2

TiES Revista de Tecnología e Innovación en Educación Superior

Publicación Semestral • Octubre de 2019 • ISNN 2683-2968

Universidad Nacional Autónoma de México
Dirección General de Cómputo y de Tecnologías de Información y Comunicación

Octubre 2019 | número de revista 2 • ISSN 2683-2968
https://www.ties.unam.mx/

TIES, REVISTA DE TECNOLOGÍA E INNOVACIÓN EN EDUCACIÓN SUPERIOR (www.ties.unam.mx) 2019, Año 1, No. 2, octubre 2019, es una publicación semestral editada por la Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, a través de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, (DGTIC), Circuito Exterior s/n, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, Teléfono: (55) 56228166, https://www.ties.unam.mx, revista.ties@unam.mx. Editor responsable: Lic. Lizbeth Luna González. Número de reserva de Derechos de Autor otorgado por INDAUTOR: 04-2019-011816190900-203 ISSN: 2683-2968, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número, Dirección General de Cómputo y de Tecnologías de Información y Comu-nicación, (DGTIC). Circuito Exterior s/n, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, fecha de la última modificación, octubre de 2019.

Octubre 2019 | número de revista 2 • ISSN 2683-2968
https://www.ties.unam.mx/

DIRECTORIO

Universidad Nacional Autónoma de México

Dr. Enrique Luis Graue Wiechers
Rector

Dr. Leonardo Lomelí Vanegas
Secretario General

Dirección General de Cómputo y Tecnologías de Información y Comunicación

Dr. Felipe Bracho Carpizo
Director de la DGTIC

José Fabián Romo Zamudio
Director de Sistemas y Servicios Institucionales

M.A.O. Miguel Ángel Mejía Argueta
Responsable de Acervos Digitales

Dr. Luis Alberto Gutiérrez Díaz de León
Director General de la Revista

Lic. Lizbeth Luna González
Directora Editorial de la Revista

CRÉDITOS

Dr. Felipe Bracho Carpizo
Director General de la Revista

Lic. Lizbeth Luna González
Directora Editorial de la Revista

Liliana Minerva Mendoza Castillo
Asistente Editorial de la Revista

José Fabián Romo Zamudio
Lic. Lizbeth Luna González
Liliana Minerva Mendoza Castillo
Diseño de Contenidos

Lic. Lizbeth Luna González
Arquitectura de la información

Mtro. Rodolfo Cano Ramírez
Formación PDF

Daniel Méndez (Becario)
Estructura HTML

Liliana Minerva Mendoza Castillo
Diana Lissete Macías Ruíz (Estudiante)
Corrección de estilo

Liliana Minerva Mendoza Castillo
Ing. Carlos Alberto Román Zamitiz
Formación HTML

Lic. Lizbeth Luna González
Ing. Carlos Alberto Román Zamitiz
Liliana Minerva Mendoza Castillo
M.A.O. Miguel Ángel Mejía Argueta
Administrador del OJS

Octubre 2019 | número de revista 2 • ISSN 2683-2968
https://www.ties.unam.mx/

COMITÉ EDITORIAL FUNDADOR

Octubre 2019 | número de revista 2 • ISSN 2683-2968
https://www.ties.unam.mx/

ÍNDICE

Octubre 2019 | número de revista 2 • ISSN 2683-2968
https://www.ties.unam.mx/

EDITORIAL

Se mide la inteligencia de un individuo por la cantidad de incertidumbres que es capaz de soportar.

Immanuel Kant

Las sociedades actuales han incrementado exponencialmente el uso de las TIC, incorporándolas a su quehacer cotidiano, pero es evidente la existencia de diferencias significativas de acuerdo a los aspectos políticos, sociales, económicos y culturales de cada región; por ello las Instituciones de Educación Superior (IES) cobran un papel importante para su generación y uso.

Las IES son organizaciones complejas y diversas de acuerdo a su situación geográfica, tamaño, estructura organizacional y visión establecida; asumen el reto de desarrollar sus funciones de docencia, investigación, vinculación y extensión en un contexto de incertidumbre y cambio, donde en el ámbito educativo tienen que formar profesionales para trabajos que aún no han sido creados.

La vigilancia tecnológica permite que las IES incorporen en sus estrategias institucionales nuevos proyectos de TIC; día a día emergen tecnologías, conceptos y estándares que están avanzando a gran velocidad en el ámbito de la industria, gobierno y otros organismos, como son: Inteligencia Artificial, Big Data, Minería de Datos, Ciberseguridad, Seguridad de la Información, entre otros.

En esta publicación digital se abordan temas que sin duda permitirán al lector reflexionar acerca de la relevancia y pertinencia de avanzar desde distintos frentes colaborando para acortar la brecha digital al generar conocimiento y presentar investigaciones y aplicaciones que otras IES han desarrollado como:

- La teoría de la mente basada en el reconocimiento de patrones que nos lleve a entender cómo funciona el cerebro y cómo genera la inteligencia, en donde la Inteligencia Artificial puede ser incorporada a las IES, no solo como parte de los programas educativos, sino también en la transformación de los procesos universitarios.
- La importancia de hacernos cuestionamientos y al mismo tiempo dar respuesta a la necesidad de concientización en materia de seguridad de la información hacia las áreas estratégicas de TIC y las IES. Se presentan contenidos que aportan argumentos que pueden ser de utilidad para justificar las inversiones y recursos destinados a la prevención, mantenimiento e implementación de sus proyectos de Ciberseguridad y de Seguridad de la Información.
- La explosión de los datos que estamos viviendo, ha dado lugar a conceptos como Big Data y Ciencia de Datos, en donde las instituciones han encontrado en la minería de datos opciones para la solución a retos que se presentan en nuestro país asociados con la deserción escolar de alumnos.

Ante un panorama que se vislumbra impredecible en el ámbito de las TIC, esta publicación contribuye a generar conocimiento para acelerar y multiplicar los procesos de asimilación tecnológica que nuestras IES demandan.

Les hacemos una atenta invitación a todos los académicos, investigadores, técnicos y especialistas de nuestras IES a compartir sus experiencias en esta revista digital, ya que es un espacio propicio para seguir construyendo una comunidad de apasionados por las TIC y por la excelencia de la calidad educativa de nuestro país.

Mtra. Carmen Díaz Novelo
Editora invitada

LA INTELIGENCIA ARTIFICIAL EN LA TRANSFORMACIÓN DE PROCESOS UNIVERSITARIOS

Yuri Sebastián Martínez
https://www.ties.unam.mx/

Fecha de recepción: 27 de junio de 2019 • Fecha de publicación: octubre de 2019

Octubre de 2019 | número de revista 2 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

LA INTELIGENCIA ARTIFICIAL EN LA TRANSFORMACIÓN DE PROCESOS UNIVERSITARIOS

Resumen

En el presente artículo se expondrá la posible aplicación de la Inteligencia Artificial (IA) en las diferentes áreas que conforman las universidades: en su comunidad (académico, investigadores, estudiantes, personal administrativo), la oferta académica y en el ámbito de investigación, de cultura deportivo y el impacto que puede tener en cada una de ellas, para la toma de decisiones, propuestas, trabajos e investigaciones. De igual manera se analizarán los pros y contras de aplicar dicha tecnología en las universidades.

Palabras clave:

Inteligencia Artificial, aprendizaje, pensamiento, automatización, software, hardware, educación, administración, procesos.

THE ARTIFICIAL INTELLIGENCE IN THE TRANSFORMATION OF UNIVERSITY PROCESSES

Abstract

In this article we will discuss the possible application of artificial intelligence in the different areas that make up the universities: their community (academics, researchers, students, administrative staff), the academic offer, research, culture and sports and the impact that in each of them, the possible application of artificial intelligence may help with each one of the decisions, proposals, works, investigations and startups of each of the activities of the universities. In the same way, the pros and cons of applying said technology in each of the processes in the universities will be analyzed.

Keywords:

Artificial Intelligence, learning, thinking, automation, software, hardware, education, administration, processes.

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

LA INTELIGENCIA ARTIFICIAL EN LA TRANSFORMACIÓN DE PROCESOS UNIVERSITARIOS

Introducción

EL ARTÍCULO PRESENTA UN PANORAMA DE QUÉ ES LA INTELIGENCIA ARTIFICIAL (IA) Y LOS ALCANCES QUE PUEDE lograr en la transformación de procesos universitarios así como sus aplicaciones y el debate que sobre ella existe. Por tanto, empezaremos diciendo que Inteligencia Artificial es *el arte de crear máquinas con capacidades de desarrollar funciones que normalmente realizan los humanos y que requieren inteligencia*, como lo definió Ray Kurzwail en su libro "La era de las máquinas inteligentes" [1], [2]

Los estudios principales de la IA son: [3], [1]

1. Razonamiento del sentido común: tiene que ver con el modo en que los humanos resolvemos problemas utilizando el sentido común, de esta manera la máquina desarrolla una forma de resolver el problema específico para la que fue creada.
2. Aprendizaje automático: se relaciona con el procedimiento para aprender las normas y reglas que nos permiten desenvolvernos en la vida a través de vivencias cotidianas.
3. Algoritmos genéticos: estudia la evolución de los algoritmos biológicos y ve la forma en que los algoritmos computacionales puedan imitar la evolución de los mismos.
4. Redes neuronales artificiales: estudia la interacción de las neuronas biológicas para generar ese comportamiento en una red artificial de neuronas.
5. Razonamiento de lógica formal: estudia cómo el ser humano toma decisiones desde un razonamiento válido para poder aplicarlo en la programación de las máquinas.

Cada uno de los puntos mencionados pueden y deben cumplirse para decir que la IA se está aplicando en algún algoritmo, software o máquina y puesto que estos sistemas requieren de grandes cantidades de datos para avanzar en el aprendizaje (o en otras palabras autogenerar códigos nuevos), cada día se nutren de mayor información.

Dentro de la categorización que existe en la IA, se encuentran los sistemas que están diseñados para emular algunas características del pensamiento o comportamiento humano como las redes neuronales artificiales, resolución de problemas y toma de decisiones. A continuación se enumeran y se da un ejemplo de estos sistemas: [4], [5]

1. Sistemas que actúan como humanos: la robótica es una de las aplicaciones de la IA, las máquinas simulan el comportamiento humano.
2. Sistemas que piensan racionalmente: se basan en el pensamiento lógico racional del ser humano y son llamados como sistemas expertos.
3. Sistemas que actúan racionalmente: imitan el comportamiento humano, como los agentes inteligentes.

Figura 1.
G. Altmann, "Inteligencia Artificial, cerebro, control," 2019. [Fotografía]. Disponible en: https://pixabay.com/es/illustrations/inteligencia-artificial-cerebro-4469138/ [Consultado en septiembre 17, 2019].

Tipos de pensamientos en la IA:

1. Inteligencia artificial simbólica deductiva: analiza formalmente y desde la estadística el comportamiento humano en diferentes problemas.
2. Inteligencia computacional, desarrollo o aprendizaje interactivo: se basa en datos empíricos.

Con base en los tipos de pensamiento, la IA ha aplicado los siguientes tipos de programación para solucionar o ayudar a las diferentes áreas de la vida (como se ejemplificará más adelante).

Las programaciones más usadas son: [2], [4], [5]

1. *Machine Learning*: viejo modelo de programación para resolver un problema donde hay una entrada y una salida. Después de muchos años, los investigadores probaron diversos estudios para crear la IA, pero en vez de programar computadoras para que fuesen inteligentes mediante rutinas de software codificadas manualmente y realizar una tarea en particular, se les dio a las máquinas acceso a una gran cantidad de datos de muestra para codificarlos y encontrar patrones o aprender por sí mismos cómo desarrollar la tarea. Por lo tanto, *Machine Learning* es un subconjunto de la Inteligencia Artificial y el enfoque principal es "aprender" en lugar de solo programar computadores. Aquí una máquina utiliza algoritmos complejos para analizar una cantidad masiva de datos, reconocer patrones y hacer una predicción, sin requerir que una persona programe instrucciones específicas en el software.
2. *Deep learnig*: la inteligencia de la computadora va aún más lejos, estas redes imitan la conectividad del cerebro humano, clasifican conjuntos de datos y encuentran correlaciones entre ellos. Con su nuevo conocimiento, adquirido sin intervención humana, la máquina puede aplicar sus conocimientos a otros

conjuntos de datos, cuantos más datos tenga la máquina a su disposición, más precisas serán sus predicciones. Esta es la tecnología que ha tenido más auge en los últimos años, por todo lo que ha podido conseguir. Utiliza los principios de los algoritmos básicos de *Machine Learning* intentando modelar abstracciones de alto nivel en datos usando arquitecturas computas. Con este aprendizaje se puede hacer prácticamente cualquier cosa, la desventaja requiere de un número superior de datos, comparado con los otros algoritmos, el mayor inconveniente es que necesita tecnología de punta para poder procesar los algoritmos, razón por la cual no ha sido hasta ahora que se ha venido desarrollando este aprendizaje.

Desarrollo

El desarrollo del pensamiento expresado a través de software o hardware que da vida a la IA tiene como objetivo ayudar en los quehaceres de la vida cotidiana y sobre todo los que tienen que ver con procesos repetitivos, de alto cálculo matemático, donde esté en riesgo la vida, en la salud, en los hábitos de consumo, en las ventas, en el mercadeo, etc., el campo de aplicación es variado, como se explica a continuación.

Algunas de las aplicaciones de la IA son: [2], [4], [5]

1. Sistemas de control avanzado: reconocimiento facial, ergonómico, físicos, conocer hábitos sociales, económicos, alimentación, asociación de la persona con vehículos (por ejemplo que reconozca si es el conductor asignado y con base en ello identificar sus hábitos de manejo y de ahí determinar sus rutas para conducir).
2. Sistemas autónomos: capacidad de las máquinas de ver y entender el mundo para tener autonomía de movimiento mediante decisiones que permitan reducir accidentes y mejorar el tráfico, en cuanto a transportes se refiere. Se reduciría el número de vehículos ya que habría la opción de su uso bajo demanda.
3. Arte: mediante algoritmos aplicados a la técnica denominada transferencia de estilos, a la imagen tomada mediante una cámara se le podrá dar el estilo de una pintura previamente aprendida por el algoritmo. De igual manera se podría copiar hasta el detalle más mínimo de una obra de arte para preservarla mediante su reproducción o mejor aún, aprender las técnicas para ser un crítico de arte. Existe también el programa *Amper Music* que crea música por medio de las redes neuronales.

Figura 2.
S. Soman, "Inteligencia Artificial, tecnología futurista," 2018. [Fotografía]. Disponible en: https://pixabay.com/es/photos/inteligencia-artificial-tecnolog%C3%ADa-3262753/ [Consultado en septiembre 26, 2019].

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

Figura 3.
E. Hubbard, "Conferencia Internacional sobre Robots" 2018. [Fotografía]. Disponible en: http://www.gaceta.unam.mx/justina-a-la-conferencia-internacional-sobre-robots/ [Consultado en septiembre 26, 2019].

