UNIVERSIDAD CATÓLICA DE CUENCA

Maestría en Ciberseguridad

Informe de Investigación previo a la obtención del título de Magíster en Ciberseguridad

Tema: "Auditoría de una aplicación de escritorio para la anonimización de información sensible de bases de datos de los hospitales de tercer nivel Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues".

Autor: Ing. Msc. Oscar Freed Carrera Pozo.

Asesores: Dr Vinicio Santillán

Cuenca, 2025

Certificación de Asesores

Se certifica que:

El informe de investigación "Auditoría de una aplicación de escritorio para la anonimización de información sensible de bases de datos de los hospitales de tercer nivel Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues". Esta auditoría fue realizada por el señor Ingeniero Oscar Freed Carrera Pozo CC: 1002140729 ecuatoriano, previo a la obtención del Título de Cuarto Nivel o Posgrado correspondiente a Magíster en Ciberseguridad, el cual cumple con la caracterización y estructura (parte protocolaria y parte expositiva) y se sujeta a la normativa pertinente exigida por el Consejo de Educación Superior, CES y la Universidad Católica de Cuenca, en consecuencia se autoriza su presentación para los trámites pertinentes.

Santa Ana de los Cuatro Ríos de Cuenca

Abril 2025

_____ _____

Dr Vinicio Santillán Asesor Científico

Ing. Juan Carlos Ortega Castro. Mg

Asesor Metodológico

Certificación de Autoría

Certifico que:

"Auditoría de una aplicación de escritorio para la anonimización de información sensible de bases de datos de los hospitales de tercer nivel Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues", es el tema del informe final de investigación de mi AUTORÍA, previo a la obtención del Título de Cuarto Nivel o Posgrado correspondiente a Magíster en Ciberseguridad, por lo que, asumo su originalidad y el uso de fuentes de terceros registrados según las normas APA vigentes.

Santa Ana de los Cuatro Ríos de Cuenca

Abril, 2025.

_____

Ing.Msc. Oscar Freed Carrera Pozo

CC: 1002140729

Agradecimiento

Quiero expresar mi más sincero agradecimiento a la **Universidad Católica de Cuenca**, institución que ha sido fundamental para mi desarrollo académico y profesional durante el transcurso de este proyecto de maestría en Ciberseguridad. Agradezco profundamente a los docentes de la universidad, quienes con su dedicación y conocimiento han aportado de manera significativa a mi formación, motivándome a enfrentar los retos con compromiso y determinación.

De manera especial, deseo reconocer y agradecer a mi asesor de proyecto, el **Dr. Vinicio Santillán**, por su guía, apoyo constante y valiosas orientaciones durante todo el desarrollo de este trabajo. Su experiencia y compromiso han sido claves para la ejecución y finalización exitosa del proyecto.

Asimismo, expreso mi gratitud a los hospitales de tercer nivel **Vicente Corral Moscoso de Cuenca** y **Homero Castanier de Azogues** por permitir la realización de la auditoría de su aplicación de escritorio para la anonimización de información sensible de sus bases de datos, contribuyendo así al avance de la seguridad de la información en el ámbito de la salud.

Este proyecto no habría sido posible sin el respaldo y la colaboración de todas estas personas e instituciones. A todos ustedes, mi más profundo agradecimiento.

Dedicatoria

Este proyecto de investigación está dedicado con todo mi amor a mis hijos **Steven y Dilan**, quienes son mi mayor inspiración y motivo para seguir adelante. Cada paso que doy y cada meta que alcanzo es pensando en su bienestar y en el ejemplo que quiero dejarles. Gracias por llenar mi vida de alegría y darme la fuerza necesaria para superar cualquier desafío.

También dedico este trabajo a mi madre, cuyo amor incondicional, sacrificio y apoyo han sido fundamentales en mi vida. Mamá, eres el pilar que me sostiene y el faro que guía mi camino. Este logro es tanto tuyo como mío, y no podría haberlo alcanzado sin tu constante fe en mí. A ustedes, mi familia, este esfuerzo y dedicación. Siempre serán mi mayor motor.

Resumen

El presente proyecto de investigaceión titulado **"Auditoría de una aplicación de escritorio para la anonimización de información sensible de bases de datos de los hospitales de tercer nivel Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues"** tiene como objetivo principal evaluar la eficacia y el cumplimiento normativo de una herramienta tecnológica diseñada para proteger la privacidad de los datos sensibles en el sector de la salud.

Contexto y Problemática

Los hospitales analizados manejan grandes volúmenes de información médica considerada crítica según la Ley Orgánica de Protección de Datos Personales (LOPDP) de Ecuador (Asamblea Nacional del Ecuador, 2021). Por tal razón, se construyó la APP de escritorio que será un gran aporte a estas instituciones con la finalidad de realizar investigaciones sobre los datos obtenidos. Debemos garantizar que los resultados obtenidos cumplan con los estándares de anonimización; para ello, se realizará la auditoría a los resultados entregados por el aplicativo (ISO, 2018).

Metodología

El estudio incluyó una auditoría a los resultados que entrega el software de anonimización, con énfasis en:

Análisis de los productos generados: Se procederá a realizar la comparación entre los datos originales y anonimizadas para identificar posibles errores y evaluar la precisión del proceso efectuado y sus resultados basados en el criterio de la no re-identificación de los datos.

Resultados Claves

- **Cumplimiento del protocolo**: El sistema de anonimización cumple con las normativas y protege eficazmente los datos sensibles procesados.
- **Eficiencia técnica**: Los datos anonimizados son útiles para análisis posteriores sin comprometer la privacidad.
- **Vulnerabilidades detectadas**: Se identificaron áreas de mejora en la verificación manual previa al procesamiento y en la capacitación del personal hospitalario.

Impacto y Contribuciones

Este proyecto refuerza la confianza en los sistemas hospitalarios al demostrar que las herramientas tecnológicas utilizadas cumplen con los estándares legales y protegen los derechos de los pacientes. Además, propone mejoras técnicas y operativas que pueden aplicarse en otras instituciones del sector salud.

Recomendaciones

1. Implementar auditorías periódicas del software.
2. Monitorear las regulaciones para mantener el cumplimiento normativo.

En conclusión, este trabajo contribuye significativamente a verificar el cumplimiento del proceso de anonimizacion de datos sensibles de salud. Las auditorias son consideradas una oportunidad de mejora en todo proceso auditado, mediante la ejecución de esta investigación se evalúa los

resultados entregados por el sistema de anonimizacion frente a los estándares en las buenas prácticas de anonimazion y lo estipulados en la ley orgánica de protección de datos personales.

Palabras claves:

Auditoría informática

Identificador

Cuasi Identificador

Anonimización de datos

Datos sensibles

Ciberseguridad

Protección de datos personales

LOPDP (Ley Orgánica de Protección de Datos Personales - Ecuador)

Re-identificación de datos

Base de datos hospitalarios

Aplicación de escritorio

Seguridad de la información

Normativa de privacidad

Ingeniería inversa

Evaluación de riesgos

Confidencialidad de la información

ISO 27001

HIPAA (Health Insurance Portability and Accountability Act)

GDPR (General Data Protection Regulation)

Cumplimiento normativo

Tercer nivel de atención hospitalaria

Software de anonimización

Abstract

The present research project, titled **"Audit of a Desktop Application for the Anonymization of Sensitive Information in Databases of Tertiary-Level Hospitals Vicente Corral Moscoso in Cuenca and Homero Castanier in Azogues,"** aims to evaluate the effectiveness and regulatory compliance of a technological tool designed to protect the privacy of sensitive data in the healthcare sector.

Context and Problem Statement

The analyzed hospitals manage large volumes of medical information classified as critical under Ecuador's **Organic Law on Personal Data Protection (LOPDP).** Despite the availability of an anonymization tool, there are associated risks, such as data re-identification, unauthorized access, and potential deficiencies in its implementation. This project addresses the need to audit the application to ensure security, functionality, and legal compliance.

Methodology

The study involved an audit of the anonymization software's outputs, focusing on:

1. **Analysis of the authorization procedure for anonymization.**
2. **Evaluation of the generated outputs:** Comparison between original and anonymized databases to identify errors and assess accuracy.
3. **Reverse engineering tests:** Assessment of system robustness against re-identification attempts.

Key Findings

- **Protocol compliance:** The anonymization system adheres to regulations and effectively protects processed sensitive data.
- **Technical efficiency:** The anonymized data remains useful for further analysis without compromising privacy.
- **Detected vulnerabilities:** Identified areas for improvement in manual verification prior to processing and in hospital staff training.

Impact and Contributions

This project strengthens confidence in hospital systems by demonstrating that the technological tools used comply with legal standards and safeguard patient rights. Additionally, it proposes technical and operational improvements that can be applied to other healthcare institutions.

Recommendations

1. Implement periodic audits of the software.
2. Improve the documentation of anonymization processes.
3. Monitor regulations to maintain compliance.

In conclusion, this study significantly contributes to the protection of sensitive personal data in the hospital sector, ensuring both privacy and usefulness for legitimate purposes.

Keywords:

IT Audit

Data Anonymization

Sensitive Data

Cybersecurity

Personal Data Protection

LOPDP (Organic Law on Personal Data Protection - Ecuador)

Data Re-identification

Hospital Databases

Desktop Application

Information Security

Privacy Regulations

Reverse Engineering

Risk Assessment

Information Confidentiality

ISO 27001

HIPAA (Health Insurance Portability and Accountability Act)

GDPR (General Data Protection Regulation)

Regulatory Compliance

Tertiary-Level Healthcare

Anonymization Software

Índice de contenidos

Contenido

Índice de tablas

Índice de Figuras

Capítulo I. Introducción

1.1 Situación problemática

Los centros hospitalarios Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues manejan grandes volúmenes de información sensible. Estos datos son considerados de alto nivel de criticidad según la Ley Orgánica de Protección de Datos Personales (LOPDP) de Ecuador, y están sujetos a estrictos requisitos de protección para su utilización mediante el mecanismo de anonimización, con el fin de prevenir la re-identificación de pacientes (Asamblea Nacional del Ecuador, 2021).

Con este contexto, se ha desarrollado una aplicación de escritorio basada en la utilización de inteligencia artificial para la selección de campos y la generación de consultas con el propósito de anonimizar datos sensibles. Surge la necesidad de evaluar la efectividad de esta herramienta en la entrega de resultados que garanticen el cumplimiento del proceso de anonimización, según se establece en las normativas vigentes (ISO, 2018).

La falta de una auditoría exhaustiva podría derivar en el incumplimiento de la normativa vigente, con posibles sanciones legales, daño reputacional y pérdida de confianza por parte de los pacientes y partes interesadas. Por lo tanto, la realización de una auditoría basada en los resultados del proceso de anonimización permitirá identificar vulnerabilidades, proponer mejoras y garantizar que la herramienta cumple su propósito sin permitir la re-identificación de los datos anonimizados (García et al., 2020).

1.2 Línea de Investigación

Tipo de línea	06 Tecnologías de la información y la comunicación (TIC)	Ciencia de los ordenadores, Analítica de datos y Algoritmos computacionales Auditoria y seguridad informática	
En desarrollo	Energía eléctrica y Tecnologías de la Información para la innovación y el desarrollo sostenible	Ciencia de los ordenadores, analítica de datos y algoritmos computacionales	☒
		Sistemas eléctricos de potencia, energía e iluminación	☐
		Modelado, automatización y control	☐

Tabla 1 línea de investigación

Este estudio permitirá la verificación de la correcta aplicación de la ley orgánica de protección de datos personales en el ecuador en el sector salud, mediante la auditoria a los resultados entregados por la APP de escritorio de anonimizacion de datos sensibles.

1.3 Objeto del estudio

El objeto de estudio es el análisis del resultado que la aplicación de escritorio diseñada para la anonimización de información sensible en las bases de datos de los hospitales de tercer nivel Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues, cumpla con los requerimientos técnicos en el tratamiento de Identificadores y cuasi identificadores.

1.4 Campo de acción

Auditoría a los resultados arrogados por parte de la aplicación de escritorio en el proceso de anonimazion.

