

**LA CIENCIA DE LOS DATOS Y SU IMPACTO
EN LA GESTIÓN UNIVERSITARIA.**

PhD. Giraldo de la Caridad León Rodríguez

Doctor en Ciencias Informáticas.

Docente de la Facultad de Sistemas y Telecomunicaciones de la Universidad Tecnológica
ECOTEC.

Coordinador de Educación Online.

gleon@ecotec.edu.ec

Recibido: 16 de diciembre de 2014.

Aceptado: 27 de enero de 2015.

RESUMEN

En la actualidad en las Instituciones de Educación Superior (IES) se capta, gestiona, almacena y publica una gran cantidad de datos relacionados con los procesos docente, investigativo, de vinculación y de apoyo en las IES. Esta gran cantidad de información se almacena de curso en curso para cada una de las carreras, proyectos, eventos, etc. El reto actual consiste en la posibilidad de analizar científicamente toda dicha información empleando los conocimientos y herramientas existentes en los campos de la estadística, inteligencia artificial, matemática computacional, etc. En el presente trabajo se expone cómo el empleo de la ciencia de los datos, nueva disciplina surgida ante la necesidad del análisis de grandes volúmenes de información, impacta en el grado de fundamentación de las decisiones que se toman. Se exponen los criterios fundamentales acerca de qué es la ciencia de los datos, las disciplinas que la conforman, así como las habilidades y cualidades de un científico de datos en el entorno universitario.

Palabras Clave: Gestión Universitaria, Ciencia de los Datos, Análisis de la Educación.

ABSTRACT

Today in Higher Education Institutions (IES) is captured, managed, stored and published a large amount of data related to teaching, research, linkage and support in the IES processes. This

wealth of information is stored from course to course for each of the races, projects, events, etc. The current challenge is the ability to scientifically analyze all this information using existing knowledge and tools in the fields of statistics, artificial intelligence, computational mathematics, etc. Shown in this paper how the use of data science, new discipline arose from the need of analyzing large volumes of information, impacts the degree of support to the decisions made. The fundamental assumptions about what is data science, disciplines that form it, as well as the skills and qualities of a data scientist in the university environment are exposed.

Keywords: University Management, Data Science, Analysis of education.

INTRODUCCIÓN

La Oficina de Cooperación Universitaria (OCU) en España publicó en “2020 Tendencias Universidad. Estudio Prospectiva” en el 2010 los resultados de una encuesta aplicada a 84 expertos de 13 países. Dichos resultados plantean aspectos fundamentales a tener en cuenta en los próximos 10 años tales como:

- Sistemas transparentes de certificación de la calidad, la eficiencia en la gestión y la financiación asociada a los resultados.
- Orientación del servicio y trámites universitarios al uso de las TIC.
- El futuro modelo universitario será de carácter transdisciplinario y con un alto contenido de responsabilidad social evaluable con indicadores específicos.
- La capacidad de innovación y transformación de los procesos propios constituye condición clave para lograr mayores resultados del aprendizaje, la satisfacción de los estudiantes y la eficiencia en la gestión.
- Las universidades exitosas serán aquellas que adecuen sus estructuras organizativas para mantenerse vigentes y competitivas, minimizando la burocracia e incrementando su eficacia en la gestión de los recursos.

Como principales factores críticos al cambio se citan en dicho documento: la capacidad de liderazgo de los dirigentes universitarios; en América Latina, el fortalecimiento de las competencias del personal de docencia, investigación y administración; la actitud frente al cambio, la transformación del marco normativo y legal. (León, Rodríguez, G. 2011)

Particularidades de las IES como objetos de gestión. Gestión universitaria.

No es posible alcanzar resultados de excelencia en las IES sin una gestión también excelente de todos sus procesos y recursos, sin un carácter más proactivo que prevea las futuras necesidades de la sociedad. Dicha gestión debe tener en cuenta las particularidades de las IES entre las que podemos mencionar:

- la complejidad en la gestión y coordinación de gran número de actividades diferentes, diversidad de estructuras, actores y jerarquías coexistentes.
- carácter multidisciplinario de muchas de las actividades relacionadas con la creación y/o difusión del conocimiento.
- diverso grado de subordinación real o metodológica de las IES.
- gran cantidad de normativas y resoluciones emitidas por los organismos rectores de las actividades que se realizan así como la gran cantidad de información que tributan las IES al entorno.