4. Redes generativas adversarias: los algoritmos crean nuevas imágenes de personas, animales u objetos, mezclando características de las mismas aprendidas previamente, es decir, si hablamos de aves se puede crear un nuevo tipo de aves con características de las existentes en la naturaleza.
5. Interfaces cerebro – máquina: por medio de equipos de resonancia magnética poder visualizar los pensamientos en una imagen. Futurizando, se puede pensar que el enlace entre una máquina y nuestro pensamiento podría crear desde un texto hasta el control de la misma.
6. Software de *IBM Watson y Sesame Workshop* para niños de preescolar: este tutorial ayuda a los niños de manera individualizada a aprender palabras en edad temprana. De igual manera el programa *IBM Watson* ayuda de manera personalizada a estudiantes de nivel universitario. Ambos programas analizan el nivel de avance del estudiante para dar una ayuda más apegada a la misma sin necesidad de que un humano le programe el nuevo nivel en el que está el estudiante. Desde la perspectiva de los profesores, el programa es una guía basada en los temas en que los alumnos tienen más dudas y de esa manera pueden ajustar sus clases.
7. IBM Proyecto Debate: IBM ha desarrollado algoritmos que permiten por medio de la AI debatir en temas diversos como lo hacen los humanos, argumentando en base a temas previamente aprendidos por la máquina, lo cual impacta en la toma de decisiones.
8. Ayuda para ciegos: pasillos virtuales a través de un sensor tipo brazalete, el cual emite una vibración si al caminar o correr en algún espacio abierto se sale del pasillo virtual. De igual manera el dispositivo conectado al GPS de un *Smartphone*, dará la ruta más segura y con menos tránsito para las personas con discapacidad visual.
9. Seguridad biométrica: por medio de algoritmos, las cámaras de seguridad pueden reconocer a una persona por medio de la lectura de su rostro, lo cual lleva a que no solo pueda detectarse algún reporte policial o para el acceso a un sitio como el hogar u oficina, sino que, se han desarrollado algoritmos que predicen por medio del análisis facial si una persona es un delincuente.
10. *Watson For Drugs Discovery*: acelera la investigación médica o científica de la cantidad de datos que existe referente a un tema, lo cual reduce los tiempos gracias a su capacidad de relacionar temas referentes a lo que se está estudiando.
11. Ventas: se está utilizando para predecir la demanda de algún artículo, optimizar precios, optimizar la cadena de suministro y recomendaciones personalizadas, como lo hacen varias cadenas de ventas por internet.
12. Mercadeo: análisis de ventas por internet, optimización de anuncios, detección de gustos del consumidor de artículos en internet y *chatbots*.
13. Salud: predicción de riesgos en pacientes, diagnóstico de enfermedades y alertas.

14. Telecomunicaciones: automatización de ayudas por medio de *bots*, mantenimiento preventivo, análisis de consumo de datos y asistentes virtuales (SIRI, GOOGLE Assitant, Amazon Alexa y Microsoft Cortana).
15. Finanzas: detección de fraudes, análisis de riesgo y puntuación crediticia.

En cada uno de los ejemplos anteriores tienen cabida las actividades que se desarrollan en las universidades como: la administración, la investigación, la economía de las universidades, la cultura, el deporte, la seguridad y sobre todo la educación.

Aplicaciones

LA PLATAFORMA DE *IBM WATSON* POR EJEMPLO, APOYA EN EL APRENDIZAJE; PODEMOS VISUALIZAR un campo lleno de expectativas para los alumnos y los profesores. Con la IA los alumnos podrán tener software especializado en su tema de estudio que les sea una guía para la asesoría en particular. Lo mismo en la ingeniería, medicina o en las ciencias sociales o humanidades. En otras palabras, tendrán un tutor personalizado que les auxilie en el estudio de un tema. Por otro lado, a los profesores les ayudaría en ver los tópicos en los cuales los alumnos tienen más dudas y redirigir sus técnicas pedagógicas a esos temas para un mejor aprendizaje.

En la investigación, otra área relevante para las universidades es el uso de plataformas como por ejemplo, *IBM Watson for Drugs*, ha ayudado a los investigadores en cuestiones médicas, búsqueda de información y análisis de la misma para tener resultados en menor tiempo.

El algoritmo de debate que desarrolló IBM, podrá auxiliar para tomar decisiones en la administración de las universidades, desde el análisis de cierto problema por medio del debate hasta el hecho de poder tener una asistente personal que realice el trabajo de oficina en cada una de las áreas de gestión de las universidades o en la parte financiera para hacer cálculo de riesgos o detección de fraudes.

Otro apoyo significativo se presenta en la difusión de las carreras, cursos o diplomados que se imparten en las universidades a través del análisis de las tendencias del mercado en cuanto a la demanda laboral, para reforzar

Figura 4.
G. Altmann, "Ciencia y Tecnología, binaria uno," 2016. [Fotografía]. Disponible en: https://pixabay.com/es/illustrations/binaria-uno-cyborg-cibern%C3%A9tica-1536651/ [Consultado en septiembre 26, 2019].

los planes de estudio en áreas específicas o la creación de nuevas carreras, cursos o diplomados que fortalezcan las habilidades y capacidades de los estudiantes para el campo laboral.

Respecto a los sofisticados programas de reconocimiento facial pueden emplearse de diversas maneras, las más usadas para la seguridad en las universidades, pero también puede usarse en procesos administrativo de inscripciones, tramites en general como los préstamos del acervo bibliotecario por medio de la seguridad biométrica.

En el área de las telecomunicaciones, servicio que se ha vuelto indispensable en las universidades, los administradores e implementadores de las Tecnologías de Información y Comunicación (TIC), con la IA pueden tener procesos automatizados para ver el análisis de consumo de datos, implementar ayudas para los usuarios de las TIC por medio de *bots* o asistentes virtuales, calcular los momentos en los cuales se pude hacer el mantenimiento preventivo o correctivo de los elementos de la red de telecomunicaciones, detección de fallas recurrentes y su autocorrección.

La mayoría de las aplicaciones de IA en las universidades se basan en la cuestión administrativa y de aprendizaje de los alumnos, como lo es el *chatbot* " Lola" implementado en la Universidad de Murcia, que ayuda a los estudiantes de nuevo ingreso durante el proceso de preinscripción y matrícula en dicha universidad. En el R.D Head Elementary School, en Lilburn Georgia Estados Unidos, se aplica el proyecto de *IBM Watson Education* en conjunto con *Sesame Workshop* llamado *Word Chow* para el aprendizaje individualizado de los niños de *kindergarden*. De igual manera, en colaboración con Pearson ha desarrollado otro programa llamado *Tutor Watson* que ayuda a profesores y alumnos en el aprendizaje incivilizado en 18 universidades en los Estados Unidos, como en Greenville Technical College, en Carolina del Sur. [4], [6]

Conclusión

LA IA TIENE MÚLTIPLES APLICACIONES EN EL QUEHACER UNIVERSITARIO, NO SOLO EN LO YA EXISTENTE sino también, en el desarrollo tecnológico y en las áreas de investigación de las universidades, lo que implicará tener recursos humanos capacitados (estudiantes, profesores, investigadores, etc.) para el uso y creación de estas nuevas tecnologías que ayuden no solo a la sociedad en general sino a las gestiones y procesos dentro de las universidades.

Dentro de la relevancia que puede tener la IA en los quehaceres humanos existen puntos de vista que plantean, previenen o ponen en alerta el uso inadecuado de la tecnología o el impacto que socioeconómicamente puede tener.

1. Se plantea por ejemplo, en el ámbito laboral, la disminución de empleos al tener mecanismos automatizados en diversos procesos, como por ejemplo los robots industriales, que realizan el trabajo de varias personas. El software con IA sustituye a los asistentes ejecutivos en los medios directivos donde llevar la agenda, contestar teléfonos, redactar oficios, sea un trabajo cotidiano para un software que está programado para hacer estas tareas desde la racionalidad, o que sustituya a personal que realiza trabajos de alto riesgo. La respuesta a estas inquietudes dependerán del tipo de sociedad que cada país decida, la IA para que sea un apoyo en las actividades humanas o bien una herramienta meramente económica. No debemos olvidar que uno de los principios de la creación de la IA es aumentar la capacidad de los humanos para resolver problemas, ser más prósperos y saludables.
2. Inteligencia general Artificial o singularidad: es el crecimiento acelerado de un fenómeno, pero en la IA se denomina singularidad cuando alcanza el nivel de razonamiento de la mente humana, se convierte en una Inteligencia Artificial Heurística, que se define como la solución de problemas en los cuales los resultados se descubren por la evaluación del progreso logrado en la búsqueda del resultado final. Esto acarrea que la máquina o el software empezarán a solucionar nuestras problemáticas antes que nosotros, lo cual puede ser lo que se llama "Final de la edad humana" o también llamado "La explosión de IA" entendiéndose por explosión como el fenómeno en donde se puede duplicar una cantidad una y otra vez (por ejemplo la explosión demográfica, donde la explosión es la cantidad de personas en un espacio geográfico) lo que conlleva a que algo pequeño se convierta

en algo muy grande, lo cual implica a que una pequeña inteligencia puede explotar hasta que supere a la inteligencia humana.

El dilema de ambos puntos de vista sobre la IA puede dejar de serlo al formar los recursos humanos en los diferentes ámbitos del quehacer humano como la ciencia, la tecnología, las humanidades y las artes. Esta formación debe estar con base en principios éticos y morales que se vean reflejados desde la concepción, diseño, comercialización, implementación y uso de las tecnologías que usen la IA. Con la visión y misión fundadas en estos principios las universidades pueden ser el espacio donde cada vez más surjan nuevas ideas para introducir las nuevas tecnologías que se basen en la IA en la vida social, económica, cultural, ecológica, de investigación y medicina de cada sociedad que conforman el mundo.

BIBLIOGRAFÍA

[1] R. Kurzweil, *La era de las maquinas inteligentes*, MIT PRESS, 1990.

[2] S. J. Russell y P. Norvig, I*nteligencia artificial un enfoque moderno*, 2.a ed., Madrid: Pearson Prentice Hall, 2004.

[3] Wantubi, Inteligencia Artificial, Ingeniería de Sistemas, Marzo 10, 2014. [En línea]. Disponible en: https://www.youtube.com/watch?v=vNZ5Gfr4SkI [Consultado en septiembre 20, 2019].

[4] Discovery Channel, Inteligencia Artificial IBM, Discovery Latinoamérica, Septiembre 26, 2018. [En línea]. Disponible en: https://www.youtube.com/watch?v=5rvZBsueMoc [Consultado en septiembre 20, 2019].

[5] J. Elias, "*Titanes de la Inteligencia Artificial - El Futuro de Ayer Hoy*," Octubre 2, 2017. [En Línea]. Disponible en: https://www.eliax.com/index.cfm?post_id=11524 [Consultado en septiembre 20, 2019].

[6] A. Pedreño, Interviewee, *La Inteligencia Artificial en las Universidades*. [Entrevista]. 24 abril 2019.

Cómo se cita:

Y.S. Martínez, "La Inteligencia Artificial en la transformación de procesos universitarios," *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 2, octubre, 2019. [En línea]. Disponible en: https://www.ties.unam.mx/ [Consultado en octubre, 2019].

PRINCIPALES ELEMENTOS PARA EL DISEÑO DE LA GOBERNANZA INSTITUCIONAL/ORGANIZACIONAL DE SEGURIDAD DE LA INFORMACIÓN

Xóchitl Díaz Pillado
https://www.ties.unam.mx/

Fecha de recepción: 29 de junio de 2019 • Fecha de publicación: octubre de 2019 Octubre de 2019 | número de revista 2 • ISSN 2683-2968

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

PRINCIPALES ELEMENTOS PARA EL DISEÑO DE LA GOBERNANZA INSTITUCIONAL/ ORGANIZACIONAL DE SEGURIDAD DE LA INFORMACIÓN

Resumen

El presente artículo tiene como objetivo brindar un panorama general de los principales elementos para el diseño de la Gobernanza Organizacional de Seguridad de la Información (SI). Para ello, en la introducción se aborda como punto de partida, los Servicios Críticos Institucionales, que deberán ser considerados en primera instancia, por ser los que constituyen la misión y visión de la organización. También se efectúa la diferencia entre *SI y Seguridad Informática*.
El breve contexto menciona distintas acciones que al respecto se han llevado a cabo internacionalmente y en México, tanto el sector público como privado.
Las consideraciones generales para el diseño de la gobernanza institucional/organizacional de SI, plantean cuestiones relativas a Arquitectura de Negocio, Arquitectura Tecnológica Empresarial, Grupo Estratégico de SI, Servicios Críticos Institucionales, Infraestructura Crítica, Gestión de Riesgos e Incidentes, entre otros.
En las Áreas de Oportunidad, se enuncian las principales para las IES, como: SGSI, SGCN, cumplimiento legal y regulatorio, oferta educativa en SI y ciberseguridad para todos los niveles y áreas del conocimiento, vinculación y fomento a la cultura de SI.
Finalmente la conclusión plantea la corresponsabilidad que atañe a toda la organización que requerirá a su vez mayor colaboración interinstitucional.

Palabras clave:

Seguridad de la información, gobernanza, continuidad del negocio, servicios críticos institucionales – infraestructura crítica.

ORGANIZATIONAL SECURITY INFORMATION GOVERNANCE DESIGN: KEY ELEMENTS

Abstract

This paper aims to show a general scope of the key elements to design the Organizational Information Security Governance. To do so, in the introduction we can see why the organization critical services must be taken into account first of all. That is because those services are essential to accomplish the organization's mission and vision: its core. In this section, we contrast Information Security and Information Technology Security concepts.
As a brief context, we can find different Information Security international actions taken in one hand, and in México on the other hand, in both public and private sectors.
In order to achieve the Organizational IS Governance Design, in the general considerations section we talk about Business Architecture, Technological Business Architecture, Information Security Committee, Organizational Critical Services, Critical Infrastructure, Risk and Incident Management, etc.
In the main areas of opportunity section we have listed several of which might be of importance to the purposes of the Universities, such as: ISMS, BCMS, Law and Regulatory Enforcement, IS and Cybersecurity Academic Programs for every level and field of knowledge, IS and Cybersecurity Research and Development, IS and Cybersecurity Collaboration agreements and IS awareness.
Finally, in the conclusion section we strongly recommend that the Information Security can no longer be considered as a single area's responsibility, but as a shared commitment across the whole organization with inter-agency coordination and collaboration.

Keywords:

Information Security, governance, business continuity, critical organizational services, critical infrastructure.

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

PRINCIPALES ELEMENTOS PARA EL DISEÑO DE LA GOBERNANZA INSTITUCIONAL/ ORGANIZACIONAL DE SEGURIDAD DE LA INFORMACIÓN

Introducción

MANTENER LA CONTINUIDAD DE LA OPERACIÓN DE SUS SERVICIOS ESENCIALES REDUNDA EN EL MAYOR INTERÉS Y BENEFICIO DE TODA INSTITUCIÓN, así como también en beneficio no sólo de sus clientes o usuarios, ya sean internos o externos, sino de todas las partes interesadas a distintos niveles. Pero, ¿cómo saber bien a bien cuáles son esos servicios esenciales?, ¿qué hacer para identificarlos? Como punto de partida deben ser considerados los que constituyan la razón de ser de la organización; es decir, aquellos que sean cruciales para cumplir su misión y proyectarla hacia su visión. Acto seguido, efectuar los siguientes cuestionamientos: de presentarse una degradación o interrupción en la prestación de dichos servicios (incidentes), ¿cuáles serían las consecuencias y cuál sería su gravedad?, ¿cuáles causas (vulnerabilidades, riesgos) podrían originar tal interrupción? Y ante ello, ¿qué implicaciones legales, regulatorias, normativas, contractuales, financieras, laborales, sociales, de reputación o confiabilidad, etc., habría que enfrentar?, ¿cuáles serían los peores escenarios que podrían presentarse?, ¿quiénes podrían conocer mejor las repercusiones que esa situación representaría y tener el panorama completo del grado de impacto que ocasionaría?, ¿qué rol juegan las áreas técnicas? Para poder responder tales preguntas y así como para poder pensar en continuidad, primero hay que considerar seriamente la Seguridad de la Información (SI) a nivel organizacional.

Aún hoy en día, al escuchar "Seguridad de la Información," la mayoría de las personas lo siguen asociando únicamente a "Seguridad Informática," refiriéndose en muchos casos a ambas como sinónimos y por ende, circunscribiéndolas al ámbito de las Tecnologías de la Información y Comunicación, o más coloquialmente, "a la gente de TI" por una parte y a las investigaciones de los académicos por otra. Sin embargo, la SI tiene un sentido mucho más amplio, cuyo fin está orientado a preservar en la mayor medida posible la *confidencialidad, integridad y disponibilidad* (Tríada CID) de la información institucional, sin perder de vista por supuesto, la trazabilidad y el no repudio de la misma desde la perspectiva de la Arquitectura de Negocio (organización), mientras que la Seguridad Informática se avoca al aspecto técnico desde el enfoque de la Arquitectura Tecnológica Empresarial. En este orden de ideas, la segunda es sólo uno de los subconjuntos que conforman a la primera. Para ilustrar lo anterior, abordaremos someramente algunos detalles.

