

VARIABLES QUE INFLUYEN EN EL RENDIMIENTO DE LOS ESTUDIANTES DE POSGRADO: UNA PERSPECTIVA DESDE LA ANALÍTICA DEL APRENDIZAJE

Argelia B. Urbina-Nájera*



<https://orcid.org/0000-0002-3700-7287>

RECIBIDO: Agosto 2020 / ACEPTADO: Octubre 2020 / PUBLICADO: Enero 2021

Como citar: Urbina-Nájera, Argelia. (2021). Variables que influyen en el rendimiento de los estudiantes de posgrado: Una perspectiva desde la analítica del aprendizaje. **Telos: revista de Estudios Interdisciplinarios en Ciencias Sociales**, 23 (1), Venezuela. (Pp.36-50).
DOI: www.doi.org/10.36390/telos231.04

RESUMEN

En la última década, el uso de la analítica de aprendizaje y la gestión de grandes volúmenes de datos han contribuido sustancialmente en la manera como las instituciones de educación superior dan seguimiento, analizan la información y predicen el desempeño de los estudiantes. El objetivo del presente trabajo fue identificar las variables que influyen en el desempeño académico de estudiantes de postgrado, mediante la aplicación de técnicas de learning analytics basado en el modelo de Chatti, et al. (2012). Se aplicaron los algoritmos selección de atributos y árboles de decisión a partir de la teoría de Witten, Frank, Hall, Pal (2016) a una muestra de datos recolectadas de 136 estudiantes de posgrado de forma aleatoria simple. Se identificó que en general prefieren estudiar por la tarde y que invierten 43.83% de su tiempo en la revisión del contenido del curso mientras se encuentran activos en la plataforma; el 10.92% del tiempo participan en foros y el 31.10% del tiempo realizan actividades. Mediante el algoritmo selección de atributos se identificaron a las cuatro variables más importantes que influyen en el desempeño, a saber: tiempo total invertido en el curso de consulta del curso, elaboración de tareas, participaciones en foros y trabajo en equipo. También, aplicando árboles de decisión se establecen 6 patrones que determinan alguna nota final, cuya variable más importante es el tiempo total invertido en la plataforma. Finalmente, se determina que las variables: tiempo invertido en la plataforma en la consulta del contenido, trabajo en equipo, tareas y actividad en foros, influyen de manera positiva en el desempeño satisfactorio del estudiante de posgrado y aquellas variables relacionadas con las consultas, hora y día de estudio no intervienen en dicho desempeño, estos hallazgos dan la pauta para centrar esfuerzos en la construcción de contenidos significativos y tareas enfocadas en el logro de los aprendizajes deseados apoyados por actividades en equipo.

* Doctora en Planeación Estratégica y Dirección de Tecnología (UPAEP). Perteneció al Sistema Nacional de Investigadores Mexicano. Sus líneas de investigación se enfocan en la aplicación de minería de datos educativa y aprendizaje computacional en el ámbito educativo, salud y negocios, así como tecnología educativa. Actualmente labora como Profesora-Investigadora de Tiempo Completo en la Facultad de Tecnologías de Información en UPAEP. Correo electrónico: abunajera@gmail.com; argeliaberence.urbina@upaep.mx



Palabras clave: analítica de aprendizaje; desempeño académico; patrones de comportamiento; estudiantes de posgrado.

Variables that influence the performance of graduate students: A perspective from the learning analytics

ABSTRACT

In the last decade, the use of learning analytics and the management of large volumes of data have contributed substantially to the way higher education institutions track, analyze information and predict student performance (Clow, 2013). The objective of this work was to identify the variables that influence the academic performance of graduate students, through the application of learning analytics techniques (Chatti, et al., 2012). The algorithms selection of attributes and decision trees (Witten, Frank, Hall, Pal, 2016) were applied to a sample of data collected from 136 graduate students in a simple random way. It was identified that in general they prefer to study in the afternoon and that they invest 43.83% of their time in the review of the course content while they are active in the platform; 10.92% of the time they participate in forums and 31.10% of the time they carry out activities. Through the algorithm of attribute selection, the four most important variables that influence performance are defined, namely: total time invested in the course of course consultation, elaboration of tasks, participation in forums and teamwork. Also, applying decision trees, 6 patterns are established that determine some final note, whose most important variable is the total time spent on the platform. Finally, it is determined that the variables: time invested in the platform in the consultation of content, teamwork, tasks and forum activity, positively influence the satisfactory performance of the graduate student and those variables related to the consultations, time and day of study do not intervene in such performance, these findings give the guideline to focus efforts on building meaningful content and tasks focused on achieving the desired learning supported by team activities.

Keywords: learning analytics; academic performance; usage patterns; postgraduate students.

Introducción

En la última década, las instituciones de educación superior han hecho uso de soluciones tecnológicas con el fin de mejorar sus procesos de enseñanza y apoyo a la formación de los estudiantes. Especialmente, el uso de la analítica de aprendizaje (*learning analytics*) y la gestión de grandes volúmenes de datos (*big data*) han contribuido radicalmente en la manera como las instituciones de educación superior dan seguimiento y predicen el desempeño de los estudiantes. Estas dos áreas, benefician la interpretación de un amplio rango de datos operacionales y administrativos, que posibilitan evaluar el desempeño y progreso institucional, sirviendo como base para predecir sucesos futuros e identificar potenciales problemas de programación académica, investigación, enseñanza y aprendizaje (Baepler, et al., 2010; Daniel, 2014).

