

Identificación del comportamiento y preferencias para usuarios del campus virtual mediante técnicas de minería web

Identification of behavior and preferences for virtual campus users using web mining techniques

Oswaldo Enrique Vélez Langs, Daniel José Salas Álvarez, Alexander Fernández Arango

Departamento de Ingeniería de Sistemas

Universidad de Córdoba

Montería, Colombia

{oswaldovelez, danielsalas, alexanderfernandez}@correo.unicordoba.edu.co

Resumen — Este trabajo busca conocer el comportamiento general de los usuarios de la plataforma virtual de la Universidad de Córdoba (Montería-Colombia), a través la de minería del uso de los datos provenientes de archivos de monitoreo de acceso. Para esto fue necesario realizar un proceso de exploración de los datos almacenados en el campus virtual en los archivos de referencia de acceso, a través de una minería de uso. Se implementaron técnicas de minería exploratoria, con lo que se identificó la manera de abordar los datos obtenidos, seguidamente se pudo consolidar los resultados, lo que permitió la realización de su contraste frente al análisis descriptivo y lo arrojado por la minería de datos web. Cabe resaltar que el comportamiento sugerido por el modelo en la agrupación de datos llega a ser similar al de los usuarios en la parte descriptiva del análisis. Se pudo concluir que la generación de modelos predictivos a través del uso de técnicas de minería web, puede ser una herramienta de gran importancia para el conocimiento de los patrones y el comportamiento de los usuarios de la plataforma virtual de la Universidad de Córdoba, así como la posibilidad de usar esto como respaldo para un proceso futuro de actualización y análisis de estos patrones y comportamientos.

Palabras Clave - Minería de datos, Modelos de predicción, visualización de datos, Herramientas de minería web, Sistemas de gestión de aprendizaje.

Abstract — This paper search for the understanding of the general behavior of users of the Learning Management System of the University of Cordoba (Montería-Colombia), using web usage mining of log files. For this purpose, it was necessary to carry out a process of exploration of the data stored in the virtual campus in the access reference files, through usage mining. Exploratory mining techniques were implemented, with which the way to approach the data obtained was identified, then the results were consolidated, which allowed for their contrast with the descriptive analysis and the results of the web data mining. It should be noted that the behavior suggested by the model in the grouping of data is like that of the users in the descriptive part of the analysis. It could be concluded that the generation of predictive models using web mining techniques can be a tool of

great importance for the knowledge of the patterns and behavior of the users of the virtual platform of the University of Cordoba, as well as the possibility of using this as support for a future process of updating and analysis of these patterns and behaviors.

Keywords - Data mining, predictive modeling, data visualization, web mining tools, learning management systems.

I. INTRODUCCIÓN

En la actualidad, el uso de plataformas virtuales para la educación ha experimentado un importante crecimiento [1]. Esto se debe a la necesidad de las instituciones de incorporar nuevas herramientas que faciliten la administración de contenidos y procesos académicos internos.

En este contexto, los Learning Management System (LMS) han adquirido una relevancia crucial como elementos de apoyo a la educación. Estos sistemas permiten administrar, controlar y distribuir las actividades académicas necesarias para la formación, facilitando la interacción entre docentes y estudiantes de manera sincrónica o asincrónica [2], [3].

En el contexto educativo, el uso de técnicas de Web Usage Mining puede ayudar a comprender cómo los estudiantes interactúan con la plataforma virtual, cuáles son las áreas de interés y las dificultades que enfrentan al realizar tareas y actividades. Esto permite a los docentes y responsables de la plataforma adaptar y mejorar el contenido y la experiencia de aprendizaje, de acuerdo con las necesidades y preferencias de los usuarios [4].

En la actualidad, gracias a la integración de diversas herramientas, es posible lograr una gran similitud entre los procesos de educación presenciales y virtuales. Una de estas herramientas son las conferencias sincrónicas, que proporcionan un ambiente virtual adecuado para la interacción entre docentes y estudiantes.

La Universidad de Córdoba en Colombia ha sido una institución que se ha adaptado a los requerimientos actuales de la educación virtual, y ha incorporado herramientas como las plataformas LMS para la interacción entre la planta docente y estudiantil. Actualmente, la Universidad de Córdoba utiliza el LMS Moodle [5],[6] como sistema de apoyo al proceso académico llevado a cabo en la institución.

