

# 41

Fecha de presentación: julio, 2022  
Fecha de aceptación: octubre, 2022  
Fecha de publicación: diciembre, 2022

## LA ANALÍTICA ACADÉMICA

Y LA MINERÍA DE DATOS EDUCACIONAL EN EL NIVEL UNIVERSITARIO: REVISIÓN SISTEMÁTICA

### ACADEMIC ANALYTICS AND EDUCATIONAL DATA MINING AT THE UNIVERSITY LEVEL: A SYSTEMATIC REVIEW

Mariana Chavarry Chankay<sup>1</sup>

E-mail: [mchavarry@usat.edu.pe](mailto:mchavarry@usat.edu.pe)

ORCID: <https://orcid.org/0000-0001-5136-7177>

Jury Yesenia Aquino Trujillo<sup>1</sup>

E-mail: [jaquino@usat.edu.pe](mailto:jaquino@usat.edu.pe)

ORCID: <https://orcid.org/0000-0003-1662-6406>

Fiorella Vanessa Li Vega<sup>1</sup>

E-mail: [fli@usat.edu.pe](mailto:fli@usat.edu.pe)

ORCID: <https://orcid.org/0000-0002-1760-4850>

Nilton César Germán Reyes<sup>2</sup>

E-mail: [ngerman@unprg.edu.pe](mailto:ngerman@unprg.edu.pe)

ORCID: <https://orcid.org/0000-0003-0232-2129>

<sup>1</sup> Universidad Católica Santo Toribio de Mogrovejo. Perú

<sup>2</sup> Universidad Nacional Pedro Ruiz Gallo. Perú

#### Cita sugerida (APA, séptima edición)

Chavarry Chankay, M., Aquino Trujillo, J. Y., Li Vega, F. V. & Germán Reyes, N. C.. (2022). La Analítica Académica y la Minería de Datos Educacional en el nivel universitario: revisión sistemática. *Revista Universidad y Sociedad*, 14(S6), 377-390.

#### RESUMEN

En estos últimos años, ha incrementado la tendencia de aplicar la analítica académica y minería de datos en el campo de la educación, abordando distintos temas como la deserción, rendimiento académico de los estudiantes, desempeño de docentes, entre otros. El objetivo del presente artículo, fue indagar por medio de una revisión sistemática las diversas formas de aplicación de la analítica académica y minería de datos educacional en las universidades. Para ello, se realizó búsquedas de artículos en revistas científicas en las bases de datos virtuales más reconocidas, mediante la declaración prisma se sistematizó, obteniéndose 22 artículos que cumplieron con los criterios de inclusión y exclusión establecidos. Como resultado se pudo identificar en qué temas del proceso enseñanza-aprendizaje se enfocan, las metodologías aplicadas en los proyectos de analítica académica y minería de datos educacionales, los programas de aplicación y los lenguajes de programación usados con mayor frecuencia, asimismo, las técnicas empleadas para la implementación de los modelos.

**Palabras clave:** Minería de datos educacional, Analítica académica, Educación universitaria, Proceso enseñanza – aprendizaje.

#### ABSTRACT

In recent years, the trend of applying academic analytics and data mining in the field of education has increased, addressing different issues such as dropout, student academic performance, teacher performance, among others. The objective of this article was to investigate, through a systematic review, the various forms of application of academic analytics and educational data mining in universities. For this, searches for articles in scientific journals were carried out in the most recognized virtual databases, through the prisma statement it was systematized, obtaining 22 articles that met the established inclusion and exclusion criteria. As a result, it was possible to identify which topics of the teaching-learning process are focused on, the methodologies applied in the projects of academic analytics and educational data mining, the application programs and the programming languages used most frequently, as well as the techniques used. for the implementation of the models.

**Keywords:** Educational data mining, academic analytic, university education, teaching-learning process

## INTRODUCCIÓN

En la actualidad todas las organizaciones generan grandes volúmenes de datos, los cuales por sí mismos no generan valor para la organización (Adi & Isnanto, 2020). Las organizaciones que no transforman los datos e información en conocimiento están desperdiciando un activo intangible y valioso que puede generar valor para la organización, ya que le permite a la organización reaccionar y adaptarse al entorno, mejorar sus procesos y productos (Sumpala & Rasyid, 2019). Hoy en día los datos deben convertirse en el activo más importante de las organizaciones, por su acceso rápido, efectivo y preciso, vital para la toma de decisiones estratégicas y la competitividad de toda organización, surgiendo así la teoría de analítica de datos. Esta teoría busca facilitar información objetiva y precisa para soportar la toma de decisiones estratégicas y que permita mantener la competitividad de las organizaciones, mediante la aplicación de modelos y métodos analíticos y predictivos que van más allá de las estadísticas descriptivas. Las universidades no son la excepción, su reto fundamental consiste en enfrentar estos cambios, desarrollando nuevas herramientas de análisis, transformar mentalidades y actitudes que se adapten a las necesidades emergentes basadas en la información y el conocimiento. Los datos se deben convertir en información y la información en conocimiento para la toma de decisiones, generando así "inteligencia accionable" para mejorar la gestión del día al día y la gestión estratégica en la educación superior con miras a mejorar los procesos. Las exigencias hacia la educación superior, fomenta a las universidades hacer uso de la analítica académica, como una nueva oportunidad para cumplir con esta responsabilidad, apoyados en las tecnologías de la información (TI) para que reexaminen sus procesos y herramientas aplicadas en la toma de decisiones (Saiful & Nur, 2020). La analítica académica se define como un proceso que se ocupa de la recopilación, análisis y visualización de las actividades de un programa académico, como cursos, investigación, asignación de recursos y gestión para generar conocimiento institucional y proporcionar a las instituciones de educación superior los datos necesarios para responder a los desafíos y toma de decisiones que enfrentan las universidades en la actualidad.

En países desarrollados existen avances importantes en analítica de aprendizaje, en Latinoamérica en general, los esfuerzos de su adopción en modelos y tecnologías aún es escasa, las instituciones de educación superior deben trabajar para promover la adopción e incorporación de estas herramientas en sus procesos.

La minería de datos es una disciplina de la Informática que abarca técnicas y tecnologías que permiten explorar,

procesar y analizar amplias cantidades de datos con el fin de obtener conocimiento en base a ellos (Riquelme et al., 2006). La aparición de la minería de datos en la década de los 90 está estrechamente ligada a que la concepción que se tenía de los datos empezó a cambiar sustancialmente. En aquel momento, los datos almacenados en las bases de datos de las organizaciones eran el soporte para las operaciones diarias. Pero, esa idea cambió y los datos comenzaron a tomarse como una primordial fuente de conocimiento siendo favorable y beneficioso.

