión sistemática de la literatura mediante un método de investigación de Ingeniería de software, mapeo sistemático de la literatura (SMS). Este permitió sistematizar la evidencia empírica la aplicación de minería de datos educacional en el Nivel de Educación Superior. Se lograron identificar los tipos de problemas que se intentan resolver con los procesos de minería de datos, se pueden mencionar el rendimiento académico de los estudiantes, la deserción y el desgranamiento, calidad de las estrategias educativas, entre otros.

Se evidenció que la metodología más utilizada para este tipo de proyecto es KDD.

Las herramientas de trabajo y lenguajes de programación empleados son Weka, Rapid Miner y Python.

Se observó que se utilizaron algoritmos para resolver problemas descriptivos como predictivos.

El desarrollo del SMS contribuyó a la definición de un área de investigación, la cual permite dar respuesta a la problemática que ocurre en las cátedras de tesis de las carreras de Informática de la Universidad de Morón. Se ha observado que el mayor inconveniente que posee el alumno al comenzar la materia es la definición del tema, ocasionando un retraso en la finalización de sus estudios, y en algunos casos el abandono de la carrera en su última materia. De esto surge la necesidad de definir: ¿Cuáles son los patrones de comportamiento de los tesis de las carreras de grado de informática en la UM (Universidad de Morón)?

Otros de los aportes del SMS han sido la definición de la metodología a emplear como así también la elección de las herramientas de trabajo. Se optó por emplear el proceso KDD (Hernández Orallo *et al.*, 2004) y como herramienta de trabajo WEKA (WEKA, 2019) y *Rapid Miner (Rapid Miner Studio, 2019)*.

Agradecimientos

La investigación que se reporta en este artículo es financiada por el Proyecto de Investigación titulado: "Aplicación de tecnologías inteligentes de explotación de información para el análisis de perfiles de tesis de grado

de carreras informáticas de la UM” (Código 17/01-MP-001) de la Secretaría de Ciencia y Tecnología de la Universidad de Morón.

Agradezco también a los tesisistas de la carrera Licenciatura en Sistemas de la UM por su colaboración en la búsqueda de artículos, Lucila Mira, Gastón Álvarez, Matías García y Gabriel Mariuz y al grupo de investigación con quien se realizó la validación del protocolo de revisión y la extracción de datos, Iris Sattolo, Javier Lafont y Nicolás Armilla.

Referencias Bibliográficas

- Chacon F., Spicer D. and Valbuena A. (2012). “Analytics in Support of Student Retention and Success”. Research Bulletin. Louisville, CO: EDUCAUSE Center for Analysis and Research. Vol 3. Available from <http://www.educause.edu/ecar>.
- Genero Bocco M., Cruz-Lemus J.A. y Piattini Velthuis M. (2014). Métodos de investigación en ingeniería del software. Madrid, España: Editorial Ra-Ma.
- Hernández Orallo J., Ramírez Quintana M.J., Ferri Ramírez C. (2004). Introducción a la Minería de Datos. Ed. Pearson Educacion S.A. Madrid.
- Kitchenham, B. and Charters, S. (2007). Guidelines for Performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.
- Kitchenham, B, Budgen, D., Brereton, P. (2016). Evidence-Based Software Engineering and Systematic Reviews. CRC Press.
- Lin, S. H. (2012). Data mining for student retention management. Journal of Computing Sciences in Colleges, 27(4), 92-99.
- Luan, J. (2002). Data mining and its applications in higher education. New Directions for Institutional Research 2002 (113): 17-36.
- Panizzi M. (2019). Establecimiento del estado sobre el proceso de implantación de sistemas informáticos. Reporte Técnico de Tesis Doctoral - RT-UNLP- III-LIDI. Publicado en SEDICI: <http://sedici.unlp.edu.ar/>
- Rapid Miner Management Team (S/A). Rapid Miner Studio. Página Web: <https://rapidminer.com/products/studio/>. Disponible on line en febrero de 2019.
- Romero C., Ventura S., Pechenizkiy M. y Baker R. S. de J. (eds.). Handbook of Educational Data Mining. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, 2010. ISBN: 1-43980-457-5.
- Sierra J.M., Vizcaíno A., Genero M., Piattini M. (2018). A systematic mapping study about socio-technical congruence. Information and Software Technology, 94: 111-129. DOI: <https://doi.org/10.1016/j.infsof.2017.10.004>.
- Weka. University of Waikato. Machine Learning Group. Página web: www.cs.waikato.ac.nz/ml/weka/downloading.html. Disponible online en febrero 2019.

(Endnotes)

- 1 RedUNCI: Red de Universidades Nacionales con carreras en Informática, sitio: <http://redunci.info.unlp.edu.ar/>
- 2 SEDICI: Repositorio digital de la Universidad Nacional de La Plata, en el cual se publican las Actas del Congreso Argentino de Ciencias de la Computación y del Congreso Argentino de Tecnología de la Educación y Educación en Tecnología, sitio: <http://sedici.unlp.edu.ar/>
- 3 KDD en Inglés Knowledge Discovery in Data bases, “Descubrimiento de conocimiento en bases de datos”.
- 4 CRoss Industry Standard Process for Data Mining.
- 5 Metodología definida por SAS Institute Inc.