1.5 Eficiencia técnica

Analizar la funcionalidad de los resultados y el rendimiento de la aplicación para procesar datos de manera efectiva sin afectar la privacidad de la información de los pacientes.

1.6 Objetivos

1.6.1 General

Auditar los resultados generados por la APP de escritorio diseñada para la anonimazion de información sensible de bases de datos de salud, mediante el análisis del proceso de anonimización, y la validación de los resultados obtenidos, asegurando el cumplimiento de la Ley orgánica de protección de datos personales.

1.6.2 Específicos

- Analizar los productos generados por la APP de escritorio, evaluando la exactitud y precisión del sistema de anonimización, valorando la calidad del proceso de identificación y anonimización de los campos sensibles en bases de datos de salud.
- Evaluar la robustez del proceso del anonimización, mediante técnicas de ingeniería inversa, utilizando un software de análisis con el fin de garantizar la privacidad de los datos anonimizados.

1.7 Problema Científico Preguntas de Investigación

El problema científico radica en determinar si la APP de escritorio utilizada para la anonimización de información sensible en las bases de datos de los hospitales Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues es capaz de garantizar un tratamiento efectivo en el proceso de anonimización. La anonimización de datos es un requisito fundamental para la protección de la privacidad y la seguridad de la información en el ámbito de la salud, conforme a la Ley Orgánica

de Protección de Datos Personales (LOPDP) de Ecuador (Asamblea Nacional del Ecuador, 2021) y los estándares internacionales establecidos por la Organización Internacional de Normalización (ISO, 2018).

El problema científico, por tanto, se centra en evaluar los resultados entregados por la herramienta desde una perspectiva técnica, normativa y operativa, con el objetivo de determinar si es capaz de garantizar la protección efectiva de los datos sensibles anonimizados. Esto implica el cumplimiento de los estándares legales y técnicos necesarios para un manejo seguro y confiable de la información, minimizando el riesgo de re-identificación de los datos anonimizados (García et al., 2020).

1.8 Preguntas científicas

- ¿Los datos anonimizados mantienen su utilidad para análisis estadísticos y médicos?
- ¿Qué nivel de resistencia tiene el software frente a técnicas de ingeniería inversa o re-identificación?

1.9 Hipótesis

1.9.1 Principal

El software de anonimización entrega un reporte que cumpla con los estándares de la Ley orgánica de protección de datos personales del ecuador.

1.9.2 Secundarias:

Los productos anonimizados mantienen una calidad adecuada para su uso en investigaciones y análisis médicos.

1.10 Variables

A continuación realizamos el análisis de la variable dependiente e independiente el mismo que se hace referencia en la Tabla 2.

1.10.1 Independiente: Proceso de anonimización empleado por el software.
1.10.2 Dependiente: Calidad, precisión, utilidad y seguridad de los datos anonimizados.

Conceptualización	Unidad de medida	Instrumento
1. Anonimización de Datos: • **Definición:** Proceso de modificar datos para que no puedan ser usados para identificar a personas individuales. Implica técnicas como k-anonimato, l-diversidad, y t-closeness. • **Objetivo:** Garantizar la privacidad de los pacientes al compartir datos para investigación, cumpliendo con normativas legales como la LOPDP.	**1. Efectividad de Anonimización:** • **Unidad de Medida:** Nivel de protección alcanzado, medido en términos de reducción del riesgo de re-identificación.	**1. Cuestionarios y Encuestas:** • **Uso**: Evaluar la percepción y la satisfacción de los usuarios con respecto a la anonimización de datos y su utilidad en aplicaciones prácticas.
2. Privacidad de Datos Médicos: • **Definición:** Protección de la información sensible contenida en los registros	**2. Calidad de Datos Anonimizados:** • **Unidad de Medida:** Grado de utilidad de los datos anonimizados para la investigación.	**2. Pruebas de Anonimización:** • **Uso:** Validar la eficacia del esquema de anonimización implementado.

médicos para evitar la re-identificación de pacientes. • **Objetivo:** Prevenir violaciones de privacidad y asegurar la confidencialidad de la información médica.	• **Ejemplo:** Precisión, integridad, y relevancia de los datos en aplicaciones prácticas.	• **Ejemplo:** Aplicación de técnicas de k-anonimato y l-diversidad a conjuntos de datos para medir la reducción del riesgo de re-identificación.
3. Cumplimiento Normativo: • Definición: Adherirse a las regulaciones establecidas por la Ley Orgánica de Protección de Datos Personales (LOPDP) en Ecuador. • Objetivo: Asegurar que las prácticas de manejo de datos sean legales y protejan los derechos de los pacientes.	**3. Cumplimiento Normativo:** • Unidad de Medida: Grado de conformidad con la LOPDP. • Ejemplo: Porcentaje de requisitos legales cumplidos según auditorías y evaluaciones de cumplimiento.	**Auditorías de Cumplimiento:** • Uso: Verificar la conformidad con la LOPDP y otros estándares legales. • Ejemplo: Revisiones documentales y auditorías realizadas por un equipo legal o de cumplimiento normativo.

Tabla 2 Análisis, de la variable dependiente e independiente

1.11 Justificación – contribuciones de la investigación

El proceso de investigación sobre la auditoría de los resultados proporcionados por la aplicación de escritorio en el manejo de anonimización de información sensible en los hospitales Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues es de gran importancia científica, técnica y social. Los resultados de esta auditoría ofrecerán una oportunidad para mejorar el proceso y garantizar la privacidad de la información, en línea con los principios de confidencialidad, integridad y prevención de reidentificación establecidos por la Ley Orgánica de Protección de Datos Personales del Ecuador (2021) y el Reglamento General de Protección de Datos (RGPD, 2016). En el ámbito social, la investigación permite la posibilidad de realizar estudios científicos sobre datos de salud, generando confianza entre los pacientes y otras partes interesadas de que su información personal está protegida contra usos indebidos o accesos no autorizados (El Emam, 2010), y que no podrán ser reidentificados a partir de los datos anonimizados (Article 29 Data Protection Working Party, 2014).

1.12Estado del arte o antecedentes

La anonimización de datos sensibles en salud es un tema crucial en la investigación médica, ya que permite el uso de información clínica de manera ética y segura para generar conocimiento científico sin poner en riesgo la privacidad de los pacientes y su re-identificación.

Ley Orgánica de Protección de Datos Personales (LOPDP), promulgada en 2021, establece un marco normativo que asegura la protección de datos personales sensibles, especialmente en sectores críticos como el de la salud. Esta ley fomenta el uso de técnicas de anonimización como un medio para tratar datos sensibles con fines legítimos, como la investigación científica en medicina, mientras se respeta la privacidad de los titulares. A nivel local, hospitales de tercer nivel, como el Vicente Corral Moscoso y el Homero Castanier, enfrentan el reto de implementar sistemas

de anonimización seguros y efectivos que permitan el uso de datos clínicos anonimizados en investigaciones científicas sin comprometer la seguridad ni violar la ley orgánica de protección de datos personales del Ecuador. La auditoría de esta herramienta es fundamental para asegurar un sistema confiable que promueva el avance científico y el respeto por los derechos de los pacientes en cuanto a la privacidad de sus datos personales.

1.13 Marco teórico referencial

El presente marco teórico aborda los conceptos fundamentales, normativas legales, y prácticas tecnológicas relevantes para el desarrollo de una auditoría a los resultados entregados por la aplicación en el proceso de anonimazion de datos sensibles de salud. Esta investigación se centra en los hospitales d Vicente Corral Moscoso de Cuenca y Homero Castanier de Azogues.

1.14 Conceptos Claves

1.14.1 Anonimización de Datos

La anonimización es el proceso mediante el cual los datos personales se transforman para que no se pueda identificar directa o indirectamente a un individuo. Este proceso es crucial en entornos hospitalarios, donde se maneja información altamente sensible, como historiales clínicos y datos demográficos de los pacientes.

1.14.2 Auditoría de Sistemas

La auditoría de sistemas es el proceso de evaluación y análisis de los sistemas de información para determinar su conformidad con normas, reglamentos, y mejores prácticas.

1.14.3 Bases de Datos Hospitalarias

Estas bases contienen información crítica sobre pacientes, tratamientos, diagnósticos, y recursos médicos. La anonimización de estos datos es esencial para proteger la privacidad de los pacientes y facilitar el cumplimiento normativo.

1.15 Marco Legal

1.15.1 Ley Orgánica de Protección de Datos Personales (LOPDP)

La LOPDP, promulgada en el Ecuador en 2021, establece los principios, derechos, y obligaciones relacionados con el tratamiento de datos personales. Los aspectos más relevantes para esta investigación incluyen:

- **Principio de minimización**: Los datos recolectados deben ser adecuados, pertinentes, y limitados a lo necesario.
- **Anonimización y pseudonimización**: Estas técnicas son promovidas como mecanismos para mitigar riesgos de privacidad.
- **Derechos de los titulares**: Derecho al acceso, rectificación, y eliminación de datos personales.

1.15.2 Normas Internacionales Relacionadas

1.15.2.1 Reglamento General de Protección de Datos (GDPR)

El Reglamento General de Protección de Datos (GDPR) es una normativa de la Unión Europea que comenzó a aplicarse el 25 de mayo de 2018, estableciendo un marco legal estricto para la protección de los datos personales de los ciudadanos europeos. La influencia del GDPR en la legislación de varios países ha sido notable, incluyendo a Ecuador. En 2021, Ecuador aprobó la Ley Orgánica de Protección de Datos Personales (LOPDP), que incorpora muchas de las disposiciones del GDPR, adaptándolas al contexto ecuatoriano. La LOPDP establece principios

similares a los del reglamento europeo y reconoce derechos equivalentes para los ciudadanos, como la autodeterminación informativa, la rectificación y la eliminación de datos personales.

Uno de los aspectos más destacados de la influencia del GDPR en la LOPDP es la creación de la Autoridad de protección de datos en Ecuador, un organismo regulador encargado de asegurar el cumplimiento de la ley, supervisar el tratamiento de datos personales y sancionar infracciones. Además, la LOPDP introduce mecanismos como la obligación de notificar violaciones de seguridad, la realización de evaluaciones de impacto en la privacidad y la imposición de sanciones significativas en caso de incumplimiento, aspectos que también son fundamentales en el GDPR.

1.15.2.2 ISO 27001

La ISO 27001 es un estándar internacional para la gestión de la seguridad de la información, publicado por la Organización Internacional de Normalización (ISO). Su propósito es establecer un Sistema de Gestión de Seguridad de la Información (SGSI) que proteja la confidencialidad, integridad y disponibilidad de los datos dentro de una organización. Para lograrlo, define un conjunto de controles de seguridad y un marco de gestión basado en el análisis de riesgos. Su implementación contribuye a prevenir incidentes de seguridad y a garantizar el cumplimiento normativo.

Por otro lado, la ISO 27701 es una extensión de la ISO 27001 que introduce directrices específicas para la gestión de la privacidad y la protección de datos personales. Su propósito es proporcionar un Sistema de Gestión de la Información de Privacidad (SGIP), alineado con regulaciones como el GDPR y otras normativas de protección de datos. Esta norma ayuda a las organizaciones a

gestionar adecuadamente la información personal que procesan, fortaleciendo la confianza y reduciendo riesgos legales.

Ambas normas son complementarias y su adopción conjunta permite a las empresas fortalecer su seguridad y privacidad, alineándose con estándares globales.

1.16 Relevancia de la Investigación

Este proyecto fue generado con el propósito de auditar los resultados del proceso de anonimizacion de los datos de salud entregados por la APP de escritorio desarrollada con la finalidad de cumplir con los estándares normativos para el tratamiento de datos sensibles.

Lo relevante de este proyecto está enfocado en dos ejes:

- El eje social que permitirá el manejo de estos datos de carácter sensible en beneficio de la ciencia al poder ejecutar proyectos científicos en el área de la medicina, como un aporte a la ciencia e investigación
- En el campo tecnológico, lo innovador es la utilización de la Inteligencia artificial en la etapa de análisis de los datos de origen donde se clasifican los encabezados y mediante una retroalimentación selecciona cuales pueden ser considerados datos sensibles de esta estructura. De igual forma se genera mediante Inteligencia artificial la consulta Sql que permita la clasificación de los datos a ser tratados. En este método se aplica Exactitud es la toma del campo correcto y precisión es que esos datos son anonimizados.