Las definiciones de gestión universitaria existentes en la literatura han ido incorporando los nuevos roles, misiones, retos y responsabilidades de estas instituciones. Este autor plantea que la gestión de las instituciones de educación superior tiene como objetivo operar y desarrollar eficientemente la docencia, la investigación y la extensión así como los recursos financieros, humanos, materiales y de información vinculados con las IES, con el fin de lograr los resultados definidos como relevantes para la institución y la sociedad cumpliendo con el marco regulatorio vigente. (León, Rodríguez, G. 2011)

Tal y como se expresa en (Estrada y Benítez, 2006), “en el caso de las universidades.... la aplicación de la gestión de la información y el conocimiento debe encaminarse tanto en la reorganización interna de todos sus procesos, como en la mejora de la docencia y la investigación, con el objetivo de facilitar el desarrollo de una universidad competitiva y adaptada a las nuevas demandas de la sociedad.”

En (Larrea, E. 2014), se catalogan como ejes de la calidad de las IES los modelos de gestión universitaria por procesos, el de organización del conocimiento e investigación, el de organización académica y de los aprendizajes, así como el de gestión de talento humano.

De igual forma se plantea que “los ejes básicos de sustentación y sostenibilidad de la calidad de los sistemas de la educación superior ecuatoriana, radican en las transformaciones de las matrices de organización del conocimiento, organización académica y organización de los aprendizajes formando plataformas que se iteran en cada uno de los procesos”. (Larrea, E. (2014:1)

Para poder dar solución a los problemas de las IES anteriormente expuestos, teniendo en cuenta a su vez la complejidad de las IES como organizaciones, se requiere contar con sistemas de información que reflejen adecuadamente el comportamiento de las variables que reflejan su desempeño. Estos sistemas de información deben garantizar la necesaria interoperabilidad entre los procesos sustantivos de las IES (docencia, investigación, vinculación y procesos de apoyo) así como garantizar el almacenamiento eficiente, fiable y seguro de la información generada por cada uno de los procesos.

Es posible decir que la gran mayoría de las IES cuentan con sistemas de información que reflejan con mayor o menor alcance y calidad el comportamiento y marcha de los procesos sustantivos. Sin embargo, muchos de estos sistemas de información presentan insuficiencias en lo referente a la capacidad de realizar análisis integrales y predicción de los indicadores que reflejen la calidad de dichos procesos,

A partir de la problemática anterior, el objetivo del presente trabajo es exponer qué áreas del conocimiento científico y cuáles competencias y habilidades se requieren introducir en la gestión universitaria para poder realizar análisis integrales y predicción de los indicadores que reflejan la calidad de los procesos sustantivos en las IES.

DESARROLLO

Para dar solución al problema y el objetivo planteado, se realizó un análisis del estado del arte de la problemática asociada al análisis de los datos en la literatura. Como resultado, pudo detectarse que en la actualidad se refleja, con una alta frecuencia en las publicaciones científicas, el empleo e introducción de nueva disciplina llamada Ciencia de los Datos surgida ante la necesidad del análisis de grandes volúmenes de información. Dicha disciplina impacta en el grado de fundamentación de las decisiones que se toman e en las organizaciones a partir de la información brindada por los sistemas de información utilizados.

A continuación se expone la definición de Ciencia de los Datos y las principales áreas del conocimiento que son objeto de su estudio.

La Ciencia de los Datos.

A diario en el mundo se generan 2.5 trillones de bytes de información. Esta información proviene de todos lados, sensores que recogen información climática, publicaciones en las redes sociales, imágenes y vídeos digitales, registros de compra y transacciones y señales de GPS de los móviles, entre otros. Toda esta información se conoce como “Big Data” y es a partir de esta fuente masiva de datos que es inminente el nacimiento de un profesional que conozca y genere un uso a esta información: el científico de datos. (Universia, 2014)

En diciembre del 2010 Michael E. Driscoll abrió una discusión en torno a esta cuestión “What is data science?”. Driscoll aporta su propia definición sobre la Ciencia de Datos: “Es la ingeniería civil de los datos. Sus acólitos poseen un conocimiento práctico de herramientas y materiales, junto con una comprensión teórica de lo que es posible”. Por otro lado, Giuseppe Paleologo define la Ciencia de los Datos como “el conjunto de prácticas sobre almacenamiento, gestión y análisis de conjuntos de datos lo suficientemente grandes que requieren de computación distribuida y los recursos de almacenamiento”.

