

# TECNOLOGÍAS PARA EL ANÁLISIS DE PROCESOS EDUCATIVOS DESDE BIG DATA: LEARNING ANALYTICS Y PROCESS MINING

## TECHNOLOGIES FOR THE ANALYSIS OF EDUCATIONAL PROCESSES FROM BIG DATA: LEARNING ANALYTICS AND PROCESS MINING

Arturo Orellana García<sup>1</sup>

<sup>1</sup> Grupo de Investigación de Minería de Procesos/ Universidad de las Ciencias Informáticas, Cuba, aorellana@uci.cu

**RESUMEN:** *Hoy en día las universidades ofrecen múltiples modelos de enseñanza, y como parte fundamental de estos la introducción de tecnologías. Los entornos en línea, aulas inteligentes y los MOOC permiten generar grandes cantidades de datos relacionados con los procesos de enseñanza-aprendizaje, de los que es posible extraer una valiosa información que puede ser usada para mejorar el desempeño del estudiante y del proceso en sí. La presente investigación pone en evidencia la utilidad de Learning Analytics & Process Mining para transformar el pensamiento y la práctica en la educación desde Big Data. Se pretende que este documento sirva como punto de partida para la elaboración de proyectos de investigación que propicie contribuir al proceso de enseñanza-aprendizaje, a partir del uso de estas tecnologías desde el big data educacional.*

**Palabras Clave:** Big data, educación, Learning analytics, process mining, proceso de enseñanza-aprendizaje.

**ABSTRACT:** *Nowadays universities offer multiple teaching models, and as a fundamental part of these the introduction of technologies. Online environments, intelligent classrooms and MOOCs allow the generation of large amounts of data related to teaching-learning processes, from which valuable information can be extracted that can be used to improve student performance and the process itself. The present research highlights the usefulness of Learning Analytics & Process Mining to transform thinking and practice in education since Big Data. It is intended that this document serves as a starting point for the development of research projects that contribute to the teaching-learning process, from the use of these technologies from the educational big data.*

**KeyWords:** Big data, education, Learning analytics, process mining, teaching-learning process.

### 1. INTRODUCCIÓN

El sistema educativo tradicional consta de un profesor haciendo una misma lección para treinta estudiantes. Sin embargo, dos o más estudiantes pueden aprender de la misma forma. Teniendo datos analizados sobre el progreso de cada estudiante dentro de la clase, residencia estudiantil, biblioteca o salas de estudio, los profesores pueden adaptar sus actividades a las fortalezas y debilidades de

cada escolar y así reforzar la confianza y motivación de cada uno para aprender [1].

La aparición de nuevas plataformas digitales está disminuyendo las brechas y promoviendo el principio de universalidad en materia educativa. La cobertura en Educación Superior en América Latina crece a un ritmo acelerado, gracias a la oferta de programas virtuales. El surgimiento y éxito de los MOOC (Cursos Online Masivos y Abiertos), creados apenas hace menos de 10 años, han creado una nueva

tendencia en la educación a distancia que, para las universidades, representa un reto pedagógico y tecnológico. La estrategia implica un valor agregado a la producción, distribución y almacenamiento de estos recursos educativos, que surgieron por la necesidad de formación, especialmente, en campos relativamente nuevos y de rápido crecimiento [2-3].

Es un hecho, la educación está cambiando. Además de hacerlo por la evidente transformación social, lo hace por la inclusión de nuevas tecnologías a las aulas. Los MOOC, se han convertido en una fuente de grandes cantidades de datos sobre comportamientos específicos de los estudiantes virtuales [3].

Learning Analytics es el uso inteligente de datos, derivados del comportamiento de los alumnos en un curso de aprendizaje en línea; el número de accesos, los materiales revisados, las participaciones y las puntuaciones, comprenden información de valor para las instituciones de Educación Superior [4-5]. A pesar de ser vistos como datos en bruto, lo cierto es que de ellos es posible extraer los recursos suficientes para entender el perfil, las necesidades y el sentir de los estudiantes, frente a los programas de formación virtual. Con esta información se está en la capacidad de tomar decisiones que contribuyan con la optimización de la gestión universitaria ya que, a diferencia del Big Data para otros sectores, está tratando con alumnos, no con clientes [6].