Este sistema es ampliamente utilizado en la educación a nivel mundial y permite la gestión de contenidos, actividades y evaluaciones de manera eficiente.

Este proceso de apoyo se realiza a través de la interacción entre docentes y estudiantes llevando a cabo en el desarrollo de los contenidos académico y como apoyo al proceso pedagógico de la Universidad.

El uso rutinario de este tipo de plataformas genera habitualmente una cantidad notoria de información relacionada, que a través de su análisis permite la comprensión de diferentes tipos de comportamiento, ubicado justamente en el proceso que se desarrolla a través de las labores académicas [6].

Precisamente es en este punto donde surge la necesidad de hallar una metodología que permita el análisis eficiente de esta cantidad de información, la cual permitirá comprender variables relacionadas con el comportamiento de uso de la plataforma para los estudiantes y para los docentes.

La Minería de Datos (MD) [7], se refiere a un conjunto de procesos, técnicas y herramientas utilizadas para extraer información útil de grandes conjuntos de datos, como las bases de datos.

En particular, la minería de datos se enfoca en la exploración y descubrimiento de patrones, tendencias y comportamientos previamente desconocidos, y se utiliza como un proceso de descubrimiento en el camino hacia la inteligencia artificial [8].

En el contexto de la educación, el EDM (Educational Data Mining) es una extensión de la minería de datos, donde hay un área concreta que se enfoca en descubrir patrones y comportamientos en los repositorios académicos para predecir el comportamiento de los estudiantes y profesores [9],[10].

El análisis de datos académicos a través del EDM, apoyado por herramientas conceptuales y tecnológicas como las Analíticas de Aprendizaje, permite medir, recopilar y presentar datos para optimizar el aprendizaje en diferentes contextos [11],[12],[13].

Además, la minería de datos en la educación también puede utilizarse para extraer información sobre los hábitos, preferencias y patrones de comportamiento de los usuarios de plataformas virtuales, lo que puede ser útil para mejorar la calidad de la educación en línea.

La organización y visualización de los datos y sus diferentes técnicas constituye en un factor esencial para mostrar los resultados producto del análisis desde los LMS, desde esta perspectiva el estudio de Gómez- Aguilar en el 2014 [14], mediante el modelo Vela (Analytical visualization in Visual e-learning) para representación datos de estudiantes recuperados a través de un LMS, están relacionadas con líneas

de tiempo en espiral, nube de etiquetas, coordenadas paralelas y gráficos de redes sociales, por otra parte, Al. Ashmoery en el 2015 [15], realiza un análisis de las técnicas de visualización, entre las que se destacan Moclog, Travis, Gismo, CourseViz, DÍAS, MTRDS, entre otras, sin embargo, en su trabajo utiliza interfases de graficas sencillas utilizando datos de estudiantes en un LMS.

Un estudio reciente realizado por González et al. En 2021 [23] utilizó técnicas de minería web para analizar el comportamiento de los estudiantes en un campus virtual. Los resultados mostraron que la mayoría de los estudiantes accedían al campus virtual en horarios nocturnos y que utilizaban principalmente dispositivos móviles para acceder a la plataforma.

Por otro lado, Aparicio et al. en 2020 [22] utilizó técnicas de minería web para analizar las preferencias de los estudiantes en cuanto a los contenidos y recursos del campus virtual. Los resultados mostraron que los estudiantes preferían los contenidos multimedia y los recursos interactivos, y que tendían a evitar los materiales excesivamente teóricos.

Además de la identificación de comportamientos y preferencias de los usuarios en un campus virtual, la minería web también puede ser utilizada para la personalización del aprendizaje en línea.

Un estudio reciente realizado por Chen et al. en 2022 [24] utilizó técnicas de minería web para analizar los patrones de navegación de los estudiantes en una plataforma de aprendizaje en línea. Los resultados mostraron que la personalización del aprendizaje basada en los patrones de navegación mejoró significativamente el rendimiento académico de los estudiantes.

Por otro lado, la minería web también puede ser utilizada para la detección de plagio en los trabajos de los estudiantes. Un análisis realizado por Mohammadi et al. en 2021 [25] utilizó técnicas de minería web para detectar plagio en los trabajos de los estudiantes de una universidad.

