



Análisis de Trayectorias Académicas en el Aprendizaje de la Física mediante Minería de Datos Educativa en estudiantes de Ingeniería

Daniel Sánchez Guzmán^a, José Antonio Hernández Camacho^a

^aInstituto Politécnico Nacional – UPIIG. Av. Mineral de Valenciana, No. 200., Guanajuato, México.

ARTICLE INFO

Received: 12 de febrero de 2022

Accepted: 22 de abril de 2022

Available on-line: 31 de mayo de 2022

Keywords: Physics on-line courses, Engineering students, Educational Data Mining, Learning trajectories, Clustering and Bayesian Classification Algorithms.

E-mail addresses: dsanchez@ipn.mx,

ISSN 2007-9847

© 2022 Institute of Science Education.
All rights reserved

ABSTRACT

The use of digital tools for most of the learning processes that are carried out today and promoted by the confinement situation caused by the SARS-CoV-2 virus (COVID-19), involves a large generation of data to be analyzed. Information such as times of use of administration platforms in learning, delivery of evidence by students, application of online evaluations, chats, and discussion forums, are just some of the activities that students must develop online but that, due to the large amount of information generated, it is difficult for teachers to provide timely and personalized follow-up to students. This paper describes the results of investigating the application of clustering techniques and Bayesian networks, proposed by Educational Data Mining, with the aim of analyzing the academic trajectories of novices' students at the engineering level. The results show that the students with better performances within the learning trajectories presented a better level in the levels of reading comprehension of scientific texts and writing work reports in Physics subjects, identifying communication skills as a fundamental part of the trajectory. Of learning for the understanding of abstract Physics concepts in students, such as: force, work, and energy. To analyze the performance, all the activities that the students developed were considered, such as homework, report grades and online exams, quantitative weightings were generated according to the index of difficulty and discrimination in the students.

El uso de las herramientas digitales para la mayoría de los procesos de aprendizaje que se realizan en la actualidad y promovidas por la situación de confinamiento provocado por el virus SARS-CoV-2 (COVID-19), involucra una gran generación de datos para ser analizados. Información como: tiempos de uso de plataformas de administración en el aprendizaje, entrega de evidencias por parte de los estudiantes, aplicación de evaluaciones en línea, chats y foros de discusión, son sólo algunas de las actividades virtuales que los estudiantes deben desarrollar en línea pero que, por la gran cantidad de información generada es complicado para los docentes el dar un seguimiento puntual y personalizado al desempeño académico los estudiantes. El presente trabajo describe los resultados de investigar la aplicación de las técnicas de agrupamiento (Clustering) y redes bayesianas, propuestas por la Minería de Datos Educativa, con el objetivo de analizar las trayectorias académicas de estudiantes de nuevo ingreso a nivel ingeniería. Los resultados presentan que los estudiantes con mejores rendimientos dentro de las trayectorias de aprendizaje presentaron un mejor nivel en los niveles de comprensión lectora de textos científicos y redacción de reportes de trabajo en las asignaturas de Física, identificando las habilidades comunicativas como parte fundamental en la trayectoria del aprendizaje para la comprensión de conceptos abstractos de Física en los estudiantes, como pueden ser: fuerza, trabajo y energía. Para analizar el rendimiento se consideraron todas las actividades que los estudiantes desarrollaron, como tareas, calificaciones de reportes y exámenes en línea, se generaron ponderaciones cuantitativas de acuerdo con el índice de dificultad y discriminación en los estudiantes.

I. INTRODUCCIÓN

Los procesos educativos presentaron un cambio radical derivado de la situación de confinamiento propiciada por la aparición del virus SARS-CoV-2 (COVID-19), la comunidad académica (estudiantes, docentes y administrativos) tuvieron que realizar una migración revolucionaria de la modalidad presencial en el aprendizaje a una modalidad virtual de aprendizaje. Dentro de este cambio, la mayoría de los docentes tuvieron que aprender, en muchos casos y de manera acelerada, a utilizar las herramientas digitales disponibles, principalmente lo que se conoce como un Sistema de Gestión en Aprendizaje (o Learning Management System - LMS, de su definición en idioma Inglés), entre los espacios más representativos se tienen plataformas como Google Classroom®, Moodle®, Blackboard®, Quizlet®, Edmodo, D2L®, Canvas, Absorb o Schoology (Fenton, 2018), entre otras. Estos espacios digitales de trabajo ya habían sido adoptados por varias universidades en su mayoría, pero el uso de la tecnología dentro de los espacios educativos, hasta antes de marzo de 2020, no representaba más del 10 % (diez por ciento), esto es, se contaba con la disponibilidad de los espacios virtuales, pero no se empleaban en su mayoría dentro de los procesos educativos.

Con el cambio de modalidad educativa, las trayectorias académicas de los estudiantes presentaron modificaciones conforme transcurrió el tiempo, se presentaron factores como la deserción escolar, cambios en el rendimiento académico de los estudiantes y un impacto en los procesos sociales, ya que tanto los estudiantes como los docentes, tuvieron que acondicionar las actividades académicas junto con las personales, al realizar sus actividades dentro de sus hogares. Este ajuste en las actividades y los tiempos de trabajo propiciaron el desarrollo de nuevas habilidades de comunicación e interacción, reduciendo, en parte, la brecha digital que existía entre docentes y estudiantes, ya que los primeros comenzaron a utilizar las redes sociales para interactuar con los estudiantes dentro de sus procesos educativos, y los segundos comenzaron a emplear las mismas redes sociales como un espacio de aprendizaje, se crearon grupos en Facebook®, Whatsapp®, Twitter®, YouTube®, Instagram®, etc. Donde de manera virtual, tanto docentes como estudiantes pudieron continuar con su proceso de formación y aprendizaje.

