



**Tecnológico
de Monterrey**



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

Inteligencia artificial avanzada para la ciencia de datos II
TC3007C.501

Reto Privacidad y Seguridad de los Datos



Rodolfo Sandoval Schipper A01720253

Arturo Garza Campuzano A00828096

Marcelo Márquez A01720588

25 oct 2023

1. Introducción

En este documento se discute acerca de la naturaleza de los datos con los que se trabajarán en el proyecto **Classroom AI**, con el fin de comprender las implicaciones legales y de seguridad que vienen asociados a los mismos. Para cumplir con el fin establecido se ofrece un enfoque sobre los siguientes temas: anonimización de datos; normativa de la industria; acceso y uso responsable de datos; y trazabilidad y auditabilidad del acceso a datos.

2. Anonimización de datos

Como se ha establecido en [Memorandum_Understanding](#), **Classroom AI** es un sistema para medir la asistencia y participación en el entorno universitario cuyos objetivos son: perder el menor tiempo posible pasando lista entre los alumnos, tomar acción cuando se detecta una baja participación entre todos los alumnos, y la participación y asistencia detectada es visible a los involucrados en el curso. Tomando en cuenta estos objetivos, se cree que es del interés de los usuarios visualizar información personal para obtener una mejor experiencia de usuario y darle un uso efectivo al sistema.

Sin embargo, en este caso, se implementa la anonimización de distinta maneras. Contamos con la anonimización de correos aleatorios para los estudiantes, donde podemos visualizar solamente el ID de un estudiante ya que los demás atributos no coinciden con ningún dato personal al crear un nuevo registro. Hemos implementado hashing para las contraseñas. Sin embargo, este método se puede utilizar para anonimizar todos los datos que se han registrado en el sistema. Tal que podemos recurrir a la visualización sin exponer ningún dato personal del estudiante o Usuario. Como solución temporal del prototipo hemos utilizado esta técnica para solo resolver la anonimización de los usuarios y mejorar la seguridad del sistema con de acuerdo al acceso de los datos en la interfaz.

Para lograr dicha anonimización de datos es necesario considerar, en primer lugar, qué información es personal. A continuación se muestran los datos que se consideran personales dentro de las tablas de la base de datos del sistema.

Estudiantes

- *Nombre ** (Se utiliza para identificar qué alumno está participando/asistiendo)
- *Apellido ** (No es necesario pero puede servir para identificar un alumno en caso de que se repita el nombre)
- *Correo ** (Es parte del usuario “Estudiantes” no es necesario mostrar el correo pero sirve como una llave ya que solo existe un correo por estudiante, verificamos esta autenticación con firebase y en la base de datos)
- *Foto del Estudiante ** (Se utiliza para alimentar el modelo de face recognition)

Usuarios

- *Correo electrónico* * (Se muestra para que el administrador pueda hacer operaciones CRUD para modificar la cuenta del usuario)
- *Contraseña* * ("One way password Hashing con el algoritmo SHA 256" esté dato no será accesible para ningún usuario)

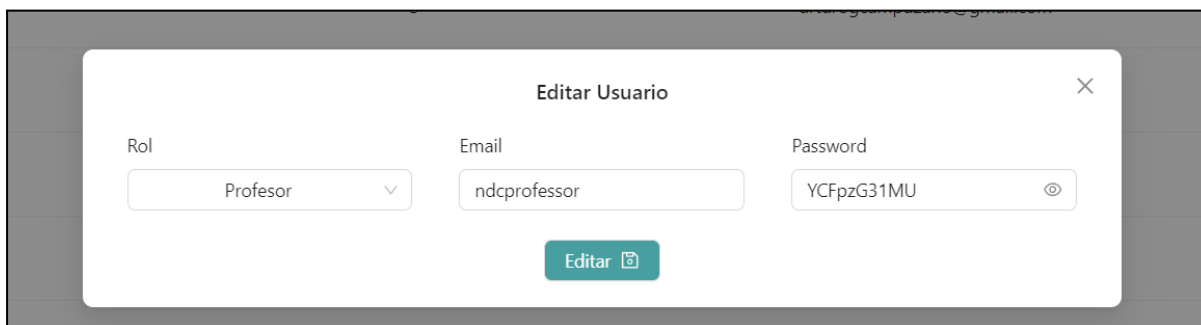
Profesores

- *Nombres de Profesores* * (Se utiliza cómo un atributo dummy para verificar la creación de un curso con de acuerdo al profesor)