Los resultados mostraron que la minería web fue una técnica efectiva para la detección de plagio en los trabajos de los estudiantes.

En este caso específico, el análisis de Web Usage Mining se está aplicando al campus virtual de la Universidad de Córdoba para identificar patrones de comportamiento y preferencias de los estudiantes y docentes al interactuar con la plataforma virtual.

Esto implica la recopilación de datos relacionados con las acciones que realizan los usuarios en el campus virtual, como las páginas que visitan, el tiempo que pasan en cada página, las acciones que realizan en la plataforma, entre otros aspectos relevantes. [26],[27].

Con esta información, se pueden descubrir patrones de comportamiento y preferencias de los usuarios, lo cual puede ser útil para mejorar la experiencia de usuario en el campus virtual y optimizar el proceso de enseñanza y aprendizaje. Además, el análisis de Web Usage Mining también puede proporcionar información valiosa para la toma de decisiones en la gestión de la plataforma virtual y en la implementación de

estrategias para mejorar la interacción entre los usuarios y la plataforma [28].

En este contexto, el Web usage mining, además, ayuda a desarrollar sistemas de aprendizaje personalizado, con el objetivo de mejorar sustancialmente el aprendizaje de los estudiantes.

Esto como parte de una posible estrategia de mejoramiento [19]. La identificación de patrones que genera la implantación del proceso llevado a través de los elementos del Web usage mining, logrando con esto la adaptabilidad de los sistemas de gestión educativos al entorno propio de la educación, lo que incrementa la efectividad en las respuestas hacia los usuarios [21].

II. METODOLOGÍA

Para llevar a cabo un análisis completo y detallado del comportamiento y las preferencias de los usuarios del campus virtual Unicordoba, es necesario aprovechar la gran cantidad de datos que se generan en esta plataforma. Para ello, se pueden utilizar técnicas de minería de datos, que permiten procesar y analizar estas variables de comportamiento.

Mediante el análisis de datos web, es posible descubrir pautas y tendencias en la utilización de un entorno virtual de aprendizaje, tales como la frecuencia de ingreso, las páginas con mayor afluencia, los recursos más solicitados, entre otros aspectos relevantes. Además, esta metodología posibilita la generación de perfiles de usuarios según sus intereses y preferencias individuales.

El objetivo principal de este análisis es comprender el comportamiento de los usuarios de la plataforma virtual de la Universidad de Córdoba, mediante el uso de técnicas de minería de datos aplicadas a los archivos de monitoreo de entradas al sistema.

La metodología que se puede utilizar para la identificación del comportamiento y preferencias de los usuarios del campus virtual mediante técnicas de minería web es el proceso KDD (Knowledge Discovery in Databases), que consta de las siguientes etapas: selección, preprocesamiento, transformación, minería de datos y evaluación [31].

La selección implica identificar y definir el conjunto de datos a analizar. En este caso, los datos a analizar serían el registro de actividades de los usuarios en el campus virtual, como las páginas visitadas, los recursos descargados, la frecuencia de acceso, etc.

El preprocesamiento se refiere a la preparación de los datos para el análisis, lo que incluye la eliminación de datos irrelevantes o duplicados, la normalización de los datos y la corrección de errores. La transformación implica la conversión de los datos preprocesados en un formato adecuado para el análisis, como tablas de frecuencia o gráficos.

La minería de datos se refiere a la aplicación de técnicas de minería web para analizar los datos y extraer patrones y tendencias en el comportamiento de los usuarios del campus virtual. Así como el uso de algoritmos de agrupamiento que permitiría identificar la forma de abordar los datos obtenidos con resultados confiables y representativos. Seguidamente es

necesario consolidar estos datos que conllevan a la identificación de falencias y posibles consideraciones de mejoras [32].

Todo esto, como se ha descrito, a través de la evaluación implica la validación de los resultados obtenidos mediante técnicas de minería de datos para asegurar que sean relevantes y útiles para el análisis [30].

Todo esto conlleva el uso de un extenso proceso de análisis descriptivo, fue posible identificar patrones de uso, grupos, tendencias, frecuencias y cantidades, al igual que tipos de mayor y menor interés.