De acuerdo a la consultora estadounidense de las tecnologías de la información Gartner Inc., “para 2015 4,4 millones de empleos en el sector a nivel global tendrán un vínculo directo con el procesamiento de Big Data”. (Universia, 2014)

Según Mike Loukides , el problema no es encontrar datos sino saber qué hacer con ellos. “La cuestión a la que se enfrenta toda empresa, hoy en día, es cómo usar los datos de forma efectiva. No sólo los datos propios sino de múltiples fuentes, todos los datos que están a disposición y son relevantes. Usar los datos de forma efectiva requiere algo diferente de las tradicionales estadísticas”. (Loukides.M, 2011)

Lo que diferencia el Big Data de las estadísticas es que la Ciencia de Datos tiene un enfoque holístico. Cada día se generan más datos en la naturaleza y en la actividad social, espiritual y productiva del hombre La Ciencia de Datos tiene que ver con la recopilación de información, su conversión a un formato factible para su procesamiento y extracción del conocimiento.

Un reciente estudio en los Estados Unidos del McKinsey Global Institute, plantea que un reto urgente e importante para el país lo constituye la escasez del talento analítico y de gestión necesario para sacar el máximo provecho al Big Data. En dicho reporte se expone que, tan solo en EEUU, 5 millones de puestos de trabajo requerirán de habilidades en el análisis de datos en el 2018. Además se necesitarán 1,5 millones de directivos y analistas con profundas habilidades analíticas y técnicas que puedan realizar las preguntas correctas a los datos y utilizar los resultados que brinde dicho análisis efectivamente (Loukides.M, 2011).

En estos momentos se vive una verdadera explosión en cuanto a la cantidad de datos a nivel mundial. IBM estima que el 90% de los datos existentes en la actualidad han sido creados tan solo en los últimos dos años. La capacidad de la humanidad para extraer el valor social y económico de los datos recientemente disponibles está limitada por su propia falta de experiencia. Se requiere de todo un conjunto de nuevas habilidades y herramientas con este fin.

El volumen de datos puede llegar a ser tan grande que no cabe en una sola computadora, es imposible ser procesado por los sistemas de gestión de bases de datos, paquetes estadísticos o sistemas de representación gráfica tradicionales. Los textos digitalizados, el audio y contenido visual, al igual que los datos de sensores y weblogs suelen estar desordenados, incompletos y no estructurados, muchas veces de origen y calidad dudosa. Además, durante el trabajo con datos provenientes de diversos y dispersos usuarios, se tocan temas relacionados con la privacidad, autenticidad y seguridad.

Tal y como se plantea en (Loukides.M, 2011) “la web está llena de aplicaciones basadas en datos (data-driven app), siempre hay una base de datos detrás de un front-end,” tal es el caso de las aplicaciones de los bancos, de comercio electrónico o de enseñanza online en una universidad. Sin embargo, no siempre cuando se usan, e incluso analizan datos, se aplica la ciencia de los datos. Una aplicación basada en datos brinda su valor añadido a partir de sus propios datos, transformándose de una aplicación que usa datos a un producto de datos, o sea, el dato como producto resultante del procesamiento.

Google es un líder mundial en la elaboración de productos de datos. Tal y como se expone en (Loukides.M, 2011), existen ejemplos tales como el algoritmo para establecer el ranking de una página web. Dicho algoritmo fue el primero que utilizó información externa a la propia página, o sea, el número de enlaces o citas realizadas a dicha página. La empresa Amazon salva las búsquedas de productos o servicios de cada usuario y las correlaciona con las búsquedas de otros usuarios obteniendo como resultado impresionantes sugerencias.

El hilo común que une la mayoría de estas aplicaciones lo constituye el hecho de que los datos recopilados a partir de la interacción con cada usuario aporta un valor añadido. Ya sean los datos con los términos de una búsqueda, una muestra de voz, o un comentario, todos los usuarios conforman un lazo de retroalimentación que tributa al completamiento y mejora de los productos de datos ofertados. Ese es el principio básico de la ciencia de datos.