Bajo estas consideraciones se ha implementado el hashing de las contraseñas para los usuarios. Este método se puede utilizar para todos los atributos, tal que se anonimiza los correos, nombres, y apellidos de los estudiantes y usuarios en el sistema. Sin embargo, por el momento contamos con el hashing exclusivamente para las contraseñas ya que se requieren hacer bastantes consultas para acceder a todos los atributos de un estudiante en el sistema. Esto nos permite verificar y hacer pruebas de acuerdo a la información guardada en nuestra base de datos. Como solución temporal en el prototipo del sistema implementamos un método para la creación de estudiantes donde solo mostramos el ID del alumno con un correo aleatorio. De esta manera podemos anonimizar los estudiantes al igual que los Usuarios con el hashing de las contraseñas.

Evidencia de la implementación

Implementación de Hashing con Argon3 →



The screenshot shows a web form titled "Editar Usuario" with a close button (X) in the top right corner. The form contains three input fields: "Rol" with a dropdown menu showing "Profesor", "Email" with the text "ndcprofessor", and "Password" with the text "YCFpzG31MU" and an eye icon for toggling visibility. Below these fields is a green button labeled "Editar" with a small icon.

Implementación de anonimización para los estudiantes y profesores →

Agregar Curso

Alumnos

Curso: Anonimizar

✖

138	g5flkpmc@tec.mx	N/A	N/A	👁	✖
139	owswm5yt@tec.mx	N/A	N/A	👁	✖
140	rdwpslfu@tec.mx	N/A	N/A	👁	✖

3. Normativa de la industria

Apartado dedicado a la consulta de la normativa actual de la industria a la que está sujeto el socio formador y a la investigación de los pasos más comunes que se toman para garantizar la privacidad de los datos en dicha industria.

3.1 Normativa de NDS Cognitive Labs

Según NDS Cognitive Labs (n.d.), en su 'Notice of Privacy', la normativa principal a la que está sujeto es la “LFPDPPP” (Ley Federal de Protección de Datos Personales en Posesión de Particulares), ley que establece las obligaciones y derechos relacionados con la protección de datos personales en México. Esta normativa incluye:

1. Responsabilidad en el tratamiento de datos personales
2. Propósito del procesamiento de datos personales
3. Opciones y medios para limitar el uso o divulgación de datos personales
4. Transferencia de datos personales a terceros
5. Consentimiento del cliente para el tratamiento de datos personales
6. Cambios en el aviso de privacidad
7. Derecho de protección de datos ante el INAI


NDS Cognitive Labs debe seguir las regulaciones de protección de datos personales, garantizando la confidencialidad y seguridad de los datos personales de sus clientes y cumpliendo con los derechos de los titulares de datos según lo establecido por la ley.

3.2 Pasos comunes para garantizar la privacidad de los datos

Para asegurar la la privacidad de los datos Rowda Mohamud recomienda seguir los siguientes pasos:

1. **Observar los datos que se están recopilando y almacenando, y el motivo por el cual se realiza:** es crucial considerar cuánta información privada realmente se requiere para atender adecuadamente a los clientes. Cuanto más datos se acumulan, más riesgos y costos de almacenamiento se generan. La gestión adecuada de datos requiere de una estrategia de recopilación y retención que se ajuste a las necesidades, regulaciones y expectativas de privacidad de los clientes pertinentes al proyecto.
2. **Definir quién debe tener acceso a los datos:** para prevenir incidentes es esencial definir roles y niveles de acceso asociados, el principio del acceso mínimo es un enfoque sólido. Hay que considerar: la vinculación de registros de usuarios con procesos de recursos humanos, establecer un procedimiento claro para controlar los permisos de acceso y revisar regularmente los derechos de acceso en aplicaciones críticas.
3. **Comprensión de los riesgos para los datos:** aunque es cierto que la plataformas en la nube cuentan con estrictas características de seguridad, es relevante que las empresas revisen los riesgos de los datos, considerando el valor de los datos, el costo de perderlos, las implicaciones de privacidad al recopilar los datos y los costos de diversas soluciones para abordar los riesgos.
4. **Capacitar a los empleados sobre la privacidad de los datos y su función:** esencial para prevenir errores humanos y malas prácticas que suelen ser causas comunes de brechas cibernéticas. Los empleados necesitan comprender los riesgos de no seguir protocolos de privacidad de datos, y para ello, herramientas como la capacitación anual en ciberseguridad, simulaciones de phishing y una comunicación constante sobre los beneficios para el negocio son fundamentales.
5. **Superar lo mínimo necesario:** cumplir con las leyes locales es esencial para evitar multas y daños a la reputación, pero enfocarse únicamente en el cumplimiento puede perder oportunidades para destacar. Al integrar conscientemente la privacidad de datos en productos, servicios y procesos comerciales, el enfoque cambia de lo permitido a lo que genera valor para los clientes.