Breve contexto

EN EL ENTORNO TANTO INTERNACIONAL COMO NACIONAL, CADA VEZ MÁS ORGANIZACIONES PERTENECIENTES AL sector público como al privado, están llevando a cabo el diseño, implementación y certificación del Sistema de Gestión de SI (SGSI) acorde a los estándares que integran la familia ISO 27000 [1]– vigente

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

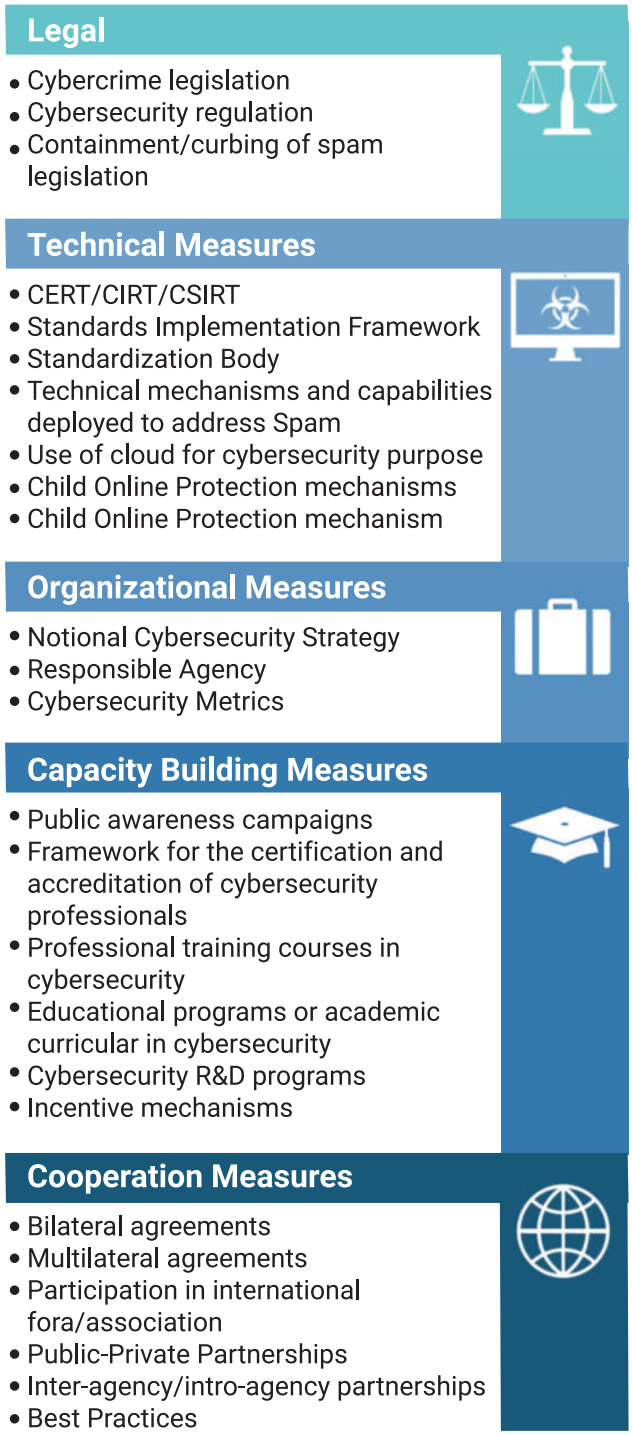

Figura 1.
" Pilares e indicadores," Global Cybersecurity Index 2018, [Fotografía]. Disponible en: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2018-PDF-E.pdf Consultado en agosto 29, 2019.]

(https://www.iso.org/isoiec-27001-information-security.html), e incluso obtienen la certificación del SGSI específicamente de conformidad a lo indicado en el estándar ISO/IEC 27001 Requerimientos del Sistema de Gestión de Seguridad de la Información [2] (SGSI) vigente.

Una vez que han obtenido la certificación de su SGSI, muchas empresas y/o instituciones dan el siguiente paso y se perfilan hacia la implementación de ISO 22317 Análisis de Impacto al Negocio [3] vigente, así como también a la implementación y certificación de ISO 22301 Requerimientos del Sistema de Gestión de Continuidad de Negocio [4] (SGCN) vigente.

Así mismo, existen diversas organizaciones internacionales que incluso desde la perspectiva del ámbito de la ciberseguridad (considerado de forma tradicional altamente técnico), han realizado estudios de relevancia con un enfoque interdisciplinario integral. Tal es el caso de la Unión Internacional de Telecomunicaciones (*International Telecommunication Union, ITU*), la cual publicó el Índice Global de Ciberseguridad (*Global Cybersecurity Index, GCI*) [5] en 2014, 2017 y 2018. El GCI es una referencia confiable que mide el compromiso de los países respecto a la ciberseguridad a nivel global [6]. México se encuentra posicionado en el lugar 63 de 175 a nivel global y en el lugar 4 en el continente Americano. El índice está compuesto por cinco pilares de los cuales sólo uno se refiere al aspecto técnico, mientras que los cuatro restantes incluyen diversos ámbitos, entre los que se encuentran los legales, organizacionales, de construcción de capacidades y de cooperación (figura 1).

Por otra parte, en México durante la Administración Pública Federal 2013 – 2018, uno de los programas que estuvieron contenidos en su Plan Nacional de Desarrollo [7], fue el "Programa para un Gobierno Cercano y Moderno," que a su vez incluyó a la Estrategia Digital Nacional [8] de la cual se desprende el Manual Administrativo de Aplicación General en las Materias de Tecnologías de la Información y Comunicaciones, y en la de Seguridad de la Información, (MAAGTICSI), cuya modificación es vigente hasta la fecha de elaboración del presente texto. Fue publicada en el Diario Oficial de la Federación el 23 de julio de 2018 (figura 2).

Uno de los nueve procesos que conforman al MAAGTICSI es precisamente el Proceso de Administración de Seguridad de la Información (ASI).

Adicionalmente, del Programa Nacional de Seguridad Pública y del Programa para la Seguridad Nacional del

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

Figura 2.
"Estrategia Digital Nacional," 2018. [Fotografía]. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/380408/MAAGTICSI_compilado__20182208.pdf [Consultado en agosto 29, 2019.]

mismo Plan, se desprendió también a su vez la Estrategia Nacional de Ciberseguridad [9] (figura 3).

En ambas Estrategias Nacionales referidas, se encuentran dispuestas entre varias cuestiones, las relativas a SI, enfoque basado en Gestión de Riesgos y Respuesta a Incidentes, a ser implementadas en las instancias de la Administración Pública Federal.

Visto el panorama descrito en la introducción y en el contexto, ¿cómo proceder?

Consideraciones generales para el diseño de la gobernanza institucional/organizacional de SI

ES MENESTER RESALTAR QUE EL ACTIVO DE INFORMACIÓN MÁS IMPORTANTE DE CUALQUIER ORGANIZACIÓN, son las personas que la integran, dado que son ellas las que poseen el conocimiento necesario, el *know how* para que aquellas operen. El tema de SI es sumamente amplio y en esta ocasión, trataremos la parte organizacional. De acuerdo con lo indicado por el estándar ISO 27001 vigente, los servicios esenciales, prioritarios y de mayor criticidad para el negocio son los que deberán ser asegurados en primera instancia, por lo que deberán ser incorporados desde el primer ciclo de implementación de un **Sistema de Gestión de Seguridad de la Información (SGSI)**. Este alcance, como la definición de la **Política Institucional de Seguridad de la Información**, deberán ser establecidos por el nivel estratégico empresarial/institucional. Esto tiene mucho sentido si consideramos que, después de todo, ¿quién mejor que la alta dirección para saber con precisión los efectos de la detención parcial o total, durante varios minutos, horas o días de las actividades neurálgicas de la institución?

Tener siempre presente la misión, la visión y los objetivos institucionales, ayudará a identificar de forma inicial, aquellos servicios que guardan alineación estratégica y por ende, los procesos de negocio que los integran. De ahí, se deberá definir cuáles son servicios esenciales y cuáles son servicios de apoyo. Con el fin de formalizar la determinación de los servicios prioritarios, marcos arquitectónicos como SOA [10] (*Service Oriented Architecture*), **TOGAF®** [11] (*The Open Group Architecture Framework*), estándares internacionales como los citados ISO/IEC 27001 vigente, ISO 22301 vigente, ISO 22317 Análisis de Impacto al Negocio – BIA, *Business Impact Analysis* –, herramientas como el enfoque de procesos, el análisis FODA, las 5 Fuerzas de Porter [12], y *Balanced Score Card* [13] son de mucha utilidad.

Es factible expresar la Arquitectura Empresarial (o de negocio), en términos sumamente simplificados en dos capas: la capa 1 involucra a los servicios institucionales/

Figura 3.
"Estrategia Nacional de Ciberseguridad," 2017. [Fotografía]. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/271884/Estrategia_Nacional_Ciberseguridad.pdf [Consultado en agosto 29, 2019.]

organizacionales (críticos o no) y a los procesos de negocio que componen a los mismos. La capa 2 contiene lo relativo a la Arquitectura Tecnológica Empresarial, la cual soporta a su vez a la capa 1. En este sentido, la Arquitectura Empresarial contempla a la organización de forma holística. Ahora, observemos cómo se gestan las interacciones entre ambas capas. Partiendo prioritariamente de la determinación de los ya citados servicios críticos, se procede a identificar aquellos activos de información en los cuales dichos servicios "viven" o están instalados. Tales activos son por añadidura de igual forma críticos o esenciales. Precisando: un grupo de integrantes de la alta dirección, llamado por ejemplo, Grupo Estratégico de SI, **GESI** [14], son quienes determinan los servicios críticos y sus procesos de negocio (capa 1). Si esos servicios son proporcionados por aplicativos, sistemas, bases de datos etc., (software), que naturalmente se encontrarán hospedados en hardware, entonces tales activos de información se denominan infraestructuras críticas/ esenciales de información. Cabe mencionar que de forma general se puede distinguir entre dos tipos principales de infraestructura crítica: de cómputo (servidores, unidades de almacenamiento, etc.), y comunicaciones (elementos que componen las redes de datos y que proveen la conectividad con otras redes locales, externas y aquellos que proveen la conexión hacia Internet, como routers, switches, etc. y la cada vez más preponderante "nube"). En este orden de ideas, tanto el hardware como el software forman parte de la Arquitectura Tecnológica Empresarial (capa 2). Para que la capa 1 pueda operar sobre la capa 2, entran en acción distintos equipos multidisciplinarios (formados tanto por los responsables de los procesos de negocio como por especialistas técnicos) que dependerán del GESI, y que por tanto lo apoyarán para que de forma conjunta, identificar tales infraestructuras, que serán contempladas para realizar una Gestión de Riesgos con apego sugerido a ISO/IEC 27005 [15] vigente, lo que permitirá averiguar cuán vulnerable es y ante qué riesgos existe exposición a efecto de diseñar un Plan de Tratamiento de Riesgos, en el que se plasmen los controles de seguridad que sean aplicables. Los controles de seguridad deberán implementarse preferentemente con base en los ya incluidos en el Anexo "A" del estándar ISO/IEC 27001 vigente, con la intención de evitar en la mayor medida posible que esos riesgos se materialicen tanto en la operación del día a día como en las ventanas de mantenimiento y que en tal caso, llegaran a convertirse en Incidentes de SI, los cuales desde luego también implican llevar a cabo una correspondiente labor de gestión.

Es también menester resaltar que es de suma importancia mantener a la vista además de los activos/ infraestructura de información críticos tangibles, los activos de información intangibles, como el prestigio, la reputación y la confianza en la institución u organización, que por un incidente dado de SI, podrían verse seriamente afectados.

La Gestión de Riesgos es un proceso vivo que no se hace una vez al año. Si llegado el caso por no haber sido contemplados ciertos controles o bien porque el control o controles implementados no hayan sido del todo adecuados o hayan resultado insuficientes, es cuando pudiera materializarse algún o algunos riesgos afectando negativamente la SI y/o provocando el incumplimiento de lo establecido en la Política Institucional de SI. Es entonces cuando los riesgos dejan de serlo y se convierten en incidentes. Por lo tanto, será momento de iniciar la Gestión de Incidentes de SI, sugiriéndose pare ello el uso de ISO/IEC 27035-1 [16] vigente, Gestión de Incidentes de Seguridad de la Información – Parte 1: Principios de Gestión de Incidentes, ISO/IEC 27035-2 [17] vigente y Gestión de Incidentes de Seguridad de la Información – Parte 2: Guía de planeación y preparación para Respuesta a Incidentes. Así mismo, será recomendable considerar el estándar ISO/IEC 27031 [18] vigente, *Guidelines for information and communication technology readiness for business continuity* – un IRBC – lo que previamente se conocía como un Plan de Recuperación de Desastres – DRP, *Disaster Recovery Plan* – y que sí, en efecto, éste elemento a diferencia de los otros, es netamente técnico –. La Gestión de Incidentes entonces proporcionará información vital que retroalimentará a la Gestión de Riesgos y a los Controles de Seguridad (también llamados mecanismos de seguridad o salvaguardas), contribuyendo así a propiciar un ciclo de mejora continua.

Cabe aclarar que es de vital importancia realizar lo descrito mediante Gestión de Cambios o Cambios Administrados con un panorama global de acuerdo con las competencias respectivas tanto en capa 1 como en capa 2 (los riesgos e incidentes se presentan tanto en los procesos de negocio como en la parte técnica).

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

Áreas de oportunidad

LAS CIRCUNSTANCIAS PREVALECIENTES EN LA ACTUALIDAD Y LA CONSTANTE EXPANSIÓN DE NUESTRA COTIDIANIDAD en el llamado "ciberespacio," han dotado más que nunca de altos niveles de prioridad a la SI. Las actividades derivadas de esta hiperconectividad constituyen una tan vasta diversidad desde de las más sencillas situaciones individuales, hasta la complejidad que implica el desarrollo en el entorno mundial.

Conscientes o no aún de ello, ningún tipo de organización puede seguir manteniéndose ajena al hecho de cada vez un mayor número de países cuentan desde hace tiempo con implementaciones del SGSI, del SGCN, así como de Estrategias Nacionales de Ciberseguridad, prioridades todas en sus agendas nacionales al más alto nivel, que en consecuencia se propagan hacia todos los sectores.

Por tanto, habiendo expuesto el panorama abordado a lo largo de estas líneas, resulta inminente que las Instituciones de Educación Superior (IES) den prioridad dentro de sus Planes de Desarrollo Institucionales, a la inclusión de un eje o línea fundamental que se avoque a la SI Institucional, involucrando así proactivamente a su cúpula directiva en el diseño de estrategias que formalicen a cabalidad la importancia que ésta verdaderamente merece, con el fin de emprender acciones con un compromiso fehaciente para su implementación en toda la institución.

Hay que reconocer que en muchas acciones de gobernanza no se empieza con un lienzo en blanco, y en ese sentido el potencial con que ya cuentan las IES es enorme. Con el patrocinio y compromiso de la alta dirección se pueden capitalizar las iniciativas ya existentes en las mismas, para dejar de operar en silos y promover la interoperabilidad, con el fin de concatenar los esfuerzos aislados hasta el momento, de forma coordinada y articulada de manera que propicien la utilización de todos los recursos disponibles. Un ejemplo de ello son las matrices de riesgos desarrolladas con las metodologías propias de la institución y con apego a la normatividad que las rige. Todo esto puede verse enriquecido por las aportaciones, innovaciones y beneficios que generan algunas de las investigaciones de la academia relacionadas y que aplicadas en conjunto con la gobernanza y las buenas prácticas, creen una sinergia que de igual forma nutran a la mejora continua.

Así mismo y de acuerdo a su misión y visión, las IES pueden diseñar planes y programas de estudio que contengan asignaturas referentes a la SI que incidan no sólo en carreras afines a las TIC, sino en todas las áreas del conocimiento, mismas que de acuerdo a sus competencias, se incorporarán a un "*cibermercado*" laboral al que deberán incursionar de facto con un bagaje mínimo indispensable en la materia y que asciende vertiginosamente.