Si se retoma lo planteado acerca de la explosión actual de los datos a nivel mundial en todo tipo de organizaciones productivas, de servicios tales como hospitales o universidades, puede observarse fácilmente que las organizaciones emplean no sólo sus propios datos o los datos que le brindan sus usuarios, sino también datos de otras fuentes como lo puede ser el empleo de mapas de zonas geográficas con vistas a analizar o visualizar los resultados de determinados indicadores. O sea, en cada caso pueden incluirse no solo los datos propios, sino todos aquellos que puedan ser relevantes ante determinado análisis o investigación.

Lo que diferencia a la ciencia de datos de las estadísticas es que la ciencia de datos tiene un enfoque holístico. Cada día se encuentran más y más datos en la naturaleza y los científicos de datos están involucrados en su recopilación, su adecuado “acondicionamiento”, de forma tal que puedan ser manejados lo mejor posible y, de ese modo, dichos datos puedan contar su historia permitiendo posteriormente a estos especialistas presentar dichas historias a los demás.

El científico de datos es una nueva profesión que hoy es considerada clave en el mundo de las tecnologías y es una de las mejor pagadas. Se trata de una persona formada en las ciencias matemáticas y las estadísticas, que domina la programación y sus diferentes lenguajes, ciencias de la computación y analítica. Este especialista debe contar con las habilidades y los conocimientos necesarios que le permita comunicar sus hallazgos, no sólo al área de tecnología sino además al sector de los negocios. Debe dominar la tecnología y las bases de datos para modificar y mejorar la orientación de los negocios de la organización a la que pertenece. “El científico de datos analiza, interpreta y comunica las nuevas tendencias en el área y las traduce a su organización con vistas a su uso y adaptación a los productos y servicios ofertados creando nuevas oportunidades de negocio. Google, por ejemplo, tiene 600 personas dedicadas al estudio del Big Data”. (Loukides.M, 2011)

Diversas fuentes coinciden en que dentro de las características innatas de los científicos de datos se incluyen la curiosidad, la creatividad, la objetividad, pensamiento lógico, el sentido común, la perseverancia, el sentido práctico, el buen juicio y atención al detalle. No hay duda de que la necesidad de los científicos de datos solo va a aumentar.

El éxito de compañías como Facebook, Google o LinkedIn no se debe solo al hecho de que hayan logrado acumular gigantescas cantidades de datos, sino a su capacidad de ser creativos con ellos, encontrándoles nuevo valor. Estamos en la “era de la información”, en la que crear valor a través de la información se ha convertido en la “nueva economía”. Y, en este nuevo escenario, surge un nuevo perfil adaptado a las nuevas necesidades: el científico de datos (data scientist).

No obstante, las áreas del conocimiento de Big Data y Data Science son tratadas de forma heterogénea por diferentes autores. Según los criterios más comunes, cualquier definición debe incluir: (Schubert, 2014)

En Big Data:

- Cómputo distribuido en múltiples servidores.
- Gestión y procesamiento de datos.
- Ir más allá de las bases de datos relacionales y data warehouses.
- Permitir resultados que no estaban disponibles con los enfoques anteriores, o que llevarían sustancialmente mucho más tiempo.

En Ciencia de Datos:

- Involucrar conocimientos de uno o más dominios.
- Tomar en cuenta aspectos computacionales.
- Incluir técnicas científicas tales como la prueba de hipótesis y la validación de resultados.
- Confiabilidad de los datos.
- Un mayor peso al empleo de métodos matemáticos y estadísticos.
- Incluir el aprendizaje automatizado (machine learning), inteligencia artificial o algoritmos de descubrimiento de conocimiento (knowledge discovery).
- Implicar la visualización y creación rápida de prototipos para el desarrollo de software.

Hoy los mercados son tan rápidos como competitivos, y ha tomado gran importancia el poder desarrollar rápidamente una idea o modelo de negocio que permita competir y dominar un nuevo segmento de mercado. Se requiere de agilidad, creatividad y de personal calificado. En este sentido, Big Data está transformando el negocio. El científico de datos ayuda a las organizaciones a maximizar la información mediante la realización de análisis de información profundos que extraen el potencial oculto en los datos y descubren sus tendencias. Estos hallazgos son predictivos y fundamentan decisiones operacionales en las organizaciones.