La minería de datos educacional (Educational Data Mining - EDM) es la disciplina en la que se desarrollan métodos para extraer información útil y conocimiento partiendo de los datos generados en los entornos educativos, y aplicarlos para su mejora. Dicho procesamiento de datos es el insumo primordial para la toma de decisiones. (Arevalo-Marín et al., 2019)

Existen estudios previos a nivel internacional, como la investigación realizada por (Lin, 2012) en donde quedó demostrado que la minería de datos puede ser aplicada para predecir a los estudiantes con riesgo de deserción, aplicó acertadamente técnicas de minería de datos para revelar a los estudiantes que podrían desistir de sus estudios. En otra investigación, se aplicó técnicas de minería de datos para optimizar los procesos para evitar que los estudiantes abandonen el centro de estudio. También, se cuenta con el estudio realizado por (Chacón et al., 2012) en cual implementó un software basado en minería de datos en la universidad Bowie State que permitió reconocer a estudiantes con alta probabilidad de desertar de sus estudios.

Como se aprecia en las investigaciones previas, se aborda con éxito estudios de aplicación de la minería de datos educacional orientado solo a la deserción estudiantil, es por ello que se pretende indagar por medio de una revisión sistemática como se está desarrollando la aplicación de la minería de datos educacional en las universidades tomando los diferentes problemas del contexto enseñanza - aprendizaje, en qué temas se enfocan, las metodologías más utilizadas, así como los programas de aplicación y lenguajes de programación que se emplean y además las técnicas aplicadas en los modelos. El objetivo de este artículo ha sido indagar por medio de una revisión sistemática como se está aplicando la analítica académica y minería de datos educacional en las universidades (Chacón et al., 2018)

Las revisiones sistemáticas son resúmenes precisos y organizados de la información dada, con la finalidad de dar respuesta a preguntas específicas. Se caracterizan por detallar el proceso de elaboración de forma clara y

estructurada para buscar, seleccionar, analizar y presentar la evidencia disponible con respecto al tema ligado a las preguntas propuestas (Moreno et al., 2018). Bajo este contexto se plantean las siguientes preguntas de investigación las cuales orientaron la revisión:

P\_01: ¿Cómo ha progresado en el tiempo las investigaciones en la cual se aplica la analítica académica y la minería de datos educacional en el nivel universitario?

P\_02: ¿Qué se pretende solucionar al aplicar la analítica académica y la minería de datos educacional en el proceso enseñanza – aprendizaje en las universidades?

P\_03: ¿Qué metodologías se utilizan para aplicar la analítica académica y la minería de datos educacional en el contexto universitario?

P\_04: ¿Qué técnicas de la analítica académica y la minería de datos se implementan?

P\_05: ¿Qué programas de aplicación y lenguajes de programación se aplican para implementar la analítica académica y la minería de datos educacional?

## DESARROLLO

### MATERIALES Y MÉTODOS

La presente investigación está orientada a realizar una revisión sistemática cuya búsqueda de información se realizó entre los meses de setiembre del 2021 y diciembre del 2021. Se detalla el método de trabajo desarrollado:

Términos de búsqueda

Se dio inicio con búsquedas previas con el fin de identificar los términos contenidos en el título, palabras claves y resumen del presente estudio, se consideraron términos principales y alternativos o sinónimos tanto en español y en inglés como se aprecia en la Tabla 1.

Tabla 1. Términos de búsqueda.

Término principal en español	Término alternativo en español	Término principal en inglés	Termino alternativo en inglés
Minería de datos educacional	Minería de datos	Educational data mining	Data mining
Nivel universitario	Educación universitaria. Educación superior. Universidad.	University level	University education. Higher education. University.
Analítica académica	Analítica de datos	Academic analytics	Data analytics

Fuente: elaboración propia

### Estrategia de búsqueda

Luego de identificar los términos principales y alternativos, se procedió a seleccionar las bases de datos virtuales más relevantes y con mayor impacto en donde se realizaron las búsquedas generales.

En el presente estudio se ha tomado como búsquedas generales aquellas que considera los términos principales y alternativos a partir del año 2010 en adelante con el fin de obtener la evolución y la tendencia de las investigaciones sobre los términos en estudio.

Para optimizar las búsquedas generales, se aplicó estrategias de búsquedas avanzadas para la cual se consideró por cada base datos virtual, combinaciones de los términos generando cadenas de búsqueda incluyendo operadores lógicos (AND / OR), además se tomaron en cuenta criterios de inclusión y exclusión tal como se presenta en la Tabla 2.

Tabla 2. Estrategias de búsqueda.

Bases de datos virtual	Variables de búsqueda con operadores lógicos / Criterios de inclusión y exclusión
Scopus	TITLE-ABS-KEY ( educational AND data AND mining ) AND TITLE-ABS-KEY ( university ) AND ( LIMIT-TO ( FREETOREAD , "all" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( SUBJAREA , "SOCl" ) OR LIMIT-TO ( SUBJAREA , "MATE" ) ) AND ( LIMIT-TO ( PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2019 ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) OR LIMIT-TO ( LANGUAGE , "Spanish" ) )  ( TITLE-ABS-KEY ( desempeño AND docente AND universitario ) OR TITLE-ABS-KEY ( analítica AND académica ) OR TITLE-ABS-KEY ( minería AND de AND datos ) ) AND ( LIMIT-TO ( PUBSTAGE , "final" ) ) AND ( LIMIT-TO ( OA , "all" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "spanish" ) )
Science Direct	("minería de datos educacional" OR "educational data mining") AND university AND education Year: 2018-2020
Ebscohost	( ( minería de datos ) AND ( FM P ) ) AND ( ( educación universitaria ) AND ( FM P ) ) Limitadores: Fecha de publicación: 20180101-20210131; Texto completo en PDF; Idioma: Spanish; Tipo de publicación: Academic Journal; Tipo de documento: Article; Tipo de publicación: Academic Journal; Tipo de documento: Article; Tipo de documento: Article Amplidores - Aplicar palabras relacionadas; Buscar también dentro del texto completo de los artículos; Aplicar materias equivalentes Modos de búsqueda - Booleano/Frase
Proquest	(minería de datos) AND (educación superior) AND universitaria Límites aplicados: Bases de datos: Coronavirus Research Database / Ebook Central / ProQuest Central La búsqueda se realizará en estas bases de datos como parte de su consulta. Limitado por: Evaluado por expertos Fecha: Desde January 2018 hasta January 2021 Tipo de documento: fuente: Revistas científicas Tipo de documento: Artículo, Artículo principal Idioma: Español Restringido por: Fecha introducida: 2019-01-01 - 2020-12-31; Artículos evaluados por expertos: Evaluado por expertos
IOPscience	"educational data mining" AND "university" Within: The last 5 years