De igual forma el proyecto cumple con lo dispuesto en la Ley Orgánica de Protección de Datos personales del ecuador.

1.17 Procedimientos Éticos

En el contexto de la auditoría de los resultados de una aplicación de escritorio destinada a la anonimización de datos sensibles de los hospitales Vicente Corral Moscoso y Homero Castanier, se deben cumplir estrictos procedimientos éticos, alineados con la Ley Orgánica de Protección de Datos Personales (LOPDP) del Ecuador y principios internacionales de protección de datos.

1.17.1 Principio de minimización

Utilizar únicamente los datos estrictamente necesarios para la investigación.

1.17.2 Confidencialidad

Firmar acuerdos de confidencialidad con los solicitantes.

1.17.3 Difusión ética

Publicar resultados de la investigación de forma anonimizada, protegiendo la identidad de los pacientes y las instituciones. Adoptar estos procedimientos asegura el respeto a los derechos de los titulares de los datos, protege su privacidad y garantiza el cumplimiento ético y legal del proyecto.

1.18 Fundamentación teórica

La protección de datos personales es un derecho fundamental que está garantizado por la Ley Orgánica de Protección de Datos Personales (LOPDP). Esta normativa regula cómo se deben tratar los datos personales, estableciendo principios como la legalidad, la transparencia, la minimización, la integridad y la confidencialidad (República del Ecuador, 2021). En el ámbito de la salud, es crucial anonimizar la información sensible para proteger los derechos de los titulares y permitir un uso ético y seguro de los datos en investigaciones.

La norma ISO/IEC 27001:2022, reconocida internacionalmente para la gestión de la seguridad de la información, ofrece un marco estructurado para identificar, evaluar y gestionar los riesgos asociados al tratamiento de datos personales. Su implementación en proyectos, como la auditoría de aplicaciones de escritorio para la anonimización en bases de datos de hospitales de tercer nivel, asegura que se cumplan los requisitos legales y técnicos, minimizando riesgos y garantizando la conformidad normativa.

Capítulo II. Diagnóstico situacional

2.1 Metodología

Análisis situacional se identificará el estado actual de los resultados del tratamiento de datos sensibles en las bases de datos de los hospitales Vicente Corral Moscoso y Homero Castanier. EL proceso de auditoria tendrá los siguientes pasos basados en la norma ISO 19100, el cual se describe en la Figura 1, donde se visualiza el flujo de pasos a ejecutar.

Se procederá a analizar los productos generados por la aplicación de escritorio, evaluando la exactitud y precisión del sistema de anonimización, valorando la calidad del proceso de identificación y anonimización de los campos sensibles en bases de datos de salud, pare ello debemos cumplir con las etapas indicadas en la Figura 1.

Figura 1 Nos ilustra la planificación del examen de auditoria basados en la metodología ISO 19001

Hay que tener tomar en cuenta que cundo nos referimos a exactitud es la selección de los campos que se considera sensible o que podrían ser claves para una re-identificación , y cuando hablamos de precisión nos referimos a que la data de estos campos no pueda ser re identificada ni individual ni grupal mente. Debiendo recordar que este proceso se lo realiza mediante la utilización de la IA (Inteligencia Artificial).

2.1.1 Análisis de los productos generados

Para realizar este análisis se procederá cumpliendo las siguientes actividades

a) **Identificar los productos**: Revisar qué genera la aplicación (Reportes anonimizados)

b) **Comparar datos**: Comparar datos de origen con la anonimizada para verificar:

   - Campos sensibles que se han modificado.
   - Si hay datos no sensibles afectados.

c) **Detectar errores**:

   - Verificar si algún dato sensible quedó sin anonimizar.
   - Comprobar si la estructura general de los datos permanece intacta.

d) **Exactitud**: Revisar si el sistema detecta correctamente todos los datos sensibles, como nombres, CC, direcciones, historial clínico.

e) **Precisión**: Asegurar que solo se anonimicen los datos sensibles necesarios.

2.1.2 Evaluar la robustez del proceso del anonimización, mediante técnicas de ingeniería inversa utilizando un software de análisis con el fin de garantizar la privacidad de los datos anonimizados.

Para lo cual debemos realizar lo siguiente:

a) **Datos iniciales**:

- Obtener la hoja de datos origen y su correspondiente reporte final con los datos anonimizados.

b) **Definición de campos sensibles**:

- Identificar los campos que deberían estar protegidos según la normativa (nombres, direcciones, identificadores personales).

2.1.2.1 Realizar técnicas de ingeniería inversa

a) **Objetivo de la ingeniería inversa**:

- Evaluar si los datos anonimizados pueden ser re-identificados o reconstruidos, ya sea parcial o totalmente.

En esta etapa se utilizara una herramienta de software y análisis físico con los datos de origen y datos del resultado.

2.1.2.2 Intento de reconstrucción directa

- Comparar la base anonimizada con datos públicos o bases externas
- Probar reconstrucciones utilizando coincidencias, como:
  - Fechas de nacimiento combinadas con iniciales.
  - Geolocalización con identificadores reducidos.

2.1.2.3 Análisis de correlaciones entre campos

- Detectar correlaciones que puedan servir para identificar registros:
    - Ejemplo: Si dos campos anonimizados están relacionados (edad y salario), deducir el valor de uno basándose en el otro.
- Evaluar si la anonimización preserva vínculos innecesarios que debiliten la protección.

2.2 Análisis situacional

El proyecto se enfoca en evaluar la efectividad y el cumplimiento de las normativas relacionadas con los resultados generados por una aplicación de escritorio que tiene como objetivo anonimizar información sensible en bases de datos de hospitales. Este enfoque es esencial en un contexto donde la protección de datos personales y la confidencialidad médica son aspectos clave y fundamentales para evitar la reidentificación de la información utilizada en los proyectos de investigación científica (Ley Orgánica de Protección de Datos Personales, 2021; Reglamento General de Protección de Datos [RGPD], 2016). La anonimización adecuada de los datos garantiza no solo el cumplimiento legal, sino también la protección de los derechos fundamentales de los titulares, especialmente en el ámbito de la salud (El Emam, 2010).

2.3 Análisis comparativo

El análisis comparativo permite contrastar diferentes aspectos relacionados con el proyecto para identificar fortalezas, debilidades y su posicionamiento frente a otros enfoques o tecnologías similares. A continuación, se presenta un análisis estructurado en torno a varios criterios clave.

2.3.1 Comparación técnica

- Enfoque: Auditoría de los resultados entregados por la herramienta específica para anonimización de datos sensibles en bases de datos hospitalarias.

- Tecnología: Aplicación de escritorio con K- anonimización, y utilización de Has en el campo historia clínica.
- Alcance: Orientada a hospitales de tercer nivel, con necesidades avanzadas de protección de datos.

Dentro de este análisis comparativo de la herramienta diseñada, con otras herramientas de corte internacional o software libre hemos fundamentado algunos criterios entre ellos el contexto hospitalario el cual se describe en la Tabla 3, este análisis también se lo llevo a cabo en referencia al componente seguridad referenciado en la Tabla 4, otro aspecto fundamental es el económico el cual se describe en la tabla 5, un factor importante dentro del análisis es la adaptabilidad es decir la facilidad de utilización de la herramienta para generar los resultados esperados el cual se describe en la Tabla 6, y por ultimo realizamos un análisis de sostenibilidad es decir que el proyecto perdure en el tiempo lo podemos mirar en la Tabla 7.

2.3.1.1 **Comparación de impacto en el contexto hospitalario**

Criterios	Proyecto actual	Soluciones comparables
Especialización	Centrada en las necesidades específicas de los hospitales ecuatorianos.	Enfoque genérico o internacional, requiere adaptación local.
Cumplimiento normativo	Diseñada para alinearse con la LOPDP y normativas internacionales	Cumplen con estándares internacionales como el RGPD.
Escalabilidad	Limitada a los hospitales seleccionados (Vicente Corral y Homero Castanier).	Mayor alcance geográfico y funcional.
Facilidad de uso	Aplicación diseñada para usuarios no técnicos.	Algunas soluciones requieren conocimientos avanzados.

Tabla 3. Se analiza las necesidades del sistema hospitalario ecuatoriano, en relación al cumplimiento de la LOPDP Comparación de impacto en el contexto hospitalario.

2.3.1.2 **Comparación en términos de seguridad**

Aspectos	Proyecto actual	Soluciones comparables
Robustez de anonimización	Dependiente del algoritmo implementado.	Uso de algoritmos avanzados en sistemas internacionales.
Riesgo de re-identificación	Evaluación enfocada en minimizar este riesgo.	Puede ser mayor si la implementación no se adapta al contexto.
Actualizaciones y parches	Podría depender del equipo local de desarrollo.	Actualizaciones frecuentes de proveedores internacionales.

Tabla 4 Otro de los pilares dentro del análisis es el relacionado con la relacionado con la seguridad enfocado al riego de re-identificación.

2.3.1.3 **Comparación económica**

Aspectos económicos	Proyecto actual	Soluciones comparables
Coste inicial	Más bajo, ya que la auditoría utiliza software existente.	Puede ser alto si se adquieren soluciones comerciales.
Mantenimiento	Depende de los recursos locales disponibles.	Requiere contratos de soporte o licencias.
Retorno de inversión (ROI)	Enfocado en mejorar la seguridad y evitar sanciones.	Similar, pero con costos recurrentes más altos.

Tabla 5 El aspecto económico es fundamental en cualquier proyecto aquí analizamos algunos puntos relevantes.

2.3.1.4 **Comparación en términos de adaptabilidad**

Aspectos	Proyecto actual	Soluciones comparables
Adaptación cultural y legal	Alta, diseñada para el contexto ecuatoriano, interfaz intuitiva y fácil de usar	Baja, requiere modificaciones para cumplir con la LOPDP.
Personalización	Posible debido a su diseño local.	Limitada en soluciones comerciales o de código cerrado.
Interoperabilidad	Compatible con sistemas hospitalarios locales.	Algunas soluciones podrían ser incompatibles.

Tabla 6 Comparación en términos de adaptabilidad, uno de los mayores retos dentro del diseño y desarrollo de software de calidad esta su fácil uso y comprensión, aquí algunos aspectos relevantes a tomar en cuenta.

2.3.1.5 Análisis de sostenibilidad

Aspectos	Proyecto actual	Soluciones comparables
Sostenibilidad a largo plazo	Alta si se asegura el soporte técnico local.	Media, dependiendo del proveedor y los costos recurrentes.
Impacto ambiental	Bajo, al ser una solución de escritorio local.	Variable, especialmente en soluciones basadas en la nube.

Tabla 7 Análisis de sostenibilidad, todo proyecto debe ser sostenible en el tiempo, y esto se fundamenta en la continuidad de la herramienta, por su utilización y fácil uso, así como los resultados de calidad que se entregan.

Capítulo III. Propuesta

3.1 Introducción a la propuesta

La presente propuesta tiene como objetivo auditar el cumplimiento los resultados del proceso de anonimización de información sensible en las bases de datos con información médica, mediante la revisión de los datos de entrada y los resultados generados por la aplicación de escritorio. Esta auditoría se llevará a cabo bajo los siguientes parámetros:

3.2 Objetivo del Plan de Auditoría

Evaluar la efectividad, fiabilidad y conformidad del software de anonimización de datos médicos con respecto a:

- La calidad de los resultados obtenidos en comparación con los datos originales.
- La verificación de que los resultados anonimizados no permitan la re-identificación de las personas.

3.2.1 Alcance

La auditoría a los resultados entregados por la aplicación de escritorio para la anonimazion de datos es de la data de los hospitales, y bajo el cumplimiento de la Ley Orgánica de protección de datos personales del ecuador.