En un mundo donde los datos están en permanente crecimiento, actualmente son muy escasos los profesionales capaces de gestionarlos y analizarlos. “La encuesta Data scientist study, realizada por EMC para el período 2012-2017, señala que la carencia de este perfil es reconocida por el 32% de las organizaciones como el mayor obstáculo para adoptar estrategias de Big Data, con un 63% que percibe que la demanda de científicos de datos supera a la oferta. A esto hay que añadir que un 83% de esas empresas considera que la contratación de este profesional en sus organizaciones genera aún mayor demanda de científicos de datos adicionales, ante la apertura de nuevas posibilidades”. (Gimenez, A., 2014).

Por otro lado, al ser un perfil muy demandado, el ciclo de vida de los científicos de datos dentro de la empresa puede verse acortado por atender a las ofertas de un mercado en ebullición. “La demanda de científicos de datos no es exclusiva de las empresas de fuerte componente comercial o competitivo, sino que puede sentirse en todo el espectro económico y social, incluyendo administraciones eficientes, ciudades inteligentes o universidades avanzadas”. (Gimenez, A., 2014).

La Ciencia de los Datos y la Gestión Universitaria.

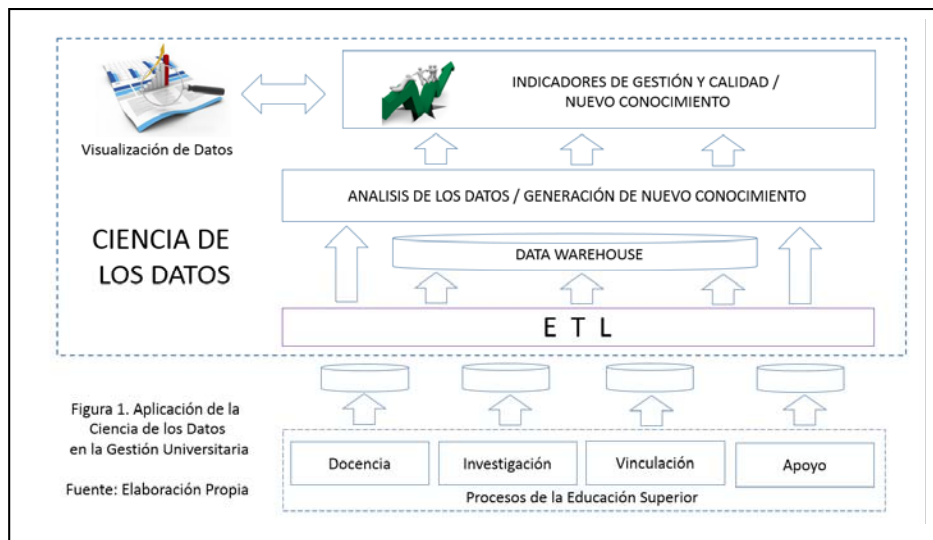
A partir de todo lo expuesto respecto a los retos actuales de las IES a nivel mundial y en el Ecuador, las características de las instituciones de Educación Superior como entidades complejas y los requerimientos de la sociedad respecto al empleo eficiente y eficaz de los recursos humanos, materiales, financieros y de información de que disponen con vistas a que los productos finales

resultantes de sus procesos sustantivos (docencia, investigación, vinculación y apoyo) cuenten con la calidad esperada, se infiere la necesidad de la introducción y empleo de la ciencia de los datos en la gestión universitaria.