Fuente: elaboración propia

### Extracción de datos

Luego de aplicar las estrategias de búsqueda se procedió a filtrar y revisar de forma sistemática las investigaciones por medio de la declaración PRISMA (Moher et al., 2014) descartando estudios por estar duplicados, por evaluación de título y resumen, excluidos tras la evaluación de texto completo, por idioma distinto y por no existir asociación estadística. De los artículos seleccionados, se analizó el contenido plasmándolo en matrices o tablas contemplando datos generales y detallados relacionados directamente a las preguntas planteadas en la presente investigación con el fin de dar respuestas concretas apoyadas en cálculos estadísticos y proceder a la respectiva discusión.

### RESULTADOS

Al realizar la búsqueda general por cada base de datos virtual se obtuvo un total de 2275 artículos, siendo necesario la optimización de dichas búsquedas lo que es comúnmente conocido como las búsquedas avanzadas en donde se aplicó las respectivas estrategias, es así que se obtuvo como resultado 187 artículos, representado un 8% del total de las búsquedas generales, el detalle se aprecia en la Tabla 3.

Tabla 3. Búsquedas generales y avanzadas por bases de datos virtual.

Bases de datos	Variables de búsqueda con operadores lógicos	General	Avanzada	%
Scopus	TITLE-ABS-KEY ( educational AND data AND mining ) AND TITLE-ABS-KEY ( university ) AND ( LIMIT-TO ( FREETOREAD , "all" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( SUBJAREA , "SOC1" ) OR LIMIT-TO ( SUBJAREA , "MATE" ) ) AND ( LIMIT-TO ( PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2019 ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) OR LIMIT-TO ( LANGUAGE , "Spanish" ) )	1090	49	4%
	(TITLE-ABS-KEY (desempeño AND docente AND universitario) OR TITLE-ABS-KEY (analítica AND académica) OR TITLE-ABS-KEY (minería AND de AND datos)) AND (LIMIT-TO (PUBSTAGE, "final" ) ) AND ( LIMIT-TO ( OA , "all" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "spanish" ) )	132	35	27%
Science Direct	("minería de datos educacional" OR "educational data mining") AND university AND education Year: 2018-2021	384	23	6%
Ebscohost	((mineríadedatos)AND(FMP))AND((educaciónuniversitaria)AND(FMP)) Limitadores - Fecha de publicación: 20180101-20210131; Texto completo en PDF; Idioma: Spanish; Tipo de publicación: Academic Journal; Tipo de documento: Article; Tipo de publicación: Academic Journal; Tipo de documento: Article; Tipo de documento: Article Ampliadores - Aplicar palabras relacionadas; Buscar también dentro del texto completo de los artículos; Aplicar materias equivalentes Modos de búsqueda - Booleano/Frase	66	25	38%
Proquest	(minería de datos) AND (analítica datos) AND universitaria Límites aplicados Bases de datos: Coronavirus Research Database / Ebook Central / ProQuest Central Labúsquedaserealizaráenestasbasesdedatoscomopartedesuconsulta. Limitado por: Evaluado por expertos Fecha: Desde January 2018 hasta January 2021 Tipo de fuente: Revistas científicas Tipo de documento: Artículo, Artículo principal Idioma: Español Restringido por: Fecha introducida: 2019-01-01 - 2020-12-31; Artículos evaluados por expertos: Evaluado por expertos	103	47	46%
IOPscience	"educational data mining" AND "university" Within: The last 5 years	500	8	2%
TOTAL		2275	187	8%

Fuente: elaboración propia

Se continuó filtrando de manera sistemática las investigaciones encontradas, labor que se realizó paso a paso según la declaración PRISMA: en la fase de identificación se tomó como base los resultados de las búsquedas avanzadas detalladas en la Tabla 3, en donde se obtuvo un total de 187 artículos identificados; en la fase de elegibilidad se descartaron las investigaciones duplicadas entre las bases de datos virtuales, siendo 24 y además analizando los títulos y resúmenes se eliminaron 80 artículos; en la fase de evaluación se excluyeron aquellos artículos sin texto completo siendo 39, también aquellos en idiomas distintos al inglés y español siendo 13. Finalmente, en la fase de inclusión el total de artículos científicos incluidos como unidades de análisis en la revisión sistemática fueron 22, como se muestra en la Figura 1.

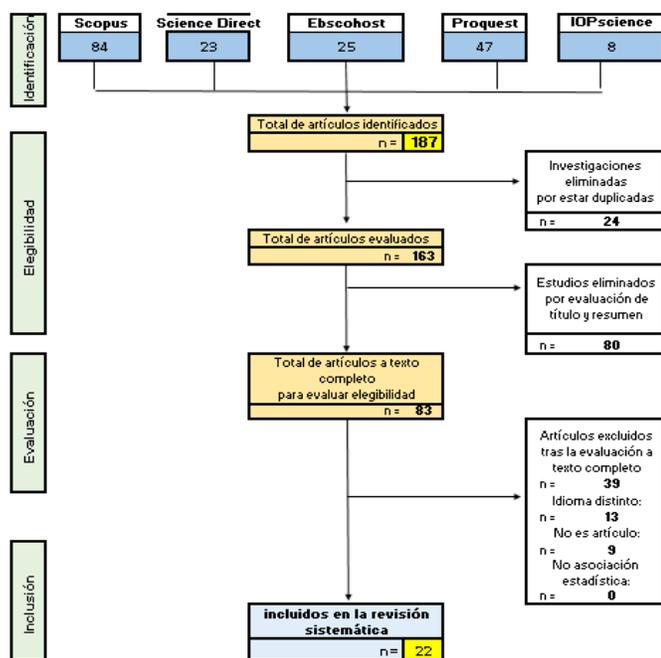


Figura 1. Diagrama de flujo a través de las fases de la declaración PRISMA.