3.2.2 Plan de auditoria

Toda auditoria para su ejecución nace con un plan de auditoria, en el cual se describen los pasos a seguir para cumplir con el objetivo previsto, esto se detalla en la Tabla 8, con las actividades y sus tiempos de ejecución.

Fase	Actividad	Responsable	Duración
Preparación	Emisión de la notificación de inicio de la auditoría.	Auditor principal	1 día
	Requerimiento de información necesaria para la auditoría.	Auditor principal	2 días
Recolección y Evaluación	Revisión de documentación técnica procesos	Auditor principal	3 días
	Comparación de resultados anonimizados con datos originales.	Auditor principal	3 días
	Evaluación del riesgo de re-identificación .	Auditor principal	2 días
Informe	Redacción del informe borrador.	Auditor principal	3 días
	Revisión de descargos y aclaraciones del auditado.	Auditor principal	2 días
	Emisión del informe final.	Auditor principal	1 día

Tabla 8 Cronograma Propuesto, para el cumplimiento de la auditoria de los resultados entregados por el software de escritorio, que permite anonimizar datos de salud.

Duración total: 17 días.

3.3 Desarrollo de la propuesta

3.3.1 Notificación de inicio de examen:

Se procede a notificar el inicio del examen al responsable del proyecto de anonimizacion de datos médicos (Anexo 1)

3.4 Sistema de anonimizacion.

Al ejecutar el sistema de escritorio que permite la anaonimazion de información médica, sensible encontraremos como primera pantalla la gráfica que se describe la Figura 2, una interface amigable y fácil de utilizar.

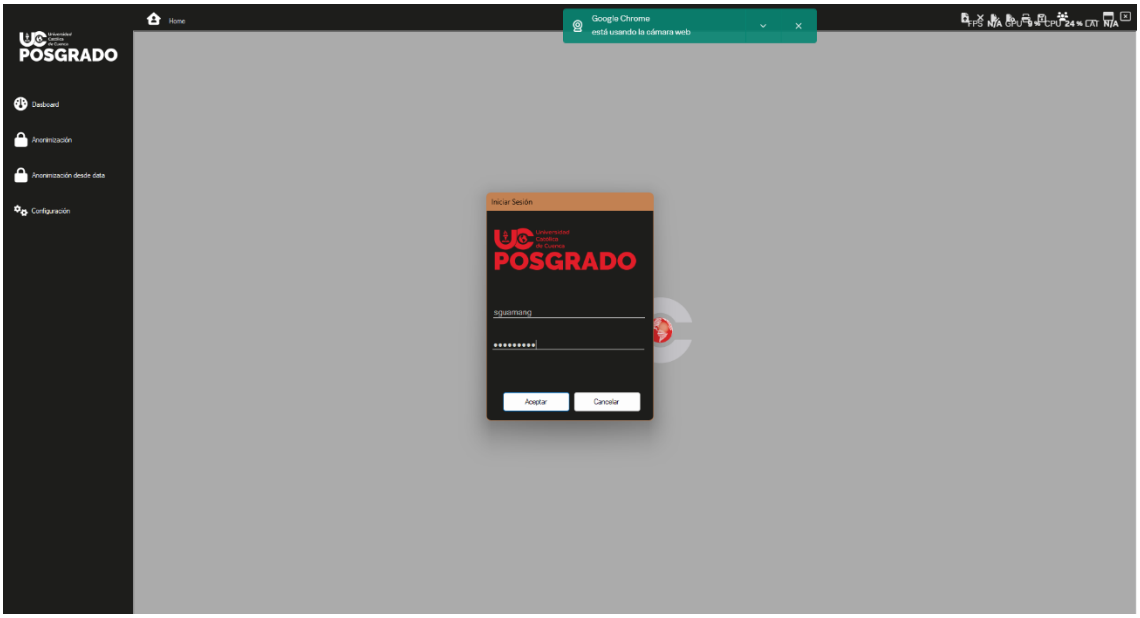


Figura 2 Pantalla de inicio de sesión del sistema de anonimizacion de datos de salud.

3.4.1 Módulo Anonimización

El módulo de anonimización está diseñado para proteger los datos sensibles mediante el método de eliminación. El programa es compatible con múltiples sistemas de gestión de bases de datos, incluyendo MySQL, SQL Server y Oracle. Cada institución puede gestionar varias bases de datos, y cada una de ellas puede contener múltiples servicios para la generación de reportes.

Para la creación de reportes, se utiliza la API de ChatGPT preentrenada, que transforma instrucciones en lenguaje natural en sentencias SQL. Antes de ejecutar estas sentencias, la consulta se limita a los primeros 100 registros para optimizar el rendimiento. La sentencia generada se envía

a la capa de negocio (Business), encargada de redirigir la consulta a la base de datos correspondiente según el servicio seleccionado.

La capa de datos se conecta a la base de datos, ejecuta la consulta SQL y devuelve los resultados en formato DataTable a la interfaz de usuario. A partir de estos datos, se extraen los encabezados para identificar posibles campos sensibles, utilizando nuevamente la API de ChatGPT. Esta operación se basa en una lista predefinida de campos sensibles y aplica reglas que permiten detectar palabras clave en diversas columnas.

Tras recibir los primeros 100 registros, se realiza una validación de los datos sensibles identificados. Los datos clasificados como sensibles se encriptan. Para mantener la integridad de la información y permitir la relación de sujetos de estudio entre diferentes reportes, se genera un código único por paciente basado en su historial clínico y una palabra secreta predefinida. Este proceso se desarrolla en los siguientes pasos:

- Creación de un hash utilizando el algoritmo *SHAH256*.
- Conversión del número de historial clínico a formato de bytes.
- Transformación del hash a un formato hexadecimal legible, en minúsculas.
- Conversión del código único del paciente a formato Base64, garantizando su consistencia en diferentes reportes.

3.4.2 Exportación de Datos

Una vez validados los datos, el programa permite exportar la información en formato .csv.
Durante este proceso, se solicitarán los siguientes datos del investigador:

- Investigador.
- Título, institución (nuevo requerimiento) y descripción del estudio Figura 6.

Además, se guardarán metadatos relacionados con el proceso de anonimización, que incluyen:

- Técnica de anonimización utilizada.
- Número de registros encontrados.
- Número total de columnas.
- Número de columnas clasificadas como sensibles.
- Listado de campos eliminados.

Antes de finalizar la exportación, el sistema solicitará al operador que seleccione la ubicación para almacenar el archivo .csv, esto lo podemos ver de forma ilustrativa en la Figura 4.

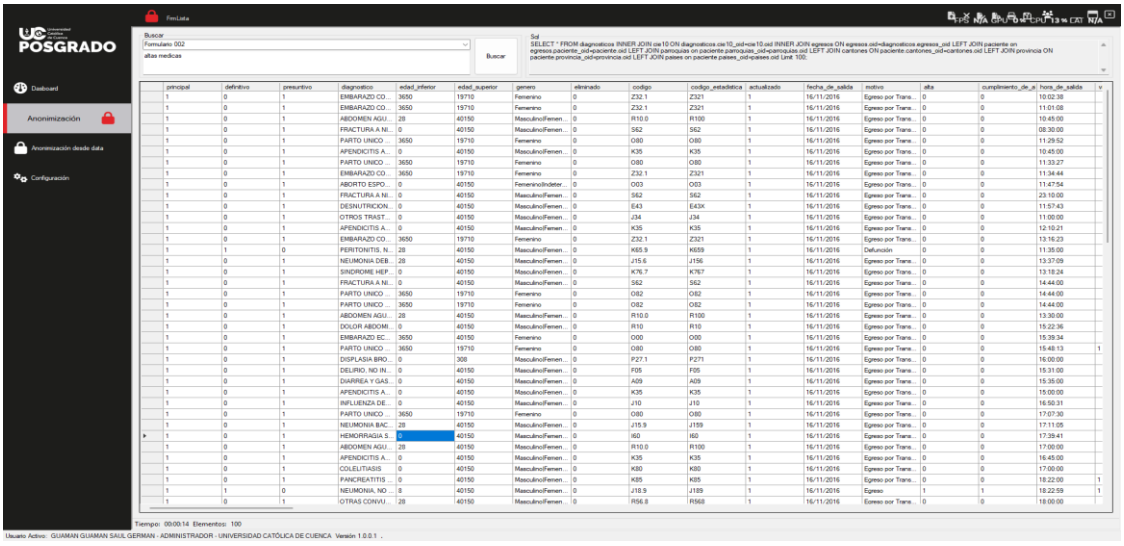


Figura 3 podemos ver como el sistema realiza una presentación de campos NO SENSIBLES, de la data cargada para el análisis.

De igual forma el sistema cuenta con una vista que indica los campos sensibles que van a ser tratados, esto lo podemos ver en la Figura 4.

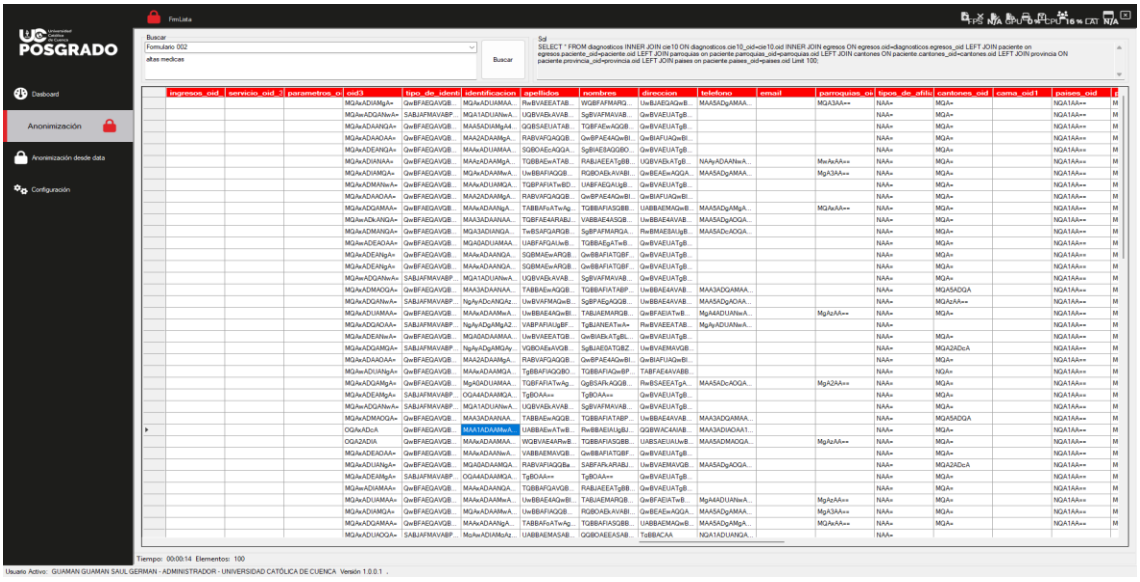


Figura 4 Lista de presentación de campos sensibles, son considerados los datos a ser tratados en el proceso de anonimizacion con la finalidad de no permitir su re-identificación.

Dentro de un conjunto de datos homogéneos y si estamos hablando de bases de datos estandarizadas se deberá contar con un campo único considerado clave primaria o único esto lo podemos ver en la Figura 5.

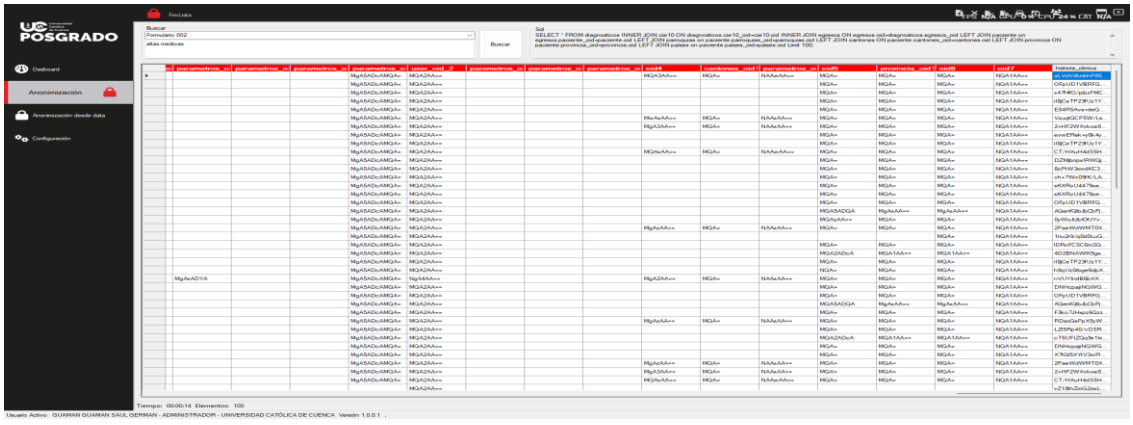


Figura 5 Campo único de Historial Clínico, este campo es considerado clave primaria el de mayor valor dentro de un proceso de re-identificación de datos, debido a que liga toda la información de un paciente.