El estudio de las trayectorias académicas tiene varios años de estudio y se han presentado diferentes enfoques para analizarlo (García Robelo & Barrón Tirado, 2011; Guevara & Belelli, 2013; Martínez Sánchez, 2018; Ortega Guerrero, López González, & Alarcón Montiel, 2015; Pérez Alcántara, 2017; Quintela Dávila, 2007), en todos es importante destacar la importancia que éstas presentan para todas las comunidades académicas, principalmente para identificar las variables y efectos, tanto positivos como negativos, de lo que puede ocurrir a los estudiantes durante su formación universitaria o de corte educativo en diferentes niveles y contextos. Un aspecto importante de los trabajos que reportan las trayectorias académicas, y que tratan de asegurar la fiabilidad de los estudios reportados, tiene que ver con la gran cantidad de información que debe ser analizada, sí bien las herramientas estadísticas son útiles, hay aspectos que no pueden ser analizados por la configuración e integración de variables, así como la naturaleza de los datos, ya que algunos son de carácter cuantitativo y hay otros de tipo cualitativo.

La Minería de Datos Educativa, es un área de las Ciencias Computacionales que utiliza la información de bases de datos, tanto cualitativa como cuantitativa, con el objetivo de realizar el descubrimiento de conocimiento asociado con las interacciones que se puedan dar en un entorno específico, como por ejemplo las trayectorias escolares, rendimientos académicos y factores socioculturales, entre otros, que impacta de manera directa e indirecta en el proceso de aprendizaje a lo largo del tiempo (Espigares, García, & Quiñones C. J., 2006; Jiménez Galindo, 2010; Rosado Gómez & Verjel Ibañez, 2017; Rico Páez, Gaytán Ramírez, & Sánchez Guzmán, 2019; Oviedo Carrascal & Jiménez Giraldo, 2019).

Como resultado de la digitalización e informatización de los diferentes procesos en la sociedad, generado por la situación de confinamiento, las bases de datos han coleccionado, a lo largo del tiempo, mucha información de los procesos educativos y actividades relacionadas al interior de las academias e instituciones educativas, con toda esa disponibilidad de información y derivado del proceso acelerado de generación de datos provocada por el confinamiento

de la sociedad para evitar los contagios del virus SARS-CoV-2 (COVID-19), es que se considera a la Minería de Datos Educativa, como una alternativa y herramienta de apoyo para el análisis de las trayectorias académicas en los estudiantes, identificando variables como rendimiento académico, comprensión de información y potenciales niveles de deserción dentro de un sistema educativo.

El presente trabajo se encuentra distribuido en la siguiente secciones, en la presente sección I, se presenta la introducción al trabajo correspondiente enunciando cuál es la problemática a resolver, en la sección II se describe lo que son las trayectorias académicas, la minería de datos en educación, así como la pregunta de investigación, hipótesis y objetivos de la investigación, en la sección III se describe la metodología implementada en la investigación y cuáles fueron las técnicas empleadas para analizar la información que se generó durante el curso de Física, la sección IV presenta el análisis de resultados y describe el comportamiento de las gráficas que describen las relaciones que se presentaron dentro de los grupos de estudio, por último la sección V describe las conclusiones generales del trabajo de investigación y trabajo futuro a realizar.

II. TRAYECTORIAS ACADÉMICAS Y MINERÍA DE DATOS EN EDUCACIÓN

II.1 Trayectorias académicas

Las trayectorias académicas han estado presentes desde la institucionalización de la educación, al momento de estructurar las secuencias de actividades y etapas que un individuo debe desarrollar para conseguir cierto perfil académico o profesional, en función de una necesidad social que sea requerida, con el objetivo de brindar, a través del estudio, una especialización respecto a una o varias tareas que se deban de realizar y para atender las necesidades de recursos humanos en un sistema económico en particular. De acuerdo con las siguientes definiciones, se considera a las trayectorias escolares como el comportamiento académico de un individuo e incluye el desempeño escolar, la aprobación, la reprobación, el promedio logrado, etcétera, a lo largo de los ciclos escolares (García Robelo & Barrón Tirado, 2011).

Para (Guevara & Belelli, 2013): “...una trayectoria académica se puede ver como un concepto complejo por las múltiples dimensiones que permite su análisis. Un abordaje comprensivo del recorrido de los estudiantes en su contexto educativo requiere de considerar la interacción de sus experiencias sociales y curriculares, situadas en un tiempo y en un espacio. Esta contextualización de la trayectoria ofrece una lectura de proceso, de discontinuidad y continuidades de una práctica particular: la estudiantil...”. Es indudable que la variable tiempo juega un papel básico en el desempeño de cualquier trayectoria, no necesariamente educativa, también en las ciencias, tanto sociales como exactas, el tiempo es una variable para considerar respecto a los experimentos o la evolución de algunos conceptos científicos. De manera adicional otros factores se presentan dentro de las trayectorias académicas en particular, que conforme se ha presentado, pueden ser variables de tipo cuantitativo, como el ingreso económico dentro del núcleo familiar, la edad, el género, el nivel educativo, etcétera, o del tipo cualitativo como nivel cultural, estilo de aprendizaje, profesión de los padres, etc.

Un impacto negativo durante la trayectoria académica de los estudiantes tiene que ver con dos fenómenos que se producen durante el tiempo de estudio, el rezago educativo (incremento en el tiempo para concluir de manera satisfactoria el plan de estudios a cursar) y la deserción escolar (abandono del plan de estudios a cursar por diferentes situaciones, tanto personales como académicas), los cuales ha sido identificados por autores como (Ortega Guerrero, López González, & Alarcón Montiel, 2015; Pérez Alcántara, 2017), si bien estos datos impactan en las instituciones educativas cuando dentro de su misión se encuentra la formación de recursos humanos, en la parte práctica es complicado evaluar cuantitativamente este tipo de situaciones porque en muchas ocasiones son identificadas una vez que se producen los eventos, como dejar de asistir a clases o reprobar la asignatura, siendo hechos que se presentan durante o al final de un

período escolar, siendo en este último punto donde se pueden contabilizar los resultados negativos, ya sea la deserción o el rezago educativo.

Para el caso particular de México, hay estudios interesantes que aportan la información respecto al sistema educativo mexicano en sus niveles de bachillerato (edades entre 14 y 18 años) y superior (edades entre los 17 y 23 años), en el caso de (Martínez Sánchez, 2018), menciona que para datos de 2017: “...*La matrícula en este nivel [bachillerato] es de 4.3 millones de alumnos; la cobertura de 69.3% para los jóvenes entre 16 y 18 años de edad. La eficiencia terminal es de 61.8% (sep, 2012). Cabe mencionar, por ejemplo, que para el nivel superior la deserción escolar en el periodo 2014-2015 fue de 9.3%, en 2015-2016 fue del 9.1 % y para el ciclo escolar 2016-2017 se espera que sea de 9 % (sep, 2016) ...*” Estos datos indican de manera general, y similar a otros contextos en Latinoamérica, que los sistemas educativos no cuentan con la información y por consecuencia los mecanismos para hacer frente a las problemáticas que se pueden presentar dentro de las trayectorias académicas en los estudiantes.