Fuente: elaboración propia

Con los artículos incluidos en la revisión sistemática, se procedió a analizarlos construyendo una matriz de extracción de datos tanto generales como aquellos relevantes para dar respuesta a las preguntas de investigación planteadas, se consideró una columna adicional ID en donde se ha asignado un código de identificación (A1, A2, A3...etc.).

Teniendo los datos necesarios de los artículos incluidos en la revisión sistemática, se continuó con un análisis cuantitativo con el fin de sintetizar la información a través cálculos y gráficos estadísticos para dar respuestas concretas y relevantes a las preguntas de investigación planteadas.

**P\_01: ¿Cómo ha progresado en el tiempo las investigaciones en la cual se aplica analítica académica y la minería de datos educacional en el nivel universitario?**

Con el fin de obtener un perfil del progreso en el tiempo de la cantidad de investigaciones en la cual se aplica la minería de datos educacional en el nivel universitario, se ha considerado el resultado obtenido en las búsquedas generales (véase Tabla 3), es decir publicaciones a partir del año 2010 en adelante detalladas por cada base de datos virtual que se consideró en el presente estudio como se ve en la Figura 2, en donde se aprecia

una tendencia de crecimiento significativo a partir del año 2017. Se tomó en cuenta obtener resultados más generales por año uniendo todas las publicaciones de las bases de datos virtual, observándose en la Figura 3 la confirmación que el crecimiento significativo del número de estas se da a partir del año 2017 con un 11% de los artículos analizados, en el 2018 con un 14%, en el 2019 con un 22% y el 2020 con un 22%. Siendo en el año 2019 y 2020 donde se da el mayor número de publicaciones.

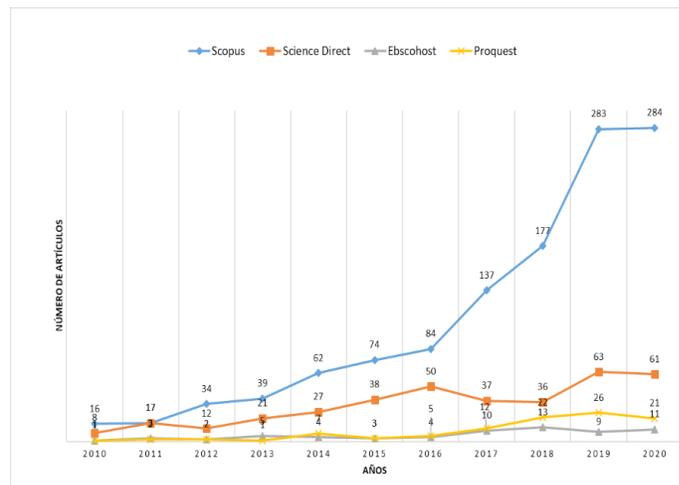


Figura 2. Artículos por años de publicación por cada base de datos virtual.

Fuente: elaboración propia

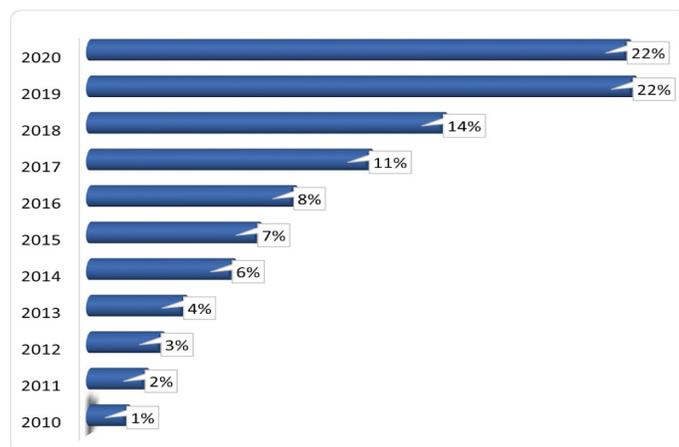


Figura 3. Artículos por año de publicación

Fuente: elaboración propia

**P\_02: ¿Qué se pretende solucionar al aplicar la analítica académica y la minería de datos educacional en el proceso enseñanza – aprendizaje en las universidades?**

Al analizar los artículos incluidos en la revisión sistemática, se pudo obtener los temas del proceso enseñanza

– aprendizajes abordados en los estudios los cuales pretenden que al aplicar la minería de datos educacional en las universidades con los resultados obtenidos explicar patrones, tendencias, predicciones que puedan dar solución a problemas relacionados a los temas en mención. En la Tabla 4, se detallan los temas considerados en los estudios, se resalta que algunos artículos consideraron como estudio a varios temas como, por ejemplo: el artículo A1 abordó en su estudio dos temas, el de rendimiento académico y patrones de comportamiento de estudiante; el artículo A9 abordó en su estudio cuatro temas, el rendimiento académico, la deserción de estudiante, la clasificación de perfiles de estudiantes y patrones de comportamiento de los estudiantes. A partir de la tabla en mención se pudo obtener los resultados plasmados en el Figura 4, en donde el tema con mayor porcentaje de investigaciones es sobre el rendimiento académico que representa el 35% de los artículos analizados, luego el segundo tema tomado en cuenta es sobre los patrones de comportamiento de estudiantes que representa el 14% de los estudios analizados, se obtuvo también que los temas sobre la deserción de estudiantes y la clasificación de los perfiles de los estudiantes se encuentran en un tercer lugar en los estudios analizados representando un 10% cada uno; con menor porcentaje estando en un rango del 7% hacia abajo se encuentran los temas como la empleabilidad de los egresados, competencias para el desarrollo sostenibles de docentes, teorías de la inteligencia, estilos de aprendizaje y la participación de estudiantes en comunidad de aprendizaje en redes sociales.

Tabla 4. Temas del proceso enseñanza - aprendizaje abordados en los artículos.