Van-Barneveld, et al., (2012) y García, et al., (2018), definen al *learning analytics* (LA) como el uso de técnicas analíticas para orientar los recursos educativos, curriculares y de apoyo, para favorecer el logro de objetivos de aprendizaje específicos; dentro de las técnicas analíticas se encuentran las algebraicas, estadísticas, sistemas de aprendizaje de reglas, minería de datos,

aprendizaje computacional, estocásticas, difusas, basadas en casos, entre otras que se aplican en el contexto educativo para analizar el comportamiento de los estudiantes (presencial, a distancia o mixto) y poder predecir algún evento futuro desde lo académico hasta lo administrativo.

Del mismo modo, Dietz-Uhler y Hurn, (2013), conciben al LA como la medición, acumulación, análisis y reportes de datos relacionados con los estudiantes y su contexto con el objetivo de entender y optimizar la enseñanza y el ambiente en que se da. Por otro lado, el *big data* se ha descrito como la capacidad de almacenar y manipular grandes cantidades de datos; la información almacenada puede descubrir patrones de desempeño estudiantil y así poder sugerir acciones para mejorar el rendimiento académico (Picciano, 2012).

Así mismo, existen diversas técnicas de LA que se pueden aplicar en múltiples contextos, que van desde definir las características de un usuario hasta las preferencias de compra. Empero, la analítica de aprendizaje es un área estrechamente ligada a la minería de datos educativa, puesto que posibilita el análisis de *big data* de los estudiantes obtenidos de diversas fuentes, para encontrar patrones ocultos que favorezcan la toma de decisiones basada en información, a fin de tomar acciones que prevengan alguna situación futura tanto en el contexto académico como en el administrativo de una institución de educación superior.

Este estudio tiene como objetivo identificar las variables que influyen en el desempeño académico de estudiantes de posgrado, mediante la aplicación de técnicas de *learning analytics*. Las técnicas aplicadas se enfocan en el uso de estadística, el algoritmo selección de atributos o selección de características (*Feature Selection*) y el algoritmo árboles de decisión (*Decisión Trees*) a una muestra aleatoria simple de datos provenientes de 136 estudiantes de posgrado de una Universidad Privada del Estado de Puebla.

De este modo, el documento se ha organizado de la siguiente manera: Una sección de trabajos relacionados que describe la forma en cómo aplicar algunas técnicas de LA a partir de trabajos recolectados de 2012 a 2018. Enseguida, se presenta la metodología que da pauta a la aplicación de técnicas estadísticas, el algoritmo de selección de atributos y el algoritmo árboles de decisión C4.5, por lo que en la sección posterior se describen con detalle los antecedentes teóricos de cada uno de ellos. Luego, se presentan los resultados obtenidos apoyados de gráficos propios de cada técnica y finalmente se presentan las conclusiones que conllevan a un trabajo a futuro, así como, valorar el uso y aplicación de las técnicas de LA mencionadas en este estudio.

Trabajos relacionados

En este apartado se hace mención de algunos trabajos relacionados a la puesta en marcha de técnicas de *Learning Analytics*. En principio, Arnold y Pistilli (2012), presentan un estudio para predecir el rendimiento de los estudiantes usando técnicas estadísticas como una forma de introducción al LA. Se basaron en las calificaciones, características demográficas, historial académico y esfuerzo de los estudiantes medido por la interacción con la plataforma Blackboard; el resultado fue entregado mediante un correo personalizado indicando un semáforo para indicar el estatus del estudiante, permitiendo también, conocer los resultados de la retención en los dos primeros años universitarios. Aseguran que el impacto de la herramienta ha sido a nivel institucional con la disminución de los índices de deserción y por consiguiente aumento en el porcentaje de retención.

También, Yu y Jo (2014), realizaron un estudio en una universidad de mujeres en Corea del Sur, con una muestra de 84 estudiantes de licenciatura presencial donde se emplea *Moodle* para descargar material académico. Analizan 6 variables: 1) frecuencia de entrada a la plataforma, 2) tiempo de estudio en la plataforma, 3) regularidad de intervalos de aprendizaje en la plataforma, 4) número de descargas de material, 5) interacción con compañeros; 6) interacción con el profesor. Mediante la aplicación de regresión lineal múltiple, se obtuvo un modelo para predecir el logro académico de cada estudiante.

Los resultados exponen que el modelo de 6 predictores fue capaz de explicar el 35.5 % de la varianza en la calificación final. Las variables tiempo de estudio en plataforma e interacción con compañeros tienen una correlación significativa con la nota final. Adicionalmente, la regularidad de intervalos de aprendizaje en la plataforma y el número de descargas de material tuvieron efectos parciales significativos en el modelo completo. Por otro lado, encontraron que la frecuencia de entrada a la plataforma y la interacción con el profesor no son útiles para predecir la nota final.

Por otro lado, Park y Jo (2015), aplicaron la visualización de datos para diseñar un tablero que muestra el patrón de comportamiento de estudiantes en línea, considerando las variables: accesos a la plataforma, visitas al tablero, tiempo total invertido, frecuencia de accesos, visitas al contenido, resumen de la actividad en línea y tiempo total en los accesos. Las pruebas se realizaron a 38 estudiantes; obteniendo un impacto significativo en los resultados de aprendizaje y nivel de comprensión de los mismos.