Con base en los planteamientos hasta aquí realizados, a continuación se enuncian una serie de rubros en las que seguramente las IES, tienen una vastedad de áreas de oportunidad:

- Acciones para iniciar con la implementación de un SGSI:
  - Realizar con alineación estratégica, la determinación de los Servicios Institucionales Críticos, Procesos de Negocio e Infraestructura Crítica.
  - Identificar los activos de información tangibles/intangibles con base en los Servicios Críticos Institucionales.
  - Establecer la Política Institucional de Seguridad de la Información, así como su alcance con base en los Servicios Críticos Institucionales determinados.
  - Designar, por parte del Titular de la Institución, al Responsable de la SI Institucional en el nivel jerárquico estratégico (*Chief Information Security Officer, CISO*).
  - Establecer por parte del Titular de la Institución en conjunto con el RSII, al Grupo (comité, equipo) Estratégico de SI Institucional que será coordinado por el RSII. Sus integrantes deberán pertenecer al nivel jerárquico estratégico y las áreas institucionales críticas.
  - Efectuar el aseguramiento de las islas locales (en unidades académicas y/o administrativas) de Infraestructura Crítica y de Desarrollo de Servicios Críticos Institucionales que se encuentren fuera de la cobertura de redundancia eléctrica y alta disponibilidad que ofrecen los Centros de Datos Institucionales.
  - Establecer Centros de Datos alternos al Centro de Datos Institucional.
  - Implementar medidas de seguridad física y lógica de la infraestructura crítica.

- Implementar repositorios seguros de configuraciones de la Infraestructura Crítica Institucional.
- Establecer marcos de referencia – *Frameworks* – Institucionales para el desarrollo seguro de sistemas, aplicaciones, etc.
- Efectuar la Gestión de Riesgos de SI.
  - Capitalizar las iniciativas y proyectos institucionales ya existentes, como las Matrices de Riesgos Institucionales.
- Efectuar la Gestión de Incidentes de SI IRBC (antes DRP), Planes de Comunicación y Escalamiento.
  - Crear *CERT's* o *CSIRT's* académicos institucionales.
- Efectuar la Gestión del Conocimiento en Ciberseguridad recabado por la institución.
- Efectuar la Gestión del Talento en Ciberseguridad formado por la institución.
- Realizar Acuerdos de Confidencialidad de SI en concordancia con las funciones y responsabilidades de directivos, responsables y comunidad en general de la institución, así como de sus distintas partes interesadas.
- Establecer, en conjunto con las áreas críticas institucionales y el área de capital humano, los procesos y procedimientos para la incorporación/separación segura del empleo.
- Realizar auditorías periódicas de SI.
- Tomar en consideración los estudios realizados para universidades (por ejemplo, el elaborado por la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), en el Estado Actual de las Tecnologías de la Información y las Comunicaciones en las Instituciones de Educación Superior / Estudio 2018 [19], cuyo capítulo cinco es relativo a SI).

– Acciones para iniciar con la implementación de un SGCN:
  - Establecer medidas para preservar en el mayor grado posible la seguridad de las personas que conforman la institución.
  - Implementar planes para la Continuidad de los Servicios Institucionales Críticos.
  - Efectuar el Análisis de Impacto al Negocio, tomando primordialmente en consideración el correspondiente a los Servicios Críticos Institucionales.
  - Realizar simulacros de ataques y ciberataques por agentes internos y/o externos (*insiders/outsiders*) a la institución.
– Acciones generales para el cumplimiento legal y regulatorio:
  - Observar leyes y reglamentos aplicables según el caso, relativos a SI (por ejemplo en cuanto a Protección de Datos Personales, clasificación y resguardo de información, preservación de documentos electrónicos, etc.)
  - Observar la Normatividad Institucional establecida con relación a la SI en caso de contar con alguna (por ejemplo en materia de gestión de riesgos, etc.)
– Acciones generales que podrían ser incorporadas en la oferta educativa a nivel Medio Superior, Superior, Posgrado e Investigación en SI y Ciberseguridad:
  - Integrar formación en SI y Ciberseguridad en planes y programas de estudios a nivel Medio Superior, Superior y Posgrado en todas las áreas de conocimiento (Médico Biológicas, Sociales Administrativas, Físico-Matemáticas, Ingenierías, etc.), acorde a dichas áreas.
  - Difundir la oferta en formación en SI y Ciberseguridad de las distintas universidades.
  - Conformar y difundir (en caso pertinente) un padrón de investigadores y proyectos de investigación en SI y Ciberseguridad.
– Acciones generales para la vinculación Academia – Investigación – Industria:
  - Generar mecanismos de vinculación, convenios de colaboración, etc., nacionales e internacionales con universidades y organismos público-privados en SI y Ciberseguridad.
– Acciones generales para establecer Programas de Fomento a la Cultura de SI
  - Crear programas de sensibilización y concientización a la comunidad en general para fomentar la cultura de la SI y Ciberseguridad Institucional.
  - Generar capacidades y formación: Cursos/Diplomados y capacitación especializada en Seguridad de la Información y Ciberseguridad.

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

¿Por dónde comenzar?

AQUÍ LOS PRIMEROS PASOS RECOMENDADOS PARA INICIAR CON EL ESTABLECIMIENTO DE LA GOBERNANZA Institucional de SI con perspectiva incremental:

Paso	Responsable	Descripción
1. Designación del Responsable de SI Institucional (RSII), *Chief Information Security Officer* (CISO) y establecimiento del Grupo Estratégico de SI.	Titular de la Institución	El RSII y los miembros del GESI deberán pertenecer al nivel jerárquico estratégico y las áreas institucionales críticas. El RSII coordinará al GESI.
2. Determinación de los Servicios Críticos Institucionales.	GESI	El RSII deberá presentarlos al Titular de la Institución para su consideración y en su caso para la aprobación del titular.
3. Definición de la Política Institucional de Seguridad de la Información y su alcance.	GESI	Con base en los Servicios Críticos Institucionales aprobados, definir la Política y presentarla al área jurídica institucional y al Titular de la Institución para su consideración y en su caso aprobación. Una vez autorizada, la Política deberá ser publicada y difundida a toda la comunidad y partes interesadas.
4. Establecimiento y coordinación de un equipo de trabajo de apoyo para el GESI.	GESI	El equipo de trabajo con base en los Servicios Críticos Institucionales, deberá determinar los procesos de negocio que los integran así como la Infraestructura Crítica que los soportan u hospedan. Tomar en consideración las interdependencias correspondientes de ambos rubros e implementar un Repositorio Seguro de Configuraciones para dichas infraestructuras. Cabe resaltar que el citado equipo deberá ser multidisciplinario, y que sus miembros deberán ser especialistas de los procesos de negocio, así como especialistas técnicos.
5. Gestión de Riesgos e Incidentes de SI	Equipo de trabajo de apoyo	Con base en los Servicios Críticos Institucionales, sus procesos de negocio y su Infraestructura Crítica, realizar ambas gestiones.
6. Evaluación y mejora continua.	GESI	Implementar un programa de evaluación para realizar la retroalimentación y mejora continua de los cinco pasos anteriores.

Tabla 1.
Seis primeros pasos recomendados para iniciar una implementación de un SGSI.
Fuente: elaboración propia.

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

Es así como en el presente documento brevemente se han abordado las áreas de oportunidad más neurálgicas. Sin lugar a dudas, existen muchas más en el amplísimo campo de acción de las IES, las cuales dependerán de las particularidades, concepción, grado de concientización, necesidades, nivel de madurez, etc., que cada una de ellas guardan con respecto a la SI y la ciberseguridad como una estrategia holística institucional.

Conclusión

NO ES SOSTENIBLE CONTINUAR DELEGANDO LA SI INSTITUCIONAL SOLAMENTE A LAS ÁREAS TÉCNICAS Y OPERATIVAS. Es un trabajo conjunto que nace de la perspectiva estratégica y que permea a todas las áreas y niveles de la organización. Siendo una corresponsabilidad que trasciende los confines de la misma al incorporarse en un frente común, mediante su participación a través de la suma de sus capacidades y talentos, en estrategias e iniciativas de colaboración interinstitucional en la que las alianzas público-privadas serán cruciales.

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

BIBLIOGRAFÍA

[1] International Organization for Standardization, "ISO/IEC 27001 Information security management" *International Organization for Standardization*, 2019. [En línea]. Disponible en: https://www.iso.org/isoiec-27001-information-security.html. [Consultado en septiembre 22, 2019].

[2] International Organization for Standardization, "ISO/IEC 27001:2013 Information technology — Security techniques — Information security management systems — Requirements" *International Organization for Standardization*, 2013. [En línea]. Disponible en: https://www.iso.org/standard/50054.html. [Consultado en septiembre 22, 2019].

[3] International Organization for Standardization, "ISO/TS 22317:2015 Societal security — Business continuity management systems — Guidelines for business impact analysis (BIA)" *International Organization for Standardization*, 2015. [En línea]. Disponible en: https://www.iso.org/standard/54534.html. [Consultado en septiembre 22, 2019].

[4] International Organization for Standardization, "ISO 22301:2012 Societal security — Business continuity management systems — Requirements" *International Organization for Standardization*, 2012. [En línea]. Disponible en: https://www.iso.org/standard/50038.html. [Consultado en septiembre 22, 2019].

[5] International Telecommunication Union, "Global Cibersecurity Index" *International Telecommunication Union*, 2019. [En línea]. Disponible en: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2018-PDF-E.pdf. [Consultado en agosto 29, 2019].

[6] International Telecommunication Union, "Global Cibersecurity Index" *International Telecommunication Union*, 2019. [En línea]. Disponible en: https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx. [Consultado en agosto 29, 2019].

[7] Gobierno de los Estados Unidos Mexicanos, "Plan Nacional de Desarrollo 2013-2018" *Diario Oficial de la Federación,* 2013. [En línea]. Disponible en: http://www.dof.gob.mx/nota_detalle.php?codigo=5299465&fecha=20/05/2013. [Consultado en agosto 29, 2019].

[8] Gobierno de los Estados Unidos Mexicanos, "ACUERDO por el que se modifican las políticas y disposiciones para la Estrategia Digital Nacional, en materia de Tecnologías de la Información y Comunicaciones, y en la de Seguridad de la Información, así como el Manual Administrativo de Aplicación General en dichas materias" *Diario Oficial de la Federación*, 2018. [En línea]. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/380408/MAAGTICSI_compilado__20182208.pdf. [Consultado en agosto 29, 2019].

[9] Gobierno de los Estados Unidos Mexicanos, "Estrategia Nacional de Ciberseguridad" *Gobierno de los Estados Unidos Mexicanos*, 2017. [En línea]. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/271884/Estrategia_Nacional_Ciberseguridad.pdf. [Consultado en agosto 29, 2019].

[10] W.R. Schulte, "Five Principles of SOA in Business and IT". Gartner, 2006.

[11] The Open Group. "The TOGAF® Standard" *The Open Group*, 2019. [En línea]. Disponible en: https://www.opengroup.org/togaf. [Consultado en agosto 29, 2019].

[12] S. Michaux, and A. C. Cadiat, *Porter's Five Forces: Understand competitive forces and stay ahead of the competition (Management & Marketing Book 1)*, 50minutes.com, 2015. [E-book] Disponible en: Amazon Kindle Edition.

[13] R. Kaplan, and D. Norton, "Balance Score Card". Harvard Business Review, 1992.

[14] Gobierno de los Estados Unidos Mexicanos, "ACUERDO por el que se modifican las políticas y disposiciones para la Estrategia Digital Nacional, en materia de Tecnologías de la Información y Comunicaciones, y en la de Seguridad de la Información, así como el Manual Administrativo de Aplicación General en dichas materias" *Diario Oficial de la Federación*, 2018. [En línea]. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/380408/MAAGTICSI_compilado__20182208.pdf. [Consultado en agosto 29, 2019].

[15] International Organization for Standardization, "ISO/IEC 27005:2018 Information technology — Security techniques — Information security risk management" *International Organization for Standardization*, 2018. [En línea]. Disponible en: https://www.iso.org/standard/75281.html. [Consultado en agosto 29, 2019].

[16] International Organization for Standardization, "ISO/IEC 27035-1:2016 Information technology — Security techniques — Information security incident management — Part 1: Principles of incident management" *International Organization for Standardization*, 2016. [En línea]. Disponible en: https://www.iso.org/standard/60803.html. [Consultado en agosto 29, 2019].

[17] International Organization for Standardization, "ISO/IEC 27035-2:2016 Information technology — Security techniques — Information security incident management — Part 2: Guidelines to plan and prepare for incident response" *International Organization for Standardization*, 2016. [En línea]. Disponible en: https://www.iso.org/standard/62071.html. [Consultado en agosto 29, 2019].

[18] International Organization for Standardization, "ISO/IEC 27031:2011 Information technology — Security techniques — Guidelines for information and communication technology readiness for business continuity" *International Organization for Standardization*, 2011. [En línea]. Disponible en: https://www.iso.org/standard/44374.html. [Consultado en agosto 29, 2019].

[19] Asociación Nacional de Universidades e Instituciones de Educación Superior, *Estado Actual de las Tecnologías de la Información y las Comunicaciones en las Instituciones de Educación Superior / Estudio 2017*. México: ANUIES, 2017.

Cómo se cita:

X, Díaz Pillado, "Principales elementos para el diseño de la Gobernanza Institucional/Organizacional de Seguridad de la Información," *TIES, Revista de Tecnología e Innovación en Educación Superior*, n.o. 2, octubre, 2019. [En línea]. Disponible en: https://www.ties.unam.mx/ [Consultado en octubre, 2019].

SISTEMAS DE GESTIÓN DE INFORMACIÓN Y EVENTOS DE SEGURIDAD (SIEM)

Manuel Ignacio Quintero Martínez
Sergio Anduin Tovar Balderas
https://www.ties.unam.mx/

Fecha de recepción: 2 de septiembre de 2019 • Fecha de publicación: octubre de 2019 Octubre de 2019 | número de revista 2 • ISSN 2683-2968

SISTEMAS DE GESTIÓN DE INFORMACIÓN Y EVENTOS DE SEGURIDAD (SIEM)

Resumen

Hoy en día, la evolución de las amenazas que existen hacia una red (en términos de seguridad de la información), han hecho que su detección se vuelva más compleja, por lo que se requiere de sistemas especializados en monitorización que ayuden a detectarlas, no sólo desde los dispositivos a los que afectan, sino a otros que pueden dar información complementaria. Un Sistema de Gestión de Información y Eventos de Seguridad (SIEM, por sus siglas en inglés) permite la integración de información a partir de eventos reportados por diversos dispositivos, correlacionando los mismos para emitir alertas y reportes que permitan tomar acciones que ayuden a proteger la confidencialidad, integridad y disponibilidad de la información en la organización. En este artículo se presentan las generalidades, conceptos y requerimientos a tomar en cuenta para la implementación de un SIEM.

Palabras clave:
Seguridad de la información, correlación, eventos de seguridad, SIEM.

SECURITY INFORMATION AND EVENT MANAGEMENT (SIEM)

Abstract

Currently, detection of information security threats in any network has been complexed, and requires specialized monitoring systems to detect them, not only from the devices of interest, also from devices that can give complimentary information. A Security Information an Event Management (SIEM) integrates information from reported events from different devices, correlating them to create alerts and reports, which can be used to take some actions to support organizational information' Confidentiality, Availability and Integrity. This paper presents some general information, concepts and requirements to be considered into a SIEM implementation.

Keywords:
Information security, correlation, security events, SIEM.

TiES Revista de Tecnología e Innovación en Educación Superior

Sistemas de Gestión de Información y Eventos de Seguridad (SIEM)
Manuel Ignacio Quintero Martínez
Sergio Anduin Tovar Balderas
Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

SISTEMAS DE GESTIÓN DE INFORMACIÓN Y EVENTOS DE SEGURIDAD (SIEM)

Introducción

ACTUALMENTE, LA INFRAESTRUCTURA TECNOLÓGICA DE CUALQUIER ORGANIZACIÓN, independientemente de su tamaño, se encuentra expuesta a múltiples amenazas a su seguridad. Estas evolucionan de forma acelerada y constante, rediseñándose para no ser detectadas, por lo que muchas técnicas de evasión suelen lograr su cometido cuando se monitorea lo que sucede en cada punto de la red organizacional individualmente. Pero ¿qué sucede cuando estas amenazas dejan evidencias que son trazables en su paso por una red, incluso antes de alcanzar su objetivo, generando alertas que permitan a los administradores tomar acciones preventivas o correctivas? Este es justamente el propósito de los sistemas de gestión de información y eventos de seguridad (SIEM, *Security Information and Event Management*, por sus siglas en inglés), los cuales buscan obtener la mayor cantidad de información útil posible para ayudar al monitoreo desde diversos frentes en una infraestructura.