Luego de realizar el proceso de anonimizacion de datos sensibles de salud, procedemos a guardar estos datos, asignado un nombre del investigador responsable, nombre del estudio, y una breve descripción de este como se muestra en la Figura 6

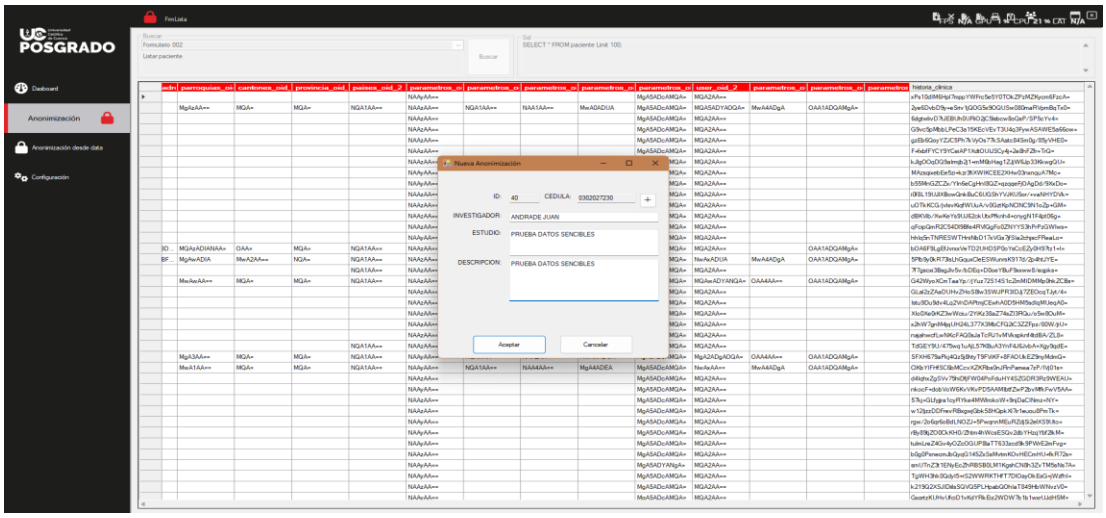


Figura 6 Selección y guardo datos de la investigación, luego del proceso de anonimizacion

El proceso de entrega de resultados de los resultados emitidos por el sistema de anonimizacionse almacena en formato csv lo cual se indica en la Figura 7

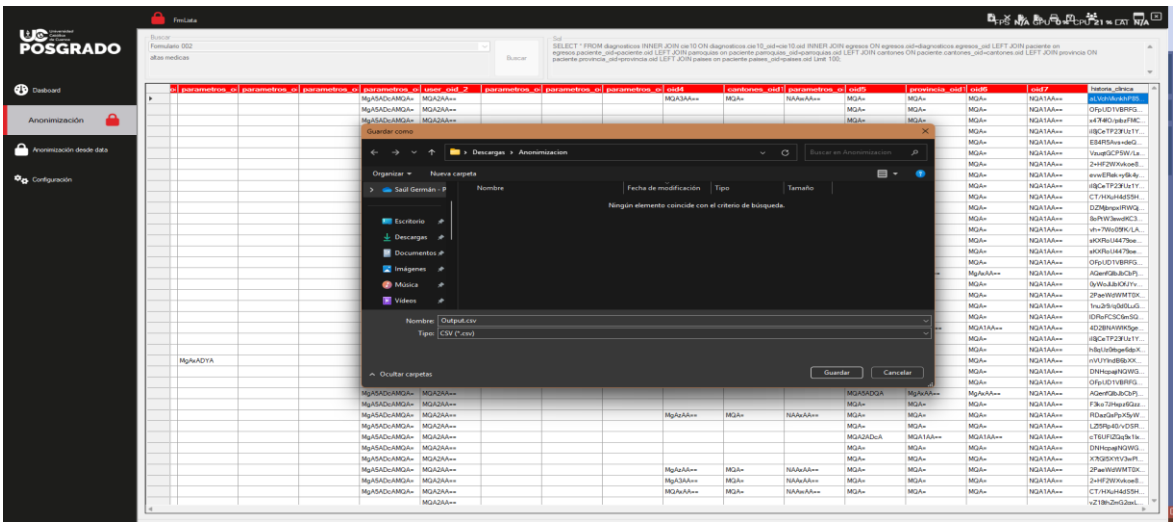


Figura 7 Se procede a guardar los resultados en formato CSV

3.5 Muestra a auditar

Debemos tomar en cuenta que toda la data con la que se generaron las pruebas son datos no reales de los hospitales pero que si cumplen con la estructura de identificadores y cuasi identificadores en la Figura 8 podemos ver una muestra de cómo se visualizan los datos sensibles a ser tratados.

Partiendo del archivo auditoria.cvs el cual tiene los registros del origen de la base de datos que será tratadas mediante el procedimiento de anonimizacion.

Debemos poner atención en los campos considerados sensibles o aquellos que son identificadores o cuasi identificadores. Sobre estos girara el análisis de auditoria a los resultados entregados por el aplicativo de escritorio de anonimazion.

3.5.1 Identificadores

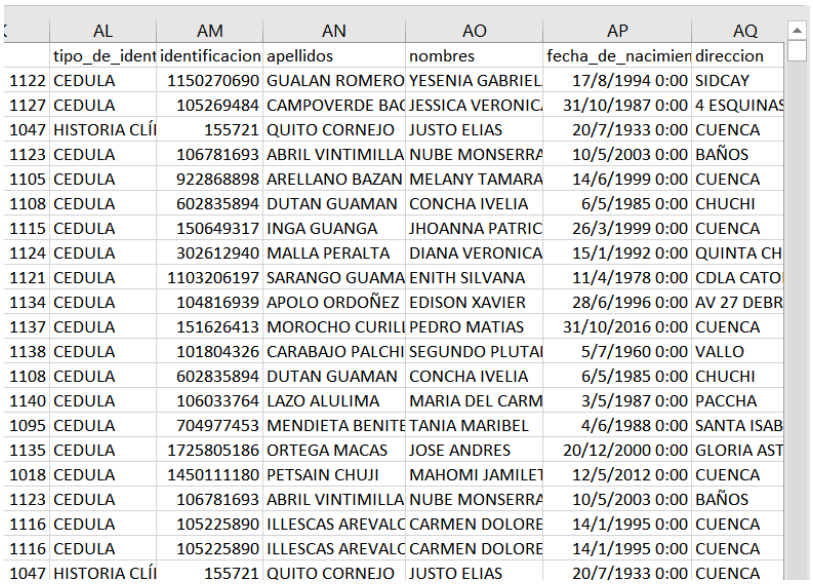
Son los campos que nos permiten de forma directa e inequívoca en la Figura 8 tenemos una muestra de este tipos de campos.

Nombre

Apellidos

Cedula

Dirección



	AL	AM	AN	AO	AP	AQ
	tipo_de_ident	identificacion	apellidos	nombres	fecha_de_nacimien	direccion
1122	CEDULA	1150270690	GUALAN ROMERO	YESENIA GABRIEL	17/8/1994 0:00	SIDCAY
1127	CEDULA	105269484	CAMPOVERDE BAC	JESSICA VERONIC	31/10/1987 0:00	4 ESQUINAS
1047	HISTORIA CLÍ	155721	QUITO CORNEJO	JUSTO ELIAS	20/7/1933 0:00	CUENCA
1123	CEDULA	106781693	ABRIL VINTIMILLA	NUBE MONSERRA	10/5/2003 0:00	BAÑOS
1105	CEDULA	922868898	ARELLANO BAZAN	MELANY TAMARA	14/6/1999 0:00	CUENCA
1108	CEDULA	602835894	DUTAN GUAMAN	CONCHA IVELIA	6/5/1985 0:00	CHUCHI
1115	CEDULA	150649317	INGA GUANGA	JHOANNA PATRIC	26/3/1999 0:00	CUENCA
1124	CEDULA	302612940	MALLA PERALTA	DIANA VERONICA	15/1/1992 0:00	QUINTA CH
1121	CEDULA	1103206197	SARANGO GUAMA	ENITH SILVANA	11/4/1978 0:00	CDLA CATO
1134	CEDULA	104816939	APOLO ORDOÑEZ	EDISON XAVIER	28/6/1996 0:00	AV 27 DEBR
1137	CEDULA	151626413	MOROCHO CURILI	PEDRO MATIAS	31/10/2016 0:00	CUENCA
1138	CEDULA	101804326	CARABAJO PALCHI	SEGUNDO PLUTAI	5/7/1960 0:00	VALLO
1108	CEDULA	602835894	DUTAN GUAMAN	CONCHA IVELIA	6/5/1985 0:00	CHUCHI
1140	CEDULA	106033764	LAZO ALULIMA	MARIA DEL CARM	3/5/1987 0:00	PACCHA
1095	CEDULA	704977453	MENDIETA BENITE	TANIA MARIBEL	4/6/1988 0:00	SANTA ISAB
1135	CEDULA	1725805186	ORTEGA MACAS	JOSE ANDRES	20/12/2000 0:00	GLORIA AST
1018	CEDULA	1450111180	PETSAIN CHUJI	MAHOMI JAMILET	12/5/2012 0:00	CUENCA
1123	CEDULA	106781693	ABRIL VINTIMILLA	NUBE MONSERRA	10/5/2003 0:00	BAÑOS
1116	CEDULA	105225890	ILLESCAS AREVALC	CARMEN DOLORE	14/1/1995 0:00	CUENCA
1116	CEDULA	105225890	ILLESCAS AREVALC	CARMEN DOLORE	14/1/1995 0:00	CUENCA
1047	HISTORIA CLÍ	155721	QUITO CORNEJO	JUSTO ELIAS	20/7/1933 0:00	CUENCA

Figura 8 Datos de origen campos sensibles que serán comparados en los resultados

3.5.2 Cuasi identificadores

Cuasi-identificador es un dato que no identifica a una persona por sí solo, pero que puede ayudar a hacerlo al combinarse con otros datos.

Fecha de nacimiento

Estatura

Sexo

3.5.3 Historia clínica

El identificador historia_clinica es considerado sensible y de alto impacto dentro del proceso de anonimazion. Para este Identificador se aplica un proceso de *Hash*.