Sclater, et al. (2016), presentan una revisión de 11 universidades que han implementado el LA para conocer el impacto que ha tenido y resaltar las oportunidades que esto representa. Resultados de dicho estudio muestran que el LA ha ayudado a identificar estrategias de enseñanza particulares, mejorar la relación entre alumnos y tutores académicos, así como también, ofrecer una retroalimentación de calidad profesor-alumno. En cuanto a la Universidad, como organización, ha ayudado a mejorar procesos en diversos aspectos del negocio institucional y a ampliar una cultura de toma de decisiones basada en datos. En cuanto a los estudiantes, se espera que, al brindarles información sobre su desempeño académico, puedan tomar mejores decisiones acerca de su educación.

Así mismo, Rossetti López; Verdugo Tapia y Bayliss Bernal (2017), recolectaron 86 registros de Moodle versión 3.2.1 para identificar factores que influyen en el desempeño académico de estudiantes presenciales. Las variables estudiadas fueron: tiempo de conexión, número de visitas al curso, número de interacciones con recursos y actividades, número de ejercicios realizados y número de aportaciones a foros de discusión. Utilizaron regresión múltiple con el fin de explicar la calificación final de los estudiantes inscritos en un curso universitario. Los resultados exponen que los estudiantes que participan de forma activa en los foros de discusión durante el curso y realizan los ejercicios obtienen un mejor desempeño que los demás. De igual modo, identificaron que el resto de variables, no tienen un impacto significativo en el desempeño académico de los estudiantes en el curso.

Finalmente, Lu, et al., (2018), muestran un estudio con datos de 33 hombres y 26 mujeres de un curso de cálculo impartido a estudiantes universitarios. Recolectaron 21 variables entre las que se encuentran: número de videos que el estudiante consulta, número de clicks, número de unidades que el estudiante estudia por semana, entre otras. Mediante regresión lineal identificaron que las variables más importantes que determinan el desempeño de un estudiante

universitario son: Número de días que un estudiante tiene actividad por semana, número de actividades por semana en la cual el estudiante se involucra, número de videos que el estudiante ve completamente, número de videos en donde el estudiante da pausa, número de veces que hace clic por semana, número de veces que hace clic en búsqueda avanzada, número de veces que el estudiante hace *click* en pausa.

Estos antecedentes dan pauta para identificar aquellas variables que han sido estudiadas y cuáles han impactado en el desempeño de los estudiantes que utilizan una plataforma como soporte a su aprendizaje.

Metodología

Diseño de la investigación: El estudio presentado es de tipo descriptivo (Hernández-Sampieri y Mendoza-Torres, 2018), puesto que se busca identificar las variables que influyen en el desempeño académico de estudiantes de posgrado, mediante la aplicación de técnicas de learning analytics.

Población y muestra: Estudiantes de maestría presencial en el área de ingeniería y negocios de una universidad privada del estado de Puebla, ofrecidos durante el ciclo escolar 2019-otoño; con una población de 150 estudiantes. La muestra fue aleatoria simple conformada por 136 estudiantes donde 73 son mujeres y 63 son hombres, pues 14 de ellos no terminaron el curso e incluir los datos podría causar sesgo en el análisis.

Conjunto de datos: Se consideran 8 cursos (Fundamentos de inteligencia de negocios, mercadotecnia, sistemas empresariales, visualización de datos, estrategias tecnológicas, estancia profesional, planeación global estratégica y seminario de actualización) (ver Tabla 1) de diversos programas de maestría presencial como mercadotecnia, planeación estratégica, ciencia de datos, tecnologías de información e ingeniería de software.

Tabla 1. Distribución de estudiantes por grupo.

NO.	CURSO	NO. ESTUDIANTES
1	Fundamentos de inteligencia de negocios	15
2	Mercadotecnia	24
3	Sistemas empresariales	32
4	Visualización de datos	20
5	Estrategias tecnológicas	19
6	Estancia profesional	3
7	Planeación global estratégica	21
8	Seminario de actualización	2

Fuente: Elaboración propia.

Aspectos éticos: Los datos son tratados bajo el concepto de privacidad de datos personales, considerando solamente datos demográficos y estadísticos.

Técnica cualitativa de elección de curso: Los cursos se eligieron en función del área al que pertenecen, en este caso, ingeniería y negocios, además de utilizar Blackboard 3.10 como repositorio y apoyo a cursos presenciales. Las **variables de estudio** se listan en la Tabla 2.

Preprocesamiento de los datos: La obtención de los datos se realizó mediante la opción de <informes del curso> proporcionada en Blackboard 3.10 (Blackboard, 2019). Los informes fueron descargados en formato *.xls uno a uno, seleccionando la fecha de inicio y fin

del periodo. En la tabla 1 se presentan las variables identificadas en el conjunto de datos. El preprocesamiento de los datos consistió en reemplazar los datos nulos por cero, se codificó al atributo <hora de acceso> bajo el esquema de periodos de tiempo <Madrugada=24:01-06:00>, <Mañana=06:00-12:00>, <Tarde=12:01-18:00> y <Noche=18:01-24:00>. También, se definió al atributo <promedio_final> como la clase del conjunto de datos, es decir, el identificador de cada instancia, entonces se asignó <Deficiente> a aquellas notas inferiores a 7.5; <aceptable> a valores entre 7.5 y 8.4; <Satisfactorio> a valores entre 8.5 y 9.4; <Notable> a valores entre 9.5 y 9.9 y <Sobresaliente> a notas igual a 10.0.

Análítica y acción: En este estudio se aplican los algoritmos selección de atributos y árboles de decisión. El primero para identificar las variables más importantes y el segundo para determinar los patrones que definen una nota final en los cursos de posgrado.

Tabla 2. Variables descriptoras del conjunto de datos.

