

UNIVERSIDAD MAYOR DE SAN ANDRES
TELEMATICA

PRIVACIDAD DEL BIG DATA EN LA EDUCACION

Participantes

Condori Huanca Freddy

Juarez Chacon Jackeline Jimena

Mamani Peralez Henry Ivan

Ramirez Ticona Blanca Soledad

Grupo No. 2

Fecha de Presentación (29/11/2017)

Revisión del Docente: _____

RESUMEN

En el presente trabajo se realizó una investigación sobre la problemática de la Falta de privacidad y Violación de identidad en los grandes datos (big data) en la educación haciendo hincapié en la privacidad de cada estudiante en cuestión a la cantidad de datos que almacena, teniendo como objetivo Analizar la ética de los grandes datos en la educación.

Indagando en conceptos básicos, como ser el concepto de Big Data, perturbación de datos, técnicas criptográficas y la minería de Datos, apoyándonos en estos conceptos para dar una propuesta de solución a este problema de la privacidad de los alumnos utilizando algunos métodos de privacidad como ser:

Anonimizacion, perturbación de datos, técnicas criptográficas que nos ayudaran a tener una mejor privacidad en la educación.

Palabras clave: base de datos, minería de datos, criptografía, ,anonimizacion, big data, perturbación de datos, criptografía,

CONTENIDO

RESUMEN

CONTENIDO

PROBLEMA

OBJETIVOS DE LA MONOGRAFIA

INTRODUCCION

DESARROLLO

CONCLUSIONES

REFERENCIAS BIBLIOGRAFICAS

INTRODUCCIÓN

Aquí se presentará el tema, enunciando el propósito de la monografía y esbozando el recorrido que el autor se propone realizar. Se presenta en página aparte y a continuación del resumen.

Aquí se explicita lo que se propone abordar, incluyendo, si es necesario, los antecedentes del tema y toda la información de contexto. Debe tenerse en cuenta que la introducción deberá redactarse cuando ya se tengan los lineamientos de lo que se quiere decir. De ninguna manera deberá ser una mera expresión de intenciones o de deseos sino el delineado de lo que constituye la propuesta del autor en el trabajo que está presentando. La introducción debe anticipar al lector lo que encontrará desarrollado en el trabajo.

PROBLEMA

Falta de privacidad y violación de identidad en los grandes datos(Big Data) en la educación.

OBJETIVOS DE LA MONOGRAFIA

Analizar la ética de los grandes datos en la educación.

DESARROLLO

Antecedentes. -

“El Big Data”. Tim Berners-Lee, director del World Wide Web Consortium (W3C), y considerado uno de los padres del internet, fue el primero en utilizar el término *datos* vinculados en 2009 y es así que surge el llamado Business Intelligence, concepto que se convirtió en una prioridad para millones de especialistas de la rama.

Desde entonces, las organizaciones en todo el mundo comenzaron a implementar nuevas tecnologías para analizar y optimizar las grandes cantidades de datos que generaban.

También inició la dependencia de la utilización de su información como un activo comercial para obtener una ventaja competitiva, ya sea en temas de marketing, experiencias personalizadas para los clientes o, simplemente, para mejorar sus procesos del día a día.

Se dice que las primeras civilizaciones que usaron la **“CRIPTOGRAFÍA”** fueron la Egiptia, la Mesopotamia, la India y la China. Los espartanos, 400 años antes de Cristo, utilizaban un sistema secreto de escritura, el cual consistía en un cilindro al cual se colocaba un papiro en forma de espiral. Se escribía entonces el texto en cada una de las vueltas del papiro, pero de arriba hacia abajo. Una vez desenrollado, sólo se podía leer una serie de letras aparentemente inconexas. Para descifrar el mensaje era necesario colocar el papiro exactamente en la misma posición en la que había sido escrito. Antiguos textos judíos fueron encriptados siguiendo el método de sustituir la primera letra del alfabeto por la última y así sucesivamente.

En la legislación española, pese a que no encontramos la palabra **“ANONIMIZACIÓN”**, se reconoce el dato disociado, como aquel que no permite la identificación de un afectado o interesado. Así mismo el artículo 3.f de la LOPD define el procedimiento de disociación como todo tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable.