Para llevar a cabo este análisis, se adquieren los datos necesarios a través de los funcionarios del centro CINTIA de la Universidad de Córdoba, quienes son responsables de la administración y gestión de la plataforma Moodle, la cual es la base fundamental de la plataforma virtual.

Estos datos son esenciales para la aplicación de técnicas de minería de datos y obtener información relevante sobre el comportamiento y preferencias de los usuarios en la plataforma. Es importante contar con la colaboración y apoyo de los expertos en el manejo de la plataforma para garantizar la calidad y precisión de los datos utilizados en el análisis.

Estos datos se encuentran en los logs de operación y gestión de la plataforma en lo relacionado a los cursos y sus usuarios cuando ingresan a esta, estos son los datos fuentes originales para el estudio. Estos se basan en los cursos de pregrados gestionados en el intervalo incluido en el primer y segundo periodo académico del año 2019.

En lo relacionado a los registros se obtuvo un archivo proveniente de los datos log de la plataforma Moodle de la Universidad de Córdoba para el intervalo en cuestión, con más de 12.000.000 registros, que conformarán el data set a utilizar, abarcando esto la operación de los usuarios en sus diversos roles al interactuar con los cursos de cada asignatura y demás procesos apropiación académica en el intervalo mencionado.

Se importan luego los datos seccionados como una base de datos disponible en MySQL, para preparar y seleccionar los campos dispuestos para el análisis descriptivo, que junto con la herramienta de hoja de cálculos se usa para dicho proceso.

Seguidamente es necesario una selección previa de los atributos para la realización del proceso de análisis descriptivo y de minería de datos. Una vez realizada la parte anterior se procedió a generar una serie de consultas que permitió la identificación inicial del comportamiento descriptivo de los usuarios según sus patrones e identificados por medio de este análisis (R. Baeza-Yates. 2009).

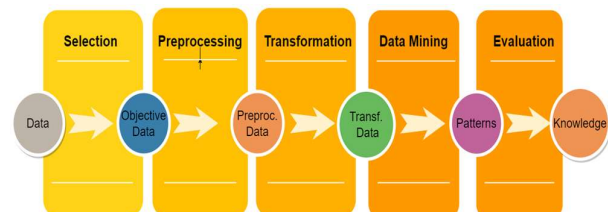


Figura 1. El proceso de descubrimiento en Base de Datos.

III. RESULTADOS

Para determinar e identificar los diferentes patrones de comportamiento y preferencias, se llevó a cabo un análisis de los datos provenientes del desarrollo diario de actividades en este LMS. Estos datos representan un seguimiento preciso de la actividad de los usuarios.

El conocimiento de dichos patrones permitió proyectar mejoras que buscan optimizar el servicio, así como identificar los procesos clave que podrían afectar negativamente la eficiencia en su uso.

La Universidad de Córdoba realiza en gran medida sus procesos académicos con la Plataforma Virtual, emplazada en una instancia de administración propia, proveída por uno de los más importantes sistemas de administración de aprendizaje como lo es Moodle LMS.

A. Recursos usados en la identificación

A través del uso de algoritmos de agrupamiento de instancias se procede a identificar las instancias con características similares, teniendo en cuenta los criterios de comparación escogidos, como son análisis por género y análisis por tipo de rol de usuario, entre los diversos atributos que definen las instancias procesadas por los algoritmos.

El algoritmo EM se origina en la estadística [34] y suele ser mucho más elaborado que su contraparte K-median [27]. Sin embargo, su uso requiere muchas más operaciones, lo que en este caso de análisis presenta una desventaja pues la cantidad de registros manejados por instancias es muy alto.

Para lograr agilidad en los procesos de cómputo y análisis en el uso del algoritmo para la gestión del modelado usando los datos previos de la plataforma LMS de la Universidad de Córdoba, fue necesario la utilización de una muestra representativa de los datos para el respectivo proceso del modelado en cuestión.

Una vez que se ha llevado a cabo el proceso mediante la realización de una muestra representativa se carga este con formato CSV [34], para proceder al análisis con la herramienta WEKA.

Sobresale, que luego de procesar el archivo a través de consultas SQL previamente desarrolladas, en donde se obtiene unos cerca de 4.000.000 millones de registros en esta fase.