Este fenómeno de deserción tiende a variar respecto a los contextos locales y nacionales de cada sistema educativo, lo que diversifica las variables a considerar en cada situación de estudio, pero hay variables que podemos describir como “universales” que se presentan en la mayoría de los sistemas educativos, como, por ejemplo: rendimiento académico, edad, género, nivel educativo y tipo de asignatura (ciencias, humanidades, profesionales, etc.), partiendo de un ejemplo, como el caso del presente trabajo, se puede hacer extensiva la inclusión de variables y consideraciones para análisis de trayectorias académicas. Partiendo de lo anterior, es necesario tomar en cuenta que para reducir estos índices negativos es necesario profundizar en el estudio de las trayectorias académicas que desarrollan los estudiantes, principalmente desde los primeros años de estudio ya que así se puede caracterizar a la población, generación o cohorte, que inicia una trayectoria académica a través de algún programa académico o plan de estudios, para darle un eficiente seguimiento al desempeño de los estudiantes conforme vayan avanzando en sus actividades académicas, propiciando soluciones preventivas o correctivas con la información del ecosistema educativo a analizar.

II.2 Minería de datos educativa

La Minería de Datos, es un área de las Ciencias Computacionales que tiene cerca de sesenta años de desarrollo, aunque con identificaciones completamente diferentes, como: “*Colección de datos (década de 1960)*”, “*Acceso de datos (década de 1980)*”, “*Almacén de datos y apoyo a las decisiones (principios de la década de 1990)*” y “*Minería de Datos Inteligente (finales de la década de 1990)*” (Cutro, 2010), por último a inicios de la década del 2000 se emplearon técnicas computacionales como el “*Dataware housing*”, con el objetivo de aplicar principios estadísticos con algoritmos de redes neuronales artificiales para la identificación de patrones en los datos y toma de decisiones con la información almacenada, esto último derivó en lo que actualmente se conoce como Minería de Datos (MD). Otro enfoque de la MD también conocida como Descubrimiento de Conocimiento en Bases de datos (sus siglas en inglés son “KDD – Knowledge Discovery in Databases”), es el campo que permite descubrir información nueva y potencialmente útil de grandes cantidades de datos. Se ha empleado en numerosos campos, incluyendo desde los ya conocidos casos de carrito de compra para una tienda en línea hasta la bioinformática o investigaciones contra el terrorismo (Jiménez Galindo, 2010).

La Minería de Datos Educativa (MDE), amplía y complementa las herramientas de indagación en comparación con los paradigmas más tradicionales de investigación educativa, como experimentos de intervención didáctica, estudios sociológicos o investigación de diseño. En particular, la creación de repositorios públicos de datos educativos ha creado una base que hace posible la MDE. En particular, los datos de estos repositorios son totalmente válidos (ya que son datos reales sobre el rendimiento y aprendizaje de estudiantes reales, en ambientes educativos de diferentes contextos y tomados en tareas de aprendizaje), son cada vez más accesibles para comenzar una investigación educativa (Jiménez Galindo, 2010). Con el apoyo de todo ese cúmulo de información, las comunidades educativas pueden incrementar su conocimiento y toma de decisiones respecto a la conducción de los programas académicos y planes de estudio,

encaminando toda esa información hacia un aspecto que combine la formación colectiva, pero con un apoyo individual para el desarrollo personal de la persona que cursa o cursará ciertos estudios en un nivel educativo específico. De acuerdo con (Jiménez Galindo, 2010), la MDE puede ayudar en el estudio de las trayectorias académicas aplicando los siguientes enfoques de la MD:

TABLA I. Principales enfoques de la Minería de Datos Educativa.

Método de MD	Objetivo del método	Aplicaciones
Predicción	Desarrollo de un modelo que infiera una variable a partir de la combinación de datos disponibles.	Detección de comportamiento en el estudiante. Modelos de dominio. Predicción y entendimiento respecto a los resultados académicos del estudiante.
Agrupamiento	Descubrimiento de conjuntos de datos que se organicen de manera natural, categorizando del conjunto completo datos relevantes.	Identificación de asociaciones curriculares en los antecedentes educativos, relación y similitudes entre diferentes centros educativos.
Minería de relaciones	Descubrimiento de relaciones directas o indirectas entre las variables y los datos que son analizados.	Identificación de las asociaciones curriculares entre las secuencias didácticas en un curso en particular o en diferentes cursos, identificación de las estrategias didácticas que mejor desempeño presentan dentro de uno o varios cursos.
Descubrimiento por modelos	Planteamiento de modelos para realizar la predicción o agrupamiento en función de los objetivos que se desean alcanzar dentro de un plan de estudios o dentro de una asignatura en particular. Se puede emplear como un componente adicional si los resultados previos abonan a los objetivos planteados al inicio del método.	Se pueden descubrir relaciones entre los comportamientos de los estudiantes y sus características tanto personales como sociales, también el contexto se emplea como una variable que influye en los resultados del método.
Destilado de datos	Los datos son reducidos a una unidad mínima de información para que se puedan identificar de manera única y unívoca dentro del proceso de integración de información para determina la contribución individual de cada instancia al momento de ser analizada.	Descubrimiento de patrones de aprendizaje en los estudiantes mediante la identificación de su desempeño al momento de realizar las actividades tanto dentro del aula como extra clase durante un periodo de tiempo. Permite la identificación explícita de variables para uso futuro dentro de la MD.

*Elaboración propia en base a (Jiménez Galindo, 2010).