Por tanto, cuando los datos pierden su condición de personal y dejan de estar vinculados a un titular concreto, quedan fuera del ámbito de aplicación de la normativa de protección de datos.

Pese a que parece tarea sencilla, disociar los datos de manera irreversible no lo es, hacen falta sistemas de gran complejidad de manera que la identificación exija esfuerzos desproporcionados y que en la práctica sea equivalente al borrado permanente.

MARCO TEORICO

¿Qué es Big Data?

Es un término que hace referencia a grandes cantidades de datos almacenados y a los procesos mediante los cuales se identifican patrones reiterativos en esta Big Data para luego realizar modelos predictivos en diversos tipos de materias; por ejemplo, identificación de enfermedades infecciosas, espionaje y tendencias en el internet y las redes sociales. Por lo tanto, existen diferentes fuentes que producen información y datos como lo son la redes sociales, blogs, páginas web, entidades bancarias y financieras, ventas en centros comerciales... etc.

Minimización de datos esta técnica implica que los datos recopilados no deben ser excesivos, sino que debe recopilarse solo la cantidad mínima para lo cual se requiere.

Al analizar el big data mientras se preserva la información surge un nuevo campo que es la preservación de la minería de datos **PPDM** (Preservación y Privacidad de la Minería de Datos).

Se desarrollaron diferentes técnicas **PPDM** para encontrar protocolos eficientes para equilibrar la privacidad, estas técnicas se dividen en tres categorías, aunque la mejor solución sería la técnica híbrida.

Principios de la Anonimización si bien se presenta como la mejor solución para tratar los datos protegiendo la privacidad de los sujetos, en los últimos años se han dado numerosos casos de re identificación de la base de datos que habían sido anonimizadas. La identificación de los sujetos mediante los datos no personales supone un debilitamiento de la anonimización como medida para asegurar la privacidad durante el tratamiento de datos.

Perturbación de datos. - una técnica para modificar datos usando un proceso aleatorio, esta técnica aparentemente distorsiona valores de los datos confidenciales el pre procesamiento de datos, se clasifica en la codificación de atributos y la obtención de un conjunto de datos codificados.

Técnicas de criptográficas. -la criptografía es una técnica a través de la cual los datos sensibles pueden estar encriptado, una técnica usada para preservar los datos y por qué proporciona seguridad a los atributos sensibles.

Datamining (Minería de datos)

El **datamining** (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo

de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Básicamente, el datamining surge para intentar ayudar a comprender el contenido de un repositorio de datos. Con este fin, hace uso de prácticas estadísticas y, en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales.

¿Qué son las bases de datos?

Se define una “**BASE DE DATOS**” como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

MARCO APLICATIVO. –

Experimentando con Big Data en las escuelas

Los pioneros de la experimentación con Big Data en educación son AltSchool, un conglomerado de escuelas que se dedica a la innovación. En concreto, en 4 pequeñas escuelas alrededor de San Francisco se está llevando a cabo un experimento para tratar datos y estudiar cómo la tecnología puede ayudar a los profesores a mejorar el aprendizaje de los alumnos.

¿Cómo consiguen esos datos? Pues de varias formas: Para empezar, cuando los alumnos llegan a la escuela deben hacer un “check-in” en una App que controla la asistencia. Además, los alumnos dedican una parte del día a trabajar de forma independiente, utilizando iPads y playlists de actividades que los profesores han seleccionado de acuerdo con sus objetivos personales. En ese momento, se almacenan los datos de los progresos de cada uno de los estudiantes para que el profesor pueda revisarlos más tarde. Por último, las clases son grabadas en vídeo para que los profesores puedan rememorar los momentos importantes con solo pulsar un botón.

La idea es que todos los datos de estas escuelas sean recopilados en un sistema inteligente y centralizado que permita a los profesores **diseñar clases efectivas y personalizadas**.

¿Qué beneficios nos aporta usar Big Data en Educación?