La presente investigación tiene como objetivo mostrar las potencialidades del Big Data, desde las tecnologías de análisis process mining [7] y Learning Analytics para contribuir al desarrollo del proceso de enseñanza aprendizaje en la Educación Superior. Así mismo, analizar las tendencias en materia tecnológica del uso de los datos a disposición del desarrollo intelectual universitario.

## 2. LA CIENCIA DE LOS DATOS

En 2011 el padre de la minería de procesos Wil M.P. van der Aalst publicó el libro "Process Mining. Discovery, Conformance and Enhancement of Business Processes" [7]. En su capítulo introductorio Aalst se refiere a la explosión de los datos expresada el tamaño del universo digital, el cual fue calculado en 1 Zettabyte (Zb) en el 2010. El IDC reportó en 2014 un universo digital de 4.4 Zb, las oportunidades que representan el análisis de esta información se expresan en los \$40 billones sobre el mercado de software para el análisis de datos y cuyo crecimiento tiene un ritmo anual del 10%. Es importante destacar que la relación de Gigabytes por profesional de las TIC se estima que crezca ocho veces hasta el 2020, cuando el universo digital alcance los 44 Zb. En ese momento el 60% del universo digital estará en las economías emergentes, incluyendo varios países de América Latina.

La tendencia se encuadra en un entorno que no suena para nada extraño: la proliferación de páginas web, aplicaciones de imagen y vídeo, redes sociales, dispositivos móviles, aplicaciones, sensores, internet de las cosas, etc. capaces de generar, según IBM, más de 2.5 quintillones de bytes al día, hasta el punto de que el 90% de los datos del mundo han sido creados durante los últimos dos años. Está en desarrollo un entorno absolutamente relevante para muchos aspectos, desde el análisis de fenómenos naturales como el clima o de datos sistomográficos, hasta entornos como salud, seguridad o, por supuesto, el ámbito educativo.

El gigante Huawei en marco de la Feria Mundial de Informática y las Comunicaciones CeBIT 2017 afirmó que, a partir de un estudio previo, por cada minuto internet se realizan 2 millones de búsquedas en google, se generan más de 83 mil dólares a través de Amazon, se efectúan más de 277 mil inicios de sesión en Facebook y 1.3 millones de reproducciones en YouTube. También se escriben alrededor de 7.000 tweets, se suben entre 700 y 800 fotos a Instagram, se realizan más de 2 mil llamadas vía Skype.

Investigaciones más recientes, ejecutadas por la compañía Cumulus Media revelan el carácter exponencial de la cantidad de contenidos que se producen constantemente. Según el informe Cumulus Media, en sólo un minuto se envían 452.000 tuits, se suben 46.200 fotos a Instagram, se crean 1,8 millones de snaps en Snapchat, se envían más de 15.000 archivos GIF a través de Facebook Messenger, se reproducen 40.000 horas de audio en Spotify, se mandan 156 millones de correos electrónicos y 16 millones de SMS, se pasan 990.000 vistazos de perfiles en Tinder, se reproducen 4,1 millones de vídeos en YouTube y se realizan aproximadamente 3,5 millones de búsquedas en Google.

### 2.1 Big Data

La tecnología de *big data* permite recolectar, almacenar y preparar grandes volúmenes de datos para analizar o visualizar la relación entre ellos. Esto inclusive a partir de datos que se estén generando en tiempo real y que provienen de redes sociales, sensores, dispositivos de diversa índole o de fuentes de audio y vídeo. De esta manera, es posible identificar y predecir con detalle las causas o efectos de eventos, sucesos o procesos complejos. En el ámbito educativo, esta tecnología ciertamente ha comenzado a brindar grandes beneficios. Por ejemplo, en la mejora de la gestión educativa, al desarrollo de nuevos métodos para la enseñanza y el aprendizaje, la creación de nuevas carreras y opciones profesionales, así como en la generación y almacenamiento de acervos digitales que constituyen el producto de años de actividad académica, docente y de investigación [8].