AU	AV	AW
historia_clinic	parroquias_oi	tipos_de_at
417479	17	
24013	13	
155721		
554398	1	
52708		
28404		
176282		
355269	31	
521999	27	
86987	34	
209700		
300128	21	
28404		
354037	11	

Figura 9 claramente podemos ver en esta grafica la columna de historial clínico, la cual es considerada un campo sensible por cuanto liga al resto de datos del paciente, esta columna será aplicado un procedimiento de has

3.6 Generación de la consulta SQL mediante la Inteligencia Artificial

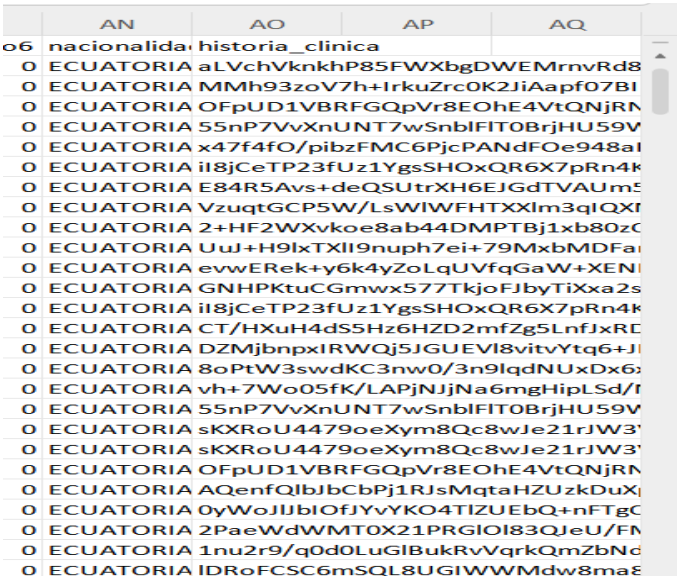
Mediante la utilización de inteligencia artificial se genera la consulta que clasificara los identificadores, cuasi identificadores y los datos sensibles en base al análisis de su encabezado y contenido, de allí se procederán a anonimizar la información.

select * from diagnosticos INNER JOIN cie10 ON diagnosticos.cie10_oid=cie10.oid INNER JOIN egresos ON egresos.oid=diagnosticos.egresos_oid LEFT JOIN paciente on egresos.paciente_oid=paciente.oid LEFT JOIN parroquias on paciente.parroquias_oid=parroquias.oid LEFT JOIN cantones ON paciente.cantones_oid=cantones.oid LEFT JOIN provincia ON paciente.provincia_oid=provincia.oid LEFT JOIN paises on paciente.paises_oid=paises.oid;

3.7 Análisis de los resultados entregados al ejecutar la consulta.

Luego de ejecutar la consulta generada por la IA y aplicar el proceso de anonimizacion vamos a ver la columna historia clínica, en donde podemos evidenciar claramente que el contenido de la

información ha cambiado en su estructura debido a la aplicación de un has como medida de seguridad para evitar la re-identificación tal como lo muestra la Figura 10



	AN	AO	AP	AQ
o6	nacionalida	historia_clinica		
0	ECUATORIA	aLVchVknkhP85FWXbgDWEMrnvRd8		
0	ECUATORIA	MMh93zoV7h+IrkuZrc0K2JiAapf07BI		
0	ECUATORIA	OFpUD1VBRFGQpVr8EOhE4VtQNjRN		
0	ECUATORIA	55nP7VvXnUNT7wSnblFlT0BrjHU59V		
0	ECUATORIA	x47f4fO/pibzFMC6PjcPANdFOe948aI		
0	ECUATORIA	il8jCeTP23fUz1YgsSHOxQR6X7pRn4k		
0	ECUATORIA	E84R5Avs+deQSUtrXH6EJGdTVAUm5		
0	ECUATORIA	VzuqtGCP5W/LsWlWFHTXXlm3qIQXI		
0	ECUATORIA	2+HF2WXvkoe8ab44DMPTBj1xb80zC		
0	ECUATORIA	UuJ+H9lxTXll9nuph7ei+79MxbMDFa		
0	ECUATORIA	evwERek+y6k4yZoLqUVfqGaW+XEN		
0	ECUATORIA	GNHPKtuCGmwx577TkjoFJbyTiXxa2s		
0	ECUATORIA	il8jCeTP23fUz1YgsSHOxQR6X7pRn4k		
0	ECUATORIA	CT/HXuH4dS5Hz6HZD2mfZg5LnfJxRD		
0	ECUATORIA	DZMjbnpxIRWQj5JGUEVl8vitvYtq6+J		
0	ECUATORIA	8oPtW3swdKC3nw0/3n9lqdNUxDx6z		
0	ECUATORIA	vh+7Wo05fK/LAPjNJjNa6mgHipLSd/I		
0	ECUATORIA	55nP7VvXnUNT7wSnblFlT0BrjHU59V		
0	ECUATORIA	sKXRoU4479oeXym8Qc8wJe21rJW3Y		
0	ECUATORIA	sKXRoU4479oeXym8Qc8wJe21rJW3Y		
0	ECUATORIA	OFpUD1VBRFGQpVr8EOhE4VtQNjRN		
0	ECUATORIA	AQenfQlbJbCbPj1RJsMqtaHZUzkDuX		
0	ECUATORIA	0yWoJlJblOfJYvYKO4TlZUEbQ+nFTgC		
0	ECUATORIA	2PaeWdWMT0X21PRGlOl83QJeU/FM		
0	ECUATORIA	1nu2r9/q0d0LuGlBukRvVqrkQmZbNc		
0	ECUATORIA	lDRoFCSC6mSQL8UGlWWMdw8ma8		

Figura 10 Verificamos los resultados entregados luego del proceso de anonimizacion de la columna historia clínica

Como se puede observar el campo historia_clinica se implementa un componente de encriptación o una cadena *Hash* como medida de seguridad para evitar su re t5identificacion.

3.7.1 EL mismo reporte en formato PDF.

El sistema nos permite generar un reporte en formato PDF como lo indica la Figura 11, este reporte es un extracto de las operaciones realizadas en el proceso de anonimizacion de datos.

Iniciar sesión

Buscar texto o herramientas

Ministerio de Salud Pública

Universidad Católica de Cuenca

UNIVERSIDAD CATÓLICA DE CUENCA

Anonymization **IMPRIME: SAUL GUAMAN SAUL GERMAN**

INVESTIGADOR: SANTILLAN VINICIO **CEN. INVESTIGACION:** UNIVERSIDAD CATOLICA DE CUENCA

ESTUDIO: AUDITORIA

DESCRIPCION: AUDITORIA

NOMBRE	SERVICIO	REGISTRO
CONSULTA SQL	CONSULTA SQL	605.522,00
NUMERO DE COLUMNAS TOTALES	NUMERO COLUMNAS	106,00
NUMERO DE COLUMNAS	NUMERO COLUMNAS ANONIMI	65,00
ELIMINADO: OID	ANONIMIZACION	0,00
ELIMINADO: INGRESOS_OID	ANONIMIZACION	0,00
ELIMINADO: CIE10_OID	ANONIMIZACION	0,00
ELIMINADO: EGRESOS_OID	ANONIMIZACION	0,00
ELIMINADO: OID1	ANONIMIZACION	0,00
ELIMINADO: SUBCAPITULO_CIE_OID	ANONIMIZACION	0,00
ELIMINADO: PARAMETROS_OID	ANONIMIZACION	0.00

Figura 11 Resultados del proceso de anonimizacion donde se genera un análisis de las operaciones efectuadas a los datos originales.

3.8 Análisis y pruebas

3.8.1 Prueba de análisis de campo

Uno de los métodos utilizados para la verificación de un posible proceso de reidentificación es la observación de datos originales y datos de salida, partiendo de tres ejes fundamentales. En primer lugar, se analiza si los datos sensibles de origen continúan presentes en los datos anonimizados, lo cual permite verificar la efectividad del proceso (Samarati & Sweeney, 1998). En segundo lugar, se revisa si campos clave como *historia_clínica* o identificadores primarios mantienen su formato o estructura original, lo que podría facilitar la correlación y posterior identificación (El Emam, 2010). Finalmente, se analiza la agrupación de los resultados, con el fin de identificar *quasi-*

identificadores que puedan formar patrones de reidentificación con los datos anonimizados entregados (Article 29 Data Protection Working Party, 2014).

3.8.2 Análisis utilizando la herramienta ARX.

Dentro del mundo de las tecnologías existen diversas soluciones para la anonimizacion y para el análisis de los datos entregados por este tipo de aplicativos. Para nuestra evaluación hemos seleccionado la herramienta ARX, la cual es de código abierto y entre sus bondades cuenta con algunas etapas de verificación y análisis sobre riesgos de re-identificación de datos anonimizados. Lo primero que procedemos a realizar es la generación de un proyecto que para nuestro caso se llamara auditoria tal como se indica en la Figura 12

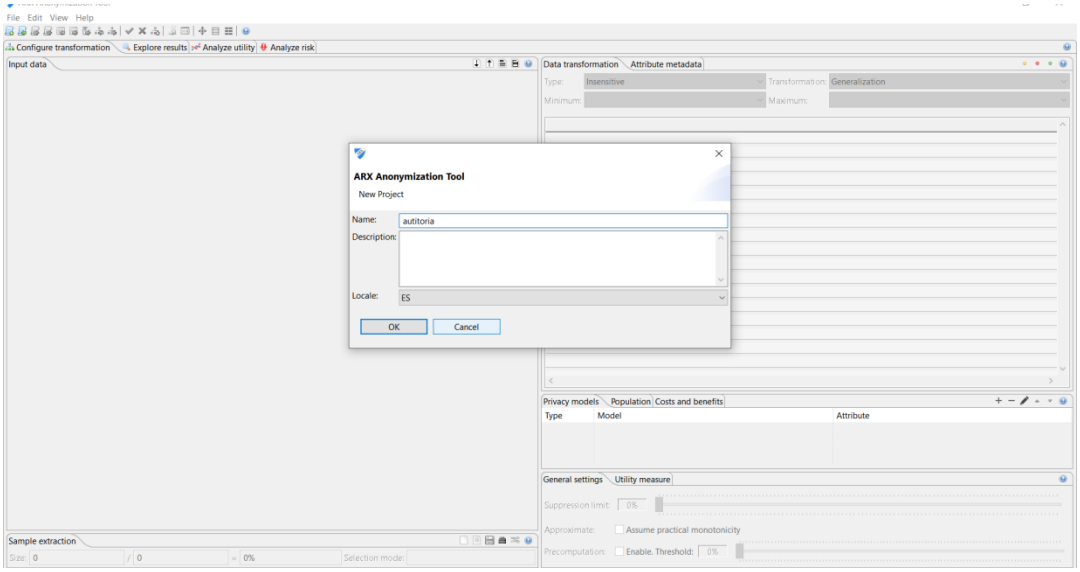


Figura 12 El análisis utilizando la herramienta ARX nace con la creación del proyecto que para nuestro análisis tendrá el nombre de auditoria.

Luego de generar el proyecto como se indica en la Figura 12, se procede a importar la data con la cual vamos a trabajar, en nuestro caso es un documento con formato cvs. El cual procedemos a cargar como se indica en la Figura 13.

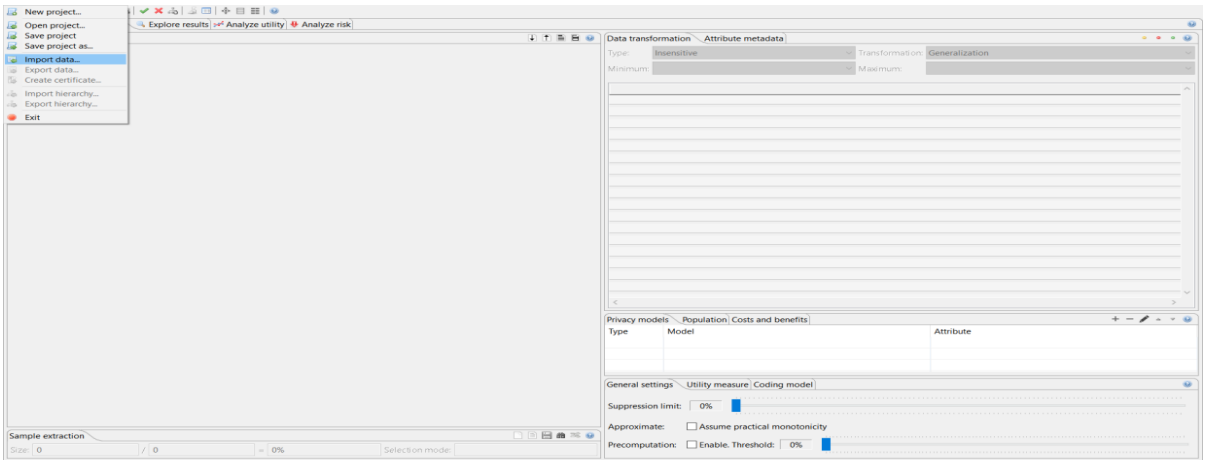


Figura 13 procedemos a importar los datos con los cuales trabajaremos.