Se obtienen varios archivos con registros de muestra representativas de entre 250.000 a 500.000 registros aproximadamente para ser procesados por la herramienta, como conjunto de datos.

B. Resultados del proceso analizado

De forma descriptiva se logran identificar inicialmente algunos patrones de uso que arrojó el procesamiento de las consultas a los registros del conjunto de datos.

Entre estas se pudo determinar que, programas con mayor cantidad de accesos en el intervalo de tiempo analizado, programas con mayor cantidad de accesos por día, programas con menor cantidad de accesos en el intervalo previsto.

Las asignaturas con mayores accesos por estudiantes en intervalo de tiempo analizado, asignaturas con mayores accesos por día, asignaturas con mayores accesos por profesores en intervalo de tiempo analizado, acceso de alumnos por día intervalo de tiempo analizado, acceso de usuarios por hora entre otros. Seguidamente se muestran algunos patrones identificados.

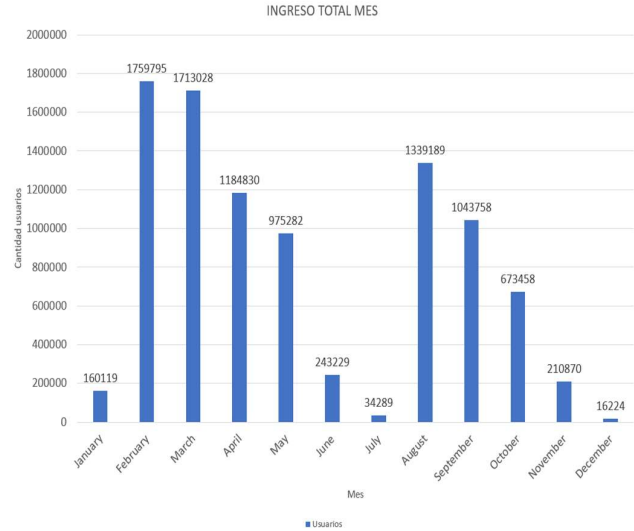


Figura 2 Ingreso total por mes.

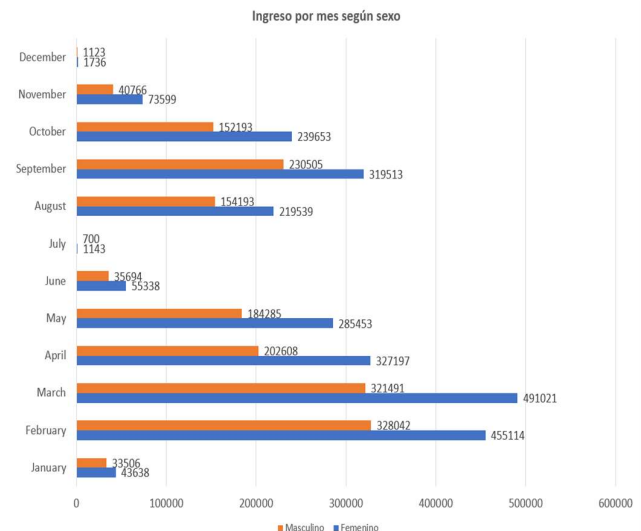


Figura 3. Ingreso por mes según sexo

Una vez que se determinó la técnica a usar para la aplicación del algoritmo escogido, se realiza la muestra representativa.

Se cargan estos con formato CSV, para proceder al análisis en la herramienta Weka [18], a través del uso del algoritmo de agrupamiento EM [34].

Los atributos quedan distribuidos en 6 instancias de agrupamiento inicial como se ve en la tabla siguiente.

TABLE I. Distribución de clústeres arrojados

Clustered	Instances	%
0	64214	13%
1	155409	31%
2	58553	12%
3	97811	20%
4	24723	5%
5	99293	20%

Las instancias de los atributos se encuentran agrupadas claramente en el Clúster número uno (1), el modelo arrojó que atributos logran una interacción justo en ese clúster. La siguiente tabla muestra la interacción de todos los clústeres en los registros del atributo mes.

TABLE II. Atributo mes, interacción del agrupamiento.