Las características del Big Data se centran en las tres v:

- **Volumen:** Grandes volúmenes de datos, a partir de TeraBytes o incluso PetaBytes.
- **Variedad:** El concepto de Big Data también suele venir acompañado de diversos tipos de fuentes de datos, ya sean estructurados o no estructurados.
- **Velocidad:** La frecuencia de las actualizaciones de estas grandes bases de datos también es un punto muy a tener en cuenta. Es por ello por lo que su procesamiento y posterior análisis también ha de realizarse prácticamente en tiempo real para poder mejorar la toma de decisiones en base a la información generada.

Actualmente, se ha trascendiendo hacia un estado en el cual se necesitan de nuevos métodos, sustentados en la tecnología, para poder hacer el seguimiento de los alumnos, mejorar sus tutorías, obtener datos objetivos de sus evaluaciones, predecir los riesgos académicos o simplemente comprender el comportamiento de los grupos escolares. En este continuo cambio los profesores están más conscientes de la necesidad de actualizarse tecnológicamente para poder ofrecer una educación más efectiva y adecuada a las necesidades de la población escolar actual [9].

El método de observación en el aula ha dejado de ser la forma más efectiva para mejorar el proceso educativo. Por su parte, la analítica del big data es el recurso clave actual para entenderlo y mejorarlo. Asimismo, la información digital, generada de manera regular por las instituciones educativas, deben ser integrada en el lago de datos y así mejorar su difusión, explotación y aprovechamiento no sólo en los procesos de analítica de big data, sino por toda la sociedad.

El análisis oportuno de big data puede revelar tendencias y estadísticas que facilitan la toma de decisiones para optimizar procesos, reducir costos o encontrar soluciones para los problemas de la vida cotidiana.

La aparición de nuevas plataformas digitales está disminuyendo las brechas y promoviendo el principio de universalidad en materia de educativa. La cobertura en Educación Superior en América Latina crece a un ritmo acelerado. En ese sentido, resulta necesaria la creación de un ambiente virtual de aprendizaje, utilizando todas las herramientas ofimáticas disponibles para generar una interacción entre el docente y el alumno. El verdadero desafío está en transformar el pensamiento y la práctica, para dar paso a los beneficios que la virtualidad trae consigo, no sólo para el estudiante, sino también para las instituciones de Educación Superior. La incorporación de las TIC a la educación exige mejo-

ras contundentes, puesto que con este el mercado educativo amplía su rango a un margen nacional e internacional [10].

Los Recursos Educativos Abiertos, también conocidos como Open Educational Resources (OER), son materiales de apoyo educativo que se encuentran en el dominio público o que han sido publicados con una licencia de propiedad intelectual.

El término nació durante el Foro Mundial de la Unesco, en el cual se trató la importancia de los cursos abiertos para la educación superior. Fue tal su impacto que, en 2001, el Massachusetts Institute of Technology (MIT) anunció la publicación de casi todos sus cursos en Internet, accesibles a todo el público, marcando así la pauta para que otras instituciones, e incluso entidades como la Unesco, dieran el paso hacia un conocimiento democratizado y globalizado.

Los recursos educativos de libre acceso proporcionan una oportunidad estratégica única para mejorar la calidad de la educación y, al tiempo, facilitar el intercambio de conocimientos y el aumento de las capacidades.