NO.	VARIABLES	VALORES DE REFERENCIA
1	Género	Masculino, femenino
2	Día de mayor acceso	Lunes a domingo
3	Hora de mayor acceso	0 horas a 23 horas
4	Tiempo total invertido en el curso	8 horas a 160.99 horas
5	Revisión del contenido del curso	170 a 703 veces
6	Correos enviados o consultados en el curso	0 a 25
7	Consulta de calificaciones	0 a 132 veces
8	Elaboración de tareas	0 a 20
9	Número de participación en foros	1 a 25
10	Trabajo en equipo	38 a 60 veces
11	Mes de mayor participación en equipo	1 a 4
12	Promedio final (clase)	7.0 a 10.0

Fuente: Elaboración propia.

Técnicas de learning analytics

De acuerdo con Chatti, et al. (2012), las técnicas comúnmente empleadas en la aplicación de LA son: estadística, visualización de la información, minería de datos y análisis de redes sociales; cada una de ellas es utilizada en función de los objetivos de la tarea de análisis que se desee realizar. En este estudio, se aplicó la técnica de minería de datos, dentro de la cual se empleó el algoritmo selección de atributos y el algoritmo árboles de decisión (Witten, Frank, Hall, Pal, 2016); el primero para encontrar las variables más relevantes que pueden describir el uso de Blackboard y el segundo para identificar los patrones que influyen en el desempeño del estudiante. Ambos algoritmos son descritos a continuación.

Selección de atributos

La selección de atributos es un proceso mediante el cual se busca automáticamente al mejor subconjunto de atributos en el conjunto de datos (Witten, Frank, Hall, Pal, 2016). Cuando se menciona el mejor se hace alusión a la más alta precisión. Brownlee (2014) sostiene que los beneficios de utilizar algoritmos de selección de atributos, recae en el hecho de obtener menos datos redundantes, menos datos con alguna inconsistencia como: vacíos, duplicados o

faltantes; lo que significa mejorar la precisión en el modelo obtenido y tener solo aquellos datos que describan de mejor manera a todo el conjunto de datos, esto implica una menor inversión de tiempo en el entrenamiento del algoritmo.

El proceso para la selección de atributos se divide en dos partes: 1) Métodos de evaluación de atributos y 2) Métodos de búsqueda. El primero, es el método que evalúa al subconjunto de atributos. Algunos ejemplos de métodos de evaluación de atributos son: CfsSubsetEval: subconjuntos de valores que se correlacionan altamente con el valor de la clase y baja correlación entre sí; ReliefFAttributeEval: evalúa el valor de un atributo muestreando repetidamente y considerando el valor del atributo dado para la instancia más cercana de la misma y diferente clase. ClassifierSubsetEval: evalúa los subconjuntos utilizando un algoritmo predictivo y otro conjunto de datos que sean especificados (Witten, Frank, Hall, Pal, 2016).

La segunda parte del proceso, es la forma estructurada en la que se navega por el espacio de búsqueda de posibles subconjuntos de atributos en función de la evaluación de subconjuntos. Algunos ejemplos de métodos de evaluación de atributos son: Ranker: Clasifica los atributos por sus evaluaciones individuales; BestFirst: Utiliza una estrategia de búsqueda de la mejor primera vez para navegar por los subconjuntos de atributos (Brownlee, 2014; Witten, Frank, Hall, Pal, 2016). En este estudio se utilizan los evaluadores CfsSubsetEval y ReliefFAttributeEval con los métodos de búsqueda BestFirst y Ranker respectivamente.

Árboles de decisión

De acuerdo a Mitchell (2000), los árboles de decisión son una técnica de clasificación fácil de interpretar y utilizar que generan reglas del tipo *Si...entonces*, por ejemplo *Si* invierto 120 horas de estudio *ENTONCES* obtengo un 10.0. Dichas reglas son representadas en forma de árbol, donde el conjunto de datos es dividido en ramas hasta obtener segmentos de similar comportamiento (nodo hoja) en función de la variable objetivo. Los árboles de decisión son popularmente utilizados en la toma de decisiones dado que son de fácil interpretación, son flexibles al explicar diversos tipos de datos, aún con datos faltantes.

Por otro lado, un algoritmo debe ser analizado para determinar el desempeño al realizar alguna tarea como: clasificar, reconocer, identificar, agrupar, categorizar, entre otros. Algunas métricas que evalúan dicho desempeño son: Precisión, exactitud, recuperación, Medida-F, matriz de confusión y similares (Witten, Frank, Hall, Pal, 2016). Para identificar el desempeño del algoritmo árboles de decisión, de este estudio, se utiliza la medida de exactitud y la matriz de confusión, también conocida como matriz de clasificación (Tabla 3).

Tabla 3. Matriz de confusión cuando se tienen dos posibles resultados de clasificación: Negativo y Positivo.

		Predicción	
		Negativo	Positivo
Actual	Negativo	a	b
	Positivo	c	d

Fuente: Witten, Frank, Hall, Pal (2016).