ATRIBUTO MES						
Mes	C. 0	C. 1	C. 2	C. 3	C. 4	C. 5
	13%	32%	12%	21%	4%	18%
January	756,9584	3734,709	873,9156	2121,2689	508,3026	1827,8476
February	11642,7058	31546,0399	10064,2366	15412,411	5153,3798	18419,2269
March	12040,4134	35036,0318	10930,6832	18774,118	12,3594	20232,3942
April	8612,0449	17076,268	8154,7364	25341,6786	3898,2555	327,0166
May	7654,0407	17620,1776	6668,1476	10485,7532	3464,5926	11549,2882
June	1119,9954	4190,445	1049,2523	1439,5456	6,4479	2893,3138
July	16,9999	106,6803	44,0151	518,4679	1,3126	2,5241
August	5976,9977	15714,5839	5021,9262	7469,4246	5,8773	10194,1904
September	8333,0755	18827,5773	7870,9383	12388,6065	4706,605	13897,1975
October	6545,974	11506,1399	6032,1677	8476,6084	3180,3462	9836,7639
November	1513,9882	3137,0532	1835,2416	2161,2264	636,23	3201,2605
December	13	64,581	32,0151	160,755	2,4165	4,2323
[total]	64226,1919	158560,287	58577,2756	104749,864	21576,1255	92385,2558

El modelado concuerda justo con la distribución de ingreso a la plataforma por parte de los usuarios en el intervalo de tiempo analizado. De igual forma los resultados arrojaron que en la muestra numero dos (02) una distribución similar en cuanto al agrupamiento de las diferentes instancias en cada atributo.

En general los atributos analizados en el modelo dan cuenta de similitudes con los procesos llevados por los usuarios de la plataforma.

C. Análisis de los resultados

El proceso muestra una evidencia de la existencia de correlación entre los patrones identificados para los diferentes usuarios de la plataforma virtual y lo que se halló luego del proceso de agrupamiento de estos como atributos en el modelado [33], [30].

En general se identificó que los usuarios femeninos, en el periodo académico utilizan más veces la plataforma. Esto se aprecia una vez es identificado y analizado desde el punto de vista en varias escalas de tiempo, es decir, diarias, mensuales,

semestrales. Así mismo, también se identificó que el conjunto de clústeres, respetan los hallazgos realizados en la parte descriptiva relacionado a lo largo del primer y segundo periodo académico.

Se pudo identificar que los objetivos de direccionamiento o uso dentro de la plataforma virtual (target), se encuentra asociado a la operación, en gran medida, por los usuarios femeninos, donde el primer direccionamiento u objetivo de uso, es realizado hacia la sección de cursos de la plataforma.

También se pudo verificar que los patrones de uso dado muestran que algunos cursos en particular son interactuados por gran cantidad de usuarios en especial en las horas de 10:00 a 16:00 dentro del intervalo de tiempo analizado en este trabajo.

Otros comportamientos encontrados están relacionados con el origen desde donde los usuarios se conectan, arrojando un alto porcentaje desde el respectivo portal web, lo demás se aleja de esto y queda reservado para las tareas administrativas.

IV. CONCLUSIONES

Este trabajo deja con mucha claridad, que el uso del algoritmo de agrupamiento maximización de esperanza (EM), provee resultados de agrupación muy consecuentes, donde los usuarios son agrupan por género y a su vez se modela como estos pueden usar la plataforma en el periodo estudiado.

Es posible entonces predecir algunos comportamientos de los usuarios, regidos por los datos analizados, provenientes de la misma plataforma virtual, en la cual se ve claramente que tipo de usuarios tiende a ingresar más, que direccionamiento o uso dentro de la plataforma y cuales cursos, así como a qué horas o días, estos tenderán a utilizar de forma regular.

Cabe destacar, que el modelo predice que los usuarios de género femenino tienden a usar más los recursos de la plataforma y en más días que los usuarios masculinos. También se se puede identificar un incremento del uso de estos recursos al final de cada periodo estudiado, concordando con las actividades académicas finales.

Esto es un indicativo que se puede usar para optimizar los procesos y actividades que la Universidad realiza en esos periodos de tiempo. El modelo también destaca que un alto número de usuario se vinculan a la plataforma virtual de la universidad a través de su portal web.