## **2.2 Métodos educativos derivados del Big Data**

Entre los principales métodos educativos derivados a partir de la analítica del big data (cuyos orígenes se remontan al año 2003), y su integración con los nuevos dispositivos inteligentes y tecnología web que hoy en día ya se están aplicando en el ámbito educativo, se tienen:

- Aprendizaje adaptativo
- Educación basada en competencias
- Aula invertida y aprendizaje combinado (flipped classroom y blended learning)
- Gamificación
- Aprendizaje móvil (mobile learning)

Estos métodos, que permiten hacer llegar la educación a una mayor población estudiantil y satisfacer las continuas demandas de los alumnos (cada vez exigen una formación más flexible y personalizada), están brindando resultados muy prometedores y han cambiado la forma en la cual los estudiantes se relacionan entre sí, con sus profesores e instituciones educativas.

### **2.2.1 Ventajas de usar los datos de las herramientas educativas**

Las tecnologías están transformando el estatus educativo y este es el verdadero desafío, pues el proceso implica, en primer lugar, el desarrollo e implementación de una infraestructura eficiente y sincronizada con las necesidades de la nueva gene-

ración de alumnos; en segundo lugar, necesita de docentes creativos, recursivos y preparados para enseñar en ambientes virtuales de aprendizaje y, por último, de estudiantes dispuestos a sumergirse en esta experiencia, que debe ser lo suficientemente satisfactoria para no desertar en el camino.

Las estadísticas y el flujo de datos que se producen en el software educativo y el ambiente online tienen gran importancia para los usuarios y su evaluación del proceso de aprendizaje, pero también son indispensables para que los encargados de dichas plataformas busquen mejoras y el sector educativo profundice en sus investigaciones.

### 3. TECNOLOGÍAS PARA EL ANÁLISIS DE PROCESOS EDUCACIONALES

#### 3.1. Learning analytics

Las soluciones de Big Data para en ámbito académico se pueden centrar en el aspecto institucional (Academic Analytics) como en el aprendizaje (Learning Analytics).

- Representan una solución eficiente y sostenible, a largo plazo, para disminuir los costos provenientes de materiales de estudio.
- Multiplican las oportunidades de innovar en el proceso de enseñanza aprendizaje y habilitan una mejor adaptación y contextualización del material proporcionado a los estudiantes. Además, permiten atender las necesidades de poblaciones estudiantiles particulares.
- Aumentan la presencia y visibilidad de las instituciones ya que las licencias obligan a que, cada vez que estos recursos sean utilizados en otros contextos, los responsables citen y redireccionen al repositorio original.
- Involucran a las comunidades académicas en el proceso, de manera directa y activa. De esta manera, se podrán evaluar y entregar contenidos de calidad.

La inmediatez y la eficacia de la analítica del aprendizaje servirán para que tanto el alumno conozca cuáles son sus propias dificultades, como para que sus maestros comprendan más a fondo las dificultades personales y grupales para enseñar. Hoy en día, la analítica del aprendizaje es usada para [11]:

- Detectar las debilidades y fortalezas de los sistemas educativos.
- Desarrollar nuevos planes de estudios y métodos educativos.
- Mejorar los cursos que imparten las instituciones educativas.

- Reflexionar sobre los logros y los patrones de comportamiento de cada alumno respecto a sus compañeros o al currículum educativo utilizado.
- Predecir en qué ámbitos específicos los estudiantes necesitarán apoyo y atención personalizada.
- Ayudar a los maestros y personal de apoyo en sus intervenciones.
- Ayudar a las instituciones de enseñanza en la gestión y la toma de decisiones.
- Adecuar los procesos de evaluación y mostrar a las necesidades cada vez más concretas y específicas de los estudiantes.
- Diseñar currículos formativos más ajustados a las realidades de las instituciones educativas.

#### 3.2. Process Mining

El concepto Minería de Procesos surgió hace menos de dos décadas [12]. La disciplina de Minería de Procesos también tiene sus raíces en el trabajo de Cook y Wolf, quienes propusieron el descubrimiento de modelos de procesos a partir de los datos contenidos en los registros de eventos de los sistemas de información [13]. Desde entonces ha sido objeto de numerosas investigaciones y, por tanto, aplicada a la mayoría de las ramas de la sociedad: ingeniería, ciencias computacionales, educación, salud y el desarrollo de software.