Los sistemas de información asociados a las universidades disponen de grandes volúmenes de información. Tan solo como ejemplo, es posible mencionar:

- Registro de estudiantes. Sus características respecto a su procedencia territorial y escolar, situación familiar, etc.
- Cumplimiento del itinerario de materias para cada carrera. Grado de la asimilación de los recursos de aprendizaje y alcance de las competencias definidas para cada asignatura y período.
- Disciplina del estudiante reflejada en la asistencia, cumplimiento de deberes,
- Participación en actividades docentes y extracurriculares, etc.
- Registro de profesores. Sus características respecto a los estudios vencidos, categoría docente y grado científico alcanzado.
- Capacitaciones recibidas tanto desde el punto de vista docente – metodológico como asociadas a la experticia que domina e imparte.
- Registro de evaluaciones de estudiantes y profesores. Relación existente entre ellas.
- Registro de proyectos de investigación.
- Participación de estudiantes y profesores en dichos proyectos, resultados alcanzados.
- Publicaciones científicas.
- Presentaciones en eventos científicos.
- Registro de trabajos de vinculación reflejando el impacto en la solución de problemas de la sociedad.
- Registro de graduados por carreras. Conjunto de datos acerca de la satisfacción de la sociedad respecto a la preparación de los graduados en las diferentes instituciones donde se emplean.
- Cumplimiento de los objetivos anuales trazados por la IES.

En la figura 1 se muestra un esquema de cómo se aplicaría la Ciencia de los Datos en la Gestión Universitaria.



Fuente: elaboración propia.

Los datos generados por cada uno de los procesos de la educación superior servirían de sustento a los almacenes de datos (data warehouse) con los indicadores definidos, así como al proceso posterior de análisis de datos y visualización de los resultados.

Los ETL son los procesos de extracción – transformación – carga de la información. A través de estos procesos se hace posible contar con los datos requeridos para el cálculo de indicadores u obtención de análisis en el formato adecuado a partir de los datos que se obtienen de cada uno de los procesos.

En toda esta etapa se emplearían métodos y herramientas asociadas a la Ciencia de los Datos.

CONCLUSIONES

El análisis de los datos asociados a los procesos sustantivos en las IES, realizados de forma integral, empleando la ciencia de los datos, permitirá fundamentar con más precisión las decisiones a tomar.

Poder analizar en detalle los datos almacenados del pasado y presente permitirá realizar las inferencias necesarias para el futuro pudiendo detectarse a tiempo cualquier anomalía o predecir nuevas tendencias y requerimientos en los procesos sustantivos de la educación superior al nivel de universidad, centro de investigación, región o país.

Deberá llevarse a cabo un proceso de capacitación y estructuración de equipos de trabajo que se encarguen de la aplicación de la ciencia de los datos en las instituciones de educación superior.

Debido a la gran demanda de los científicos de datos a nivel mundial y en el país, se hace necesario diseñar las estrategias y acciones requeridas para la retención y motivación de dicho personal en cada institución.

La introducción de la ciencia de los datos en las IES elevará su competitividad, adaptabilidad y velocidad de respuesta ante los nuevos requerimientos que se plantean a las universidades en la actualidad.

REFERENCIAS BIBLIOGRÁFICAS

Estrada, V. y Benítez F. (2006) “La gestión del conocimiento en la nueva universidad cubana”. En “La Nueva Universidad Cubana y su contribución a la universalización del conocimiento”, Editorial “Félix Varela”, ISBN 959-258-971-2.

Gimenez, A. (2014). “Científicos de datos. Tan atractivos como escasos”. Harvard Deusto Número 121 de Márketing y Ventas. Disponible en <http://www.harvard-deusto.com/articulo/Cientificos-de-datos-tan-atractivos-como-escasos> Consultado feb 2014.

Larrea de Granados, E. (2014): “El currículo de la educación superior desde la complejidad sistémica “ Ecuador

León Rodríguez, G. (2011): “Modelo de gestión del conocimiento para las áreas económicas del sistema de instituciones del Ministerio de Educación Superior”, Tesis de Doctorado, Universidad de La Habana, La Habana, Cuba.

Loukides, M. (2011). “What is Data Science?”. Ebook Kindle Edition.

Schubert, E.C (2014) “Definiendo Ciencia de Datos”. Disponible en <http://sg.com.mx/revista/43/definiendo-ciencia-datos#.VOdCEizLroY> Consultado en febrero 2015.

Universia – España. (2014) “El científico de datos: una novedosa y necesaria profesión”.

Disponible en <http://noticias.universia.es/ciencia-nt>

[tt/noticia/2014/05/06/1095994/cientifico-datos-novedosa-necesaria-profesion.html](http://noticia/2014/05/06/1095994/cientifico-datos-novedosa-necesaria-profesion.html)

Consultado en junio del 2014.