Este resultado entrega luces que lleven a la preparación contra posibles incidentes que estén relacionados con la seguridad, capacidad o con la estabilidad del uso de estos recursos, pues la plataforma se gestiona en instalaciones propias y es necesario esta preparación, teniendo en cuenta los patrones mostrados por el análisis modelado.

Es importante destacar que, en el caso presente, el uso de las técnicas de minería web, facilitó en gran medida el desarrollo de los diferentes análisis descritos en este trabajo, pues como es el caso, la magnitud del volumen de los registros arrojados por el uso continuo de la plataforma, hace casi imposible poder obtener patrones y previsiones generados por estos.

Este estudio se desarrolló previo a los acontecimientos actuales y muestra una visión clara de cómo se desarrollaba la actividad académica en ese momento previo. Sin embargo, la utilización de este análisis brinda un sustento para realizar nuevos análisis y ajustes, llevados a la actualidad cuando en el mundo los procesos de aprendizaje virtuales se han afianzado.

El impacto de estos acontecimientos incrementa la necesidad de hallar nuevas y mejores técnicas que incidan en la optimización de estos procesos destinados a mantener la academia en funcionamiento, además de brindar la solidez, robustez y confiabilidad que ofrecía el desarrollo académico tradicional, pero con el alcance indiscutible de los procesos llevados de forma virtual.