4. Acceso y uso responsable de datos

Por el momento, **Classroom AI** se encuentra en la fase de prototipo, por lo que se trabaja con un conjunto de datos reducido. La prioridad de los **Caballeros de Camelot** es cumplir con los requerimientos establecidos en el  SRS ,

garantizando la funcionalidad deseada por el cliente. Finalizado el prototipo se podría establecer un proceso más claro sobre cómo se puede trabajar con el set de datos futuro y especificar aspectos como: dónde se podrá almacenar, en qué tipo de redes estará disponible, quién lo podrá ver y cuáles son los documentos o normas que se deben firmar antes de poder acceder a los datos. A continuación se realiza una propuesta de dicho proceso considerando que el **responsable** del sistema en producción es NDS Cognitive Labs y su **cliente** es el Tecnológico de Monterrey Campus Monterrey.

El **objetivo** de este proceso es garantizar la seguridad y privacidad de los datos confidenciales de los estudiantes y profesores del Tecnológico de Monterrey Campus Monterrey. Por otro lado, las **responsabilidades** propuestas para este proceso son las siguientes:

- **Equipo de datos confidenciales:** equipo designado por NDS Cognitive Labs para gestionar y proteger los datos confidenciales de los miembros de la institución educativa.
- **Usuarios autorizados:** sólo los empleados de NDS Cognitive Labs que ofrezcan el producto y personal externo autorizado, en este caso los profesores, podrían acceder a los datos confidenciales.

Tomando en cuenta este objetivo y responsabilidades, se propone el siguiente proceso de gestión de datos confidenciales:

- **Almacenamiento de Datos:** los datos confidenciales deben ser almacenados exclusivamente en los servidores designados por NDS Cognitive Labs. No se debe permitir el almacenamiento en dispositivos personales o en servicios de una nube no autorizada.
- **Redes Aprobadas:** los datos confidenciales solo pueden estar en redes internas seguras de NDS Cognitive Labs. El acceso a través de redes públicas o no seguras está estrictamente prohibido.
- **Acceso Autorizado:** sólo los usuarios autorizados tienen permitido acceder a los datos confidenciales. El acceso se gestiona a través de cuentas y contraseñas únicas.
- **Documentación Requerida:** antes de acceder a los datos confidenciales, los usuarios autorizados deben firmar un acuerdo de confidencialidad y completar la formación obligatoria sobre la seguridad de datos.
- **Monitorización y Auditoría:** monitorización continua de las actividades relacionadas con los datos confidenciales y auditorías periódicas para garantizar el cumplimiento de la política de seguridad de datos.

5. Trazabilidad y auditabilidad del acceso a datos

Con el fin de incorporar un mecanismo que permita establecer registros sobre quién y cuándo tuvo acceso a los datos y bajo qué esquema se crea una nueva tabla en la base de datos llamada “Logs”, donde utilizamos el ID del usuario Estudiantes y Usuarios (Profesores) y los atributos LogInTime DATETIME y LogOutTime DATETIME.

Creando esta tabla lo que se quiere es registrar el momento en que un usuario entra o sale del sistema, tal que:

- UserID FK que referencia UsersProfessors (UserID) y Student ID FK que referencia Students (StudentID) identifica el tipo de usuario que inicia la sesión
- LogInTime registra la hora en la que se inicia una sesión
- LogOutTime registra la hora en la que se cerró la sesión

Cada vez que un usuario inicie la sesión en el sistema, se inserta un registro en esta tabla con la hora de inicio de sesión; esto se logra incorporar por medio de una consulta en SQL. También se crea un SELECT donde es posible consultar el tiempo conectado de ese usuario. Para calcular el tiempo de un usuario conectado se puede hacer una consulta que reste la hora de inicio con la hora de cierre para obtener la duración de la sesión; utilizar DATEDIFF en segundos del LogInTime y LogOutTime. Esta información se desplegará en un componente del front-end donde solo cuentas administradoras podrán observar el tiempo en el que los usuarios estuvieron dentro del sistema, así para llevar un registro de la información de utilización de la plataforma.