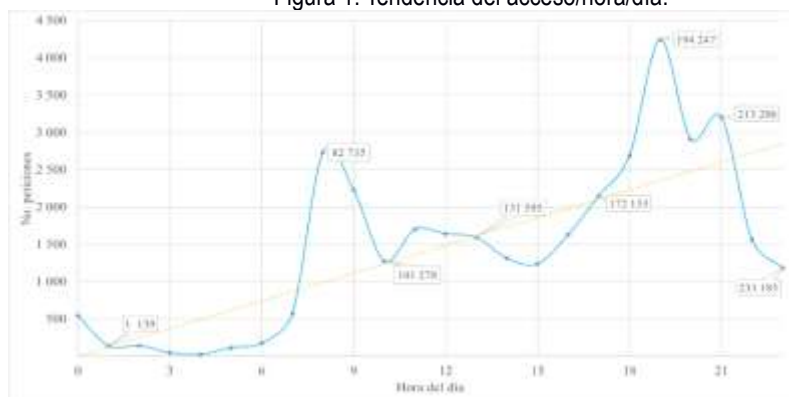
La medida de exactitud calcula la proporción del número total de predicciones que son correctas (Mitchell, 2000), mientras que, la matriz de confusión es una herramienta de visualización que se emplea para obtener información sobre las clasificaciones reales y predicciones realizadas por un sistema de clasificación, en donde los casos bien clasificados se encuentran en la diagonal de la matriz (Tabla 3, celdas [a,d]) (Bird, Klein y Loper, 2009; Witten, Frank, Hall, Pal, 2016; Berlingerio, 2019).

Resultados y discusión

Como se ha mencionado, en este estudio se busca identificar las variables que influyen en el desempeño académico de estudiantes de postgrado, mediante la aplicación de técnicas de *learning analytics* usando los atributos mostrados en la Tabla 2. Es preciso comentar que se incluyen las imágenes más representativas de algunas variables encontradas.

En principio, para identificar estas variables, se ha hecho un análisis estadístico de los datos con el fin de identificar algún patrón o tendencia. En la Figura 1 se observa que, el día en que hay menos accesos es el día sábado. Respecto a la hora de acceso de preferencia denotada por la línea punteada se identifica una ascendencia a las 8 horas y a las 19 horas; la tendencia se registra a la 1, 10, 13, 17 horas y unos minutos pasados de las 21 horas.

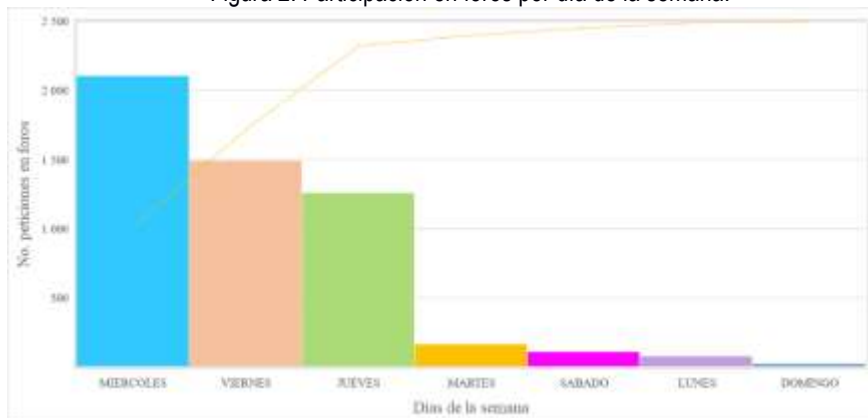
Figura 1. Tendencia del acceso/hora/día.



Fuente: Elaboración propia.

Por otro lado, en la Figura 2 se presenta la participación en foros reflejada en su mayoría en los días miércoles, actividad, que no está ligada con el total de accesos a la plataforma ni con el día en que mayor acceso que tienen los estudiantes de posgrado; pues una vez que se han entrado al foro se pueden hacer diversas participaciones mientras se encuentran conectados a la plataforma. De igual modo, se observa que el día que menos participación hay en foros es el día domingo; dato que tampoco está relacionado con los días de menor acceso en general.

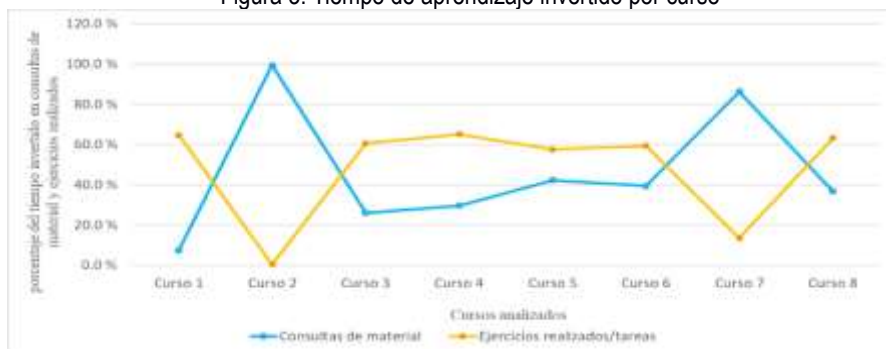
Figura 2. Participación en foros por día de la semana.



Fuente: Elaboración propia.

Respecto al tiempo de aprendizaje por alumno en la plataforma, se identifica que está directamente relacionado con el tiempo que ha invertido en consultar el material y los ejercicios realizados. Estos últimos pueden ser diversas actividades como: participación en foros/blogs, tareas, evaluaciones, trabajo en equipo, y en todas aquellas actividades que favorezcan el aprendizaje de los contenidos de forma práctica en cada curso (Yáñez, 2016).

Figura 3. Tiempo de aprendizaje invertido por curso



Fuente: Elaboración propia.