Un SIEM permite recopilar y correlacionar eventos e incidentes de diversas fuentes (como servidores, dispositivos de red, dispositivos de seguridad perimetral, entre otros) para poder realizar predicciones a través de la identificación de una secuencia de acciones elementales que permitan reconocer alguna estrategia de ataque [1], para así prevenir o reaccionar rápidamente ante la materialización de una amenaza de seguridad [2], ayudando a reducir las intrusiones de seguridad en la infraestructura de 76% de las organizaciones encuestadas por CyberSecurity Insiders [3].

Origen y uso de la información en una infraestructura tecnológica

REGULARMENTE LOS ADMINISTRADORES DE UNA RED CONOCEN EL DISEÑO Y FUNCIONAMIENTO de esta; la información sobre el desempeño y qué sucede dentro de la misma, podría permitirles reaccionar ante una posible falla o intrusión de seguridad por lo que, deberían de conocerse en el menor tiempo posible algunos aspectos como:

- Cuántos dispositivos importantes existen en la red y cuáles deberían ser monitorizados.
- Cuánto se sabe acerca de lo que sucede en cada uno de esos dispositivos.
- De las alertas que puede generar un dispositivo, cuántas son realmente monitoreadas por el personal a cargo.
- Cuánta información generada puede ayudar a mejorar la seguridad.

Responder a estos aspectos resultaría un serio inconveniente para cualquier administrador de red con tantos servicios, conexiones, usuarios, clientes, etc. Podría parecer imposible conocer lo que sucede en ellos, incluso, ¿qué sucede si esa información ya es recopilada

Figura 1.
G. Altmann, "Protección a la información en línea," 2016. [Fotografía]. Disponible en: https://pixabay.com/es/illustrations/binaria-castillo-protecci%C3%B3n-1538721/ [Consultado en septiembre 18, 2019].

de forma central en un repositorio y se generan alarmas sólo cuando se cumple con un parámetro establecido? Es por ello que se vuelve poco viable pensar que esta implementación podrá informar sobre cualquier incidente de seguridad, ya que las amenazas actuales son capaces de evadir las formas de detección más comunes y conocidas.

Esto enfrenta a los administradores a las siguientes complicaciones:

- La información obtenida podría ser insuficiente por sí misma para detectar la mayor cantidad de incidentes de seguridad.
- La información aislada puede no ser óptima e incompleta para un dispositivo en específico o incluso, manipulada por un adversario.

Estos dilemas se vuelven cada vez más complejos y conforme se han desarrollado diversas formas de afrontarlos, la manera más efectiva de hacerlo es a través de la automatización (*operación, monitorización y revisión*) del seguimiento de eventos de seguridad en la infraestructura tecnológica, siendo los SIEM, los que han mostrado mejores resultados [2].

Funciones de un SIEM

LOS SIEM INTEGRAN LAS CAPACIDADES DE DOS SISTEMAS PRECEDENTES:

Algunas de las aplicaciones de la IA son: [2], [4], [5].

- Sistema de Gestión de Información de Seguridad (SIM, *Security Information Management*, por sus siglas en inglés), encargado de la recolección de eventos de seguridad para realizar trazas y reportes sobre estos.
- Sistema de Gestión de Eventos de Seguridad (SEM, *Security Event Management*, por sus siglas en inglés), encargado de la monitorización de los eventos en tiempo real, así como la gestión de incidentes derivados de estos aunque permitiendo la respuesta sólo de aquellos eventos con patrones preestablecidos[4].

Los SIEM integran las capacidades de los SIM y SEM, siguiendo el principio básico en donde la información que puede considerarse relevante para una infraestructura no procede de una sino de múltiples fuentes y esta

información puede concentrarse y correlacionarse para encontrar patrones más robustos que permitan detectar posibles problemas de seguridad en tiempo real, o al menos en periodos muy cortos [2], pudiendo variar en su mayoría entre minutos y segundos [3]. Estos sistemas son utilizados actualmente en todos tipos y tamaños de organizaciones y existen en general, tres opciones que permiten implementar un SIEM para monitorizar una red. En términos generales son:

- Versiones propietarias en sitio o administradas por un tercero en la nube.
- Versiones gratuitas con restricciones sobre el uso.
- Productos de distribución libre.

Por otro lado, algunos de los SIEM que existen actualmente en el mercado son los siguientes (se menciona su sitio web y tipo de licenciamiento):

- Solarwinds https://www.solarwinds.com/, Licencia comercial.
- Splunk Enterprise Security https://www.splunk.com, Licencia comercial y versión gratuita con restricciones.
- LogRhythm NextGen SIEM https://logrhythm.com, Licencia comercial y versión gratuita con restricciones.
- IBM Qradar https://www.ibm.com/security/security-intelligence/qradar, Licencia comercial.
- Alienvault Unified Security Management https://www.alienvault.com/products, Licencia comercial.
- Alienvault OSSIM https://www.alienvault.com/products/ossim, Licencia libre, con restricciones de uso.

Es importante tomar en cuenta que un SIEM (como muchas otras herramientas de monitoreo), requieren una etapa de aprendizaje; es decir, en un principio comenzarán a recopilar datos de sus diferentes fuentes para poder reconocer lo que deberán tratar como un comportamiento normal de los equipos que monitorea y posteriormente comenzar a emitir alertas de patrones que podría considerar anómalos.

Desarrollo

Implementación de un SIEM

UN DISPOSITIVO GENERA BITÁCORAS EN SU PROPIO FORMATO Y ESTAS NO SON IGUALES ENTRE FABRICANTES, DE tal manera que el trabajo más importante antes que el SIEM comience a correlacionar eventos, es poner estos en un formato común para finalmente operarlos como unidades básicas, tras las cuales todas tengan la misma estructura básica. Este proceso se llama *normalización* [5].

Los eventos pueden proceder de múltiples fuentes: Montesino, Perurena, Baluja y Porvén [6] mencionan que pueden recopilarse por cuatro medios principales:

- Recepción de una cadena de datos en formato *syslog*[1] proveniente de la fuente de datos.
- Aplicaciones agentes instaladas directamente en los dispositivos a monitorear.
- Invocación de la interfaz de línea de comandos de los sistemas monitoreados.
- Interfaces de programación de aplicaciones (API) provistas por los desarrolladores de los sistemas monitoreados.

El proceso de normalización puede ser realizado por el mismo SIEM pues muchas implementaciones ya cuentan con los protocolos adecuados para procesar los eventos de los sistemas operativos o dispositivos de red, más comúnmente usados desde un conector desarrollado por el mismo proveedor. Inclusive, estos conectores pueden ser creados por el usuario para ser adaptados a sistemas específicos que los administradores requieran para la monitorización de eventos o sistemas específicos.

Con las bitácoras normalizadas, el siguiente paso se centra en el almacenamiento de la información procedente de cada dispositivo, lo que se realiza generalmente en el mismo sistema o también fuera de este, pero no es una práctica común y ocurre regularmente por requerimientos específicos de la organización que lo implementa o de regulaciones a las que se encuentra sujeta.

Aunque comúnmente los SIEM ya cuentan con un registro de posibles comportamientos anómalos en una red o de amenazas bien conocidas, es importante que este sistema se adapte de forma específica a la infraestructura que monitorizará; además deberá tener un periodo de aprendizaje al que suele llamarse "modo monitor." En este punto es importante entender que para todo proceso de aprendizaje se requiere de un refinamiento, regularmente tiene que ser realizado por personal que conoce el entorno en el que se encuentra el equipo y pueda determinar que alertas corresponden a falsos positivos.

1 Servicio de recopilación de eventos sin ser estandarizados o normalizados en un servidor local.

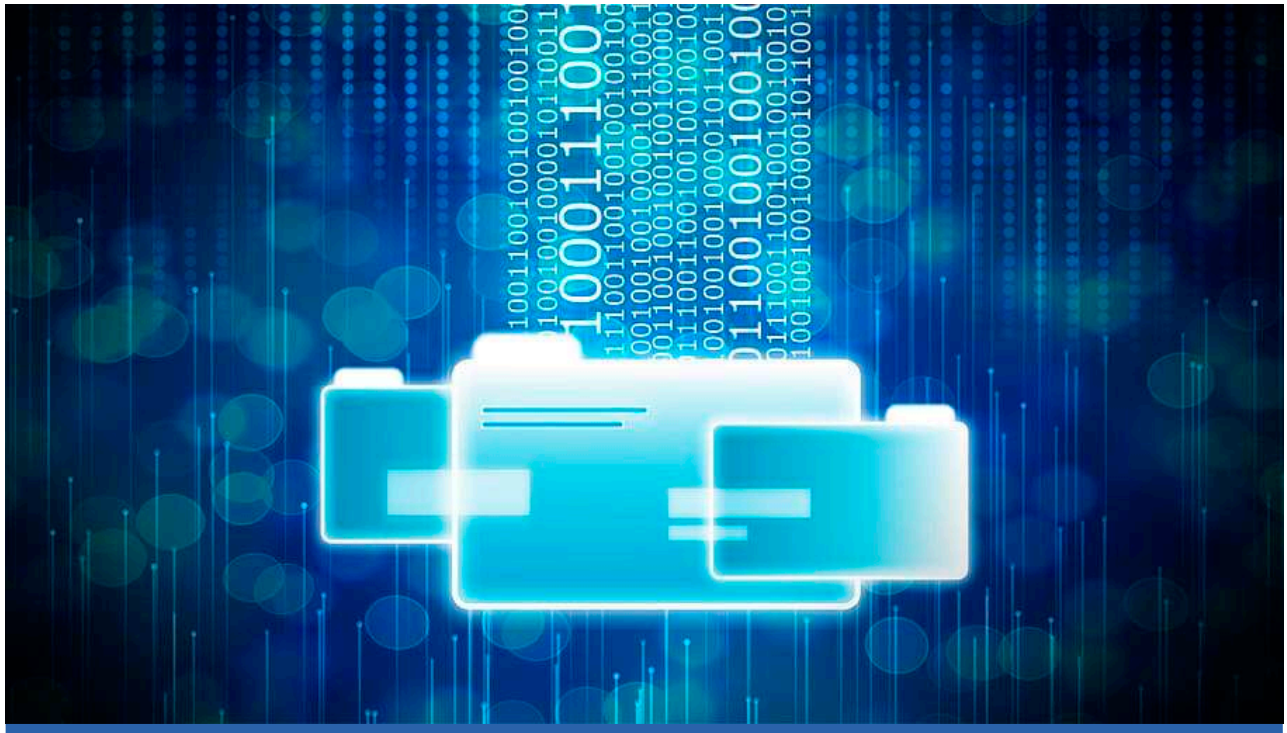

Figura 2.
"Carpeta de Datos," [Fotografía]. Disponible en: https://www.piqsels.com/en/public-domain-photo-zbgvd [Consultado en septiembre 20, 2019].

Por ejemplo, si un equipo genera tráfico de escaneo para la búsqueda de puertos abiertos hacia una red externa, regularmente generará una alerta. Sin embargo, si dentro de la organización se realizan pruebas de penetración regulares como parte de los procesos de aseguramiento de la red, este comportamiento debería ser aceptable desde ciertos equipos o direcciones hacia otro grupo o segmento de red, así como en horarios preestablecidos para evitar falsos positivos.

Tras algún tiempo en este modo con los datos recopilados y afinados, el SIEM será capaz de reconocer el comportamiento regular de la red en la que se ha implementado[7], por lo que posteriormente (y no desde el inicio) podrá comenzar a generar alertas de actividades anómalas.

En este punto, el SIEM puede comenzar a correlacionar la información a partir de patrones conocidos y aún más importante, de otros aprendidos o en algunos casos, también heurísticos y emitir alertas por medio de correo electrónico o mensajes de texto, entre algunas opciones.

Mas allá de las alertas

AL CONTAR CON INFORMACIÓN GENERAL DE LOS ACTIVOS IMPORTANTES, UN SIEM TAMBIÉN PUEDE AYUDAR a reconocer la vida general de los dispositivos que se encuentran dentro de una misma red. Para ello, es común que estos sistemas incluyan una interfaz gráfica que permita monitorear en tiempo real o de forma histórica el desempeño de algunas métricas establecidas de interés para los administradores por medio de la generación de reportes e incluso para la dirección de la organización, pudiendo en algunos casos, personalizar el tipo de panel o paneles al que cierto usuario puede acceder.

También es importante pensar que algunos estándares, como PCI DSS[2] o HIPAA[3], requieren ciertas métricas de monitoreo para los sistemas e infraestructura dentro

2 PCI DSS (*Payment Card Industry Data Security Standard*), es un estándar internacional de protección a datos tarjetas de pago. Más información en https://www.pcisecuritystandards.org
3 HIPAA (*Health Insurance Portability and Accountability Act*) es una legislación estadounidense que regula el resguardo de la información médica de pacientes, así como su privacidad y seguridad. Más información en https://www.hhs.gov/hipaa/index.html

del alcance de cada uno. La mayor parte de los fabricantes de SIEM incluyen de forma predeterminada herramientas o complementos que permiten la implementación de estos controles dentro de sus sistemas.

Factibilidad de la implementación de un SIEM

Sin duda, es difícil establecer si esta factibilidad existe, ya que depende de cada organización de forma individual, pero podríamos decir en general que contar con un SIEM brinda la posibilidad de poder contar con información en la mayor parte de los casos, de forma casi instantánea sobre lo que sucede en la red. Sin embargo, es importante establecer cuál será la intención de generar esta información; si no se cuenta o se planea tener una forma de responder a las alertas, a los eventos o la forma de generar y usar los reportes que los SIEM ofrecen, este sistema no ayudará a la mejora de la seguridad en la organización.

De hecho, la respuesta a las alertas deben darse tan rápida y eficientemente como sea posible. Como lo señala 451 Research [8], uno de los puntos clave para el éxito en la implementación de un sistema de correlación de eventos es la integración con flujos de trabajo automatizado, ayudando así a la mejora de la seguridad de la información.

Figura 3.
Pixabay, "Unidad de Control Computadora," 2018. [Fotografía]. Disponible en: https://www.pexels.com/photo/airport-business-cabinets-center-236093/ [Consultado en septiembre 18, 2019].

Eficiencia de un SIEM

Como cualquier herramienta o técnica, un SIEM no previene o evita la totalidad de los problemas de seguridad existentes, por lo que su efectividad se puede medir en términos de un porcentaje, el cual depende no sólo de la herramienta técnica utilizada, sino también de cómo se haya configurado la misma. La experiencia en la atención de incidentes del equipo que la ha puesto a punto y asumir que siempre existirán nuevas amenazas que no sean conocidas o detectables en la red. Por ejemplo, Morteza Zeinali [2] menciona que algunos sistemas eficientes pueden reportar eventos en algo muy cercano al tiempo real, con una eficiencia del 90% dentro de los 60 segundos desde que se generó el incidente de seguridad.

Sin embargo, un estudio de *Cybersecurity insiders*, mostró que 86% de las organizaciones que tienen un SIEM como parte de su estrategia de seguridad se encuentran satisfechos con su efectividad, mencionando que esta satisfacción se centra en tres elementos [3]:

- Detección y respuesta de incidentes más rápida.
- Mayor eficiencia en las operaciones de seguridad.
- Mejora de la visibilidad de amenazas.

A su vez, esta misma encuesta mostró los tipos de ataques que fueron eficientemente detectados:

- Acceso no autorizado (46%).
- Amenazas avanzadas persistentes / ataques dirigidos (42%).
- Ataques internos (maliciosos o por descuido 37%)

Por otro lado, en cuanto a la penetración de esta tecnología en instituciones superiores en México, la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES) en su encuesta ANUIES-TIC 2018 [9] reporta que los SIEM sólo forman parte de mecanismos de protección de la infraestructura en 24% de ellas, y aunque aún es un porcentaje bajo en comparación con otros sectores, mostró un crecimiento del 17% del año anterior, lo que muestra que este tipo de tecnología ha comenzado a ser parte de la estrategia de seguridad.