REFERÊNCIAS BIBLIOGRÁFICA

- [1] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz, and J. C. Núñez, "Students' LMS interaction patterns and their relationship with achievement: A case study in higher education," *Comput. Educ.*, vol. 96, pp. 42–54, 2016.
- [2] G. I. Putri, N. Nuryadin, R. E. Indrajit, E. Dazki and H. Santoso. (2021), "Analysis of Teacher and Student Responses to the Use of a Web-based Learning Management System (LMS) during COVID-19 Pandemic," Sixth International Conference on Informatics and Computing (ICIC), Jakarta, Indonesia, 2021, pp. 1-5, doi: 10.1109/ICIC54025.2021.9632927.
- [3] E. Dans, "Plataformas Educativas Y El Dilema De La Apertura," *Distance Educ.*, vol. 6, 2009.
- [4] Romero, C., & Ventura, S. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2010, 40(6), 601-618.
- [5] H. Moreno Trujillo, M. Pintor Chávez, M. Gómez Zermeño, and M. Georgina Gómez Zermeño, "Uso de plataformas de libre distribución (LMS) para educación básica," *TE Rev. Iberoam. Tecnol. en Educ. y Educ. en Tecnol.*, vol. no. 17, no. 17, pp. 95–103, 2016.
- [6] J. Dhika, H. Fitriana, D. Musa Jaya, "Data Mining Approach for Learning management system," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, 2021.
- [7] I.-H. Jo, D. Kim, and M. Yoon, "Analyzing the Log Patterns of Adult Learners in LMS Using Learning Analytics," in *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 2014, pp. 183–187.
- [8] H. Jimenez, A.; Alvarez, "Minería de Datos en Educación," *ACM Int. Conf. Proceeding Ser.*
- [9] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4, Part 1, pp. 1432–1462, 2014.
- [10] M. W. Rodrigues, S. Isotani, and L. E. Zárate, "Educational Data Mining: A review of evaluation process in the e-learning," *Telemat. Informatics*, vol. 35, no. 6, pp. 1701–1717, 2018.
- [11] T. Elias, "Learning Analytics : Definitions , Processes and Potential," *Learning*, p. 23, 2011.
- [12] I.-H. Jo, D. Kim, and M. Yoon, "Analyzing the Log Patterns of Adult Learners in LMS Using Learning Analytics," in *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 2014, pp. 183–187.
- [13] M. V. Torre, E. Tan, and C. Hauff, "EdX Log Data Analysis Made Easy: Introducing ELAT: An Open-Source, Privacy-Aware and Browser-Based EdX Log Data Analysis Tool," in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 2020, pp. 502–511.
- [14] D. A. Gómez-Aguilar, F. J. García-Peñalvo, and R. Therón, "Análítica visual en e-learning," *Prof. la Inf.*, vol. 23, no. 3, pp. 236–245, 2014.
- [15] Y. Al-Ashmoery, R. Messoussi, and R. Touahni, "Analytical tools for visualisation of interactions in online e-learning activities on LMS and semantic similarity measures on text," *J. Theor. Appl. Inf. Technol.*, vol. 73, no. 1, pp. 102–118, 2015.
- [16] G. Ng, SK;Thriyambakam, K; McLachlan, "The EM algorithm. Handbook of computational statistics," Springer, p. 139,172, 2012.
- [17] K. Chen, "A constant factor approximation algorithm for k-median clustering with outliers.," *Proc. Ninet. Annu. ACM-SIAM Symp. Discret. algorithms.*, 2008.
- [18] I. Russell and Z. Markov, "An Introduction to the Weka Data Mining System (Abstract Only)," vol. 18, no. 3, pp. 742–742, 2017.
- [19] A. Jaramoillo, H. Patricio Paz, "Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje". *Revista Tecnológica-ESPOL*, 2015, vol. 28, no 1.
- [20] R. Baeza-Yates, "Tendencias en minería de datos de la Web." , *Profesional de la Información*, Vol. 18.1 p. 5-10, 2009.
- [21] Aparicio, M., Bacao, F., & Oliveira, T. (2020). Understanding e-learning continuance intention in higher education: a holistic approach based on structural equation modelling. *Journal of Educational Technology & Society*, 23(1), 18-33.
- [22] González, J., Gómez, M., & García, F. (2021). A web mining approach to understand user behavior in a virtual campus. *Telematics and Informatics*, 60, 101549.
- [23] Chen, K., Chen, C. H., & Yeh, Y. T. (2022). Personalized Learning Recommendation Based on Navigation Patterns and Semantic Modeling. *IEEE Transactions on Learning Technologies*, 15(1), 25-38.
- [24] Mohammadi, H., Sayyadi, H., Rahimi, A., & Naseri, M. (2021). A web mining-based framework for plagiarism detection in higher education. *Education and Information Technologies*, 26(2), 2097-2118.
- [25] Gao, F., Zhang, J., & Franklin, T. (2013). Applying web usage mining to an online diagnostic system: A case study. *Computers & Education*, 61, 76-86.
- [26] Li, X., Zhang, Y., Cheng, H., Li, M., & Yin, B. (2022). Student achievement prediction using deep neural network from multi-source campus data. *Complex and Intelligent Systems*, 8(6), 5143-5156. doi:10.1007/s40747-022-00731-8
- [27] Li, Y., Zhang, Y., Tang, Q., Huang, W., Jiang, Y., & Xia, S. -. (2021). T-K-means: A robust and stable K-means variant. Paper presented at the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, , 2021-June 3120-3124. doi:10.1109/ICASSP39728.2021.9414687 Retrieved from www.scopus.com
- [28] Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618.
- [29] Jovanović, J., Gašević, D., Dawson, S., & Pardo, A. (2017). Learning analytics: Challenges and future research directions. *Journal of Educational Computing Research*, 55(8), 1068-1081.
- [30] H. Escobar, M. Alcivar, C. Marquez, "Implementación de minería de datos en la gestión académica de las instituciones de educación superior", *Didáctica y Educación*, Vol. 8 no. 3, pp.203-212, 2017.
- [31] LAEL, Tri Ahmad Djabalul; PRAMUDITO, Deskha Akmal. Use of Data Mining for The Analysis of Consumer Purchase Patterns with The Fpgrowth Algorithm on Motor Spare Part Sales Transactions Data. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 2023, vol. 4, no 2, p. 128-136.
- [32] Carrión, C. (2022). Data analysis of short - term and long - term online activities in LMS. *TEM Journal*, 11(2), 497-505. doi:10.18421/TEM112-01.
- [33] S. Rath, Y. Deshpande, S. Nagaral, A. Narkhede, R. Sajwani and V. Takalikar, "Analysis of User's Learning Styles and Academic Emotions through Web Usage Mining," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 159-164, doi: 10.1109/ESCI50559.2021.9397037.
- [34] Hirschberger, F., Forster, D., & Lucke, J. (2022). A variational EM acceleration for efficient clustering at very large scales. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12), 9787-9801. doi:10.1109/TPAMI.2021.3133763.