A pesar de que los resultados definitivos no podrán verse hasta dentro de algunos años, lo cierto es que de momento el uso del Big Data se traduce en una **enseñanza más personalizada** y centrada en la **mejora constante** de los métodos de aprendizaje. A día

de hoy, AltShcools tiene previsto abrir 4 escuelas más, e incluso vender el acceso al sistema a otras escuelas.¹

Principios de la Anonimización.- Por ejemplo, la fecha de nacimiento de los estudiantes se puede generalizar al mes de nacimiento. Los datos sensibles se identifican y luego se reemplazan por valores desconocidos² Por ejemplo, los valores reales se pueden ocultar reemplazando '1' por '0' o '0' por '1' o con cualquier valor desconocido. Esto proporciona protección de datos confidenciales contra el acceso no autorizado (Taneja et al., 2014). En los datos de los registros de los estudiantes, en lugar de identificar a los estudiantes con los números de CI, el uso de las ID de los estudiantes aumentaría el nivel de protección de los otros registros de los estudiantes.

Perturbación de datos. - Por ejemplo, en un caso simplificado de datos de registros de estudiantes, un usuario legítimo del sistema puede no tener acceso a la calificación de un alumno en particular, pero ese mismo usuario podría determinar la calificación promedio de los estudiantes en una clase.

Técnicas de criptográficas. - Utilizando técnicas de cifrado, primero se cifraron los datos, luego se transmitieron a otro extremo y se descifraron³ Un enfoque es la seudonimización, que rompe el vínculo entre la información personal y la información del registro del estudiante, y proporciona uno o más identificadores para cubrir la identidad original.

El cifrado es una técnica bien establecida para construir seudónimos⁴ Proporciona una forma de anonimato rastreable de la salud del estudiante y los registros de disciplina del estudiante. En lugar de eliminar completamente la información de identificación personal de la salud del estudiante y los datos de disciplina del estudiante, la información de identificación se transfiere a una información (es decir, un seudónimo) que no se puede asignar a un estudiante sin conocer un cierto secreto.

¹ Mireia More , <http://www.iebschool.com>

² (Parmar, Rao y Patel, 2011).

³(Panackal & Pillai, 2013; Taneja et al., 2014).

⁴(Panackal y Pillai, 2013).

CONCLUSIÓN

En el presente trabajo se buscó dar una propuesta a la gran cantidad de problemas que se tiene en cuestión al big data en la educación, se tomó como espacio de investigación la privacidad de la gran cantidad de datos en la educación, dando así solución mediante métodos como ser: la anonimización, las técnicas criptográficas y la perturbación de datos, la cual nos ayudara a controlar la gran cantidad de datos de educación que existe en el big data.

REFERENCIAS

- Fernando Paul Lara Galicia, (2017, Mayo), Conoce la historia del big data ,En Linea Disponible en : <https://cl.godaddy.com/blog/conoce-la-historia-del-big-data/>
- Instituto Internacional Español de Marketing Digital, (2017, Junio), Que es el Big Data [En Linea], Disponible en: <https://iiemd.com/big-data/que-es-big-data>
- Sinnexus, (2016, Agosto) , Datamining(Minería de datos) [En Linea], Disponible en: http://www.sinnexus.com/business_intelligence/datamining.aspx
- Damian Perez Valdes (2007, Octubre), ¿Qué son las bases de Datos?, [En Linea], Disponible en: <http://www.maestrosdelweb.com/que-son-las-bases-de-datos/>
- Mireia More (2015, Septiembre), Los beneficios de usar Big Data en La educacion [En Linea], Disponible en:<http://www.iebschool.com/blog/usar-big-data/>
- Elena Gil Gonzales (2015,Abril), Big Data, Privacidad y protección de Datos, [En Linea], Disponible en: https://www.agpd.es/portaIwebAGPD/canaIdocumentacion/publicaciones/common/pr emios_2015/Big_Data_Privacidad_y_proteccion_de_datos.pdf
- <https://cl.godaddy.com/blog/conoce-la-historia-del-big-data/>
- <https://iiemd.com/big-data/que-es-big-data>
- http://www.sinnexus.com/business_intelligence/datamining.aspx
- <http://www.maestrosdelweb.com/que-son-las-bases-de-datos/>