Temas del proceso enseñanza - aprendizaje	ID de artículo	Total
Rendimiento académico	A1, A5, A6, A7, A9, A10, A13, A16, A20, A22	10
Deserción de estudiantes	A2, A9, A18	3
Clasificación de perfiles de estudiantes	A9, A12	2
Empleabilidad de egresados	A3,	1
Patrones de comportamiento de estudiantes	A1, A9, A12, A21	4
Competencias para el desarrollo sostenibles de docentes	A8,	1
Teorías de la inteligencia	A4,	1
Estilos de aprendizaje	A4, A15	2
Participación de estudiantes en comunidad de aprendizaje en redes sociales	A11	1

Planificación académica	A14	1
Desempeño docente	A17, A19, A20	3

Fuente: elaboración propia

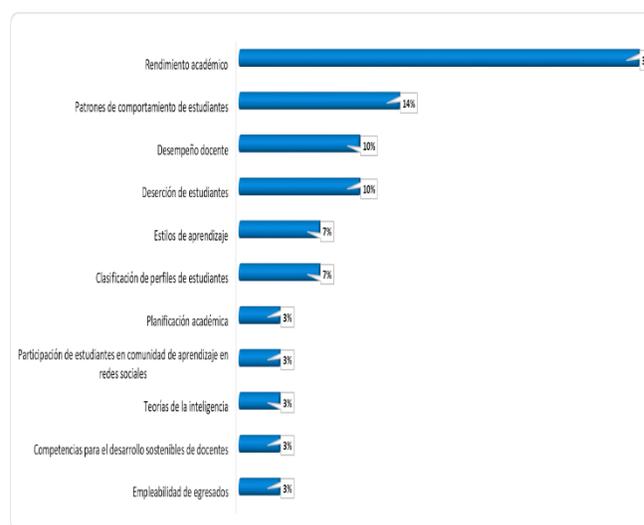


Figura 4. Temas del proceso enseñanza - aprendizaje abordados en los artículos

Fuente: elaboración propia

### P\_03: ¿Qué metodologías se utilizan para aplicar analítica académica y minería de datos educacional en el contexto universitario?

Del análisis de los artículos incluidos en la revisión sistemática, se aprecia en la Tabla 5 que la aplicación de metodologías de minería de datos propiamente dichas es mínima, centrándose en solo la aplicación de algunos procesos, además resalta la no aplicación de metodología ni de algún proceso. A partir de esta tabla se obtuvo los resultados mostrados en el Figura 5, en donde el 23% de los artículos en estudio no aplica ninguna metodología o proceso de minería de datos, luego también resultó con notoriedad que un 18% de las investigaciones aplican la metodología CRISP-DM. También se obtuvo que en un 14% de los artículos se aplican algunos procesos de minería de datos. Con lo que respecta a la aplicación de otros métodos de minería de textos, etc., solo se dio en un rango del 9% hacia abajo.

Tabla 5. Metodologías / Procesos de minería de datos utilizados

Metodología / Proceso (Minería de datos y/o Analítica de datos)	ID de artículo	Total
Procesos de minería de datos	A1, A3	2

CRISP-DM	A4, A15, A16, A18	4
SMART	A6	1
Procedimientos del DRSA	A8	1
Análisis del conjunto de datos. Preprocesamiento de datos de entrada. Análisis de desempeño de predicción	A5, A9, A10	3
Método de minería de texto	A11	1
MIDANO	A14	1
Proceso KDD	A17	1
Proceso de analítica académica	A20, A22	2
Análisis de contenido	A21	1
No aplica	A2, A7, A12, A13, A19	5

Fuente: elaboración propia

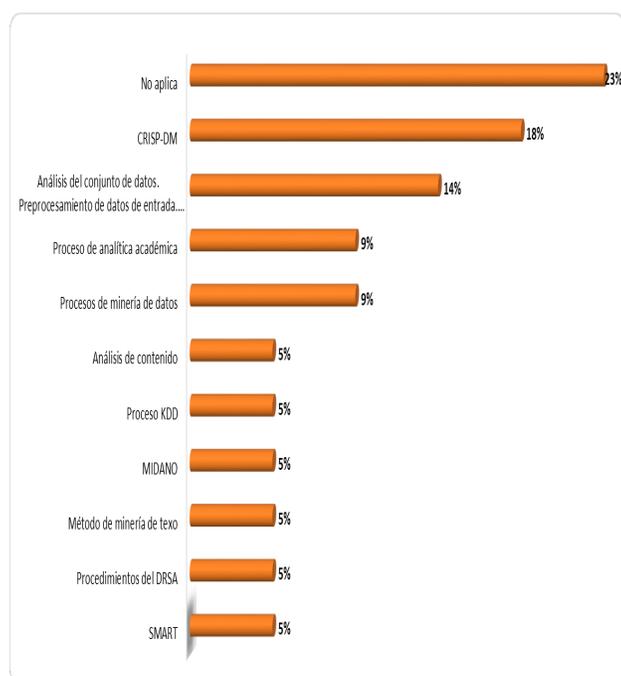


Figura 5. Metodologías / Procesos de minería de datos utilizados.

Fuente: elaboración propia

#### P\_04: ¿Qué técnicas de analítica académica y minería de datos se implementan?

Para una mejor comprensión sobre la clasificación de las técnicas de minería de datos, estas se han dividido de

acuerdo a los modelos y tareas de minería de datos. Es así que Riquelme et al. (2006), detalla que las tareas de minería de datos son las descriptivas que permiten obtener patrones o describe lo existente y las predictivas que permite predecir el o los valores desconocidos; estas tareas se clasifican en los modelos, siendo para la tarea descriptiva los modelos de agrupamiento (clustering) y asociación, para la tarea predictiva los modelos de clasificación y regresión. Los modelos se implementan por medio de sus propias técnicas o conocidos como los algoritmos de minería de datos.

Luego, al analizar los artículos incluidos en la revisión sistemática y la clasificación de las técnicas de minería de datos en modelos y tareas se obtuvo la Tabla 6, como se observa los artículos han implementado en sus estudios varias técnicas de minería de datos como por ejemplo el artículo A10 implementó en su investigación 4 técnicas: Árboles de decisión para clasificación, Naive Bayes, Redes neuronales artificiales, Máquina de vector de soporte (SVM). Resalta que en las investigaciones apuntan en gran mayoría hacia las tareas predictivas, siendo el modelo de clasificación el más utilizado. El detalle de los resultados obtenidos se muestra en la Figura 6, en donde la técnica de minería de datos K – means se implementó en 8 estudios siendo la más utilizada como modelo de agrupamiento dentro de las tareas descriptivas; resultó también que dentro de las tareas predictivas, la técnica de minería de datos Árboles de decisión para clasificación se implementó en 7 estudios siendo la más utilizada como modelo de clasificación, así continúan las técnicas de Redes neuronales artificiales y Máquina de vector de soporte (SVM) que se implementaron en 5 estudios cada una para proseguir con las técnicas de Regresión Logística y Naive Bayes que se implementaron en 4 estudios cada una, el resto de técnicas utilizadas como modelos de clasificación solo se implementaron en un artículo cada una. En cuanto a las técnicas para implementar modelos de regresión la más aplicada fue Regresión lineal. En la Figura 7, se puede visualizar en resultados porcentuales el total de técnicas empleadas por modelo de minería de datos, siendo las técnicas del modelo de clasificación que alcanzan el 57% siendo las más utilizadas, luego en un segundo lugar las técnicas del modelo de agrupamiento alcanzando el 19% de utilización, las técnicas del modelo de regresión son utilizadas en un 15%, siendo las técnicas del modelo de asociación que menos se utilizan con un 9%.