El enfoque de Clustering ha sido empleado en (Oviedo Carrascal & Jiménez Giraldo, 2019), con el objetivo de identificar los rendimientos académicos Aplicando la metodología de minería de datos CRISP-DM, realizaron un estudio basado en los resultados obtenidos en las pruebas Saber-Pro de estudiantes de ingeniería en el departamento de Antioquia (Colombia). En el estudio se consideraron 108 variables académicas, económicas y socio demográficas para realizar 3 modelos analíticos: 1) agrupación de los tipos de estudiantes, 2) selección de los factores que más influyen en el desempeño de las pruebas, y 3) predicción del desempeño en las pruebas a partir de las variables seleccionadas, como resultado encontraron que algunas de las variables más influyentes sobre el resultado de las pruebas son: el número de personas a cargo de sus estudios, el método de enseñanza, si el hogar es permanente o arrendado, el carácter académico de la institución entre pública o privada, así como las facilidades económicas como tener ciertas facilidades de electrodomésticos o medios de transporte, entre los que se encuentran el contar con un horno micro gas y/o tener una motocicleta para su desplazamiento (Oviedo Carrascal & Jiménez Giraldo, 2019).

Respecto al empleo de Redes Bayesianas (Rico Páez, Gaytán Ramírez, & Sánchez Guzmán, 2019), realizaron un estudio para complementar el rendimiento académico en estudiantes de ingeniería El objetivo de este trabajo fue

presentar la construcción, evaluación y aplicación de un modelo predictivo del rendimiento académico de estudiantes universitarios por medio de la técnica de minería de datos conocida como algoritmo Naïve Bayes. En este trabajo se recabaron datos de 107 estudiantes como entrenamiento para el algoritmo y se aplicó el modelo para predecir el rendimiento académico de 71 estudiantes. Los resultados mostraron que el modelo predictivo, además de obtener predicciones del rendimiento académico, también identifica los factores que más influyen en él. Este tipo de estudios permitió a los profesores diseñar estrategias de prevención e identificar estudiantes que son vulnerables a reprobación, partiendo de las oportunidades que ofrece la MDE, se observa la oportunidad de complementar estos estudios a través del análisis de trayectorias más particulares, como el análisis en el tiempo de los estudiantes durante un semestre y en específico en el aprendizaje de ciencias básicas como es el caso de Física (Rico Páez, Gaytán Ramírez, & Sánchez Guzmán, 2019).

II.3 Pregunta de investigación, hipótesis y objetivos

Para el presente trabajo de investigación se planteó la siguiente pregunta de investigación, hipótesis y objetivos, los cuales delinearon todas las actividades realizadas y los procesos aplicados a los datos recolectados.

- **Pregunta de investigación:** ¿cómo se puede emplear la Minería de Datos Educativa para identificar las trayectorias académicas en estudiantes que cursan un programa académico de ingeniería mediante la información en los cursos de Física que cursaron?
- **Hipótesis:** La aplicación de las técnicas de Minería de Datos: Clustering y Redes Bayesianas, permitirán identificar las trayectorias académicas de estudiantes que cursan un programa académico de ingeniería y en base a la información que proveen las actividades en un curso de Física.
- **Objetivo General:** Implementar las técnicas de Minería de Datos Educativa, Clustering y Redes Bayesianas en los datos proporcionados por los cursos de Física en estudiantes que cursan un programa académico de ingeniería y validar las trayectorias académicas mediante sus rendimientos escolares e información adicional que permita la aplicación de los algoritmos seleccionados.
- **Objetivos particulares:** a) Recolectar la información de los cursos de Física y la evidencia generada por los estudiantes como un ecosistema educativo, b) Aplicar las técnicas de Minería de Datos Educativa para encontrar o descubrir la información que no es evidente solamente por registro de datos del curso, c) Recolectar la información e integrar todo en los reportes de resultados para analizar la información.

III. METODOLOGÍA DE INVESTIGACIÓN

La metodología implementada es relacionada con un experimento mixto, que emplea información tanto cuantitativa como cualitativa de recolección de datos y se apoya en la programación de las técnicas de Minería de Datos de Clustering y Redes Bayesianas para ser implementadas en base a la información proporcionada por los docentes que impartieron un curso de Física en estudiantes de reciente ingreso a un programa académico de Ingeniería Farmacéutica. La información se recolectó de tres grupos de estudiantes de nuevo ingreso (1FM1, 1FM2 y 1FV1) con una población total de $N = 107$ estudiantes, dentro del curso denominado Física del Movimiento Aplicada, el cuál abarca los temas de Cinemática, Fuerzas, Teorema de Energía y Mecánica de Fluidos. Se cubrieron todas las actividades desarrolladas por parte de los estudiantes, la siguiente tabla muestra los datos considerados para estructurar la base de datos que servirá para la aplicación de las técnicas: Clustering y Redes Bayesianas.

TABLA II. Variables cuantitativas y cualitativas que considerar dentro del espacio de trabajo.

Tipo de variable	Descripción	Duración en tiempo	Escala
Cuantitativa	Evaluación diagnóstica (Pre-test)	1 hora	0-10 de calificación.
	Tareas durante parciales	1.5 a 3 horas	0-10 de calificación.
	Exámenes realizados	1.5 horas	0-10 de calificación.
	Prácticas virtuales de laboratorio	27 horas/semestre @ 1.5 horas/semana	0-10 de calificación.
	Trabajo en plataforma virtual	54 horas/semestre @ 3 horas/semana	Tiempo.
	Clases síncronas virtuales	81 horas/semestre @ 4.5 horas/semana	Tiempo.
	Tiempo dedicado a la escuela	243 horas/semestre @ 13.5 horas/semana	Tiempo.
	Género		Femenino, masculino.
	Comprensión lectora ¹		Literal, inferencial, crítica, apreciativa y creativa.
	Nivel de redacción de textos académicos ²		Claridad y orden, originalidad, estilo adecuado, corrección y propiedad, uso de léxico técnico adecuado, etapas
Nivel educativo de los padres		Educación básica, educación secundaria, educación media superior/bachillerato, educación superior/profesional, educación posgrado/Maestría y/o doctorado.	
Nivel económico.		Muy bajo, bajo, medio, medio alto, alto.	
Región geográfica origen		Noroeste (Baja California, Baja California Sur, Chihuahua, Durango, Sinaloa, Sonora), Noreste (Coahuila, Tamaulipas, Nuevo León), Oeste (Jalisco Michoacán, Colima, Nayarit), Este (Hidalgo, Puebla, Tlaxcala, Veracruz), Centro Norte (Aguascalientes, Guanajuato, Querétaro, San Luis Potosí, Zacatecas), Centro Sur (Ciudad de México, Estado de México, Morelos), Suroeste (Chiapas, Guerrero, Oaxaca), Sureste (Campeche, Quintana Roo, Tabasco, Yucatán)	
Tipo de vivienda		Propia, Rentada	
Zona de vivienda		Rural, Urbana	

*Fuente: Elaboración propia.