Luego de realizar la carga de los datos como se indica en el anterior paso, visualizamos los datos cargados tal como se indica en la Figura 14.

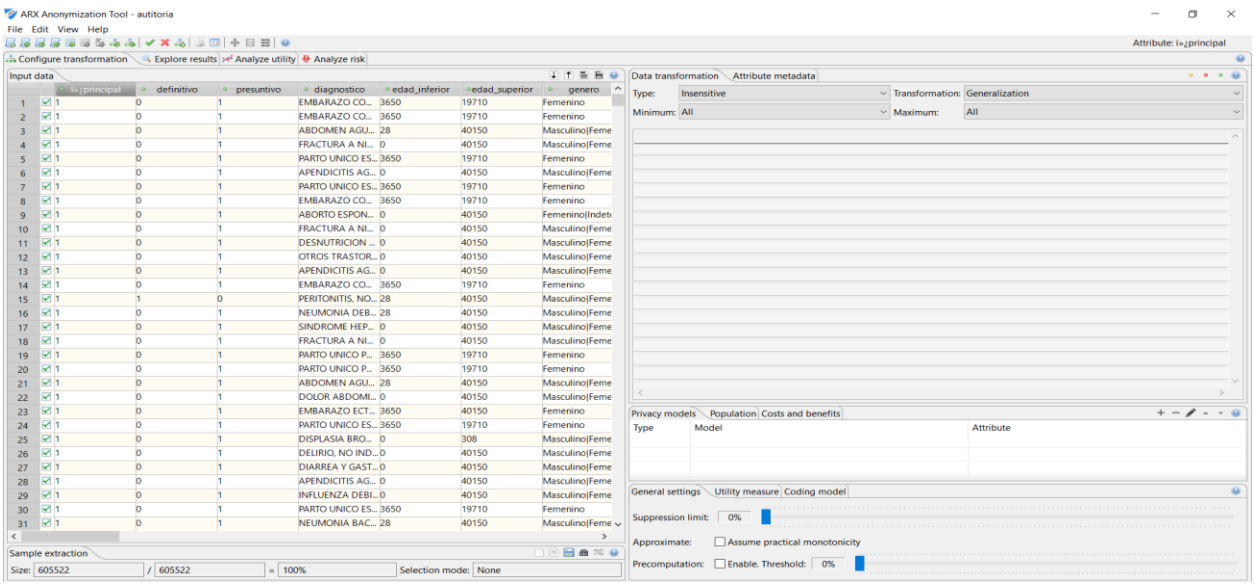


Figura 14 Visualización de datos para realizar el análisis

Uno de los análisis que me permite esta herramienta está basado en el modelo HIPAA Ley de

Responsabilidad y portabilidad lo cual se indica en la Figura 15.

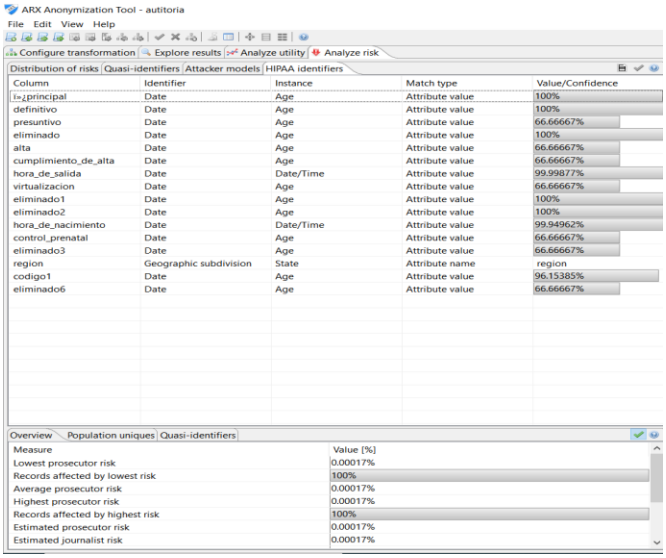


Figura 15 Análisis de los datos mediante la herramienta ARX, en el módulo HIPAA Ley de responsabilidad y portabilidad.

Procedemos a realizar el análisis de cuasi identificadores donde tenemos algunos parámetros de analizamos, en base a los resultados entregados, como se indica en la Figura 16.

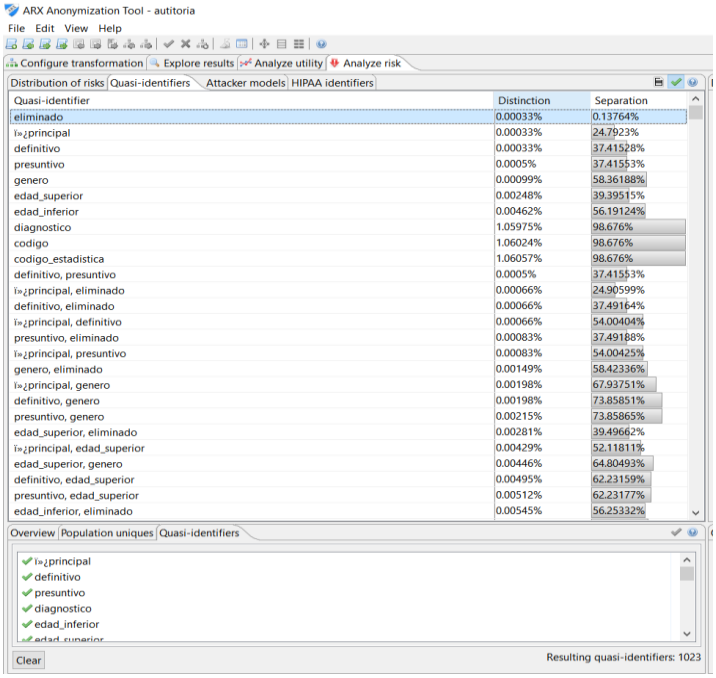


Figura 16 Análisis de los cuasi identificadores

3.9 Interpretación de datos.

El riesgo más bajo Medida Valor [%] Riesgo fiscal más bajo es: 0,00017 %. Esto no da como pauta que es un riesgo muy bajo para como factor de re-identificación.

El número de registros afectados por el riesgo más bajo es el 100%, esto nos indica que el 100% de la data están en el rango de un riesgo bajo de poder ser re identificado con los cuasi Identificadores.

El Riesgo fiscal más alto: 0,00017 %, es igual al riesgo fiscal más bajo lo cual es un indicador de satisfacción del proceso de anonimizacion de datos excelente, indicándonos que contamos con mínimas posibilidades de re-identificación de datos.

Otro análisis que se realiza mediante la herramienta ARX es el riesgo de re-identificación como se muestra en la Figura 17.

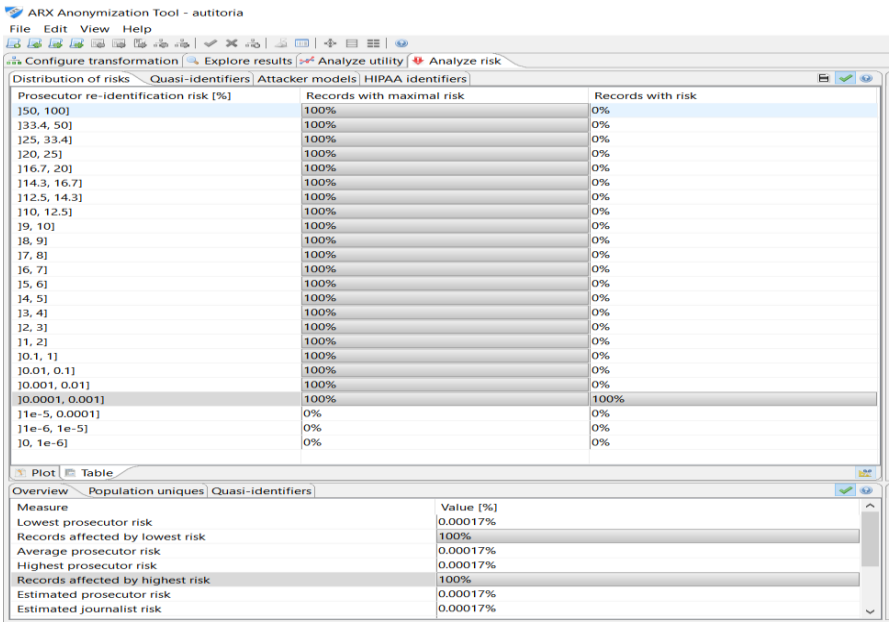


Figura 17 En esta grafica podemos analizar los registros con máximo riesgo de re-identificación, del 100% con riesgo 0%.

CAPITULO IV

4.1 Discusión

Esta investigación evaluó la efectividad del proceso de anonimización de datos sensibles en bases de datos hospitalarias utilizando una aplicación de escritorio, donde la generación de las consultas se las realiza mediante inteligencia artificial. Para ello, se analizaron los resultados generados por la herramienta, que anonimizó los identificadores y cuasi identificadores usando un hash de 256 bytes. A excepción del campo de historia clínica, que fue encriptado usando el número de historia clínica de cada paciente y un número aleatorio para obtener un código alfanumérico único por paciente. Esto con la finalidad de identificar individuos sin la posibilidad de acceder a información que permita identificar a los pacientes. Por lo tanto, esta auditoría evaluó la exactitud y precisión del sistema de anonimización de los campos sensibles de las bases de datos de los hospitales. Adicionalmente, se verificó la robustez del proceso del anonimización, mediante técnicas de ingeniería inversa, estableciendo el grado de privacidad de los datos anonimizados en referencia al riesgo de re-identificación. Se encontró que la herramienta de anonimización cumple con el proceso de identificar los campos sensibles (identificadores y cuasi identificadores), y anonimizar correctamente cada uno de estos campos. Además, se encontró que el campo de historia clínica tiene una seguridad robusta, lo cual dificulta el proceso de desencriptado. Finalmente, esta auditoría concluye que el producto obtenido por el app de escritorio, presenta un alto grado de confiabilidad y el riesgo de re-identificación mediante técnicas de ingeniería inversa es bajo.

Los resultados obtenidos indican que el riesgo fiscal más bajo es de 0,00017 %, siendo este también el valor del riesgo fiscal más alto, lo que sugiere una uniformidad en la seguridad de los datos procesados y un riesgo mínimo para la re-identificación. Además, el 100 % de los registros analizados se encuentran dentro de este nivel de riesgo, lo que representa un indicador positivo de

la eficacia del método de anonimización empleado. Según estudios previos, una tasa de riesgo de re-identificación inferior al 0,05 % es considerada altamente segura para el uso de datos en investigaciones científicas (El Emam et al., 2011). La baja probabilidad de re-identificación obtenida en este estudio reafirma la conformidad del sistema con la Ley Orgánica de Protección de Datos Personales del Ecuador (2021).

Por lo tanto, esta herramienta es altamente confiable en comparación con otros métodos de anonimización utilizados en entornos médicos, como la generalización y la permutación de datos (Sweeney, 2002). La aplicación de escritorio evaluada en este estudio logra una mayor protección sin comprometer la utilidad de los datos anonimizados. Esto en línea con investigaciones previas que han demostrado que la aplicación de técnicas basadas en IA para la detección de identificadores y cuasi-identificadores mejora significativamente la precisión en la anonimización de los datos de salud (Dankar et al., 2019).

Sin embargo, es importante considerar que la eficacia del sistema de anonimización también depende de la naturaleza de los datos procesados y de los ataques potenciales de re-identificación (CITA). Estudios recientes han demostrado que incluso con una baja probabilidad de re-identificación, ataques de vinculación con bases de datos externas pueden comprometer la privacidad de los pacientes (Rocher et al., 2019). Por lo tanto, se recomienda continuar con auditorías periódicas y la implementación de medidas adicionales, como la supresión de atributos altamente identificables y el uso de técnicas diferenciales de privacidad (Dwork, 2006).