Tabla 6. Técnicas de minería de datos implementadas.

Tareas de minería de datos	Modelos de minería de datos	Técnicas de minería de datos	ID de artículo	Total por técnica	Total por modelo	Total por tarea
Descriptivas	Agrupamiento	K-means	A1, A4, A6, A9, A12, A15, A17, A21	8	10	15
		Análisis factorial	A19	1		
		Coefficiente de correlación de Pearson	A6	1		
	Asociación	Apriori	A4, A14, A15, A17, A22	5	5	
Predictivas	Clasificación	Árboles de decisión para clasificación	A2, A10, A11, A15, A16, A17, A18	7	31	39
		Longitud de camino mínimo	A3	1		
		Naive Bayes	A3, A10, A11, A18	4		
		Redes neuronales artificiales	A5, A8, A10, A13, A18	5		
		Enfoque de conjunto aproximado (RSA)	A8	1		
		Máquina de vector de soporte (SVM)	A8, A9, A10, A11, A13	5		
		Regresión logística	A9, A10, A11, A13	4		
		Modelo logístico de efectos mixtos	A9	1		
		K-Vecinos más cercanos (k-NN)	A10	1		
		Bosque aleatorio	A13, A18	2		
	Regresión	CO-training REGresors (COREG)	A7	1	8	
		Árboles de decisión para regresión	A10	1		
		Regresión bayesiana	A10	1		
		Serie de tiempo	A11	1		
		Regresión lineal	A10, A14, A15, A20	4		

Fuente: elaboración propia

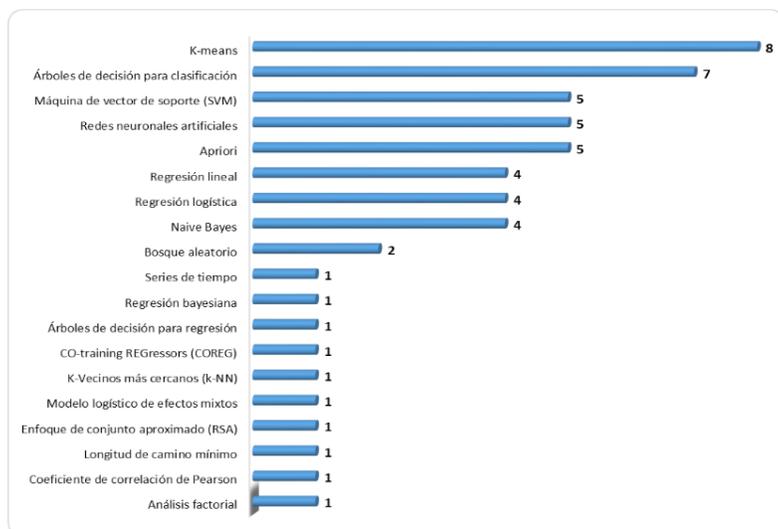


Figura 6. Técnicas de minería de datos implementadas.

Fuente: elaboración propia

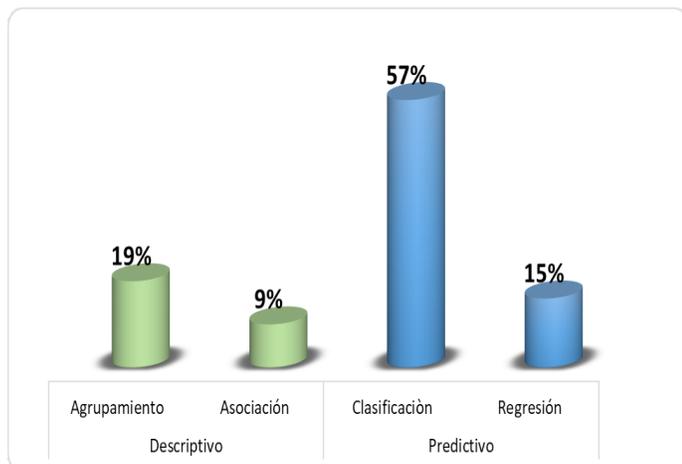


Figura 7. Total, por modelos de minería de datos utilizados

Fuente: elaboración propia

**P\_05: ¿Qué programas de aplicación y lenguajes de programación se aplican para implementar la analítica académica y la minería de datos educacional?**

Al analizar los artículos incluidos en la revisión sistemática, se pudo obtener tal como se muestra en la Tabla 7 los programas que se aplicaron para implementar minería de datos educacional en las investigaciones, los cuales se ordenaron de acuerdo al tipo de programa para su mejor comprensión, se evidencia que los programas de aplicación se impusieron ante los lenguajes de programación. De acuerdo a la tabla se elaboró el gráfico de la Figura 8 en donde se visualiza que el programa de aplicación más utilizado en las investigaciones ha sido Weka aplicado en 5 artículos, el lenguaje de programación más utilizado fue Python aplicado en 3 artículos, también se aprecia que en 4 artículos no se especifica el programa aplicado y el resto de programas no denotan una aplicación significativa puesto que son utilizados en 2 a menos artículos. En la Figura 9, se puede observar que en general los programas de aplicación han sido los más utilizados para implementar la minería de datos educacional representando un 57% de los artículos analizados, luego los lenguajes de programación representan el 21% y se aplicó el lenguaje de marcas y un manejador de base de datos en un artículo respectivamente representando el 4% cada uno.

Tabla 7. Programas aplicados en los artículos.