De la tabla anterior se pueden identificar los tipos de las variables, partiendo del origen, como puede ser cuantitativo o cualitativo, las actividades académicas están relacionadas tanto con las evidencias presentadas, como son las tareas, las evaluaciones, las prácticas virtuales y el tiempo dedicado a trabajar dentro de la plataforma educativa, se considera la duración de cada actividad durante la semana y el acumulado de tiempo invertido (para estas variables, los valores son estimados respecto a las planeaciones didácticas que se deben de reportar al área académica y estos tiempos pueden variar, sólo se estima un promedio de trabajo de manera autónoma por parte de los estudiantes), las clases de manera síncrona junto con el docente. Estas evidencias se complementan con información del tipo cualitativo, como el género, nivel de estudios de los padres, las regiones geográficas de procedencia, el nivel económico de la familia, tipo de vivienda y ubicación de la vivienda, dos variables cualitativas que permiten conectar el efecto del rendimiento académico tienen que ver con la comprensión lectora y la redacción de textos académicos, son habilidades fundamentales para el desarrollo de este tipo de actividades y más adelante se discutirá parte del aporte en estas dos variables cualitativas.

¹ (Guerrero Hernández, 2020)

² (Ñañez Silva & Lucas Valdez, 2017)

Lo que permite identificar aspectos personales de los estudiantes en función a su rendimiento académico y poder identificar si existen relaciones que no están estrechamente vinculadas entre sí, realizando una conversión cuantitativa a cada variable cualitativa se pueden establecer valores de relación directa y medir la diferencia o relación que existe entre cada dato contenido dentro de la variable para establecer posibles coincidencias respecto al desempeño académico y los antecedentes de los estudiantes respecto a su situación familiar y de vivienda. La siguiente imagen muestra una captura de pantalla en una hoja de cálculo electrónica (por ejemplo, Microsoft Excel™), donde se puede observar parcialmente las variables descritas anteriormente.