En conclusión, los resultados de la auditoría realizada a la aplicación de anonimización indican que el sistema cumple con los estándares de protección de datos personales y garantiza un bajo riesgo de re-identificación. La aplicación utilizada para la detección de identificadores y

cuasi-identificadores ha demostrado ser una estrategia eficaz, alineándose con las mejores prácticas en el ámbito de la privacidad de los datos de salud. No obstante, es necesario mantener un monitoreo continuo y adaptar las técnicas de anonimización a medida que evolucionan los riesgos y desafíos en la protección de la información médica.

Esta investigación evidencia que el proyecto cumple con los principios fundamentales establecidos tanto en la Ley Orgánica de Protección de Datos Personales del Ecuador como en el Reglamento General de Protección de Datos de la Unión Europea (RGPD). La correcta identificación y anonimización de datos sensibles, especialmente los relacionados con la salud, demuestra la aplicación efectiva del principio de minimización y confidencialidad, reduciendo significativamente el riesgo de re-identificación. Además, la seguridad reforzada en el campo de historia clínica refleja el cumplimiento del principio de integridad y confidencialidad que exige la normativa europea. El bajo nivel de riesgo encontrado respalda la adopción de medidas técnicas adecuadas y proporcionales, conforme al enfoque de prevención y responsabilidad proactiva contemplado en ambas leyes. En conjunto, estos hallazgos confirman que el tratamiento de los datos en este proyecto se alinea con estándares legales nacionales e internacionales de protección de datos personales.

4.2 Conclusiones

1. **Cumplimiento del principio de minimización y confidencialidad de los datos sensibles**

   Se verificó que la herramienta de anonimización implementada identifica correctamente los campos sensibles, incluyendo los identificadores directos y cuasi-identificadores, conforme lo establece el artículo 25 de la LOPDP. La anonimización aplicada asegura que la información de salud considerada como dato sensible según el artículo 5 de la ley no pueda ser asociada directa o indirectamente a una persona natural identificada o identificable.

2. **Mecanismos robustos de protección de historias clínicas**

   El campo correspondiente a la historia clínica fue objeto de especial atención, constatándose que cuenta con mecanismos de seguridad criptográfica robustos. La complejidad del cifrado implementado incrementa significativamente la dificultad de un posible proceso de desencriptado no autorizado, alineándose con el principio de seguridad contemplado en el artículo 26 de la LOPDP.

3. **Evaluación del riesgo de re-identificación**

   Luego de la auditoría, se concluye que el producto generado por la aplicación de escritorio tiene un alto grado de confiabilidad. El riesgo de re-identificación a través de técnicas como la ingeniería inversa se considera bajo, lo cual refuerza el cumplimiento del deber de prevención de vulnerabilidades establecido en el artículo 33 de la normativa LOPDP.

4.3 Recomendaciones

1. **Mantener un esquema de mejora continua en la anonimización**

   Aunque el riesgo de re-identificación es bajo, se recomienda implementar una política de revisión periódica de la eficiencia del manejo de la Inteligencia artificial en el proceso de

consultas en el proceso de anonimización, especialmente frente al avance de técnicas de correlación de datos y machine learning que podrían comprometer la anonimidad.

2. **Establecer controles de auditoría interna y trazabilidad**

   Registrar cada proceso de anonimización y acceso a la información sensible, conforme al principio de responsabilidad proactiva y rendición de cuentas (art. 14 LOPDP). Esto contribuirá a la transparencia en el tratamiento de datos personales.

3. **Capacitación continúa al personal**

   Es fundamental continuar con la capacitación técnica y legal del personal responsable del tratamiento de datos personales, para fortalecer la cultura de protección de datos y garantizar el uso adecuado de las herramientas de anonimización.

4. **Evaluación de impacto en protección de datos (EIPD)**

   Se recomienda que, en futuras versiones de la herramienta o nuevas aplicaciones que manejen datos sensibles, se realice una Evaluación de Impacto en Protección de Datos conforme al artículo 44 de la LOPDP, a fin de anticipar riesgos y aplicar medidas preventivas eficaces.

Bibliografía

- General Data Protection Regulation (GDPR). (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Retrieved from https://eur-lex.europa.eu.
- ISO/IEC 27001:2013. (2013). Information technology — Security techniques — Information security management systems — Requirements. International Organization for Standardization.
- Ley Orgánica de Protección de Datos Personales del Ecuador. (2021). Registro Oficial Suplemento 459 de 26 de mayo de 2021. Asamblea Nacional del Ecuador. Retrieved from https://www.asambleanacional.gob.ec.
- Universidad Católica de Cuenca. (2024). Propuesta de directrices de ciberseguridad para redes de comunicación de infraestructuras críticas en la distribución eléctrica del Ecuador. Cuenca, Ecuador: Autor.
- Phoenix Contact. (n.d.). IEC 62443: Norma de ciberseguridad industrial. Retrieved from https://www.phoenixcontact.com.
- Quiroz Tascón, D., Rosas, M., & Medina, J. (2020). La ciberseguridad en sistemas de control industrial: un análisis de vulnerabilidades. Revista Técnica Electrónica, 28(2), 113–125. https://doi.org/10.1234/rte.2020.28.2.113
- Torres Valero, F. (2020). Metodologías de trabajo y mejores prácticas para la seguridad en infraestructuras críticas. Journal of Cybersecurity Research, 14(1), 45–58.

- Dankar, F. K., El Emam, K., Neisa, A., & Roffey, T. (2019). Estimating the re-identification risk of clinical data sets. *Journal of Biomedical Informatics, 92*, 103135.
- Dwork, C. (2006). Differential privacy. *International Colloquium on Automata, Languages, and Programming*, 1-12.
- El Emam, K., Jonker, E., Arbuckle, L., & Malin, B. (2011). A systematic review of re-identification attacks on health data. *PLOS ONE*, *6*(12), e28071.
- Rocher, L., Hendrickx, J. M., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*(1), 1-9.
- Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*(5), 557-570.
- El Emam, K. (2010). *Guide to the De-Identification of Personal Health Information*. CRC Press.
- Ley Orgánica de Protección de Datos Personales, Registro Oficial Suplemento No. 459, 26 de mayo de 2021 (Ecuador).
- Reglamento General de Protección de Datos (RGPD). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Diario Oficial de la Unión Europea, L 119, 1-88.

ANEXOS

Anexo (1)

UNIVERSIDAD CATOLICA DE CUENCA

26-Enero-2025

Dr. Dr. Vinicio Santillán
Director del proyecto de anonimizacion de datos de salud

Asunto: Notificación Formal de Inicio de Examen de Auditoría

Estimado Dr. Dr. Vinicio Santillán:

Por la presente, se le notifica formalmente del inicio del examen de auditoría sobre los resultados entregados por el sistema de anonimización de datos de salud. Esta auditoría tiene como objetivo evaluar la correcta aplicación de los procedimientos de anonimización y la calidad de los resultados generados por el software utilizado en base a la Ley orgánica de protección de datos personales del ecuador.

Para cumplir con los objetivos de esta auditoría, se requiere que, en un plazo no mayor a **72 horas** a partir de la recepción de esta notificación, se entregue la siguiente documentación e información:

1. **Resultados emitidos por el software de anonimización:**
    - Resultados completos en formato digital y físico
    - Copias impresas en físico de los resultados correspondientes.

El auditor responsable designado para este examen es el Ing. Msc. Oscar Carrera Pozo, quien estará disponible para resolver cualquier consulta o requerimiento adicional relacionado con esta auditoría.

Se solicita su colaboración para garantizar el cumplimiento de los plazos establecidos y el adecuado desarrollo de esta actividad. Para cualquier consulta o para coordinar la entrega de la información solicitada, puede comunicarse directamente con el auditor responsable al siguiente correo electrónico: freedcar@yahoo.com o al número telefónico: 0968975081.

Agradeciendo de antemano su atención y pronta respuesta, quedo a su disposición para cualquier duda o comentario.

Atentamente,

Ing. Msc, Oscar Freed Carrera Pozo

Auditor Informático

Anexo (2)

INFORME DE AUDITORÍA IFORMATICA

EXAMEN DE CUMPLIMIENTO NORMATIVO **EVALUACIÓN A LA EFECTIVIDAD A LOS RESULTADOS DEL PROCESO DE ANONIMIZACION DEL SOFTWARE**			
TIPO INFORME:	**GESTIÓN**	No. INFORME:	**AINF-001**
COMPONENTE:	**EVALUACIÓN DE RIESGOS**	SUBCOMPONENTE:	**RIESGO TECNOLOGICO**
TIPO DE PROCESO:	**OPERATIVO**	FECHA DE CORTE:	**30.11.2024**
FECHA INICIO:	**11.12.2024**	FECHA FIN:	**29.01.2025**
FECHA INFORME PRELIMINAR:	**19.12.2024**	FECHA INFORME FINAL:	

Antecedentes

La ley orgánica de protección de datos personales del ecuador indica Los Artículos 31 y 32 de la LOPDP establecen la necesidad de anonimizar los datos de salud antes de su tratamiento con fines de investigación. Sin embargo, un aspecto crítico a considerar es la armonización de estas normativas con regulaciones internacionales como el GDPR**,** lo que podría fortalecer la interoperabilidad de los datos para estudios globales.

Objetivos

Analizar los productos generados por la aplicación de escritorio, evaluando la exactitud y precisión del sistema de anonimización, valorando la calidad del proceso de identificación y anonimización de los campos sensibles en bases de datos de salud.

Evaluar la robustez del proceso del anonimización, mediante técnicas de ingeniería inversa, con el fin de garantizar la privacidad de los datos anonimizados.

Alcance

Analizar el cumplimiento de la norma de anonimizacion de datos personales de la LOPDP.

Limitantes

Sin limitantes que informar

Base normativa:

- Ley orgánica de protección de datos personales

Procedimientos

- Plan de auditoria
- Análisis de los requerimientos de información solicitados al área de tecnologías.
- Presentación del informe preliminar y solicitud de descargos
- Análisis de descargos
- Presentación del informe final

Desarrollo

Información requerida.

Por la presente, se le notifica formalmente del inicio del examen de auditoría sobre los resultados entregados por el sistema de anonimización de datos de salud. Esta auditoría tiene como objetivo

evaluar la correcta aplicación de los procedimientos de anonimización y la calidad de los resultados generados por el software utilizado en base a la Ley orgánica de protección de datos personales del ecuador.

Para cumplir con los objetivos de esta auditoría, se requiere que, en un plazo no mayor a **72 horas** a partir de la recepción de esta notificación, se entregue la siguiente documentación e información:

Resultados emitidos por el software de anonimización:

- Resultados completos en formato digital y físico
- Copias impresas en físico de los resultados correspondientes.

El auditor responsable designado para este examen es el Ing. Msc. Oscar Carrera Pozo, quien estará disponible para resolver cualquier consulta o requerimiento adicional relacionado con esta auditoría.

Se solicita su colaboración para garantizar el cumplimiento de los plazos establecidos y el adecuado desarrollo de esta actividad. Para cualquier consulta o para coordinar la entrega de la información solicitada, puede comunicarse directamente con el auditor responsable al siguiente correo electrónico: freedcar@yahoo.com o al número telefónico: 0968975081.

Agradeciendo de antemano su atención y pronta respuesta, quedo a su disposición para cualquier duda o comentario.

I.ANÁLISIS DE LOS REQUERIMIENTOS

Mediante reunión mantenida con el responsable del proyecto de anonimizacion se procede a realizar la verificación de los datos orígenes y resultados entregados por el sistema creado con un componente Inteligencia artificial.

Hallazgo

Condición: Luego de las pruebas realizadas a los datos entregados, y del análisis efectuado se determina una oportunidad de mejora con la implementación de inteligencia artificial para la generación de consultas, se determina que se cumple con lo establecido en la Ley orgánica de protección de datos personales y el manual

Recomendación: Se debe ampliar la utilización de la Inteligencia artificial con la finalidad de mejorar los tiempos en el desarrollo y las consultas, de igual forma realizar auditorías de control.

Atentamente,

AUDITOR INFORMÁTICO

Ing. Oscar Freed Carrera Pozo, Msc.