Will van der Aalst define la Minería de Proceso como: "(...) la disciplina de investigación que permite descubrir, monitorear y mejorar los procesos reales a través de la extracción de conocimiento de los registros de eventos ampliamente disponibles en los actuales sistemas de información (...)" [7].

Las plataformas Moodle son ampliamente utilizadas en el entorno educativo para la administración, documentación y seguimiento de cursos online. En este tipo de tecnología es posible el registro de los datos de eventos que se generan en la interacción profesor-estudiante con la plataforma.

- Secuencia ordenada de eventos.
- Se pueden agrupar en trazas.
- Las trazas corresponden al comportamiento de los usuarios.
- Se pueden añadir atributos para completar la información de los eventos: marcas de tiempo, recurso accedido, etc.

La figura 1 representa un ejemplo de registros de eventos de una plataforma Moodle.



Timestamp	User id	Moodle Element	Resource	Moodle Event	Resource id
< 09/11/15 11:29,	123456,	Archive,	Lesson 3,	view resource,	934626 >
< 10/11/15 14:34,	123456,	Forum,	New forum,	view a discussion,	650215 >
< 10/11/15 15:32,	654321,	Assignment,	Assignment 2,	submit a file,	283223 >
< 10/11/15 15:32,	654321,	Assignment,	Assignment 2,	upload a file,	283223 >

Figura 1. Fragmento de registro de eventos de Moodle

Con esta información, al aplicar minería de proceso es posible:

Analizar los registros de un LMS (Learning Management Systems- Moodle) [14] utilizando técnicas de model checking.

- Lógica temporal (LTL) para comprobar patrones de comportamiento más frecuentes.

Verificación del alineamiento del plan docente

Análisis de cómo utilizan los estudiantes los recursos educativos a través del LMS

- Definición de patrones para estudiar las preguntas más frecuentes
- Comportamiento de los estudiantes al afrontar la ejecución de un proyecto.
- Diferencias entre el comportamiento esperado por los profesores y el que han mostrado realmente los estudiantes
- Correlación entre el comportamiento y las calificaciones obtenidas

Análisis del comportamiento de los estudiantes en relación al plan docente

- Acciones correctivas
- Retroalimentación para modificar y mejorar el diseño del curso
- Monitorización del comportamiento de los estudiantes durante el curso

Experiencia dirigida al estudiante y al docente

- Mejora del rendimiento académico
- Mejora de la temporalidad del curso
- Mejora del aprovechamiento del material docente

La aplicación de la minería de procesos en el ámbito educativo es una línea de investigación emergente que centra su atención en modelar, generar y evaluar los procesos que los estudiantes y profesores llevan a cabo durante el desarrollo de una unidad o módulo docente. En este contexto la granula-

ridad de lo que se entiende por proceso abarca desde el comportamiento de los estudiantes cuando realizan una actividad educativa hasta el flujo de trabajo que coordina la ejecución de todas las actividades de una unidad docente.

Actualmente, la minería de procesos está siendo usada para extraer el comportamiento de los estudiantes a la hora de navegar por las preguntas de un test, obtener las actividades realizadas por los estudiantes que crean un documento de texto de forma colaborativa, o identificar el flujo de trabajo que sigue un grupo de estudiantes cuando colaboran entre sí a través de una herramienta de comunicación determinada colaborativa [15-17]. Sin embargo, no se identifica en la literatura ningún trabajo que aborde la generación y/o evaluación automática del flujo de trabajo asociado a las actividades educativas.

#### 4. CONCLUSIONES

Las tecnologías de análisis de datos constituyen un factor importante y decisivo en la era de las comunicaciones, a partir de las cuales, la educación puede desempeñar un rol destacado en la transformación y modernización de sus procesos sustantivos.