Particularmente, en la Figura 3 se observa que el tiempo promedio que los estudiantes de posgrado invierten en su aprendizaje es de 52.56 %, llevando a la práctica lo aprendido, mientras que invierten un 43.82 % en leer el contenido temático de cada curso que puede incluir guías de aprendizaje, contenido, lecturas, material complementario, entre otros. Estos datos son casi equiparables con los obtenidos por López, et al. (2019), en donde se afirma que los participantes aprenden hasta 5 veces más sin incrementar el tiempo invertido en su formación comparada con la educación presencial (40 % en línea y 60 % presencial).

Del mismo modo, se observó que, respecto a los hábitos de estudio, es notorio que los estudiantes de posgrado prefieren estudiar los días miércoles y los días martes. El horario que mayormente adoptan para estudiar es a las 19:00 horas, seguido con menor frecuencia a las 21:00 horas; y finalmente, a las 08:00 horas (Figura 1). En el estudio presentado por Muñoz-Díaz, (2017), se reportó que los jóvenes universitarios prefieren estudiar de madrugada, situación que discrepa con los estudiantes de posgrado analizados en este estudio, que optan por hacerlo durante la noche (37.36 %) y tarde (34.96 %), muy cercano a la mañana (24.78 %), y de madrugada solamente un 2.90 %. Estas preferencias se observan en la Figura 1, representada en los picos de la gráfica denotadas por el mayor número de accesos en los lapsos de hora señalados.

En cuanto al uso por sección, se descubrió que, dentro de las 52.56 % de horas que los estudiantes de progrado invierten en hacer ejercicios y tareas, solamente el 10.92 % corresponde a la participación en foros (Figura 2), un 10.54 % en actividades grupales y el porcentaje restante que equivale al 31.10 % realizan actividades individuales. En tanto que, invierten casi la mitad del tiempo (43.83 %) en revisar/leer el contenido del curso (Figura 3). Por último, se determinó que existe una relación entre el tiempo invertido y el desempeño obtenido en el curso.

En general, se registran en promedio 57 accesos para consultar el material de estudio. De este modo, se identificó que el 10.47 % obtuvo una nota final en el curso entre 7.0 y 8.0, mientras que el 5.76 % obtuvo una calificación final entre 8.1 y 9.0; el resto consiguió un promedio entre 9.4 y 9.9. Estas cifras permiten deducir que las personas que tuvieron mayor número de accesos al material durante el periodo tienen mayor desempeño en el curso, lo que concuerda con los resultados presentados por Yu y Jo (2014) y Rossetti López; Verdugo Tapia y Bayliss Bernal (2017).

Por otro lado, aplicando el algoritmo selección de atributos con el evaluador CfsSubsetEval y el método de búsqueda BestFirst, se obtiene que *el atributo más relevante es el tiempo total invertido en el curso*, el cual debe superar 53.4 horas de estudio en la plataforma LMS, es decir, consultando el material. De este modo, Yu y Jo (2014), afirman que el tiempo de estudio en la plataforma y la interacción con compañeros están directamente relacionados con el desempeño del estudiante, hallazgo que concuerda con este estudio en la primera variable, en tanto que la interacción con compañeros se encuentra en el lugar 4 y 5 del ranking obtenido (Tabla 3). Mientras que, Rossetti López; Verdugo Tapia y Bayliss Bernal (2017), indican que la actividad en foros y ejercicios son las variables más importantes que influyen en el desempeño, resultados que discrepan con los hallazgos de este estudio al obtener que la elaboración de tareas y participación en foros se encuentran en el lugar 3 y 4 respectivamente del ranking (Tabla 4).

Tabla 4. Lista de los atributos más importantes

NO.	ATRIBUTOS	RANKED
1	Tiempo total invertido en el curso	0.718 4
2	Elaboración de tareas	0.105 8
3	Número de participación en foros	0.101
4	Trabajo en equipo	0.031 1
5	Revisión del contenido del curso	
6	Hora de mayor acceso	

7	Día de mayor acceso
8	Mes de mayor participación en equipo
9	Correos enviados o consultados en el curso
10	Consulta de calificaciones
11	Género

Fuente: Elaboración propia.

Al mismo tiempo, empleando el evaluador ReliefAttributeEval con el método de búsqueda Ranker se obtiene la lista de atributos ordenados por su importancia tal como se muestra en la Tabla 4; observándose que las principales variables que determinan una tendencia en el uso de Blackboard son: tiempo total invertido en el curso, elaboración de tareas, número de participaciones en foros y trabajo en equipo. Se observa con claridad, que de acuerdo a la literatura y resultados obtenidos tras aplicar el algoritmo selección de atributos, hay una concordancia de 3 de 4 variables, exceptuándose el trabajo en equipo que, conforme a la revisión hecha, ningún autor considera al trabajo en equipo como variable de interés para su estudio o factor en el desempeño del estudiante al usar Blackboard.

Patrón identificado

Por otra parte, para identificar el patrón de comportamiento que determina una nota final cuando un estudiante de posgrado usa Blackboard, se aplicó el algoritmo árboles de decisión, cuyo resultado se observa en la Figura 6 obteniendo una exactitud de clasificación del 79.10%, porcentaje que se traduce en 107 instancias correctamente clasificadas y 29 incorrectamente clasificadas.

Figura 6. Árbol de decisión que define una nota final cuando un estudiante de posgrado usa Blackboard.



Fuente: Elaboración propia.

De la Figura 6 se derivan 6 reglas o patrones que definen cierta nota final, mismas que se detallan en la Tabla 5 en donde se indica que para obtener una nota final sobresaliente (10.0) se debe seguir la regla 6, mientras que una nota deficiente ocurre cuando se cumple la regla 1 y la regla 2.