6. Reflexión Individual

Rodolfo Sandoval Schipper Para la entrega de privacidad y datos, decidimos implementar métodos de anonimización, trazabilidad, y asegurar la comunicación entre las distintas aplicaciones que se utilizan para mandar/recibir datos y permitir la funcionalidad en el sistema. Primero, se implementó un nuevo rol de Usuario en el que podemos esconder todos aquellos datos que consideramos personales. Datos como el Nombre de un Profesor/Estudiante, Apellidos, y Correos. El motivo principal se enfoca en ocultar información sensible sin afectar la funcionalidad del sistema. Nosotros también contamos con los servicios de Azure para mantener la Base de datos en un servidor de producción. Azure por default cuenta con encriptación SSL para verificar el acceso al servidor que almacena la base de datos con de acuerdo a la IP y cuenta utilizada para ingresar al portal. Se pueden implementar servicios como Azure Bastión o modificar el método de

autenticación para prevenir ataques al servidor. Sin embargo, consideramos que los servicios de Azure pueden mantener una seguridad estable en su entorno en la nube.

Con de acuerdo a la trazabilidad, hemos implementado una nueva tabla que se enfocara en tomar el tiempo en el que un registro de Usuario ingresa a la plataforma. Es decir, creamos una nueva tabla donde restamos el tiempo Ingreso y el tiempo de en el que se cierra la sesión de un Usuario. Todo esto con el fin de que una cuenta administradora pueda monitorear el acceso del sistema e identificar quién está utilizando el sistema. Esto no solo mejoraría la seguridad con dicho monitoreo, sino que también ayudaría a escalar el sistema de acuerdo a la cantidad de usuarios que se estén llevando a cabo en el registro.

Marcelo Márquez Para esta entrega y el desarrollo del proyecto, la importancia de la protección de datos y privacidad siempre fue de vital importancia. Eso incluye la anonimización de datos, mantener una infraestructura segura y monitoreo. Esta diferenciación no solo cumple con las normativas de protección de datos, sino que también minimiza el riesgo de mal uso de la información. Dado a esta estrategia, se preserva la operatividad, permitiendo que tareas como la creación de cursos se realicen sin comprometer la seguridad.

Como pilar para nuestra infraestructura de base de datos, optamos por utilizar los servicios de Azure. Esta plataforma provee encriptación SSL de serie, lo que garantiza una capa de seguridad robusta durante la autenticación y transmisión de datos. La utilización de múltiples capas de seguridad busca minimizar la superficie de ataque y reducir el riesgo de exposición de datos.

Arturo Garza Campuzano Durante el desarrollo de esta entrega, nos centramos en la anonimización de datos. Nuestro proyecto implica la aseguración de datos y debemos lograr este propósito que hemos establecido por medio de un enfoque en varios temas clave. Primero nos enfocamos en la anonimización de datos por medio de visualización en la interfaz. Para llevar a cabo la anonimización primero identificamos los datos personales en las tablas de la base de datos. Estos datos incluyen el nombre, apellidos, y correos.

Por el momento estamos en la fase de prototipo para Classroom AI. Sin embargo, queremos proponer un proceso de gestión de datos. Tal vez podamos llegar a un acuerdo para garantizar la seguridad de la información que se maneja en el sistema. Ya sea con de acuerdo al almacenamiento de datos, redes aprobadas, acceso autorizados, y monitorización. Por ejemplo el uso de una infraestructura segura en la nube y métodos de trazabilidad en el sistema. Para este caso creamos una nueva tabla donde llevamos a cabo y tomamos en cuenta el tiempo en el que

un Usuario inicia y cierra sesión. De esta manera podemos mejorar varios aspectos de monitoreo dentro del sistema y podemos llevar a cabo este tipo de registro para identificar los Usuarios que han ingresado a la plataforma.

7. Referencias bibliográficas

1. NDS Cognitive Labs. "Notice of Privacy." Ndsognitivelabs.com, ndscognitivelabs.com/notice-of-privacy/.
2. Mohamud, Rowda. "5 Steps to Ensuring Data Privacy in Your Business." BDC.ca, 20 Oct. 2022, www.bdc.ca/en/articles-tools/blog/5-steps-improving-data-privacy-in-business.