La minería de procesos y las analíticas de aprendizaje son tecnologías de marcado desarrollo en las investigaciones y su puesta en práctica para solucionar problemas han demostrado su aplicabilidad para mejorar los procesos docentes, tanto administrativos como de enseñanza-aprendizaje.

La creación de proyectos de investigación, alrededor de estas tecnologías beneficiarán en gran medida a las universidades y centro de investigación, a partir del uso eficiente de datos obtenidos de plataformas usadas para apoyar y guiar los procesos educativos.

## 5. REFERENCIAS BIBLIOGRÁFICAS

1. **Tedesco, J. C. (2011).** Educación y sociedad. XXV Semana Monográfica de la Educación. Madrid: Santillana.

2. **Matías, H, et al. (2014).** Los Cursos en Línea Masivos y Abiertos (MOOC) como alternativa para la educación a distancia.

3. **Vera, M., Calatayud, V., Espinosa, M. (2017).** Los MOOC y la evaluación del alumnado: revisión sistemática (2012-2016). @ tic. revista d'innovació educativa, no 18, p. 65.

4. **Siemens, G., B, R. SJ. (2012).** Learning analytics and educational data mining: towards communication and collaboration. En Proceedings of the 2nd international conference on learning analytics and knowledge. ACM. p. 252-254.

5. **Baker, R., Inventado, P. (2014).** Educational data mining and learning analytics. En Learning analytics. Springer New York, 2014. p. 61-75.

6. **Walker, S. (2014).** Big data: A revolution that will transform how we live, work, and think.

7. **Van der aalst, Wil MP. (2011).** Data Mining. En Process Mining. Springer Berlin Heidelberg. p. 59-91

8. **Manyika, J, et al. (2011).** Big data: The next frontier for innovation, competition, and productivity.

9. **Provost, F., Fawcett, T. (2013).** Data science and its relationship to big data and data-driven decision making. Big Data. vol. 1, no 1, p. 51-59.

10. **Area, M. (2010).** ¿Por qué formar en competencias informacionales y digitales en la Educación Superior? Disponible en: <http://hdl.handle.net/10609/2601>

11. **Vidal, M. J., et al. 2013.** Recursos educativos abiertos. Educación Médica Superior. vol. 27, no 3, p. 307-320.

12. **Agrawal, R., Gunopulos, D. & Leymann, F. (1998).** Process Mining Models from Workflow Logs, Sixth International Conference on Extending

Database Technology, p. 469-483.

13. **Cook, J. E., & Wolf, A. L. (1999).** Software process validation: quantitatively measuring the correspondence of a process to a model. ACM Transactions on Software Engineering and Methodology (TOSEM), Vol. 8 no. 2, p. 147-176.

14. **Dahlstrom, E; Brooks, D. Ch; Bichsel, J. (2014).** The current ecosystem of learning management systems in higher education: Student, faculty, and IT perspectives. Research report. Louisville, CO: ECAR, September 2014. Available from <http://www.educause.edu/ecar>. 2014 EDUCAUSE. CC by-nc-nd.

15. **Bogarín, A, et al. 2014.** Clustering for improving educational process mining. En Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. ACM. p. 11-15.

16. **Romero, C, et al. 2016.** EDUCATIONAL PROCESS MINING. Data Mining and Learning Analytics: Applications in Educational Research. p. 1-28.

17. **Bogarín, A; Cerezo, R; Romero, C. 2017.** A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

## 6. SÍNTESIS CURRICULARES DE LOS AUTORES

**Arturo Orellana García:** Graduado de Ingeniero en Ciencias Informáticas en la Universidad de las Ciencias Informáticas UCI, La Habana, Cuba en el año 2012. Obtuvo el grado de Máster en Informática Aplicada en 2015 y Doctor en Ciencias Técnicas en 2016. Investigador de la Minería de Procesos con resultados en componentes de software donde se aplican estas tecnologías en Sistemas de Información de Salud. Coordinador de proyectos de investigación en Cuba y Europa, Conferencista en varios eventos internacionales y posee más de 40 publicaciones de Inteligencia de Negocios.