La figura anterior, permite identificar de manera general que la cantidad de información obtenida de un curso es lo suficientemente amplia como para poder generar un núcleo al cuál se le pueda aplicar alguna técnica de Minería de Datos Educativa, esto es útil ya que los docentes que trabajan con la información de sus cursos les permite de manera general llevar un control de sus cursos, pero al tratar de manipular está información en un software de ofimática como es una hoja electrónica de cálculo se complica ya que esta no tiene la capacidad de analizar la información de acuerdo al tipo de las variables que se tienen almacenadas o de acuerdo a las relaciones que se puedan establecer en dichas variables, por lo que se requiere de técnicas más complejas que permita esta interacción y estudio de las mismas. Con la información recolectada se obtuvieron las siguientes instancias de valores:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA			
GRUPO	ZONA_VIVIENDA	TIPO_VIVIENDA	REGION	ECONOMIA	NIVEL	GENERO	EVALUACION	TAREA 1	TAREA 2	TAREA 3	TAREA 4	TAREA 5	TAREA 6	TAREA 7	TAREA 8	TAREA 9	TAREA 10	TAREA 11	TAREA 12	TAREA 13	TAREA 14	TAREA 15	TAREA 16	TAREA 17	TAREA 18	TAREA 19	TAREA 20		
2	GRUPO	ZONA_VIVIENDA	TIPO_VIVIENDA	REGION	ECONOMIA	NIVEL	GENERO	EVALUACION	TAREA 1	TAREA 2	TAREA 3	TAREA 4	TAREA 5	TAREA 6	TAREA 7	TAREA 8	TAREA 9	TAREA 10	TAREA 11	TAREA 12	TAREA 13	TAREA 14	TAREA 15	TAREA 16	TAREA 17	TAREA 18	TAREA 19	TAREA 20	
3	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	10	5.5	6.25	6.25	4.56	3.5	1.4	5.96	5	10	10	10	10	2.5	10	8	8	5.0142857	9.13	3.652		
4	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	POSGRADO	F	0	8.2	3.75	5	6.25	3.024	6.6	2.64	5.664	5	10	10	10	10	0	0	0	0	0	4.2857143	7.95	3.18	
5	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	BACHILLERATO	F	10	7	6.25	6.25	5	4.14	7.1	2.84	6.98	7	10	10	7.5	10	10	10	10	8	8	5.6142857	8.83	3.532	
6	IPMA1	URBANA	RENTADA	CENTRO NOR MEDIO ALTO	SUPERIOR	F	10	9.4	4.31	1.25	5	3.5952	3.7	1.48	5.0752	5	10	10	2.5	0	5	8	8	8	3.5571429	6.75	2.7		
7	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	9.4	6.38	8.75	3.75	4.9956	4.9	1.96	6.5556	7	10	10	0	0	0	10	8	8	6.3857143	0	0		
8	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	M	10	4	4.75	6.25	1.25	3.15	5.2	2.08	5.23	5	10	10	6.25	7.5	10	0	8	8	5.1214286	8.25	3.3		
9	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	SUPERIOR	M	10	9.4	6	0	0	3.048	0	0	3.048	3	0	0	0	0	0	0	0	0	0	0	0	0	
10	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	POSGRADO	F	8	8.8	4.81	3.75	3.75	3.7532	4.8	1.92	5.6552	5	10	10	6.25	6.67	10	8	8	8	4.0217143	6.48	2.592		
11	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	8.8	4.85	7.5	6.25	4.464	7.2	2.88	7.344	7	10	10	8.75	7.5	10	10	10	10	10	5.6785714	10	4	
12	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	0	0	0	0	1.25	0.15	0	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	M	10	7	4.9	10	0	2.828	9.1	3.64	7.488	7	10	10	10	7.5	10	10	10	10	10	5.7671429	7.33	2.932	
14	IPMA1	RURAL	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	1.1	4.75	10	0	3.052	6.3	3.52	6.432	6	10	10	8.75	10	6.67	8	8	8	5.2645714	7.05	3.06		
15	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	7	5.15	1.25	0	2.808	0.9	0.36	3.168	3	10	10	5	7.5	5	8	8	8	4.5857143	7.35	2.94		
16	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	4.6	3.66	5	2.5	3.0912	5.8	2.32	5.4112	5	10	10	7.5	2.5	8.33	10	8	8	4.8285714	10	4		
17	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	10	2.75	8.75	2.5	4.08	3.4	1.36	5.44	5	10	10	6.25	7.5	3.33	8	8	8	4.54857143	7.93	3.132		
18	IPMA1	URBANA	RENTADA	CENTRO NOR MEDIO	SUPERIOR	M	0	0	0	0	0	0	1.3	0.52	0.52	0	10	10	0	0	0	0	0	0	0	1.7142857	1.18	0.472	
19	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	SUPERIOR	M	10	8.8	6.15	6.25	5	4.344	8.6	3.44	7.784	8	10	10	10	7.5	10	10	10	10	10	5.7857143	8.83	2.532	
20	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	7.6	7.16	2.5	2.5	3.9712	5.8	2.32	5.8912	5	10	10	10	10	6.67	8	8	8	5.5431429	6.5	2.6		
21	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	SUPERIOR	F	10	7.6	2	7.5	3.75	3.702	2	0.8	4.502	4	10	10	7.5	2.5	8.33	6	8	8	4.4854286	10	4		
22	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	BACHILLERATO	F	10	7.6	8	6.25	3.13	4.1976	8.6	3.44	7.8376	8	10	10	10	5	3.33	6	10	10	4.656871	9.7	3.88		
23	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	M	10	8.2	5.98	8.75	3.75	4.0216	6.9	2.76	7.1616	7	10	10	7.5	5	6.67	8	8	8	4.728871	9.13	3.652		
24	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	6.4	4.15	10	8.75	4.116	4.9	1.96	6.076	6	10	10	10	5	6.67	10	8	8	5.1145714	8.53	3.412		
25	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	BACHILLERATO	F	10	5.2	7.32	7.5	2.5	3.9024	3.3	1.32	3.2214	3	10	10	7.5	2.5	3.33	6	8	8	4.056871	7.33	2.94		
26	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	POSGRADO	F	10	7.6	4.08	7.5	2.5	3.9016	4.3	1.72	5.3216	5	10	10	7.5	5	6.67	10	8	8	4.9002857	7.05	2.82		
27	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	BACHILLERATO	F	0	0	9.88	6.25	4.38	1.7932	4	1.6	3.9332	3	10	10	5	0	1.67	0	2	2	2.4574286	7.65	3.06		
28	IPMA1	RURAL	RENTADA	CENTRO NOR MEDIO	BASICO SECUNDARIA	F	10	8.2	4	8.75	7.5	4.614	4.3	1.72	6.336	4	10	10	10	2.5	10	10	10	10	10	6.50142857	9.7	3.88	
29	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	BASICO SECUNDARIA	F	10	8.2	3.33	2.5	0	2.7636	3.4	1.36	4.1236	4	10	10	7.5	2.5	10	8	8	8	4.8	5.88	2.352		
30	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	BASICO SECUNDARIA	F	0	8.8	5.4	3.75	1.25	2.964	4.06	1.624	3.938	3	10	10	8.75	10	5	10	8	8	5.1142857	8.25	3.3		
31	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	BACHILLERATO	M	0	5.2	4.4	9.88	6.25	3.0276	7.4	2.96	5.9876	5	10	10	7.5	10	10	10	10	10	8	5.6142857	9.1	3.64	
32	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	8.8	7.15	7.5	2.5	4.314	4.9	1.96	6.274	6	10	10	5	0	10	10	10	10	10	8	4.5428571	6.75	2.7
33	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	M	10	3.5	6.68	3.75	0	2.8476	0.6	0.24	3.0876	3	10	10	5	5	3.33	4	4	4	3.5425714	5.3	2.12		
34	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	9.4	4.48	1.25	3.75	3.4656	6.3	2.52	5.9856	5	10	10	7.5	5	6.67	6	6	6	4.386	8.25	3.3		
35	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	0	0	5.15	8.75	6.25	2.418	8.6	3.44	5.858	5	10	10	10	5	3.33	6	0	0	3.7997143	6.5	2.6		
36	IPMA1	URBANA	PROPIA	CENTRO SUR MEDIO	SUPERIOR	F	10	3.4	6.4	0	0	2.376	5.7	2.28	4.656	4	10	10	10	0	0	0	0	0	0	2.9142857	3.23	1.292	
37	IPMA1	URBANA	PROPIA	CENTRO SUR MEDIO	SUPERIOR	F	10	6.5	5.25	8.75	1.25	3.81	8.8	3.52	7.33	7	10	10	10	5	3.33	8	8	8	4.4854286	7.63	3.052		
38	IPMA1	URBANA	RENTADA	CENTRO NOR ALTO	SUPERIOR	F	10	8.2	4.5	5	1.25	3.474	3.3	1.32	4.794	4	10	10	10	7.5	6.67	10	8	8	5.328871	8.25	3.3		
39	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	0	3.4	0	1.25	1.25	1.088	6.9	2.52	3.588	3	10	10	7.5	7.5	6.67	10	10	10	5.286	4.7	1.88		
40	IPMA1	URBANA	PROPIA	SURCOESTE MEDIO	SUPERIOR	F	10	6	3.58	5	8.75	3.9996	4.3	1.72	5.1196	5	10	10	2.5	0	3.33	4	10	10	3.414	5.3	2.12		
41	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	M	0	7	4.15	2.5	5	2.238	4.3	1.72	3.958	3	10	10	7.5	5	0	6	8	8	3.8142857	0	0		
42	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	8.2	7.9	7.5	6.25	4.782	6.6	2.64	7.422	7	10	10	10	10	10	10	10	10	10	5.1428571	7.08	2.832	
43	IPMA1	URBANA	PROPIA	CENTRO NOR MEDIO	SUPERIOR	F	10	5.2	5.35	2.5	2.5	3.042	2.3	0.92	3.962	3	10	10	5	5	5	10	8	8	4.5428571	4.38	1.752		
44	IPMA2	URBANA	PROPIA	CENTRO NOR MEDIO ALTO	POSGRADO	F	10	8.8	5.25	7.5	7.5	4.686	7.7	2.88	7.786	8	10	10	10	7.5	10	0	10	10	10	4.9285714	7.95	3.18	
45	IPMA2	RURAL	PROPIA	CENTRO NOR MEDIO	BACHILLERATO	F	0	10	5.15	2.5	0																		

información se realizó la aplicación de las técnicas de MDE para identificar patrones o relaciones no evidentes dentro del curso de Física antes descrito, los resultados y discusión se presentan en la siguiente sección.