Tabla 5. Patrones de comportamiento (reglas) de un estudiante de posgrado

NO.	REGLA
1	SI (tiempo_total_invertido <= 53.4 h) ENTONCES (Notal_Final =Deficiente)
2	SI (tiempo_total_invertido >53.4 h) y (Tareas <=15) y (participación_en_foros <=11) ENTONCES (nota_final =DEFICIENTE)
3	SI (tiempo_total_invertido >53.4 h) y (Tareas <=15) y (participación_en_foros >11) ENTONCES (nota_final =ACEPTABLE)
4	SI (tiempo_total_invertido >53.4 h) y (Tareas >15) y (trabajo_en_equipo <=45) ENTONCES (nota_final =SATISFACTORIO)
5	SI (tiempo_total_invertido >53.4 h) y (Tareas >15) y (trabajo_en_equipo >45) y (participación_en_foros <=18) ENTONCES (nota_final =NOTABLE)
6	SI (tiempo_total_invertido >53.4 h) y (Tareas >15) y (trabajo_en_equipo >45) y (participación_en_foros >18) ENTONCES (nota_final =SOBRESALIENTE)

Fuente: Elaboración propia.

En la Tabla 6 se presenta la matriz de clasificación obtenida también como métrica del desempeño del algoritmo árboles de decisión, considerando que en la diagonal (marcado con verde) se encuentran las instancias correctamente clasificadas. Dicha matriz, puede interpretarse de la siguiente manera: Hay 24 instancias identificadas con una nota deficiente, pero el algoritmo ha clasificado correctamente a 23 de ellas, mientras 1 fue clasificada incorrectamente como satisfactorio; en tanto que, la clase aceptable ninguna instancia ha sido clasificada correctamente pues el algoritmo las ha catalogado como satisfactorio.

De igual modo, en la clase notable ninguna instancia se ha clasificado de forma correcta; el algoritmo ha clasificado a 8 instancias como satisfactorio y a 3 como sobresalientes.

Tabla 6. Matriz de clasificación obtenida mediante el árbol de decisión.

a	b	c	d	e	Clasificado como
23	1	0	0	0	a = deficiente
0	69	0	4	0	b = Satisfactorio
0	2	0	0	0	c =Aceptable
0	9	0	15	2	d =sobresaliente
0	8	0	3	0	e =Notable

Fuente: Elaboración propia.

Estas clasificaciones incorrectas pueden deberse a los pocos ejemplos analizados según cada clase (Tabla 6), es decir, que el conjunto de entrenamiento es demasiado pequeño como para ser una muestra representativa de la verdadera función objetivo (Witten, Frank, Hall y Pal, 2016). Este inconveniente se llama sobreajuste de los datos (*overfitting*) que ocurre cuando el algoritmo se ajusta a aprender los casos particulares que le enseñen y será incapaz de reconocer nuevos datos de entrada.

El problema de *overfitting* puede resolverse con algunas estrategias como: procurar que los datos sean balanceados (es decir, la misma o similar cantidad de datos en cada clase), reducir o eliminar la cantidad de atributos (variables) a analizar para encontrar el modelo, dividir el conjunto de datos (80%-20%) en datos de entrenamiento y datos de prueba o bien, realizar podas antes que el árbol crezca demasiado (Witten, Frank, Hall y Pal, 2016). Por lo que, aplicar

alguna de esas estrategias no fue posible en este estudio debido a las características del conjunto de datos, descritas con anterioridad.

Conclusiones

La aplicación de técnicas de LA cómo el algoritmo selección de atributos ha permitido encontrar las 4 variables más importantes que determinan una nota final: tiempo total invertido en la plataforma Blackboard, número de tareas, participación en foros y trabajo en equipo; variables que explican una nota final aprobatoria; así mismo, los atributos relacionados a la consulta de información o comunicación como envío de correos, día u hora de acceso o incluso el género, no influyen en el desempeño del estudiante y tampoco en la obtención de alguna nota final aprobatoria. Estas 4 variables permitirán enfocar los esfuerzos en fortalecer este tipo de actividades para lograr un mejor desempeño que favorezca el logro de las habilidades y competencias requeridas en el curso.

Del mismo modo, usando árboles de decisión, se han identificado las reglas que determinan una nota final, en la cual, es notorio que para obtener notas superiores a 9.4 es preciso cumplir con los límites mínimos que se estipulan para cada una de las principales variables obtenidas por el algoritmo selección de atributos, además, ha permitido conocer los patrones a seguir para obtener una nota final aprobatoria. Los algoritmos de selección de atributos y árboles de decisión son un referente para ampliar el estudio construyendo un conjunto de datos vasto que posibilite la aplicación de otras técnicas de *learning analytics*, y faciliten la identificación del comportamiento del estudiante de posgrado, con el fin de construir mecanismos que apoyen no solamente un buen desempeño, sino también un aprendizaje significativo.

Finalmente, se ha identificado que para conocer con profundidad los patrones de comportamiento de estudiantes no solo de posgrado, es preciso enriquecer el conjunto de datos ampliando su extracción a más cursos de diversas áreas de conocimiento, de manera que con las reglas que se pudieran obtener se generen contenidos significativos que favorezcan el aprendizaje a lo largo de la vida, así como estrategias pedagógicas que contribuyan a mejorar el desempeño en cada uno de los cursos y con esto ampliar la aplicación de *learning analytics* en otros contextos educativos como el aprendizaje y trabajo colaborativo.