Conclusión

Una frase común en seguridad de la información, dice que no se trata de si habrá un problema de seguridad (es casi un hecho de que éste ocurrirá en algún momento en cualquier organización), sino cómo se

Octubre 2019 • Número 2 • ISSN 2683-2968
https://www.ties.unam.mx/

responderá cuando ocurra. Es aquí donde un SIEM puede cobrar relevancia para cualquier organización. Si se tiene una implementación exitosa, permitirá contar con información ágil y certera para la detección y resolución de incidentes.

El SIEM sin duda, requerirá de un proceso de planeación para su implementación, así como del aprendizaje y refinamiento que le permita identificar de forma puntual que comportamiento es normal para poder emitir alertas sobre conductas que salgan de este, lo cual puede requerir una inversión no sólo de recursos financieros, sino de tiempo y conocimientos previos de los administradores de la infraestructura tecnológica. Sin duda redituará en la detección, respuesta e inclusive la posible prevención de incidentes de seguridad.

Sin embargo, es importante pensar que este SIEM no puede ser pensado como un medio para eliminar problemas de seguridad, sino como el disparador para iniciar métodos que los solucionen, ya sea por algún otro proceso automatizado, la intervención de un administrador de los recursos, o incluso la intervención de un equipo de respuestas a incidentes, así como de otros relacionados a la mejora continua y optimización de los recursos tecnológicos.

BIBLIOGRAFÍA

[1] E. Anumol, "Use of machine learning algorithms with SIEM for attack prediction," *Intelligent Computing, Communication and Devices Springer*, 2015.

[2] S. M. Zeinali, "Analysis of security information and event management (SIEM) evasion and detection methods," Tesis de Maestría, Universidad Tecnológica de Tallinn, Estonia, 2014.

[3] Cybersecurity Insiders, "2019 SIEM Report," Cybersecurity Insiders, 2019.

[4] A. Sapegin, D. Jaeger, F. Cheng, *et al*., "Towards a system for complex analysis of security events in large-scale networks," *Computers and Security*, no. 67, pp. 16–34, 2017.

[5] S. Bhatt, P. Manadhata, y L. Zomlot, "The operational role of security information and event management systems," *IEEE Security and Privacy Magazine*, vol. no. 5, pp. 35–41, 2014.

[6] R. M. Perurena, W. B. García y J. P. Rubier, "Gestión automatizada e integrada de controles de seguridad informática," *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 34, no.1, pp. 40-58, 2013.

[7] T. Liy, L. Yan, "SIEM Based on Big Data Analysis," *Cloud Computing and Security*, 167-175, ICCCS 2017

[8] 451 Research, "Security Analytics. Transforming Enterprise Security Strategy," 451 Research Advisory, Abril 2018.

[9] Asociación Nacional de Universidades e Instituciones de Educación Superior, "Estado actual de las tecnologías de la información y las comunicaciones en las instituciones de educación superior en México, Estudio 2018," ANUIES, 2018.

Cómo se cita:

M.I. Quintero Martínez y S.A. Tovar Balderas, "Sistemas de Gestión de Información y Eventos de Seguridad SIEM," *TIES, Revista de Tecnología e Innovación en Educación Superior*, n.o. 2, octubre, 2019. [En línea]. Disponible en: https://www.ties.unam.mx/ [Consultado en octubre, 2019].

MINERÍA DE DATOS: IDENTIFICANDO CAUSAS DE DESERCIÓN EN LAS INSTITUCIONES PÚBLICAS DE EDUCACIÓN SUPERIOR DE MÉXICO

Fredy Jesús López Pedraza
Ma. del Consuelo Macías González
Edgar R. Sandoval García
https://www.ties.unam.mx/

Fecha de recepción: 29 de junio de 2019 • Fecha de publicación: octubre de 2019 Octubre de 2019 | número de revista 2 • ISSN 2683-2968

MINERÍA DE DATOS: IDENTIFICANDO CAUSAS DE DESERCIÓN EN LAS INSTITUCIONES PÚBLICAS DE EDUCACIÓN SUPERIOR DE MÉXICO

Resumen

La deserción escolar es un grave problema al que tienen que hacer frente las Instituciones Públicas de Educación Superior, lograr que todos los alumnos concluyan sus estudios es una tarea compleja; cuando se aplica de manera adecuada el proceso de minería de datos a la información de las Instituciones Educativas, permite determinar patrones e identificar las causas de deserción del alumnado, para evitarlas y conseguir que un mayor número de estudiantes termine su formación; por lo anterior es necesario entender las diferentes metodologías que existen, qué técnicas son las que mejor se adaptan al problema, así como utilizar herramientas informáticas que brinden resultados acordes al objetivo de los proyectos.
Conocer las investigaciones realizadas en Instituciones públicas de México, la forma en que fueron implementadas y los logros alcanzados, proporciona un panorama de la situación del problema; no obstante, el sector educativo es un campo muy extenso por explorar, encontrar nuevas variables que incidan en la deserción escolar permitirá implementar acciones concretas para evitarla.

Palabras clave:
Minería de datos, deserción escolar, educación superior.

DATA MINING: IDENTIFYING CAUSES OF STUDENT DROPOUT IN THE PUBLIC HIGHER EDUCATION INSTITUTIONS OF MEXICO

Abstract

Student dropout is a serious problem faced by public higher education institutions, to make that all students complete their studies is a complex task; when the data mining process is properly applied to the information of the Educational Institutions, allows to determine patterns and identify the causes of student dropout, to avoid them and get a greater number of students to finish their studies; so it is necessary to know the different methodologies that exist, which techniques are best adapt to the problem, as well as to use computer tools that provide results according to the objective of the project.
To know the research made in public institutions in Mexico, the way in which they were implemented and the results achieved, gives an overview of the current situation of the problem; however, the educational sector is a very extensive field to explore, to find new variables that affect the student dropout will allow institutions implement concrete actions to avoid it.

Keywords:
Data Mining, Student Dropout, Higher Education.

MINERÍA DE DATOS: IDENTIFICANDO CAUSAS DE DESERCIÓN EN LAS INSTITUCIONES PÚBLICAS DE EDUCACIÓN SUPERIOR DE MÉXICO

Introducción

LOS ESTUDIOS REALIZADOS, ENFOCADOS AL ABANDONO ESCOLAR SON DIVERSOS, LA MAYORÍA TRATA DE identificar las causas que lo originan, implementando estrategias para abatir el fenómeno, algunas con mejores resultados que otras, pero al final ninguna ha sido suficiente.

En los últimos 20 años, el sector educativo mexicano se ha vinculado de manera más estrecha con otras ramas del conocimiento, particularmente con las tecnologías computacionales, aprovechando las ventajas que estas representan en cuanto a procesamiento masivo de información, minería de datos y aprendizaje automatizado, a fin de encontrar información inédita que ayude a comprender mejor las causas por las que los estudiantes no concluyen sus estudios. Los resultados obtenidos dan mayor claridad sobre el problema y generan conocimiento para ser empleado en el establecimiento de nuevas acciones con miras a erradicarlo.

Desarrollo

LA DESERCIÓN ESCOLAR HA SIDO Y ES UNO DE LOS GRANDES PROBLEMAS QUE AFECTAN A LAS INSTITUCIONES de educación superior, con el objeto de encontrar una solución que la disminuya, las universidades han explorado nuevas opciones, principalmente en el área de minería de datos; en el desarrollo de este artículo se presenta la información relacionada con el tema, iniciando con los conceptos básicos, sus metodologías, técnicas y algoritmos, así como las herramientas de software más empleadas por los profesionales del sector, para concluir con el análisis de las investigaciones realizadas en las Instituciones Públicas de Educación Superior de México y los resultados alcanzados por éstas.

Educación Superior en México

LA *LEY GENERAL DE EDUCACIÓN DE MÉXICO*, EN SU ARTÍCULO 37, ESTABLECE TRES TIPOS DE EDUCACIÓN: básica, media superior y superior; el tipo superior está integrado por el técnico superior universitario, licenciatura y posgrado [1]; a continuación se muestra la matrícula del ciclo 2017-2018 para conocer los indicadores registrados.

La SEP[1] determina que los indicadores a medir en Educación Superior son la *absorción y cobertura*. El primero se define como el número de alumnos de nuevo ingreso al grado inicial de un nivel educativo, por cada cien egresados del nivel y ciclos inmediatos anteriores [3]; para el ciclo 2017-2018, se alcanzó el 74% [2]. La *cobertura* es el número total de alumnos inscritos en un nivel educativo al inicio del ciclo escolar, por cada cien del grupo de población con la edad reglamentaria para cursar ese nivel [3], es decir, ¿qué porcentaje de

1 Secretaría de Educación Pública

Figura 1.
Matrícula del ciclo 2017-2018.
Fuente: Adaptada de [2].

personas con la edad requerida para estudiar determinado nivel realmente lo están haciendo?; en el ciclo de referencia, la matrícula escolarizada y mixta, incluyendo posgrado, fue de 3,864,995 estudiantes, lo cual equivale al 29.5% del total de la población de 18 a 23 años, dicho porcentaje representa la cobertura del nivel [2].

Deserción en Educación Superior

POR OTRA PARTE, LA *DESERCIÓN* ES EL NÚMERO DE ALUMNOS QUE ABANDONAN SUS ESTUDIOS ANTES DE TERMINAR algún grado o nivel educativo [3]. Si bien se ha definido el término, no es indicador del tipo superior. Según [4], hay dificultades para explicar la deserción, ya que presenta variantes particulares que la complican por su carácter longitudinal. Para reforzar el dicho, el autor plantea *¿A partir de qué momento se considerará desertor?, ¿Son desertores los alumnos que lo hacen desde que se inscriben, o debe ajustarse el dato de inscripción?*, las cuales a casi dos décadas de su planteamiento no han sido resueltas.

Aunque no hay una información oficial que permita confrontar la deserción con otros países, estudios de la OECD[2] señalan que México tiene la proporción más baja de adultos con un título de educación superior, logrando apenas el 17%, cifra inferior al promedio que es del 37%, y debajo de otros países de la región donde la media es del 21%; también se reconocen avances en el nivel superior, señalando que en los últimos 16 años el porcentaje de adultos jóvenes que finalizaron educación superior pasó del 17% al 23% y prevé que en el futuro el 26% de los jóvenes mexicanos cuenten con título de nivel superior [5].

La ANUIES[3] , en [6], propone como meta para 2024 reducir al 6%[4] la tasa nacional de abandono en licenciatura y técnico superior, aunque no precisa cómo se determinan dichas cifras ni cómo alcanzarlas.

Sin duda, la *deserción* se mantiene como una situación de alarma por resolver para las instituciones de educación superior. Para atenderla, es preciso conocer

2 Organización para la Cooperación y el Desarrollo Económicos.

3 Asociación Nacional de Universidades e Instituciones de Educación Superior.

4 En el ciclo 2017-2018 la línea base fue del 8.3%

las causas que la originan e implementar acciones que la minimicen. Conseguir lo anterior no resulta sencillo, a pesar de que se han implementado variadas estrategias; por ello, en fechas relativamente recientes, las Instituciones han considerado el conocimiento que se extrae de los grandes volúmenes de datos y la información de valor que de ella se genera.

Actualmente, las instituciones exploran alternativas para atender la deserción, buscando respuestas en grandes volúmenes de datos, sin embargo es tal la cantidad de información que el proceso de analizarla, e interpretarla de manera manual, resulta muy tardado y costoso. A continuación se presentan los procesos para apoyar e identificar patrones o causas de deserción.

Big Data

EL TÉRMINO SE TRADUCE COMO DATOS MASIVOS, Y SE REFIERE A TAL CANTIDAD DE INFORMACIÓN QUE, el proceso de analizarla e interpretarla de manera manual, resulta muy tardado y costoso [7]. En [8] se señala respecto a Big Data, que en un principio los datos habían aumentado tanto, que aquellos que se examinaban ya no cabían en la memoria de los ordenadores para poder procesarlos, por lo que hubo que modernizarlos.

En [9] precisan que no solo se refiere al volumen de la información, sino también a la *variedad* del contenido y a la *velocidad* con la que se genera, almacena y analiza, lo cual se conoce como las 3V. En [10] refieren que varían según las características de las organizaciones, para unas, prima el *volumen*; para otras es la *velocidad*; y otras consideran mejor la *variabilidad* de las fuentes. Es claro, que representa un activo de valor, por lo que es más frecuente almacenar datos. Sin embargo, el beneficio se obtiene cuando se procesan adecuadamente, se identifican patrones, tendencias y limitantes de la información. SAS Institute señala que los datos fluyen de todas partes a velocidades y volúmenes nunca vistos, pero tomar decisiones eficientes no depende de la cantidad, de hecho, tener tantos, puede ser un obstáculo [11].

Data Mining[5]

ES UN PROCESO DE SELECCIÓN, EXPLORACIÓN, MODIFICACIÓN, MODELIZACIÓN Y VALORACIÓN DE LOS DATOS con el objetivo de descubrir patrones desconocidos o no detectados a través de procesos manuales, incluso se pueden utilizar para predecir comportamientos futuros [11]. Por su parte en [12] refiere que descubre relaciones, tendencias, desviaciones, comportamientos atípicos, patrones y trayectorias ocultas, con el propósito de soportar los procesos de toma de decisiones con mayor conocimiento; [10] indica que es un proceso que utiliza técnicas estadísticas, matemáticas, inteligencia artificial y de aprendizaje automático para extraer e identificar información útil que se convierte en conocimiento a partir de grandes bases de datos; además de que puede realizar dos operaciones básicas: predecir tendencias y comportamientos y/o identificar patrones desconocidos.

Metodologías

ES NECESARIO IMPLEMENTAR METODOLOGÍAS APROPIADAS AL OBJETIVO DEL PROYECTO Y ACORDES CON los datos a manipular para minarlos. Su utilización permite realizar el proceso en forma sistemática y no trivial, al proveer una guía para la planificación y ejecución del proyecto, estableciendo fases, tareas a realizar y cómo llevarlas a cabo [13]. En 2014, KDnuggets[6] realizó una encuesta [14] donde preguntó a profesionales de datos ¿qué metodología principal utilizaron en el último año para realizar sus proyectos de análisis, minería de datos o ciencia de datos?, los resultados se muestran en la siguiente imagen, posteriormente se describen las tres más elegidas.