IV. ANÁLISIS DE RESULTADOS Y DISCUSIÓN

Para analizar los resultados obtenidos, se aplicaron las técnicas de K-vecinos más cercanos, considerando el agrupamiento de las instancias en función de los promedios finales y tomando como referencia los niveles de comprensión lectora y de comunicación escrita en los estudiantes ya que son habilidades constantes de desarrollo y aplicación durante el curso de Física que realizaron los estudiantes, en la siguiente figura (Fig. 2), se puede observar el comportamiento en el rendimiento académico de los estudiantes en función de los grupos de estudiantes y su correspondiente ajuste conforme a su situación académica, calificaciones departamentales (3 por semestre), ingresos por parte de los padres y resultados en las calificaciones obtenidas. En todos los casos, los porcentajes corresponde a la relación entre mujeres y hombres para tomar la decisión para cada nodo del árbol y los porcentajes se van distribuyendo conforme las variables definidas en la Tabla II.

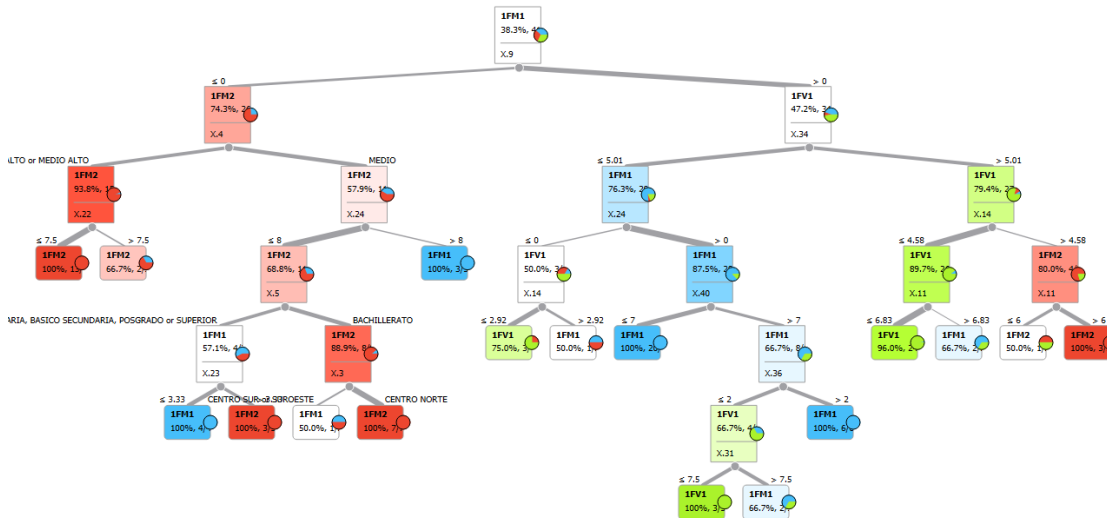


FIGURA 2. Árbol de instancias en función del género, nivel de ingreso en casa y promedios finales.

En la figura anterior se puede observar que el comportamiento es positivo para el aprovechamiento, cuando los estudiantes tienen mejores habilidades de comunicación, de manera más específica y filtrando la información para focalizar los agrupamientos en función solamente de la comunicación media se generó el siguiente árbol (ver Figura 3).

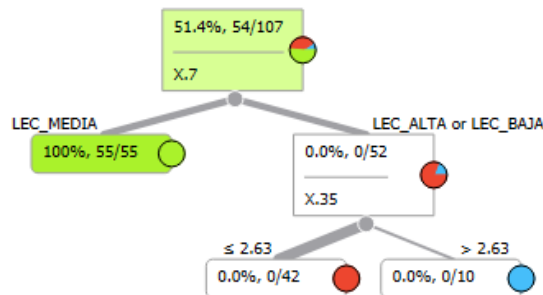


FIGURA 3. Árbol de decisiones en función de las habilidades de comunicación en los estudiantes.

De la figura anterior, se puede observar que mientras los estudiantes tengan desarrolladas las habilidades de comunicación a nivel más alto, es más confiable y probable que los mismos puedan acreditar el curso de Física, esto es

importante ya que el desarrollo de los estudiantes en estas habilidades serán de suma importancia para su formación como profesionistas en el ámbito laboral, considerando que los cargos de nivel intermedio de directivo hacia puestos más altos, implican un mayor desarrollo en las comunicaciones, tanto verbales como escritas. Para el caso de los comportamientos en función del género (Masculino [M] o Femenino [F]), se observaron los siguientes comportamientos conforme a los resultados de rendimiento académico, véase la Figura 4.

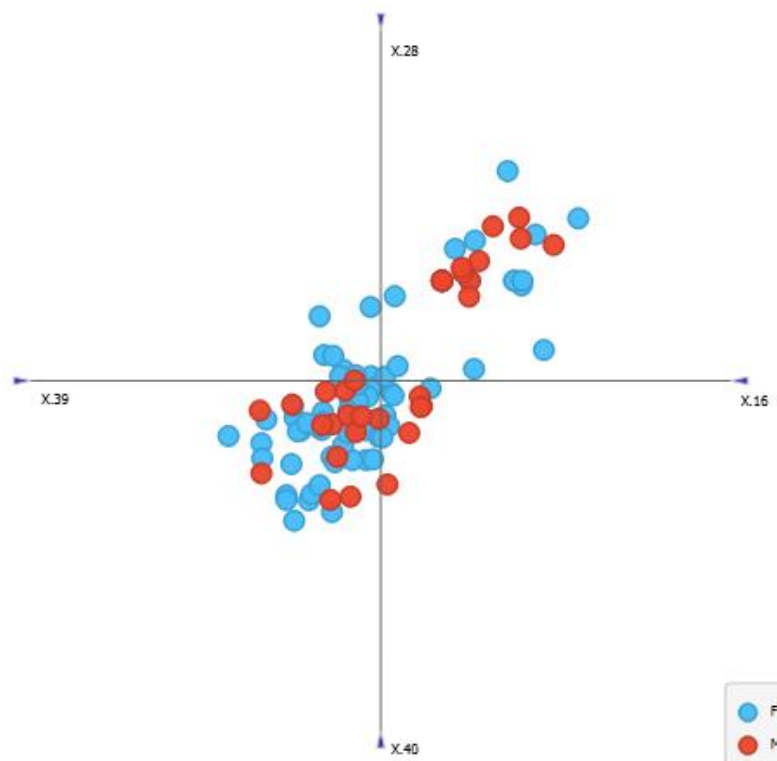


FIGURA 4. Rendimientos académicos en función del género por parte de los estudiantes (Masculino [M] o Femenino [F]).