Tipo de programa	Programa	ID de artículo	Total por programa	Total por tipo de programa
De aplicación	Excel	A1, A17	2	16
	SPSS	A1, A18	2	
	RapidMiner	A1, A2	2	
	Orange Data Mining	A3	1	
	Weka	A4, A14, A15, A16, A17	5	
	SHAP	A7	1	
	SmartPLS	A19	1	
	DQ_Analizer	A15	1	
	JAMF	A8	1	
Lenguaje de marcas	HTML	A9	1	1
Lenguaje de programación	R	A5	1	6
	Python	A6, A9, A13	3	
	Javascript	A9	1	
	MatLab	A10	1	
Manejador de base de datos	MySQL	A22	1	1
No especifica	No especifica	A11, A12, A20, A21	4	4

Fuente: elaboración propia

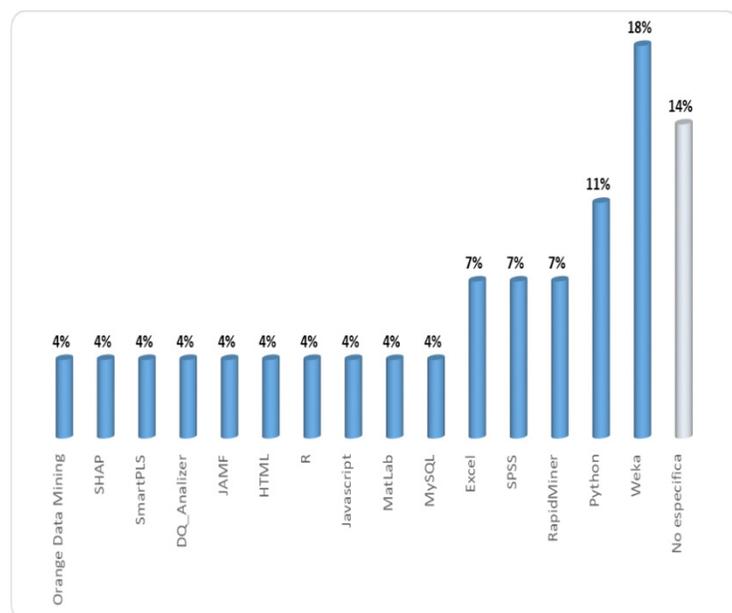


Figura 8. Programas aplicados en los artículos.

Fuente: elaboración propia

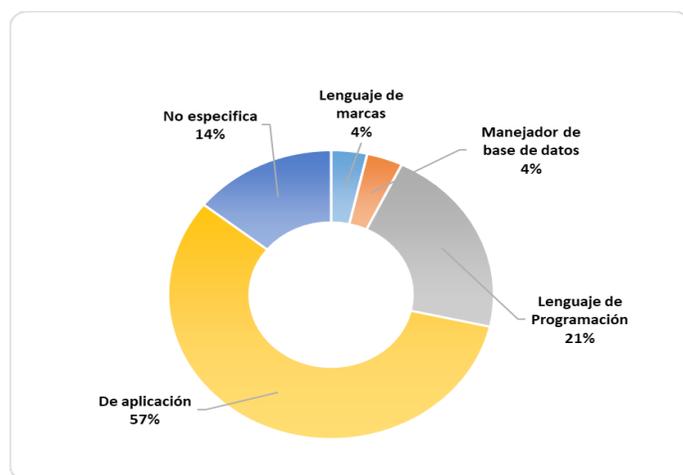


Figura 9. Total, por tipo de programa utilizados en los artículos.

Fuente: elaboración propia

## DISCUSIÓN DE RESULTADOS

Se llevó a cabo una revisión sistemática de artículos de investigación con el fin de presentar la evidencia disponible sobre cómo o de qué manera se aplica y aporta la analítica académica y la minería de datos educacional en el proceso de enseñanza – aprendizaje para el nivel universitario de acuerdo a las preguntas planteadas en el presente estudio.

El progreso en el tiempo de las investigaciones sobre la aplicación de la analítica académica y la minería de datos educacional en el nivel universitario (véase Figura 3) es creciente significativamente a partir del año 2017, continuando esta tendencia en los siguientes años; lo que demuestra que el interés en la comunidad educacional sobre este tema o disciplina ha ido en aumento por su importancia en la generación de conocimiento.

Se pudo identificar los diversos temas del proceso enseñanza – aprendizaje que han sido abordados en los artículos (véase Figura 4) sobresaliendo con un 35% lo que respecta al rendimiento académico, tal como indica Reyes et al. (2019) en su investigación Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje la minería de datos educacional desarrolla modelos con el fin de explorar los datos existentes de los entornos de aprendizajes educativos para detectar patrones que permitan predecir variables de interés, en donde obtuvo una precisión alta de su modelo de predicción lo que demuestra la efectividad de los predictores del rendimiento académico. Con el análisis realizado se aprecia que los temas relacionados al proceso enseñanza – aprendizaje son diversos, que

los resultados de los estudios son favorables en cuanto a mejora de los procesos y toma de decisiones más rápidas y acertadas, lo que indica que queda por explorar aún mucho más sobre este campo profundizando en los temas ya identificados, pero con mínima cantidad de estudios o descubriendo más temas en base a patrones identificados.

En cuanto a las metodologías que se utilizan para aplicar la minería de datos educacional en el contexto universitario (véase Figura 5) quedó en evidencia que en los estudios analizados en un 23% no aplican ninguna metodología relacionada a la minería de datos, además el 18% aplican la metodología CRISP-DM y un 14% aplican algunos procesos de minería de datos. La utilización de la metodología CRISP-DM en el desarrollo de proyectos de minería de datos organiza el trabajo en seis etapas y cada una de estas se divide en varias tareas, que podría verse como un paso a paso idealizado de hechos que en la práctica varias de las tareas pueden realizarse en distinto orden e inclusive retornar si es necesarios varias veces a alguna de ellas como ha quedado demostrado en los artículos analizados. Es decir, si es necesario la aplicación de procesos o metodologías que direccionen el trabajo a realizar, pero éstos pueden ir adaptándose y acoplándose según se presenta la realidad.