- *CRISP-DM*[7] , tiene un proceso de seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implantación. La sucesión de fases no es necesariamente rígida, cada una de ellas es descompuesta en varias tareas generales de segundo nivel que se proyectan a tareas específicas [13]. Esta metodología no solo garantiza la adecuada planeación sino una mayor efectividad de los resultados [15].
- *SEMMA*[8] , creada por SAS Institute, se define como el proceso de selección, exploración y modelado de grandes volúmenes de datos para descubrir patrones de negocio desconocidos, se enfoca especialmente en aspectos técnicos, excluyendo actividades

5 Minería de Datos

6 Sitio web líder en inteligencia artificial, análisis de datos, minería de datos, ciencia de datos y aprendizaje automatizado

7 Cross Industry Standard Process for Data Mining

8 Sample, Explore, Modify, Model and Assess

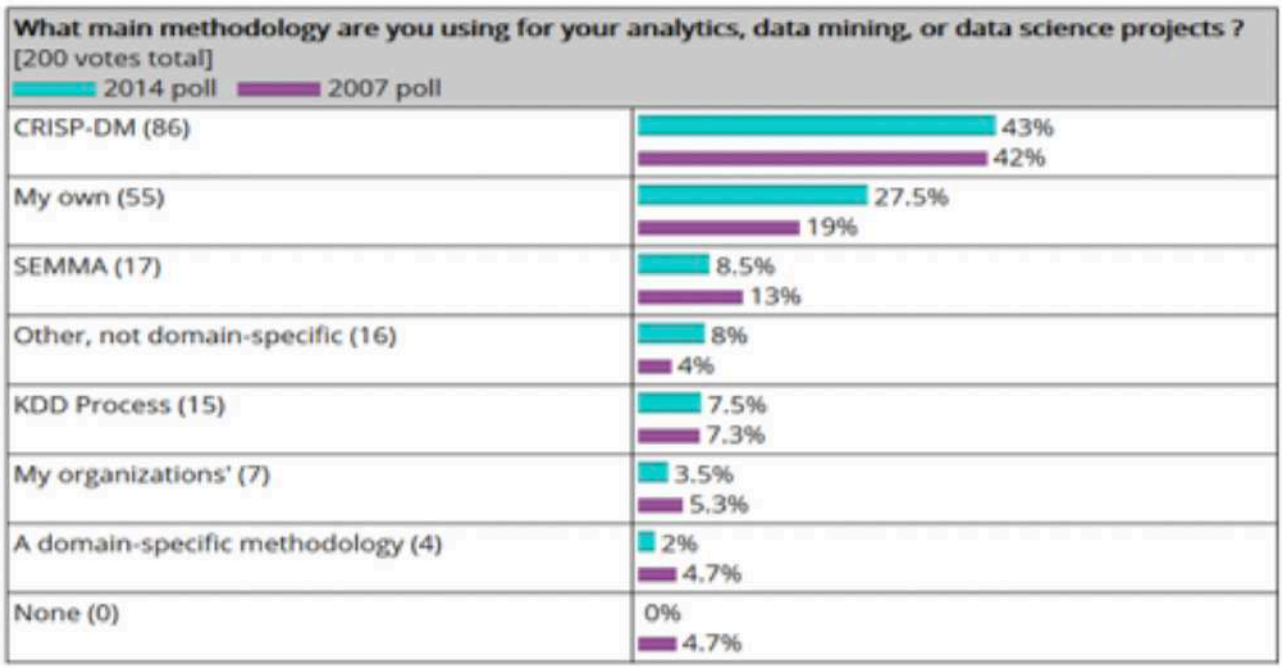

Figura 2.
Metodologías más utilizadas en 2014 para realizar proyectos de análisis, minería de datos o ciencia de datos.
Fuente: Tomada de [4].

de análisis y comprensión del problema que se está abordando [13]. Sus fases son selección, exploración, limpieza, transformación, minería de datos, evaluación y difusión [16].

- *KDD*[9] , tiene sus orígenes hace más de tres décadas. Ya que fue la primera en aparecer, es común que se utilice como sinónimo de minería de datos por lo que se utiliza mayormente para hacer referencia al proceso completo de descubrimiento de conocimiento [13]. El proceso de *KDD* consiste en transformar información de bajo nivel en conocimiento de alto nivel, es interactivo e iterativo, considera las etapas de comprensión del dominio de aplicación, extracción de los datos objetivo, preparar los datos, minería de datos, interpretación y utilización del conocimiento descubierto [17].

Técnicas y Algoritmos

LAS TÉCNICAS Y ALGORITMOS INTENTAN OBTENER MODELOS O PATRONES A PARTIR DE LOS DATOS recopilados, constituyendo el enfoque conceptual para extraer la información y ser implementadas por algoritmos [9]; en [18] señala de que establecen modelos utilizando datos de ejemplo o experiencias pasadas. De este modo identifican patrones o regularidades y se construyen buenas aproximaciones al problema.

La elección de la técnica viene determinada por dos condicionantes: el tipo de datos y el objetivo que se quiera lograr [17]. En [9] se clasifica en dos categorías[10] : supervisadas o predictivas y no supervisadas o descriptivas. Las predictivas se utilizan para prever el valor de un atributo de un conjunto de datos, denominado etiqueta, conocidos otros atributos, a partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos, con esta relación se predicen datos donde la etiqueta es desconocida. En contraparte, las descriptivas ayudan a la comprensión, tratando de ordenar los ejemplos en determinado orden, según las regularidades en la distribución de los pares atributo-valor sin la guía del atributo especial

9 Knowledge Discovery in Databases

10 Considerando que hay modelos predictivos que también pueden ser descriptivos y los modelos descriptivos también pueden emplearse para realizar predicciones, esta clasificación principalmente señala el propósito para el que son más utilizadas estas técnicas

clase. Este es el proceder de sistemas que realizan *clustering* conceptual y de los que adquieren nuevos conceptos como los de asociación.

Los algoritmos permiten desarrollar las técnicas paso a paso, por esto es indispensable un entendimiento de alto nivel, para comprender sus parámetros y características para preparar los datos a analizar [9]. En [19] precisan que un algoritmo es un conjunto de heurísticas, cálculos y operaciones que permiten crear un modelo a partir de determinados datos, a través de un gran número de iteraciones se determinan los parámetros óptimos para crear el modelo.

En la encuesta 2016 de [20], se preguntó ¿Qué algoritmos utilizaron en los últimos 12 meses para una aplicación real relacionada con ciencia de datos? Los resultados se muestran en la siguiente imagen.

Los algoritmos de regresión logística son un tipo de análisis estadístico orientado a la predicción de una variable categórica en función de otras variables consideradas como parámetros predictores [21]; destaca que el valor a predecir es numérico de tipo dicotómico; la regresión lineal, se utiliza frecuentemente con variables cuantitativas, y son considerados algoritmos de predicción.

Los algoritmos de *clustering*, agrupamiento o segmentación, parten de una medida de proximidad entre individuos y a partir de ahí, buscan los grupos más parecidos entre sí, según una serie de variables mesuradas [21]. Los árboles de decisión, considerados de clasificación, permiten dividir datos en grupos basados en los valores de las variables; determinan las variables más significativas para un elemento dado, el mecanismo base consiste en elegir un atributo como raíz y desarrollar el árbol según esas variables [16].

Software

En 2018, KDnuggets realizó la 19ª encuesta anual para conocer que software de análisis, minería de datos y ciencia de datos usaron los expertos en el desarrollo de sus proyectos, participaron más de 2,300 votantes de todo el mundo, quienes eligieron de una lista de más de 90 programas. El reporte en [22] destaca los 11 programas más utilizados que se muestran en la siguiente imagen.

- *Python* es un lenguaje de programación de alto nivel, interpretado y multipropósito [23]; cuenta con facilidades para la programación orientada a obje-

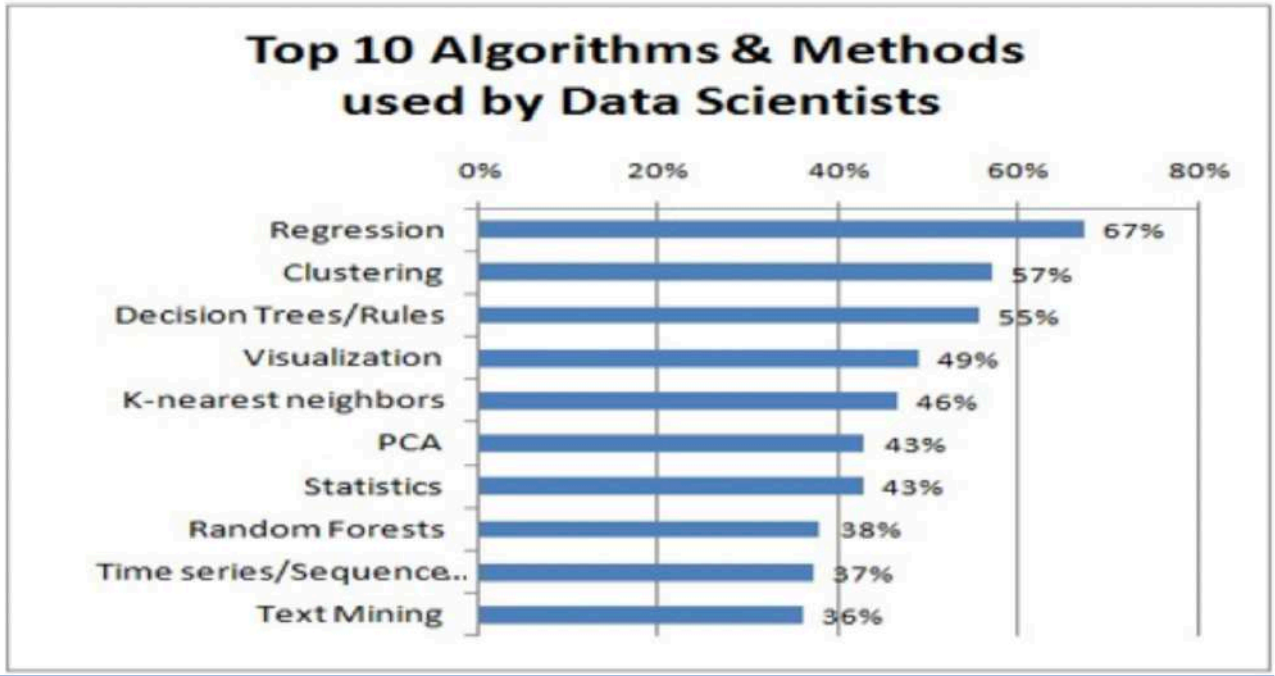

Figura 3.
Algoritmos más utilizados en 2015 para una aplicación real relacionada con ciencia de datos.
Fuente: Tomada de [20].

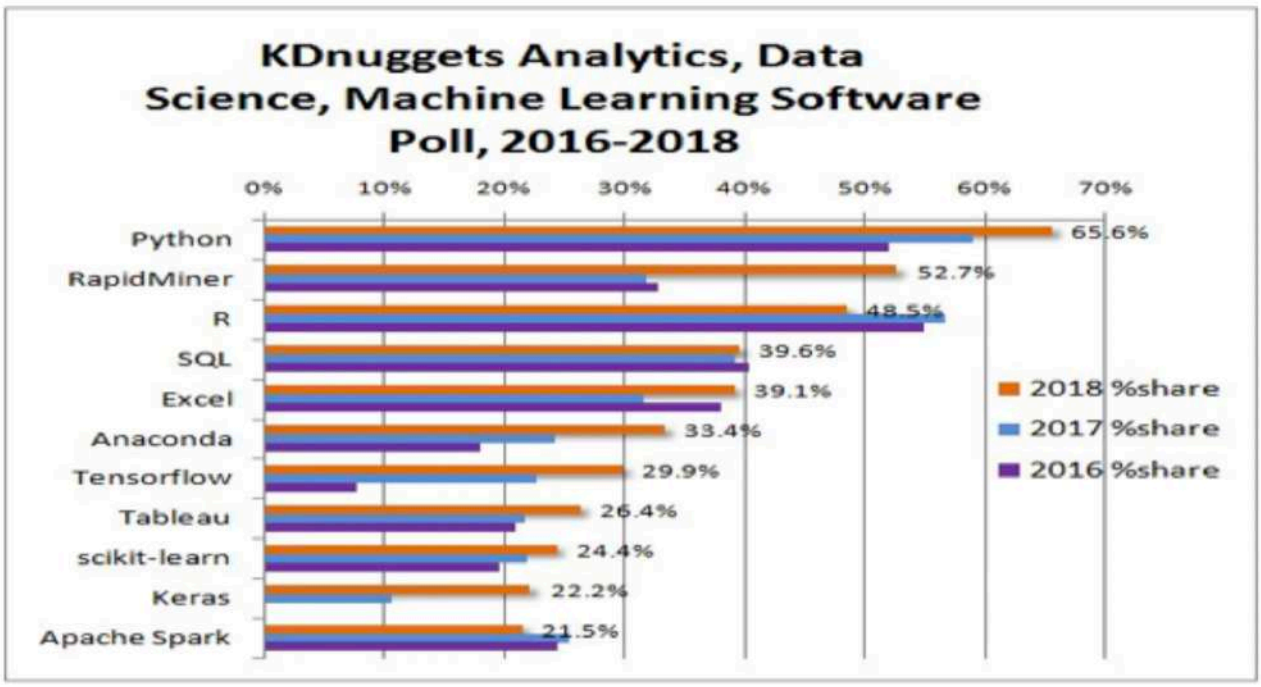

Figura 4.
Principales programas de análisis, minería de datos y ciencia de datos que usaron los expertos en el desarrollo de sus proyectos en 2018.
Fuente: Tomada de [22].

tos, imperativa y funcional, por lo que se considera un lenguaje multiparadigmas [24]. Los motivos principales del creciente uso de *Python* son las numerosas librerías con las que cuenta y su integración con aplicaciones como *MongoDB, Hadoop* o *Pentaho* [25].

- *RapidMiner* es un entorno que contiene procedimientos de *data mining* y aprendizaje automático, el proceso puede hacerse mediante operadores arbitrariamente anidados, descritos en ficheros XML y creados con la interfaz gráfica de usuario, también integra esquemas de aprendizaje y evaluadores de atributos del entorno Weka[11] y esquemas de modelización estadística de *R-Project* [9].
- *R* es un lenguaje de programación de código abierto con un entorno de programación apto para cálculos estadísticos y gráficos. Es muy popular para manipular algoritmos con datos no estructurados, ofrece gran variedad de estadísticas y técnicas gráficas [10]. Dispone de muchos paquetes para la creación de gráficos que le aportan capacidades avanzadas en la visualización de datos y resultados del análisis [25]. Abarca multitud de campos y permite combinar diferentes funciones para análisis más complejos [26].

Estudios realizados

La minería de datos ha tenido presencia en diferentes estudios con la finalidad de encontrar causas de la deserción escolar. En el caso particular de las Instituciones de Educación Superior públicas de México son distintas las investigaciones y muy variados los resultados alcanzados. A continuación, se abordan algunos de los más representativos.

- El primero a analizar es el caso del Instituto Tecnológico de Roque, publicado en el documento *Prototipo de Minería de Datos en la detección oportuna de Estudiantes en riesgo de Abandono Escolar GUÍA (Gestión Universitaria Integral del Abandono)* [27]. Su objeto es el desarrollo de un prototipo web que permita predecir de manera temprana que estudiantes se en-

11 Waikato Environment for Knowledge Analysis, programa computacional

cuentran en mayor riesgo de abandonar sus estudios. Cabe precisar que esta investigación aún se encuentra en desarrollo y lo referido en este documento son los avances y las acciones a realizar en el futuro. Para identificar las causas de abandono escolar, los autores utilizan el análisis estadístico desarrollado en el proyecto Alfa GUÍA[12] , y lo sistematizaron por medio del prototipo web para aplicarlo a los estudiantes del primer año de la carrera de Ingeniería en Tecnologías de la Información y Comunicaciones en la Institución antes referida. La encuesta consta de 34 preguntas que miden 82 variables las cuales se clasifican en cinco categorías, denominadas factores, individual, académico, sociocultural, económico e institucional. A pesar de que el estudio no está concluido, presentan los resultados obtenidos del análisis estadístico, y determinan que dentro de la muestra (no se especifica el tamaño), existe una probabilidad de deserción escolar del 24%. También señala que los factores que más inciden en la deserción son el académico, el económico y el individual. Por último, los investigadores precisan que la siguiente fase de su investigación considera la utilización de técnicas de minería de datos para determinar patrones que pudieran no haber sido identificados mediante el análisis estadístico.