De la figura anterior se puede observar que no hay una diferencia significativa respecto al género que participó en el estudio correspondiente, esto es útil ya que las estrategias didácticas implementadas no influyeron respecto al pensamiento de género, sólo en función de las habilidades de comunicación oral o escrita, concordando con los resultados obtenidos en las figuras anteriores, dónde es el desempeño individual y grupal, así como los habilidades de transmitir y comunicar las ideas para que el rendimiento en los estudiantes sea favorable o positivo para su aprendizaje. En la siguiente sección se presentan las conclusiones del trabajo desarrollado.

V. CONCLUSIONES

El análisis de la información proporcionada por las actividades académicas que desarrollaron los estudiantes permitió identificar el comportamiento en el rendimiento académico, habilidades a tomar en cuenta por los estudiantes tienen que ver con los aspectos de comunicación, tanto oral como escrita, este tipo de habilidades son básicas para el desarrollo y aprendizaje de conceptos complejos como es el caso de Física.

Respecto a la pregunta de investigación planteada al inicio del trabajo, así como la contrastación de hipótesis se puede concluir lo siguiente: “¿cómo se puede emplear la Minería de Datos Educativa para identificar las trayectorias académicas en estudiantes que cursan un programa académico de ingeniería mediante la información en los cursos de Física que cursaron?” La hipótesis propuesta en el trabajo permite la ratificación de manera positiva de la misma, así

como la respuesta la pregunta de investigación también planteada en la sección II del presente trabajo, la cuál describe que: “*La aplicación de las técnicas de Minería de Datos: Clustering y Redes Bayesianas, permitirán identificar las trayectorias académicas de estudiantes que cursan un programa académico de ingeniería y en base a la información que proveen las actividades en un curso de Física*”.

Adicionalmente los objetivos planteados se cumplieron de manera satisfactoria, considerando que las actividades y las etapas en el experimento propuesto permitieron identificar el comportamiento de los datos, así como la naturaleza de estos para poder emplearlos en el descubrimiento de nueva información. Se recomienda continuar ampliando el uso de la Minería de Datos a grupos más grandes y analizar las trayectorias de nuevas generaciones en una situación postpandemia, para identificar nuevos posibles comportamientos o rendimientos académicos.

AGRADECIMIENTOS

Se agradece al Instituto Politécnico Nacional de México por las facilidades otorgadas para el análisis de la información dentro de los cursos en línea ubicados en la dirección electrónica: <http://virtual.upiig.ipn.mx>, también se agradece el apoyo de los programas EDI-SIP-IPN, SIBE-COFAA-IPN y BEIFI-COFAA-IPN, así como los fondos a través de los proyectos SIP-2021-1052 y SIP-2022-0796.

REFERENCIAS

- Cutro, A. (02 de 2010). *DataPrix Knowledge is the Goal*. Obtenido de 1.7.1 Evolución Historia de la Minería de. <https://www.dataprix.com/es/mineria-datos-aplicada-encuesta-permanente-hogares/171-evolucion-historia-mineria-datos>
- Espigares, M., García, R., & Quiñones C. J. (2006). *Minería de datos educativos en plataformas de teleformación*. Sevilla: Universidad de Sevilla.
- Fenton, W. (2018). *PC Magazine Online*. de The Best (LMS) Learning Management Systems: <https://www.pcmag.com/picks/the-best-lms-learning-management-systems#:~:text=While%20Blackboard%20remains%20the%20most,and%2027%20percent%20of%20enrollments>.
- García Robelo, O., & Barrón Tirado, C. (2011). *Un estudio de la trayectoria escolar de los estudiantes de doctorado en Pedagogía*. Perfiles Educativos, XXXIII(131), 94-113.
- Guevara, H., & Belelli, S. (2013). *Las trayectorias académicas: Dimensiones personales de una trayectoria estudiantil. Testimonio de un actor*. Revista del Instituto de Investigaciones Socio Económicas, 4(4), 45-56.
- Jiménez Galindo, A. (2010). *Minería de Datos en la Educación*. Madrid: Universidad Carlos III de Madrid.
- Martínez Sánchez, A. (2018). *Trayectorias escolares en el nivel superior, un análisis desde algunos indicadores académicos*. Atenas.
- Ortega Guerrero, J., López González, R., & Alarcón Montiel, E. (2015). *Trayectorias escolares en educación superior. Propuesta metodológica y experiencias en México*. Universidad Veracruzana - Instituto de Investigaciones en Educación.

- Oviedo Carrascal, A., & Jiménez Giraldo, J. (2019). *Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas SABER-PRO*. Revista Politécnica, 128-140.
- Pérez Alcántara, B. D. (2017). Análisis de trayectorias escolares de estudiantes de licenciatura en geografía de la UAEMEX. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 7(14), 1-23.
- Quintela Dávila, G. (2007). *Trayectorias académicas y origen social*. Barcelona: Universitat Autònoma de Barcelona.
- Rico Páez, A., Gaytán Ramírez, N., & Sánchez Guzmán, D. (2019). *Construcción e implementación de un modelo para predecir el rendimiento académico de estudiantes universitarios mediante el algoritmo de Naïve-Bayes*. *Dialogos sobre educación. Temas actuales en investigación educativa*, 10(19), 3-18.
- Rosado Gómez, A., & Verjel Ibañez, A. (2017). *Aplicación de la minería de datos en la educación en línea*. *Revista Colombia de Tecnologías de Avanzada*, 1, 92-98.