Referencias Bibliograficas

- Arnold, Kimberly; Pistilli, Matthew. (2012). Course signals at Purdue: Using learning analytics to increase student success. **2nd International Conference on Learning Analytics and Knowledge, ACM**. Vancouver, BC. Canadá. Extraído de <https://bit.ly/3koVTw0>
- Baepler, Paul; James-Murdoch, Cynthia. (2010). Academic Analytics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*. 4(2), USA. (Pp. 1-9). Extraído de <https://bit.ly/2F3cbd1>
- Berlingerio, Michelle; Bonchi, Francesco; Gärtner, Thomas; Hurley, Neil; Ifrim, Georgiana. (2019). Machine Learning and Knowledge Discovery in Databases: **European Conference, Ecm1 Pkdd 201**. Dublin, Ireland.
- Bird, Steven; Klein, Ewan; Loper, Edward. (2009). **Natural Language Processing with Python**. O'Really Media, Inc. USA.
- Blackboard. (2019). Ayuda sobre la aplicación Blackboard. Extraído de <https://bit.ly/3n4dav1>

- Brownlee, Jason. (2014). Feature Selection to Improve Accuracy and Decrease Training Time. Extraído de <https://bit.ly/3a3GXP2>
- Chatti, Mohamed Amine; Dyckhoff, Anna Lea; Schroeder, Ulrik; Thüs, Hendrik. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*. 4(56), Switzerland. (Pp. 318-331). Extraído de <https://bit.ly/3koB03X>
- Clow, Doug. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6), England. (Pp. 683-695). <https://doi.org/10.1080/13562517.2013.827653>
- Daniel, Ben. (2014). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*. 46(5),UK. (Pp. 904-920). <https://doi.org/10.1111/bjet.12230>
- Dietz-Uhler, Beth; Hurn, Janet. (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. *Journal of Interactive Online Learning*. 12(1), USA. (Pp. 17-26). Extraído de <https://bit.ly/2C8XZ1H>
- García Herrero, Jesús; Berlanga de Jesús, Antonio; Patricio Guisado, Miguel Ángel; Padilla, Washintong. (2018). **Ciencia de datos: Técnicas analíticas y aprendizaje estadístico en un enfoque práctico** (Primera ed.). Alfaomega. Colombia.
- Hernández-Sampieri, Roberto; Mendoza-Torres, Christian Paulina (2018). **Metodología de la investigación: Rutas cuantitativa, cualitativa y mixta** (Sexta ed.). McGraw Hill. México.
- López, Jorge; Pazmiño Maji, Rubén; García Peñalvo, Francisco. (2019). Entornos de aprendizaje en la Escuela Superior Politécnica de Chimborzo, transformación mediante moodle y Google Analytics. **Identidad Bolivariana**, Volumen Edición Especial. Bolivia. (Pp. 1-10). Extraído de <https://bit.ly/2XFfQEY>
- Lu, Owen; Huang, Anna; Huang, Jeff; Lin, Albert; Ogata Hiroaki; Yang, Stephen. (2018). Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning. *Educational Technology and Society*, 21(2), Japón. (Pp. 220-232). Extraído de <https://bit.ly/3a3LP6O>
- Mitchell, Tom (2000). Decision Tree Learning. Extraído de <https://bit.ly/2GqgYNq>
- Muñoz-Díaz, Joaquín Ignacio. (2017). **Estudio de la relación entre la actividad online y el rendimiento académico de estudiantes de ingeniería y educación en dos sistemas de aprendizaje**. Tesis de pregrado. Ingeniería industrial. Universidad de Chile. Chile
- Park, **Yeonjeong**; Jo, **Il-Hyun**. (2015). Development of the Learning Analytics Dashboard to Support Students' Learning Performance. *Journal of Universal Computer Science*, 21(1), Austria. (Pp. 110-133).
- Picciano, Anthony. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Network*. 16(3), España. (Pp. 9-20). Extraído de <https://bit.ly/31ya9tw>
- Rossetti López, Sergio; Verdugo Tapia, María; Bayliss Bernal, Daniel. (2017). Learning analytics para determinar la relación entre uso de un learning management system y rendimiento académico. **XXII Congreso Internacional de Contaduría, Administración e Informática**. Extraído de <https://bit.ly/30ZVe9P>
- Slater, Niall; Peasgood, Alice; Mullan, Joel. (2016). Learning Analytics in Higher Education: A review of UK and international practice. Extraído de <https://bit.ly/2kbgMUC>

Variables que influyen en el rendimiento de los estudiantes de posgrado: Una perspectiva desde la analítica del aprendizaje

- Van-Barneveld, Angela; Arnold, Kim; Campbell, John. (2012). Analytics in Higher Education: Establishing a Common Language. **EDUCAUSE: Learning Initiative**. 2012(1). USA. (Pp 1-11). Extraído de <https://bit.ly/2PApvbC>
- Witten, Ian; Frank, Eibe; Hall, Mark; Pal, Christopher. (2016). **Data mining: Practical Machine Learning Tools and Techniques**. 4th ed. Morgan Kauffman. USA.
- Yáñez, Patricio. (2016). El proceso de aprendizaje: fases y elementos fundamentales. Revista San Gregorio. 1(11), Ecuador. (Pp. 70-81). Extraído de <https://bit.ly/3gQ9IRC>
- Yu, Taeho; Jo, Il-Hyun. (2014). Educational technology approach toward learning analytics: relationship between student online behavior and learning performance in higher education. **Proceedings of the Fourth International Conference on Learning Analytics and Knowledge**. Extraído de <https://bit.ly/33DLPt2>