- Por otra parte, en el Centro Universitario UAEM[13] Valle de México, realizaron un *Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar* [28] a fin de determinar qué algoritmo de clasificación obtiene mejores resultados para predecir la deserción escolar. En esta investigación, mediante la metodología KDD, se analizaron las calificaciones por asignatura obtenidas durante el primer año de estudios de los alumnos de Ingeniería en Sistemas y Comunicaciones, de las generaciones 2008, 2009 y 2010. Los algoritmos comparados fueron árboles de decisión y bayesianos, particularmente, ID3[14] , C4.5[15] (J48), Naive Bayes Tree, Naive Bayes, y Redes Bayesianas[16] , con apoyo del software Weka; las pruebas se realizaron en dos bloques, uno para los datos de tipo nominal y otro para los datos de tipo numérico. En ambos casos se hicieron 3 experimentos variando los datos, todos los modelos fueron evaluados mediante validación cruzada de 10 pliegues, que divide el conjunto de datos en diez partes, utilizando nueve partes para entrenamiento y una para prueba. Los resultados obtenidos concluyen que es posible obtener un modelo de predicción de la deserción escolar fiable, utilizando las calificaciones obtenidas por los alumnos en el primer año de sus estudios. Los mejores algoritmos son Naive Bayes Tree y J48, ya que el primero es, desde el punto de vista cuantitativo, el que tiene menor error al momento de clasificar; si se considera aspecto cualitativo, el árbol generado por el algoritmo J48 provee mejor información y de mayor utilidad pues precisa las asignaturas que influyen en mayor medida al abandono de los estudios por parte del alumno.
- Otro caso a considerar es el realizado en una Institución de Educación Superior del Estado de México, no se precisa el nombre de la misma, donde un grupo de investigadores realizó el *Diseño de un Modelo predictivo aplicando Minería de Datos para identificar causas de Deserción Estudiantil Universitaria* [29], llamado PredATIS, el cual se basa en reglas de clasificación y selección de atributos, y cuyo objetivo fue identificar patrones relacionados con los aspectos de mayor influencia en la deserción estudiantil. La investigación se realizó mediante un análisis exploratorio, correlacional y explicativo, a partir del cual se creó un modelo de minería de datos; los datos empleados para el entrenamiento es una muestra de 170 estudiantes de un programa educativo (no se especifica cual), del tercer período escolar del año 2017, en la cual consideraron 39 variables agrupadas en 6 factores: personales, vocacionales, académicos, socioeconómicos, de salud y otros. El entrenamiento del modelo se realizó con el software Weka, empleando algoritmos de árboles de decisión J48 y REPTree[17] , así como reglas de clasi-

12 Proyecto de la UNESCO para la mejora de los índices de permanencia de los estudiantes de Enseñanza Superior

13 Universidad Autónoma del Estado de México

14 Es utilizado en la construcción de árboles de decisión, principalmente para aspectos de inteligencia artificial

15 Es empleado para generar un árbol de decisión de forma recursiva, su uso primordial es en técnicas de clasificación

16 Son clasificadores estadísticos que determinan la probabilidad de que una instancia pertenezca a una clase determinada

17 Permite construir un árbol de decisión, considerado de aprendizaje de decisión rápida

ficación basadas en JRIP[18] , OneR[19] y ZeroR[20] ; los resultados evidencian que el algoritmo J48 permitió identificar de mejor manera las causas que más influyen en la deserción, precisando que un estudiante está en riesgo de baja si tiene planes de matrimonio, ha presentado exámenes extraordinarios de las asignaturas técnicas del perfil, su mayor tiempo libre lo ocupa en realizar actividades culturales, deportivas o de entretenimiento, así como dedicar poco tiempo en estudiar o visitar bibliotecas, adicional a que la carrera elegida no fue su primera opción y es obligado por sus padres o por la cercanía a la universidad. Finalmente, cuando se compararon los resultados con el reporte real de la Institución se obtuvo un porcentaje de 90% de clasificaciones correctas.

- En el Tecnológico de Estudios Superiores de Jocotitlán se realizó una investigación presentada en el documento *Minería de datos aplicada para la identificación de factores de riesgo en alumnos* [30], cuyo objetivo fue implementar un sistema que realiza más eficiente el proceso de tutorías. Para esto los investigadores analizaron información de 831 estudiantes de las generaciones de 2008 a 2013 de la carrera de Ingeniería en Sistemas Computacionales. Tomando como base la metodología KDD, emplearon técnicas de *clustering* y reglas de asociación, particularmente los algoritmos K-Means[21] y A priori[22] . Para determinar las reglas de asociación utilizaron datos sobre práctica de deporte, problemas económicos, si trabajan o no, promedio de bachillerato, interrupciones en sus estudios y el campo de acción de la carrera; para el agrupamiento emplearon datos como la carrera elegida, tratamientos médicos, dependencias económicas, si están casados o tienen hijos y conocimiento de programas de becas. Posteriormente realizaron el preprocesamiento de datos, la aplicación de técnicas de minería de datos hasta llegar a la interpretación y evaluación de los resultados. La investigación indica que con las reglas de asociación obtuvieron un panorama general de la situación de los alumnos estudiados y las causas probables por las que abandonan sus estudios, destacando que: los alumnos que practican algún deporte tienen problemas económicos; la mayor parte de los estudiantes que desertan tienen problemas económicos; los desertores obtuvieron buenos resultados en niveles de estudio previos, remarcando que la mayoría de los que desertan nunca interrumpieron sus estudios en los niveles básico y medio superior. Con la aplicación del algoritmo de K-Means señalan algunas premisas de los resultados, mas no concluyen si las mismas son factores en el abandono de los estudios.
- Por último, está la investigación realizada en la Universidad Tecnológica de Izúcar de Matamoros (UTIM) donde en el documento *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos* [31]exponen los resultados obtenidos, que concluyen en una herramienta que permite calcular la probabilidad de deserción de cada uno de los estudiantes de la UTIM, apoyando de esta manera el proceso de tutorías de dicha institución. Si bien los autores no especifican explícitamente la metodología que siguen, señalan las fases implementadas, mismas que corresponden a la metodología KDD; utilizando como fuente de datos los resultados del EXANI–II[23] de los alumnos inscritos y de los alumnos que causaron baja en la universidad, llevaron a cabo la selección, limpieza y transformación de los datos. Posteriormente realizaron la clasificación de los mismos, donde emplearon un árbol de decisión mediante el algoritmo C4.5 y el método de aprendizaje de los *k* vecinos más cercanos. A continuación generaron los modelos para los dos algoritmos y con la ayuda del software Weka se probaron ambos con diferentes variables, obteniendo que el modelo del árbol de clasificación tuvo una mejor precisión con un porcentaje del 98.98%, mientras que los *k* vecinos más cercanos apenas superó el 70%. Los autores refieren en los resultados que los alumnos desertan por tres causas: la edad, ya que tiene que ver con la madurez y perspectiva de futuro de los estudiantes; los ingresos familiares, para alum-

18 Admite la poda incremental reducida, para producir reducción de errores

19 Basado en el algoritmo ID3, clasifica desde el conjunto de datos de entrenamiento

20 Predice la clase mayoritaria, si es nominal, o el valor promedio, si es numérico

21 Algoritmo de clasificación no supervisada, que agrupa objetos en k grupos basándose en sus características

22 Permite encontrar eficientemente conjuntos de ítems frecuentes que sirven de base para generar reglas de asociación

23 Examen Nacional de Ingreso a la Educación Superior

nos de 18 años o menos, puesto que dependen de los ingresos familiares; la tercera causa es el nivel de inglés, para alumnos mayores de 18 años. Para minimizar estas causas, se implementó una aplicación web, a través de la cual los tutores determinan el factor de riesgo de manera oportuna, y dan seguimiento a aquellos estudiantes vulnerables. En este punto vale la pena precisar que, si bien en todas las investigaciones analizadas sugieren estrategias para minimizar las causas de la deserción escolar, en su mayoría programas de tutoría o de seguimiento académico, en ninguna se refieren los resultados alcanzados de dichas acciones, tampoco señalan si las acciones sugeridas realmente están logrando disminuir los índices de abandono escolar. Esta falta de conclusiones puede deberse a que tres de los cinco estudios analizados fueron publicados recientemente, en los dos últimos años para ser exactos, otro más en 2013 y el más antiguo es de 2010.

Conclusión

La educación superior en México es amplia y compleja, presenta problemas de distintas magnitudes, unas simples y otras más complicadas cómo la cobertura, la absorción, la reprobación y la deserción. No obstante, es obligación del Estado garantizar el derecho a la educación de los mexicanos. A partir de la recopilación realizada, se concluye que hay muchas posibilidades para implementar minería de datos en el sector educativo; existen diferentes metodologías, técnicas, algoritmos y herramientas de software para ser empleadas en el proceso que en conjunto facilitan el análisis de la información y generan conocimiento que da mayor certeza a la toma de decisiones y reduce los problemas.

Respecto a la deserción escolar, desde siempre se ha sabido que es un problema multifactorial, esta premisa cobra mayor fuerza cuando se comparan los resultados de los estudios y se observa que los mismos determinan los factores que lo originan, encontrándose que principalmente son de tipo personal, académico y socioeconómico.

Adicionalmente, es importante destacar que ninguno de los casos analizados precisa resultados de las acciones implementadas después de haber identificado las causas de origen de la deserción; esta ausencia de información, aunque no se señala, bien puede deberse a que las investigaciones principalmente son realizadas por docentes, quienes carecen de la autoridad para implementar estrategias institucionales al respecto, otro factor puede ser el alcance de la investigación, y en todos los casos las investigaciones se limitan a identificar causas de deserción y proponer acciones que pueden disminuirla, sin llegar a la implementación y menos aún evaluar los resultados de su aplicación.

Otro punto importante a comentar es referente al enfoque que se da a los estudios, encontrándose que en todos los casos, las variables analizadas estuvieron directamente relacionadas con los estudiantes, lo que induce a pensar que en la mayoría de los casos, los orígenes del problema involucran a los alumnos; sin embargo, habrá que preguntarse si esto siempre es cierto o hay factores no relacionados con el alumno que influyen en su deserción.

Para reafirmar o desmentir lo anterior, los que escriben desarrollan una investigación de grado, cuyo objetivo es determinar, mediante técnicas de minería de datos, si las características, perfil y desempeño de los docentes influyen en la deserción de los estudiantes de educación superior. La investigación antes referida está en la fase de desarrollo, y aun no cuenta con resultados que puedan comentarse en este documento. A referencia se comenta para identificar una arista más del problema de deserción y sentar el precedente de la investigación que se realiza.

BIBLIOGRAFÍA

[1] Gobierno de los Estados Unidos Mexicanos, *Ley General de Educación*, México, 2018.

[2] Dirección General de Planeación, Programación y Estadística Educativa, "Sistema Educativo de los Estados Unidos Mexicanos, Principales Cifras 2017-2018," *Secretaría de Educación Pública*, México, 2018

[3] Dirección General de Planeación y Programación, "Glosario. Términos utilizados en la Dirección General de Planeación y Programación," *Secretaría de Educación Pública*, México, 2008.

[4] F. Martínez Rizo, "Estudio de la eficiencia en cohortes aparentes," *Deserción, rezago y eficiencia terminal en las IES. Propuesta metodológica para su estudio*, Serie Investigaciones, México, ANUIES, 2001.

[5] OECD, "Higher Education in Mexico: Labour Market Relevance and Outcomes," *OECD* Publishing, Paris, 2019.

[6] ANUIES, "Visión y Acción 2030, Propuesta de la ANUIES para renovarla educación superior en México," *Publicaciones ANUIES*, México, 2018.

[7] IBM, "¿Qué es Big Data? Todos formamos parte de ese gran crecimiento de datos," junio 18, 2012. [En línea]. Disponible en: https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html. [Consultado en abril 9, 2019].

[8] V. Mayer-Schönberger y K. Cukier, *Big Data. La revolución de los datos masivos*, Madrid: Editor digital Titivillus, 2013.

[9] J. García, J. M. Molina, A. Berlanga, *et al., Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Bogotá: Alfaomega Colombiana, 2018.

[10] L. Joyanes Aguilar, *Big Data: Análisis de grandes volúmenes de datos en organizaciones*. México: Alfaomega Grupo Editor, 2013.

[11] SAS Institute, "La Minería de Datos de la A a la Z: Cómo Descubrir Conocimientos y Crear Mejores Oportunidades," 2015. [En línea]. Disponible en: https://www.sas.com/es_mx/whitepapers/data-mining-from-a-z-104937.html. [Consultado en abril 22, 2019].

[12] B. Beltrán Martínez, *Puebla: Benemérita Universidad Autónoma de Puebla*. Facultad de Ciencias de la Computación, 2019.

[13] S. Gordillo, A. S. Haedo y J. M. Moine, "Estudio comparativo de metodologías para minería de datos," *XIII Workshop de Investigadores en Ciencias de la Computación*, Argentina, 2011.

[14] G. Piatetsky, "CRISP-DM, still the top methodology for analytics, data mining, or data science projects," *KDnuggets*, 2014. [En línea]. Disponible en: https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html. [Consultado en febrero 3, 2019].

[15] J. L. Cendejas Valdez, M. Á. Acuña López, G. Cortes Morales y G. Bolaños Jiménez, "El uso de modelos y metodologías de minería de datos para la inteligencia de negocios," *Revista de Sistemas Computacionales y TIC's*, vol. 3, no. 8, pp. 54-63, junio 2017.

[16] M. Pérez Marqués, *Minería de datos a través de ejemplos*. Madrid: RC Libros, 2014.

[17] J. C. Riquelme, R. Ruiz y K. Gilbe, "Minería de Datos: Conceptos y Tendencias", *Revista Iberoamericana de Inteligencia Artificial*, vol. 10, no. 29, pp. 11-18, 2006.

[18] J. T. Palma Méndez y R. Marín Morales, *Inteligencia Artificial: Métodos, técnicas y aplicaciones*. España: McGraw-Hill / Interamericana de España, S.A.U., 2008.

[19] Microsoft, "Algoritmos de minería de datos (Analysis Services: Minería de datos)," abril 4, 2018. [En línea]. Disponible en: https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017 [Consultado en abril 04, 2019].

[20] G. Piatetsky, "Top Algorithms and Methods Used by Data Scientists," *KDnuggets*, 2016. [En línea]. Disponible en: https://www.kdnuggets.com/2016/09/poll-algorithms-used-data-scientists.html. [Consultado en febrero 23, 2019].

[21] T. Aluja, "La Minería de Datos entre la Estadística y la Inteligencia Artificial," *QUESTIIO*, vol. 25, no. 3, pp. 479-478, 2001.

[22] G. Piatetsky, "Python eats away at R: Top Software for Analytics, Data Science, Machine Learning in 2018: Trends and Analysis," *KDnuggets*, 2018. [En línea]. Disponible en: https://www.kdnuggets.com/2018/05/poll-tools-analytics-data-science-machine-learning-results.html. [Consultado en abril 3, 2019].

[23] A. Fernández Montoro, *Python 3 al descubierto*. Madrid: RC Libros, 2012.

[24] I. Challenger Pérez, Y. Díaz Ricardo y R. A. Becerra García, "El lenguaje de programación Python," *Ciencias Holguín*, vol. XX, no. 2, pp. 1-13, 2014.

[25] P. Rochina, "Python vs R para el análisis de datos," noviembre 16, 2016. [En línea]. Disponible en: https://revistadigital.inesem.es/informatica-y-tics/python-r-analisis-datos/. [Consultado en mayo 20, 2019].

[26] L. A. Vargas, J. H. Farfán, F. Aramayo, *et al.*, "Comparación de las principales herramientas de Data Mining y Análisis de Sábanas Telefónicas," *II Segunda jornada Argentina de Tecnología, Innovación y Creatividad*, 2016.

[27] Y. D. Guzmán Islas, E. Ramos Ojeda y A. Guzmán Zazueta, "Prototipo de Minería de Datos en la detección oportuna de estudiantes en riesgo de abandono escolar GUÍA (Gestión Universitaria Integral del Abandono)," *Pistas Educativas*, vol. 39, no. 129, pp. 64-79, 2018.

[28] M. Quintana López, J. C. Trinidad Pérez, S. J. Morales Escobar, *et al.*, "Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar," *Research in Computing Science*, vol. 67, pp. 13-23, 2013.

[29] P. N. Maya Pérez, J. R. Aguilar C., R. A. Zamora R., *et al.*, "Diseño de un Modelo predictivo aplicando Minería de Datos para identificar causas de Deserción Estudiantil Universitaria," *Strategy, Technology & Society*, vol. 7, no. 2, pp. 11-39, 2018.

[30] A. Reyes-Nava, A. Flores-Fuentes, R. Alejo, *et al.*, "Minería de datos aplicada para la identificación de factores de riesgo en alumnos," *Research in Computing Science*, no. 139, pp. 177-189, 2017.

[31] S. Valero Orea, A. Salvador Vargas y M. García Alonso, "Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos," *Recursos digitales para la educación y la cultura, volumen Kaambal*, Mérida, Yucatán, Universidad Tecnológica Metropolitana, Mérida, Yucatán, México y Universidad de Cádiz, Andalucía, España: 2010, pp. 33-39.

Cómo se cita

F.J. López, M.C. Macías y E.R. Sandoval, "Minería de datos: identificando causas de deserción en las instituciones públicas de educación superior de México," *TIES, Revista de Tecnología e Innovación en Educación Superior*, n.o. 2, octubre, 2019. [En línea]. Disponible en: https://www.ties.unam.mx/ [Consultado en octubre, 2019].