Se obtuvieron de los artículos en estudio, las técnicas de minería de datos implementadas en el contexto de la educación en el nivel universitario (véase Tabla 6, Figura 6 y 7), estas técnicas pueden tener un direccionamiento predictivo o descriptivo, con tareas específicas, lo que hace que en las investigaciones se obtengan resultados diferentes. Las técnicas más resaltantes son las que generan predicciones en un 72% de los estudios por medio de modelos de clasificación que representan un 57% entre ellas están Árboles de decisión para clasificación, Redes neuronales artificiales, Máquina de vector de soporte (SVM), Regresión logística y Naive Bayes. En su investigación Prada et al. (2020) inicialmente para obtener patrones centra su investigación en la implementación de cluster con la técnica K-means, luego estos grupos obtenidos los consigue clasificar implementando técnicas de máquina de vectores de soporte (SVM), regresión logística y modelo logístico de efectos mixtos, generando modelos predictivos en el desempeño, comportamiento, causas de deserción y perfiles de estudiantes. Con los resultados obtenidos en el análisis de las investigaciones, al implementar en un mismo estudio varias técnicas de minería de datos permite realizar comparaciones de resultados de estas generando precisiones cada vez más altas en los modelos.

En el mercado actual existe una gran variedad de programas de aplicación orientados al desarrollo de minería de datos facilitando la labor de quienes pretenden implementar proyectos de minería de datos, no es necesario contar con grandes conocimientos de computación basta con ingresar los datos necesarios y el programa generará los resultados esperados, también existen los lenguajes de programación de propósito general y específico en la cual es posible el desarrollo de proyectos de minería de datos pero a diferencia de los de aplicación, en este caso si es necesario contar con conocimientos avanzados de computación específicamente de programación, en cuanto a la revisión sistemática realizadas se encontró en los artículos analizados (véase Tabla 7, Figura 8 y 9) la utilización de una amplia gama de programas para implementar la minería de datos destacando los programas de aplicación representando el 57% de las investigaciones y los lenguajes de programación en un 21%.

## CONCLUSIONES

La importancia de la revisión sistemática se centró en precisar la existencia de investigaciones en el área de la analítica académica y la minería de datos educacional aplicados a los procesos de enseñanza - aprendizaje en el nivel universitario. Se encontraron estudios significativos que con ellos se pudo dar respuesta a las interrogantes planteadas. Por la gran variedad de técnicas de minería de datos implementadas en las investigaciones analizadas no permite establecer con precisión el uso específico de algunas de ellas. Por lo que se concluye implementar las comparaciones entre las técnicas más empleadas para precisar resultados.

Existen aún temas del contexto enseñanza – aprendizaje con muy pocas evidencias de estudio o temas que aún faltan explorar que pueden ser abordados por la analítica académica y minería de datos educacional.

## REFERENCIAS BIBLIOGRÁFICAS:

- Adi, K., & Isnanto, R. R. (2020). Rice crop management expert system with forwarding chaining method and certainty factor. In *Journal of Physics: Conference Series* (Vol. 1524, No. 1, p. 012037). IOP Publishing. <https://iopscience.iop.org/article/10.1088/1742-6596/1524/1/012037/pdf>
- Arevalo-Marín, P., Cabrera-Piedra, Y., Cabrera-Sarango, M., Caicedo-Rodriguez, J., Camacho-Veliz, A., Cartuche-Morocho, S., ... & Roman-Sanchez, M. (2019). Minería de datos educativa para identificar la relación entre cociente intelectual, estilos de aprendizaje, inteligencia emocional e inteligencias múltiples de estudiantes de ingeniería. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E17), 48-63.
- Chacón, F., Spicer, D., & Valbuena, A. (2012). Analytics in support of student retention and success. *Research Bulletin*, 3, 1-9. <https://library.educause.edu/-/media/files/library/2012/4/erb1203-pdf.pdf>
- Chacón, P. E. C., González, A. Z., Domínguez, V. H. M., & Herrera, P. J. C. (2018). Análisis del desempeño del profesorado universitario en el uso de MOODLE a través de técnicas de minería de datos: propuestas de necesidades formativas. *Revista de Educación a Distancia (RED)*, (58), 1-41. <https://revistas.um.es/red/article/view/351411/251801>
- Lin, S. (2012). Datamining for student retention management. *Journal of Computing Sciences in Colleges*, 27(4), 92-99. [https://www.researchgate.net/profile/Linda-Werner-3/publication/262412335\\_Using\\_Scrum\\_in\\_a\\_quarter-length\\_undergraduate\\_software\\_engineering\\_course/links/564deaec08ae4988a7a51f5d/Using-Scrum-in-a-quarter-length-undergraduate-software-engineering-course.pdf#page=103](https://www.researchgate.net/profile/Linda-Werner-3/publication/262412335_Using_Scrum_in_a_quarter-length_undergraduate_software_engineering_course/links/564deaec08ae4988a7a51f5d/Using-Scrum-in-a-quarter-length-undergraduate-software-engineering-course.pdf#page=103)
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA, G. (2014). Ítems de referencia para publicar revisiones sistemáticas y metaanálisis: la Declaración PRISMA. *Revista Española de Nutrición Humana y Dietética*, 18(3), 172-181. <https://dialnet.unirioja.es/descarga/articulo/4821653.pdf>
- Moreno, B., Muñoz, M., Cuellar, J., Domancic, S., & Villanueva, J. (2018). Revisiones Sistemáticas: definición y nociones básicas. *Revista clínica de periodoncia, implantología y rehabilitación oral*, 11(3), 184-186. <https://www.scielo.cl/pdf/piro/v11n3/0719-0107-piro-11-03-184.pdf>
- Prada, M., Domínguez, M., Lopez, J., Vara, P., Barbu, M., Podpora, M., Spagnolini, U. Varanda, M. y Vilanova, R. (2020). Educational Data Mining for Tutoring Support in Higher Education: A Web-Based Tool Case Study in Engineering Degrees, *IEEE Access*, 8, 212818-212836. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9272294>

- Reyes, N., Morales, J., Moya, J., Terán, C., Rodríguez, D., & Altamirano, G. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E17), 258-266. <https://media.proquest.com/media/hms/PFT/1/qJDZ8?s=ixio rEKpss87bGaT8u173c9dGa8%3D>
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, 10 (29), 11-18. <https://idus.us.es/bitstream/handle/11441/43290/Miner%c3%ada%20de%20datos.pdf?sequence=1&isAllowed=y>
- Saiful, M., & Nur, A. M. (2020). Application of Expert System with Web-Based Forward Chaining Method in Diagnosing Corn Plant Disease. In *Journal of Physics: Conference Series*, 1539(1),
- Sumpala, A. T., & Rasyid, R. (2019). Expert system for corn plant disease diagnosis with the breadth-first search method. In *IOP Conference Series: Earth and Environmental Science* 382,(1) p. 012001). IOP Publishing. <https://iopscience.iop.org/article/10.1088/1755-1315/382/1/012001/pdf>